# 1 JEDNOROZMĚRNÁ POPISNÁ STATISTIKA – 1. ČÁST

***Pojem statistika*** lze chápat v zásadě ve třech pojetích:

1. číselné nebo slovní údaje (data) a jejich souhrny o hromadných jevech
2. činnost spočívající ve sběru, zpracování a vyhodnocování dat o hromadných jevech
3. teoretická disciplína (věda), která zkoumá zákonitosti hromadných jevů, resp. souhrn vědeckých metod sběru, zpracování a analyzování dat.

## 1.1 Základní statistické pojmy

***Hromadné jevy***
- takové skutečnosti, které se vyskytují mnohokrát a mohou se znovu opakovat
- jevy, které se vyskytují v masovém měřítku u velkého počtu prvků (běžně alespoň 30)
- pouze v případě hromadnosti lze předpokládat, že to, co je ve zkoumaných vlastnostech prvků podstatné a pravidelné převáží nad tím, co je individuální, náhodné atd.

***Statistický soubor***
- množina prvků s přesně stanovenými shodnými vlastnostmi (např. množina osob, organizací, atd.).

***Statistická jednotka***
- prvek statistického souboru
- individuální nositel vlastností daného statistického souboru.

***Rozsah statistického souboru***
- počet jednotek statistického souboru (symbolické značení – ***n, N)***.

Existují dvě možnosti přístupu ke statistickému souboru, jejich chápání je přitom relativní a závisí na konkrétní situaci.

- ***Základní soubor (populace):*** statistický soubor všech jednotek, které jsou předmětem zkoumání, obvykle velmi rozsáhlý, rozsah značíme ***N.***

- ***Výběrový soubor (výběr):*** vzorek ze základního souboru, pořízený tak, že se určitým způsobem vyberou pouze některé jednotky, rozsah značíme ***n***.

***Statistický znak***
- označení (odraz) určité vlastnosti, kterou má v té či oné míře každá jednotka daného statistického souboru
- u souboru osob např. věk, váha, výška, atd.

***Hodnota statistického znaku (pozorování)***
- míra dané vlastnosti (statistického znaku) u každé jednotky statistického souboru.

*Počet hodnot (pozorování) = rozsah souboru (n).*

***Obměna (varianta) statistického znaku***
- hodnota ve smyslu vyjádření různého stupně dané vlastnosti.

*Počet variant (k) ≤ rozsah souboru (n).*

***Statistický znak shodný:*** v daném statistickém souboru nabývá pouze jedné varianty.
***Statistický znak proměnný:*** v daném statistickém souboru nabývá více než jedné varianty. Ekvivalentní označení: ***statistická proměnná***.

## 1.2 Druhy proměnných

Klasifikace proměnných může být prováděna z několika různých hledisek. Správné určení druhu proměnných je nezbytné pro volbu adekvátních metod jejich zpracování a analýzy.

### 1. Způsob vyjádření hodnot proměnné

- ***slovní (kategoriální, alfabetické, kvalitativní):*** jsou vyjádřeny slovy
- ***číselné (numerické, kvantitativní):*** jsou vyjádřeny čísly**.**

### 2. Typ vztahů mezi obměnami a hodnotami proměnné

- ***nominální (jmenné, názvové):*** slovní proměnné, jejichž obměny nelze hierarchicky uspořádat, tzn., že nelze jednoznačně stanovit, která je nižší a která vyšší. O jejích obměnách lze pouze konstatovat, zda jsou stejné nebo různé. *Např.: pohlaví, jméno, rodinný stav, atd.*

- ***ordinální (pořadové):*** slovní i číselné proměnné, jejichž obměny lze jednoznačně seřadit od nejnižší k nejvyšší nebo obráceně. Jejich obměny lze porovnávat rozdílem, ale ne podílem. *Např.: nejvyšší dokončené vzdělání, hodnosti v armádě, pořadí v soutěži, atd.*

- ***metrické (měřitelné):*** vždy číselné, jsou udány v určitých měrných jednotkách – vyjadřují tedy velikost měřených vlastností. Nabývají jak kladných, tak nekladných hodnot. Lze změřit o kolik je jedna obměna větší či menší než druhá. Obměny lze porovnávat rozdílem, někdy také podílem (ne vždy – pokud jsou některé obměny záporné či nulové, není to možné). *Např.: teplota vzduchu, zisk podniku, atd.*

- ***kardinální (stěžejní):*** ty metrické proměnné, které nabývají pouze kladných hodnot, jejich obměny lze porovnávat jak rozdílem, tak podílem. Je tedy možno změřit, o kolik měrných jednotek je jedna obměna větší či menší než druhá a také kolikrát je jedna obměna větší či menší než druhá. Např.: věk, váha, výška, atd.

### 3. Počet variant, kterých proměnné nabývají

- ***alternativní:*** nabývají pouze dvou obměn
- ***množné:*** nabývají více než dvou obměn.

**4. Počet hodnot, kterých proměnné nabývají**

- ***diskrétní (nespojité):*** nabývají spočetně mnoha hodnot z konečného či nekonečného intervalu. *Např.: počet dětí v rodině, atd.*

- ***spojité (kontinuální):*** nabývají všech hodnot z konečného či nekonečného intervalu. *Např.: spotřeba elektrické energie, příjem, atd.*

# 2 JEDNOROZMĚRNÁ POPISNÁ STATISTIKA – 2. ČÁST

Ke statistickému zkoumání je třeba mít k dispozici statistická data (údaje, pozorování), která získáváme prostřednictvím statistických šetření. *Šetření v sociálně-ekonomické oblasti* přitom vykazují určitá specifika, takže data z nich pocházející je třeba zpracovávat za pomoci vhodných statistických metod a postupů.

## 2.1 Zpracování dat

- *hodnoty proměnných = data = údaje = pozorování*
- většinou jde o větší množství údajů, které jsou více či méně nepřehledné
- prvním krokem při zpracování dat je proto jejich zpřehlednění (setřídění)
- zpřehlednění provádíme za pomoci *statistických tabulek a grafů*
- cílem tohoto procesu je, aby vynikly charakteristické rysy a zákonitosti souboru.

### 2.1.1 Statistické tabulky

***Tabulka prostého rozdělení četností***

- je vhodná při zpracování diskrétní proměnné s nepříliš velkým počtem obměn.

Tabulka 2.1

| Obměna proměnné $x_i$ | Četnost | | Kumulativní četnost | |
|---|---|---|---|---|
| | absolutní $n_i$ | relativní $p_i$ | absolutní | relativní |
| $x_1$ | $n_1$ | $p_1$ | $n_1$ | $p_1$ |
| $x_2$ | $n_2$ | $p_2$ | $n_1 + n_2$ | $p_1 + p_2$ |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |
| $x_k$ | $n_k$ | $p_k$ | $n$ | 1 |
| Celkem | $n$ | 1 | × | × |

$$p_i = \frac{n_i}{\sum_{i=1}^{k} n_i} = \frac{n_i}{n} \quad ; \qquad \sum_{i=1}^{k} n_i = n \quad ; \qquad \sum_{i=1}^{k} p_i = 1$$

***Tabulka intervalového rozdělení četností***

- je vhodná v případě zpracování diskrétní proměnné s mnoha obměnami nebo spojité proměnné
- klíčovým problémem je určení optimálního počet intervalů ($k$), na které rozdělíme variační rozpětí ($R$)
- k tomu slouží různá pravidla (např. Sturgesovo pravidlo: $k \approx 1 + 3{,}3 \log_{10} n$ )
- při výpočtech lze pak každý interval zastoupit jeho středem
- výsledky takovýchto výpočtů jsou samozřejmě pouze přibližné.

## 2.2 Grafy

Další možností zpřehlednění statistických údajů je jejich grafické znázornění. Existuje mnoho druhů grafů, vždy je třeba vybrat takový, který odpovídá charakteru dat.

***Polygon četností***
- spojnicový graf
- vhodný pro znázornění prostého rozdělení četností.

***Histogram četností***
- sloupkový graf
- vhodný pro znázornění intervalového rozdělení četností.

***Výsečový graf (piechart)***
- vhodný pro znázornění rozdělení četností nominální proměnné.

***Sloupkový graf (barchart)***
- vhodný pro znázornění rozdělení četností nominální proměnné.

***Graf stem-and-leaf***
- vhodný pro znázornění numerické proměnné s velkým množstvím obměn.

***Krabicový graf***
- vhodný pro znázornění extrémních hodnot a kvartilů
- užitečný pro srovnání určitého numerického znaku ve dvou či více souborech.

# 8 MATEMATICKÁ STATISTIKA – 1. ČÁST

- na základě výběrových dat usuzujeme na obecnější skutečnosti, týkající se celého *základního souboru, tzv. populace*
- provádíme zevšeobecňující neboli induktivní úsudek
- induktivní usuzování pomocí matematicko-statistických metod se nazývá *statistická indukce*
- induktivní uvažování s sebou vždy nese riziko nesprávného úsudku, jinak řečeno riziko omylu
- aby bylo možno správně aplikovat metody a postupy matematické statistiky, musí být výběrová data pořízena náhodným výběrem, což znamená, že o tom, zda určitá jednotka ZS bude vybrána nebo ne rozhoduje pouze náhoda.

V praxi se používají *různé druhy náhodného výběru*:

- prostý náhodný výběr
- oblastní (stratifikovaný) výběr
- dvoustupňový (vícestupňový) výběr
- výběr skupin atd.

Výběr jednotek je vždy třeba provádět tak, aby nebyla narušena náhodnost vybírání, tedy *použít vhodnou techniku výběru:*

- losování
- výběr pomocí náhodných čísel
- systematický výběr atd.

*Matematická statistika zahrnuje dvě oblasti:*
- teorii odhadu
- testování statistických hypotéz.

## 8.1 Teorie odhadu

- souhrn metod a postupů, kterými lze z napozorovaných hodnot NV získat co nejlepší odhady neznámých parametrů rozdělení NV.

### 8.1.1 Bodový odhad

- spočívá v nahrazení neznámé hodnoty parametru základního souboru (dále jen ZS) hodnotou vhodné výběrové charakteristiky, která bude sloužit jako dobrá náhrada neznámého parametru
- vhodnost jednotlivých odhadů posuzujeme podle jeho vlastností.

***Vlastnosti bodového odhadu:***

1. nevychýlenost (nestrannost, nezkreslenost)
2. konzistence
3. vydatnost
4. výběrová charakteristika má být tzv. postačující.

***Symbolika:***

- parametry základního souboru značíme obecně $\Theta$ (konkrétně např. $\mu, \sigma, ...$)
- výběrové charakteristiky značíme obecně *t* (konkrétně např. $\bar{x}, s_x, ...$)
- *výběrová chyba:* $t - \Theta$
- *symbolický zápis bodového odhadu:* $est\ \Theta = t$ nebo $t \sim \Theta$.

### 8.1.2 Intervalový odhad

- spočívá v konstrukci náhodného intervalu, od něhož se zvolenou pravděpodobností $P = 1 - \alpha$ očekáváme, že bude obsahovat skutečnou hodnotu neznámého parametru
- dovoluje, abychom uvažovali pravděpodobnost, s níž lze očekávat, že odhad je správný.

# 9 MATEMATICKÁ STATISTIKA – 2. ČÁST

## 9.1 Intervalový odhad

***Spolehlivost odhadu* $\mathbf{1-\alpha}$**

- je to pravděpodobnost; $0 < \alpha < 1$
- volíme vždy číslo blízké 1, nejčastěji 0,95 (event. 0,99 nebo 0,9)
- čím vyšší spolehlivost požadujeme, tím je za jinak stejných podmínek interval spolehlivosti (dále jen IS) širší.

***Riziko odhadu* $\boldsymbol{\alpha}$**

- udává, v kolika případech ze 100 (tj. v jakém procentu případů) nebude IS pokrývat odhadovaný parametr $\Theta$.

*Intervaly spolehlivosti mohou být konstruovány jako:*

1. oboustranné, kde $\Theta_d < \Theta < \Theta_h$
2. jednostranné: pravostranné, kde $\Theta < \Theta_h$
   levostranné, kde $\Theta > \Theta_d$.

$\Theta_h$ je horní mez, $\Theta_d$ je dolní mez.

### 9.1.1 Odhad parametru $\mu$ (střední hodnoty) normálního rozdělení

***1. Bodový odhad***

Bodovým odhadem střední hodnoty $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ je *výběrový průměr* $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Je to nevychýlený odhad střední hodnoty $\mu$.

***2. Intervalový odhad***

*Při konstrukci IS pro parametr μ rozlišujeme následující situace:*

- **Velký výběr z normálního rozdělení se známým rozptylem $\sigma^2$:**

  ***Oboustranný IS:***

$$P\left(\bar{x} - u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

  ***Pravostranný IS:***

$$P\left(\mu < \bar{x} + u_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

***Levostranný IS:***

$$P\left(\bar{x} - u_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} < \mu\right) = 1 - \alpha$$

$\Delta = u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ je *přípustná chyba odhadu.*

- **Velký výběr z normálního rozdělení s neznámým rozptylem $\sigma^2$:**

Při řešení praktických úloh obvykle neznáme rozptyl ZS $\sigma^2$, a proto ho odhadujeme pomocí *výběrového rozptylu* $s_x^2$:

$$s_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}.$$

***Oboustranný IS:***

$$P\left(\bar{x} - u_{1-\frac{\alpha}{2}} \cdot \frac{s_x}{\sqrt{n}} < \mu < \bar{x} + u_{1-\frac{\alpha}{2}} \cdot \frac{s_x}{\sqrt{n}}\right) = 1 - \alpha$$

***Pravostranný IS:***

$$P\left(\mu < \bar{x} + u_{1-\alpha} \cdot \frac{s_x}{\sqrt{n}}\right) = 1 - \alpha$$

***Levostranný IS:***

$$P\left(\bar{x} - u_{1-\alpha} \cdot \frac{s_x}{\sqrt{n}} < \mu\right) = 1 - \alpha$$

- **Malý výběr z normálního rozdělení s neznámým rozptylem $\sigma^2$:**

Kvantily normovaného normálního rozdělení ve vzorcích z předcházející situace nahradíme kvantily Studentova rozdělení s $\nu = \mathrm{n} - 1$ stupni volnosti.

***Oboustranný IS:***

$$P\left[\bar{x} - t_{1-\frac{\alpha}{2}}(n-1) \cdot \frac{s_x}{\sqrt{n}} < \mu < \bar{x} + t_{1-\frac{\alpha}{2}}(n-1) \cdot \frac{s_x}{\sqrt{n}}\right] = 1 - \alpha$$

***Pravostranný IS:***

$$P\left[\mu < \bar{x} + t_{1-\alpha}(n-1) \cdot \frac{s_x}{\sqrt{n}}\right] = 1 - \alpha$$

***Levostranný IS:***

$$P\left[\bar{x} - t_{1-\alpha}(n-1) \cdot \frac{s_x}{\sqrt{n}} < \mu\right] = 1 - \alpha$$

### 9.1.2 Odhad parametru $\sigma^2$ (rozptylu) normálního rozdělení

***1. Bodový odhad***

Bodovým odhadem rozptylu $\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$ je *výběrový rozptyl* $s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$.
Je to nezkreslený a konzistentní odhad $\sigma^2$ .

***2. Intervalový odhad***

Při konstrukci IS pro parametru $\sigma^2$ rozlišujeme teoreticky dva případy: buď druhý parametr rozdělení ($\mu$) známe nebo neznáme. V praxi je daleko častější případ, kdy parametr $\mu$ neznáme, proto se na něj v následujícím textu zaměříme.

***Oboustranný IS:***

$$P\left[\frac{(n-1)s_x^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} < \sigma^2 < \frac{(n-1)s_x^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right] = 1-\alpha$$

***Pravostranný IS:***

$$P\left[\sigma^2 < \frac{(n-1)s_x^2}{\chi^2_{\alpha}(n-1)}\right] = 1-\alpha$$

***Levostranný IS:***

$$P\left[\frac{(n-1)s_x^2}{\chi^2_{1-\alpha}(n-1)} < \sigma^2\right] = 1-\alpha$$

### 9.1.3 Odhad parametru $\pi$ (relativní četnosti) alternativního rozdělení

Je třeba mít k dispozici výběr dostatečně velkého rozsahu; to je zajištěno splněním podmínky $n\pi(1-\pi) > 9$.

***1. Bodový odhad***

Bodovým odhadem relativní četnosti $\pi = \frac{M}{N}$ je *výběrová relativní četnost* (výběrový podíl) $p = \frac{m}{n}$.

$M$ ... počet jednotek se sledovanou vlastností v ZS
$N$ ... rozsah ZS
$m$ ... počet jednotek se sledovanou vlastností ve výběrovém souboru
$n$ ... rozsah výběru.

*2. Intervalový odhad*

***Oboustranný IS:***

$$P\left[p - u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}} < \pi < p + u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}}\right] = 1-\alpha$$

$\Delta = u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}}$ je *přípustná chyba odhadu*.

***Pravostranný IS:***

$$P\left[\pi < p + u_{1-\alpha} \cdot \sqrt{\frac{p(1-p)}{n}}\right] = 1-\alpha$$

***Levostranný IS:***

$$P\left[p - u_{1-\alpha} \cdot \sqrt{\frac{p(1-p)}{n}} < \pi\right] = 1-\alpha$$

**9.2 Stanovení minimálního rozsahu výběru**

Pokud při stanovení minimálního rozsahu výběru, potřebného k dodržení zadaných podmínek, vycházíme ***ze vzorce přípustné chyby odhadu parametru $\mu$***, jeho úpravou dostaneme:

$$n \geq \frac{u^2_{1-\frac{\alpha}{2}} \cdot \sigma^2}{\Delta^2}.$$

Pokud neznáme rozptyl základního souboru $\sigma^2$, použijeme jeho bodový odhad $s_x^2$.

Budeme-li vycházet ***ze vzorce přípustné chyby odhadu parametru $\pi$***, jeho úpravou dostaneme:

$$n \geq \frac{u^2_{1-\frac{\alpha}{2}} \cdot \pi(1-\pi)}{\Delta^2}.$$

Pokud neznáme relativní četnost základního souboru $\pi$, použijeme její bodový odhad $p$.

# 10 MATEMATICKÁ STATISTIKA – 3. ČÁST

## 10.1 Testování statistických hypotéz

- *hypotéza* je určitý předpoklad (tvrzení) o základním souboru
- jestliže se hypotéza týká neznámého parametru ZS, jde o tzv. parametrické testy
- jestliže se hypotéza týká vlastností ZS, jde o tzv. neparametrické testy
- testování hypotéz je postup, sloužící k ověření předpokladů o ZS (hypotéz) na základě výběrových dat
- testovací postup umožňuje rozhodnout, zda určitou hypotézu zamítneme či ne, a to s malým, předem zvoleným *rizikem* $\alpha$.

***Symbolika a základní pojmy:***

$H_0$ ... nulová (testovaná) hypotéza
$H_1$ ... alternativní hypotéza

***Testové kritérium (t):*** je to vhodná charakteristika, která má při platnosti $H_0$ známé pravděpodobnostní rozdělení (dále jen TK).

***Kritický obor (W):*** je tvořen hodnotami TK, které jsou při platnosti $H_0$ tak extrémní, že pravděpodobnost jejich výskytu je velmi malá.

***Obor přijetí (V):*** je tvořen všemi hodnotami TK mimo kritický obor.

***Hladina významnosti ($\alpha$) = pravděpodobnost chyby 1. druhu:*** je to pravděpodobnost, že zamítneme $H_0$, ačkoli platí.

***Pravděpodobnost chyby 2. druhu ($\beta$):*** je to pravděpodobnost, že nezamítneme $H_0$, ačkoli neplatí.

***Síla testu ($1-\beta$):*** je to pravděpodobnost správného zamítnutí $H_0$, tj. schopnost testu zamítnout neplatnou $H_0$.

**Standardní testovací postup:**

- jedná se o obecný postup
- je nezávislý na konkrétním typu testu
- provádí se v několika krocích.

1. ***Formulace hypotéz $H_0$ a $H_1$.***

2. ***Volba testového kritéria:*** zvolíme vhodnou charakteristiku, jejíž rozdělení při platnosti $H_0$ známe.

3. ***Vymezení kritického oboru:*** je omezen kvantily rozdělení TK při platnosti $H_0$ (tzv. kritické hodnoty).

4. ***Výpočet hodnoty TK z výběrových dat.***

5. ***Formulace závěru o výsledku testu:*** existují *pouze dvě možnosti*, jedna z nich tedy musí nastat.

- *TK leží v kritickém oboru (TK ∈ W):* zamítáme $H_0$, tzn. prokázali jsme $H_1$.
- *TK neleží v kritickém oboru (TK ∉ W):* nezamítáme $H_0$, tzn. neprokázali jsme $H_1$.

### 10.1.1 Některé parametrické testy

❖ ***Test parametru μ (střední hodnoty) normálního rozdělení***

**1. Formulace hypotéz**

$H_0 : \mu = \mu_0$

*a)* $H_1 : \mu \neq \mu_0$ *oboustranná alternativní hypotéza*

*b)* $H_1 : \mu > \mu_0$ *pravostranná alternativní hypotéza*

*c)* $H_1 : \mu < \mu_0$ *levostranná alternativní hypotéza*

**2. Volba testového kritéria**

*Rozlišujeme tři případy:*

*a)* známe rozptyl ZS $\sigma^2$

$U = \dfrac{\bar{x} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$, při platnosti $H_0$ má rozdělení $N(0;1)$

b) neznáme rozptyl ZS $\sigma^2$ a výběr má malý rozsah

$t = \dfrac{\bar{x} - \mu_0}{\dfrac{s}{\sqrt{n}}}$, při platnosti $H_0$ má rozdělení $t(n-1)$

c) neznáme rozptyl ZS $\sigma^2$ a výběr má velký rozsah

$U = \dfrac{\bar{x} - \mu_0}{\dfrac{s}{\sqrt{n}}}$, při platnosti $H_0$ má rozdělení $N(0;1)$

**3. Stanovení kritického oboru**

*Pro případy a) a c) a různé typy alternativních hypotéz:*

*a)* $W \equiv \left\{ u; u \leq u_{\frac{\alpha}{2}} \ a \ u \geq u_{1-\frac{\alpha}{2}} \right\}$

*b)* $W \equiv \left\{ u; u \geq u_{1-\alpha} \right\}$

*c)* $W \equiv \left\{ u; u \leq u_{\alpha} \right\}$.

*Pro případ b) a různé typy alternativních hypotéz:*

$a)\ W \equiv \left\{ t; t \le t_{\frac{\alpha}{2}}(n-1)\ a\ t \ge t_{1-\frac{\alpha}{2}}(n-1) \right\}$

$b)\ W \equiv \left\{ t; t \ge t_{1-\alpha}(n-1) \right\}$

$c)\ W \equiv \left\{ t; t \le t_{\alpha}(n-1) \right\}.$

### ❖ *Test parametru π (relativní četnosti) alternativního rozdělení*

Je třeba mít k dispozici výběr dostatečně velkého rozsahu; to je zajištěno splněním podmínky $n\pi(1-\pi) > 9$.

**1.** $H_0 : \pi = \pi_0$

$a)\ H_1 : \pi \ne \pi_0$

$b)\ H_1 : \pi > \pi_0$

$c)\ H_1 : \pi < \pi_0$

**2.** $U = \dfrac{p - \pi_0}{\sqrt{\dfrac{\pi_0(1-\pi_0)}{n}}}$, při platnosti $H_0$ má rozdělení $N(0;1)$

**3.** $a)\ W \equiv \left\{ u; u \le u_{\frac{\alpha}{2}}\ a\ u \ge u_{1-\frac{\alpha}{2}} \right\}$

$b)\ W \equiv \left\{ u; u \ge u_{1-\alpha} \right\}$

$c)\ W \equiv \left\{ u; u \le u_{\alpha} \right\}$

### ❖ *Test parametru σ² (rozptylu) normálního rozdělení*

**1.** $H_0 : \sigma^2 = \sigma_0^2$

$a)\ H_1 : \sigma^2 \ne \sigma_0^2$

$b)\ H_1 : \sigma^2 > \sigma_0^2$

$c)\ H_1 : \sigma^2 < \sigma_0^2$

**2.** Rozlišujeme dva případy – buď parametr $\mu$ ZS známe nebo ne.
V praxi je daleko častější případ, kdy parametr $\mu$ neznáme, proto se na něj omezíme.

$\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$, při platnosti $H_0$ má rozdělení $\chi^2(n-1)$

**3.** $a)\ W \equiv \left\{ \chi^2; \chi^2 \le \chi^2_{\frac{\alpha}{2}}(n-1)\ a\ \chi^2 \ge \chi^2_{1-\frac{\alpha}{2}}(n-1) \right\}$

$b)\ W \equiv \left\{ \chi^2;\ \chi^2 \ge \chi^2_{1-\alpha}(n-1) \right\}$

$c)\ W \equiv \left\{ \chi^2; \chi^2 \le \chi^2_{\alpha}(n-1) \right\}$

**Test hypotézy o shodě dvou průměrů**

| Východisko | Testové kritérium | Rozdělení TK při platnosti $H_0$ | Parametry rozdělení | Alternativní hypotéza | Kritický obor |
|---|---|---|---|---|---|
| Známe $\sigma_1^2$ a $\sigma_2^2$ | $U = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ | N | $\mu = 0$<br>$\sigma^2 = 1$ | $H_1 : \mu_1 \neq \mu_2$ | $U \le u_{\frac{\alpha}{2}}$ a $U \ge u_{1-\frac{\alpha}{2}}$ |
| | | | | $H_1 : \mu_1 > \mu_2$ | $U \ge u_{1-\alpha}$ |
| | | | | $H_1 : \mu_1 < \mu_2$ | $U \le u_{\alpha}$ |
| Neznáme $\sigma_1^2$ a $\sigma_2^2$<br>Předpokládáme:<br>$\sigma_1^2 = \sigma_2^2$ | $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{(n_1-1)s_1'^2 + (n_2-1)s_2'^2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ | t | $\nu = n_1 + n_2 - 2$ | $H_1 : \mu_1 \neq \mu_2$ | $t \le t_{\frac{\alpha}{2}}(n_1+n_2-2)$<br>a $t \ge t_{1-\frac{\alpha}{2}}(n_1+n_2-2)$ |
| | | | | $H_1 : \mu_1 > \mu_2$ | $t \ge t_{1-\alpha}(n_1+n_2-2)$ |
| | | | | $H_1 : \mu_1 < \mu_2$ | $t \le t_{\alpha}(n_1+n_2-2)$ |
| Neznáme $\sigma_1^2$ a $\sigma_2^2$<br>Předpokládáme:<br>$\sigma_1^2 \neq \sigma_2^2$ | $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1'^2}{n_1} + \dfrac{s_2'^2}{n_2}}}$ | t | $\nu = \dfrac{\left(\dfrac{s_1'^2}{n_1} + \dfrac{s_2'^2}{n_2}\right)^2}{\dfrac{1}{n_1-1}\left(\dfrac{s_1'^2}{n_1}\right)^2 + \dfrac{1}{n_2-1}\left(\dfrac{s_2'^2}{n_2}\right)^2}$ | $H_1 : \mu_1 \neq \mu_2$ | $t \le t_{\frac{\alpha}{2}}(\nu)$ a $t \ge t_{1-\frac{\alpha}{2}}(\nu)$ |
| | | | | $H_1 : \mu_1 > \mu_2$ | $t \ge t_{1-\alpha}(\nu)$ |
| | | | | $H_1 : \mu_1 < \mu_2$ | $t \le t_{\alpha}(\nu)$ |

# 11 MATEMATICKÁ STATISTIKA – 4. ČÁST

## 11.1 Některé neparametrické testy

### *$\chi^2$- test dobré shody*

- slouží k ověření shody mezi teoretickým a empirickým rozdělením
- je použitelný pouze v případě velkých výběrů
- předpokladem testu je možnost roztřídit výsledky náhodného výběru do určitého počtu ($k$) disjunktních tříd podle zvoleného znaku
- v případě, že nemáme k dispozici výběr dostatečného rozsahu, použijeme místo tohoto testu ***Kolmogorovův-Smirnovův test***.

**Požadavky na rozsah výběru:**

Je nutné, aby rozsah výběru zajistil, že bude dostatečné obsazení ve všech skupinách, do nichž je soubor roztříděn, tj. $n\pi_{0,i} > 5$ pro $i = 1, 2, \dots, k$.
Někdy bývá tato podmínka formulována mírněji: ve všech třídách musí platit $n\pi_{0,i} > 1$ a alespoň v 80 % tříd musí platit $n\pi_{0,i} > 5$.

*Nejsou-li výše uvedené podmínky splněny, je možno sloučit některé třídy (např. sousední či věcně příbuzné).*

### *$\chi^2$- test dobré shody se používá ve 2 situacích:*

1. $H_0$ udává proporce četností v jednotlivých skupinách (což může být formulováno např. intuitivně).
2. $H_0$ přepokládá, že ZS má rozdělení určitého typu:

   - Pokud $H_0$ udává typ rozdělení i jeho parametry, jedná se o *úplně specifikovaný model.*
   - Pokud $H_0$ udává pouze typ rozdělení bez specifikace parametrů, jde o *neúplně specifikovaný model.*

**Ad 1.**

**1.** $H_0 : \pi_i = \pi_{0,i}$, pro $i = 1, 2, \dots, k$
$H_1 : non\ H_0$

**2.** $G = \sum_{i=1}^{k} \frac{\left(n_i - n\pi_{0,i}\right)^2}{n\pi_{0,i}}$, při platnosti $H_0$ má rozdělení $\chi^2(k-1)$

$n_i$ je empirická (pozorovaná, výběrová) četnost
$n\pi_{0,i}$ je teoretické (hypotetická) četnost, tj. teoretické obsazení i-té třídy

**3.** $W \equiv \left\{G; G \geq \chi^2_{1-\alpha}(k-1)\right\}$

**Ad 2.**

- *Úplně specifikovaný model (příklad)*

**1.** $H_0 : Po(2)$
$H_1 : non\ H_0$

Další postup viz případ 1.

- *Neúplně specifikovaný model (příklad)*

**1.** $H_0 : Po(\lambda)$
$H_1 : non\ H_0$

**2.** $G = \sum_{i=1}^{k} \frac{(n_i - n\pi_{0,i})^2}{n\pi_{0,i}}$, při platnosti $H_0$ má rozdělení $\chi^2(k-p-1)$
$p$ je počet parametrů rozdělení, které odhadujeme

**3.** $W \equiv \{G; G \geq \chi^2_{1-\alpha}(k-p-1)\}$

**Závěr testu:** Pokud $TK \in W$, zamítáme $H_0$ (tzn., že přijímáme $H_1$). V tom případě není rozdělení, specifikované nulovou hypotézou, vhodným modelem pro empirická data. Shoda teoretického a empirického rozdělení se na hladině významnosti $\alpha$ nepotvrdila.

# Statistické charakteristiky (míry)

- shrnují informaci, obsaženou v datech (vyjadřují ji v koncentrované formě);
- charakterizují základní rysy zkoumaného souboru dat;
- umožňují porovnávání více souborů.

***4 skupiny statistických charakteristik :***

1. charakteristiky polohy (úrovně),
2. charakteristiky variability,
3. charakteristiky šikmosti,
4. charakteristiky špičatosti.

***Dva způsoby konstrukce statistických charakteristik:***

a) Charakteristiky, které jsou funkcí všech hodnot dané proměnné:
   - jsou ovlivněny případnými extrémy;
   - výpočet podle určitého funkčního předpisu.

b) Charakteristiky, které nejsou funkcí všech hodnot dané proměnné:
   - nejsou ovlivněny extrémy;
   - jsou to konkrétní hodnoty proměnné, vybrané podle určitého kritéria.

## *1. Charakteristiky polohy*

- charakterizují střed, kolem něhož hodnoty kolísají;
- charakterizují úroveň (velikost, hladinu) proměnné;
- používá se pro ně rovněž pojem ***střední hodnoty***.

### a) Charakteristiky, které jsou funkcí všech hodnot - průměry

***Aritmetický průměr***

- prostý: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

- vážený: $\bar{x} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i}$

! Používá se tam, kde má informační smysl součet hodnot proměnné.

***Harmonický průměr***

- prostý: $\bar{x}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$

- vážený: $\bar{x}_H = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$

! Používá se tam, kde má smysl součet převrácených hodnot proměnné. Např. k výpočtu průměrné doby potřebné ke splnění úkolu, kdy jednotky plní úkoly současně.

***Geometrický průměr***

- prostý: $\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$

- vážený: $\bar{x}_G = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \ldots \cdot x_k^{n_k}} = \sqrt[n]{\prod_{i=1}^{k} x_i^{n_i}}$

! Používá se tam, kde má smysl součin hodnot proměnné. Např. k výpočtu průměrného koeficientu růstu v časových řadách.

***Kvadratický průměr***

- prostý: $\bar{x}_K = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$

- vážený: $\bar{x}_K = \sqrt{\frac{\sum_{i=1}^{k} x_i^2 n_i}{\sum_{i=1}^{k} n_i}}$

! Používá se tam, kde má smysl součet čtverců hodnot proměnné. Např. tehdy, jestliže jednotlivé hodnoty jsou již samy odchylkami původních hodnot od aritmetického průměru, odchylkami od normy apod.

**Vztah mezi průměry**

Jsou -li výše uvedené 4 typy průměrů vypočítány z týchž kladných hodnot proměnné, platí pro ně vztah:
$\bar{x}_H \leq \bar{x}_G \leq \bar{x} \leq \bar{x}_K$ .

**b) Charakteristiky, které nejsou funkcí všech hodnot**

- patří sem především modus a kvantily;
- jejich výhodou je, že nejsou ovlivněny odlehlými pozorováními.

***Modus***

- varianta s největší četností (typická hodnota);
- vrchol rozdělení četností;
- označení symbolem $\hat{x}$ .

***Kvantily***

- hodnoty, které rozdělují uspořádaný statistický soubor na určitý počet stejně obsazených částí;
- hodnoty menší resp. stejné tvoří určitou stanovenou část rozsahu souboru (určitý podíl, určité procento).

*Uspořádaný statistický soubor:* hodnoty proměnné jsou seřazeny do neklesající řady.

*Obecné označení kvantilů:*

$x_p$ , kde $p$ je relativní četnost

$\tilde{x}_{100p}$ , kde $100 \cdot p$ je relativní četnost vyjádřená v %

**Druhy kvantilů :**

- **medián** - $\tilde{x}, \tilde{x}_{50}, x_{0,5}$ - prostřední hodnota uspořádaného statistického souboru.

  Člení statistický soubor na dvě stejně četné části, existuje tedy 50 % hodnot menších (nebo stejných) a 50 % hodnot větších (nebo stejných).

  ***Výpočet*** :

  a) rozsah souboru *n* je liché číslo

  $\tilde{x} = x_{\left(\frac{n+1}{2}\right)}$ , kde výraz $\frac{n+1}{2}$ udává pořadí mediánu v dané neklesající řadě hodnot.

  Při lichém rozsahu souboru je mediánem konkrétní prvek.

  b) rozsah souboru *n* je sudé číslo

  $$\tilde{x} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+2}{2}\right)}}{2}$$

  Při sudém rozsahu souboru existují 2 prostřední hodnoty a medián je jejich aritmetickým průměrem.

*Pozn.:* Kvantily < $\tilde{x}$ se nazývají *dolní kvantily*, kvantily > $\tilde{x}$ se nazývají *horní kvantily*.

- **tercily** - $\tilde{x}_{33,\bar{3}}\ (x_{0,\bar{3}})$ , $\tilde{x}_{66,\bar{6}}\ (x_{0,\bar{6}})$

  - jsou to 2 kvantily, které rozdělují uspořádaný statistický soubor na 3 stejně četné části.

- **kvartily** - $\tilde{x}_{25}\ (x_{0,25})$, $\tilde{x}$ , $\tilde{x}_{75}\ (x_{0,75})$

  - jsou to 3 kvantily, které rozdělují uspořádaný statistický soubor na 4 stejně četné části.

- **kvintily** - $\tilde{x}_{20}\ (x_{0,2}), \tilde{x}_{40}\ (x_{0,4})$, $\tilde{x}_{60}\ (x_{0,6}), \tilde{x}_{80}\ (x_{0,8})$

  - jsou to 4 kvantily, které rozdělují uspořádaný statistický soubor na 5 stejně četných částí.

- **sextily** – 5 kvantilů, 6 částí
- **septily** – 6 kvantilů, 7 částí
- **oktávily** – 7 kvantilů, 8 částí
- **nonily** – 8 kvantilů, 9 částí
- **decily** – 9 kvantilů, 10 částí
- **percentily** – 99 kvantilů, 100 částí atd.

***Výpočet pořadového čísla kvantilu :***

$$n \cdot p < m_p < n \cdot p + 1$$

n ..... rozsah statistického souboru
p ..... relativní četnost
$m_p$ ..... pořadové číslo příslušného kvantilu

## 2. Charakteristiky variability

- udávají rozptýlení (kolísání) hodnot kolem zvoleného středu (např. kolem nějaké střední hodnoty);
- variabilita = měnlivost = kolísavost = odlišnost.

### a) Míry absolutní variability

***Variační rozpětí***

$$R = x_{\max} - x_{\min}$$

***Kvantilová rozpětí***

- kvartilové rozpětí: $R_q = \tilde{x}_{75} - \tilde{x}_{25}$
- decilové rozpětí: $R_d = \tilde{x}_{90} - \tilde{x}_{10}$ atd.

***Kvantilové odchylky***

- kvartilová odchylka: $Q = \frac{\tilde{x}_{75} - \tilde{x}_{25}}{2}$
- decilová odchylka: $D = \frac{\tilde{x}_{90} - \tilde{x}_{10}}{8}$ atd.

***Průměrná absolutní odchylka***

- prostá: $\bar{d} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$
- vážená: $\bar{d} = \frac{\sum_{i=1}^{k} |x_i - \bar{x}| \, n_i}{\sum_{i=1}^{k} n_i}$

***Rozptyl***

- prostý (klasický): $s_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$
- vážený (klasický): $s_x^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 n_i}{\sum_{i=1}^{k} n_i}$

**Výpočtový tvar rozptylu**

- prostý: $s_x^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left( \frac{\sum_{i=1}^{n} x_i}{n} \right)^2 = \overline{x^2} - \bar{x}^2$

- vážený: $s_x^2 = \frac{\sum_{i=1}^{k} x_i^2 n_i}{\sum_{i=1}^{k} n_i} - \left( \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i} \right)^2 = \overline{x^2} - \bar{x}^2$

***Směrodatná odchylka***

- kladná odmocnina z rozptylu, tj. $s_x = \sqrt{s_x^2}$ ;
- udává, jak se v průměru liší jednotlivé hodnoty znaku od aritmetického průměru v obou směrech (±);
- vhodná pro interpretaci, je udána v daných měrných jednotkách.

Pokud pracujeme s výběrovým souborem, počítáme ***výběrový rozptyl*** a ***výběrovou směrodatnou odchylku***:

- prostý: $s_x'^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

- vážený: $s_x'^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{n-1}$

**Rozklad rozptylu**

Skládá-li se statistický soubor z $k$ dílčích podsouborů, v nichž známe jednotlivé dílčí rozptyly $s_i^2$, dílčí průměry $\bar{x}_i$ a četnosti $n_i$, pak rozptyl celého souboru $s_x^2$ můžeme rozložit na součet 2 rozptylů, z nichž jeden charakterizuje variabilitu mezi skupinami a druhý variabilitu uvnitř skupin: $s_x^2 = s_{\bar{x}_i}^2 + \overline{s_i^2}$.

***Rozptyl skupinových průměrů***: $s_{\bar{x}_i}^2 = \frac{\sum_{i=1}^{k}(\bar{x}_i - \bar{x})^2 n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k}\bar{x}_i^2 n_i}{\sum_{i=1}^{k} n_i} - \left(\frac{\sum_{i=1}^{k}\bar{x}_i n_i}{\sum_{i=1}^{k} n_i}\right)^2$

***Průměr skupinových rozptylů***: $\overline{s_i^2} = \frac{\sum_{i=1}^{k} s_i^2 n_i}{\sum_{i=1}^{k} n_i}$

**b) Míry relativní variability**

***Variační koeficient***

- je bezrozměrné číslo;
- umožňuje porovnávat variabilitu souborů s různou úrovní či různými měrnými jednotkami;
- obecně může nabývat hodnot z intervalu $(-\infty, \infty)$, pro kardinální proměnnou z intervalu $\langle 0, \infty)$.

$$V_x = \frac{s_x}{\bar{x}}$$

## 3. Charakteristiky šikmosti

- šikmost = asymetrie;

***Míry šikmosti***

- prostá: $\alpha = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n s_x^3}$

- vážená: $\alpha = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^3 n_i}{n s_x^3}$

- jednoduchá charakteristika (Cyhelského míra šikmosti): $\alpha' = \frac{n' - n''}{n}$

kde $n'$ je počet podprůměrných hodnot
$n''$ je počet nadprůměrných hodnot.

***Interpretace:***
- v **symetrickém** rozdělení α = 0; obvykle platí vztah: $\bar{x} = \hat{x} = \tilde{x}$; počet podprůměrných hodnot je stejný jako počet hodnot nadprůměrných;
- v **kladně sešikmeném** rozdělení α > 0; obvykle platí vztah: $\hat{x} < \tilde{x} < \bar{x}$; více hodnot podprůměrných než nadprůměrných;
- v **záporně sešikmeném** rozdělení α < 0; obvykle platí vztah: $\bar{x} < \tilde{x} < \hat{x}$; více hodnot nadprůměrných než podprůměrných.

## 4. Charakteristiky špičatosti

- špičatost = exces;
- špičatost spočívá ve větší nahuštěnosti hodnot prostřední velikosti ve srovnání se stupněm nahuštěnosti ostatních hodnot resp. všech hodnot proměnné;
- špičatější rozdělení má výraznější vrchol (tzn. že vrchol více vystupuje).

***Míra špičatosti***

- prostá: $\beta = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n s_x^4} - 3$

- vážená: $\beta = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^4 n_i}{n s_x^4} - 3$

***Interpretace:*** - vyšší hodnota míry znamená větší špičatost, tzn. špičatější je to rozdělení, které má $\beta$ vyšší;
- základem pro srovnání je normované normální rozdělení (viz další výklad).

## *Charakteristiky nominální proměnné*

**1. polohy**:

***Modus*** $\hat{x}$

**2. variability**:

***Míra mutability M***
- Mutabilita = variabilita hodnot nominální proměnné;
- udává podíl dvojic jednotek s různou obměnou z celkového počtu všech možných dvojic jednotek;
- lze ji vyjádřit v %.

$$M = \frac{n^2 - \sum_{i=1}^{k} n_i^2}{n(n-1)} \qquad M \in \langle 0,1 \rangle$$

***Nominální variance***

- používá se v případě, že jsou známy pouze relativní četnosti a není znám rozsah souboru;
- skutečný stupeň variability podhodnocuje.

$$NOMVAR = 1 - \sum_{i=1}^{k} p_i^2 \quad ; \quad NOMVAR \in \langle 0,1 \rangle$$

# 4 PRAVDĚPODOBNOST – 1. ČÁST

***Pravděpodobnost:***
- matematická disciplína, která se zabývá studiem zákonitostí, jimiž se řídí hromadné náhodné jevy
- vytváří pravděpodobnostní modely, pomocí nichž se snaží postihnout procesy, ovlivněné náhodou.

***Náhodné pokusy:*** procesy, jejichž výsledek nelze předem jednoznačně určit (je nejistý); závisí jednak na daných podmínkách, při kterých je prováděn, jednak na náhodě. Teorie pravděpodobnosti se zabývá pouze náhodnými pokusy, které jsou za stejných podmínek opakovatelné a u nichž je variabilita výsledků podstatná a vykazuje určitou zákonitost.

***Hromadné náhodné jevy:*** výsledky opakovatelných náhodných pokusů (symbolické značení – *A, B, C, …*).

***Pravděpodobnost náhodného jevu:*** pravděpodobnost náhodného jevu *A* je číslo *P(A)*, které lze interpretovat jako míru možnosti nastoupení náhodného jevu.

**Axiomatická teorie pravděpodobnosti:** pravděpodobnost je funkce, která každému náhodnému jevu přiřazuje reálné číslo, přičemž musí být splněny základní axiomy.

1. $P(A) \geq 0$
2. $P(A_1 \cup A_2 \cup ...) = P(A_1) + P(A_2) + ...$ (pro neslučitelné jevy)
3. $P(E) = 1$.

**Klasická definice pravděpodobnosti:** pravděpodobnost jevu *A* se rovná podílu případů příznivých nastoupení jevu *A* a počtu všech případů možných, jsou-li všechny stejně pravděpodobné.

$$P(A) = \frac{m}{n}$$

kde $m$ je počet případů příznivých
$n$ je počet případů možných.

**Statistická definice pravděpodobnosti:** pokud platí, že při rostoucím počtu opakování náhodného pokusu ($n$) kolísá relativní četnost $\frac{m}{n}$ ve stále užších mezích kolem určitého čísla, můžeme toto číslo považovat za pravděpodobnost jevu *A*.

relativní četnost jevu $A = \frac{m}{n}$

kde $m$ je počet nastoupení jevu *A*
$n$ je počet opakování pokusu.

- pravděpodobnosti náhodného jevu odhadujeme na základě výsledků, získaných při mnohonásobném opakování náhodného pokusu (aposteriorní charakter definice).

**Pravidla pro počítání s pravděpodobnostmi**

***Podmíněná pravděpodobnost:*** $P(A/B)$ je podmíněná pravděpodobnost jevu $A$ vzhledem k jevu $B$, tj. pravděpodobnost nastoupení jevu $A$ za předpokladu, že nastal jev $B$.

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \text{ pro } P(B) > 0$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)}, \text{ pro } P(A) > 0$$

***Pravidlo o násobení pravděpodobností:***
Pravděpodobnost současného nastoupení jevů $A$ a $B$ (tzn. jejich průniku) je rovna součinu nepodmíněné pravděpodobnosti jednoho jevu a podmíněné pravděpodobnosti druhého jevu vzhledem k prvnímu jevu.

$$P(A \cap B) = P(B) \cdot P(A/B)$$
$$P(A \cap B) = P(A) \cdot P(B/A)$$

*Zobecnění pravidla o násobení pravděpodobností pro dva a více jevů*:

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) \cdot \ldots \cdot P(A_n/A_1 \cap \ldots \cap A_{n-1}).$$

***Nezávislost jevů:***
Jestliže $P(A/B) = P(A)$, pak jev $A$ nezávisí na jevu $B$.
Jestliže $P(B/A) = P(B)$, pak jev $B$ nezávisí na jevu $A$.

*Nutná a postačující podmínka nezávislosti dvou jevů:* $P(A \cap B) = P(A) \cdot P(B)$.

*Zjednodušení pravidla o násobení pravděpodobností pro nezávislé jevy:*

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot \ldots \cdot P(A_n).$$

***Pravidlo pro sčítání pravděpodobností:***
Pravděpodobnost sjednocení jevů $A$ a $B$ je rovna součtu pravděpodobností těchto jevů, zmenšené o pravděpodobnost jejich průniku.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

***Disjunktní jevy:*** Jestliže $P(A \cap B) = 0$, pak jevy $A$ a $B$ jsou disjunktní.

*Zjednodušení pravidla pro sčítání pravděpodobností pro disjunktní jevy:*

$$P(A \cup B) = P(A) + P(B).$$

**Operace s náhodnými jevy**

- vztahy mezi náhodnými jevy graficky znázorňují tzv. ***Vennovy diagramy***.

**1.** $\boldsymbol{A \subset B}$
Jev *A* je částí jevu *B*; z jevu *A* plyne jev *B* (implikace); nastoupení jevu *A* má vždy za následek nastoupení jevu *B*.

**2.** $\boldsymbol{A = B}$
Jevy *A* a *B* jsou si rovny; $A \subset B$ a současně $B \subset A$.

**3.** $\boldsymbol{C = A \cap B}$
Jev *C* je průnik jevů *A* a *B* (logický součin); jev *C* nastane právě tehdy, nastane-li současně jev *A* i jev *B*.

**4.** $\boldsymbol{C = A \cup B}$
Jev *C* je sjednocení jevů *A* a *B* (logický součet); jev *C* nastane právě tehdy, nastane-li alespoň jeden z jevů *A* a *B*.

**5.** $\boldsymbol{C = A - B}$
Jev *C* je rozdíl jevů *A* a *B*; jev *C* nastane právě tehdy, když jev *A* nastane a současně jev *B* nenastane.

**6.** ***E*** je jev jistý, tj. jev, který musí nastat vždy.
∅ je jev nemožný, tj. jev, který nastat nemůže.

**Kombinatorika**

***Permutace***

$$P(n) = n!$$

***Variace bez opakování***

$$V_k(n) = \frac{n!}{(n-k)!}$$

***Variace s opakováním***

$$V_k'(n) = n^k$$

***Kombinace***

$$C_k(n) = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

# 5 PRAVDĚPODOBNOST – 2. ČÁST

***Náhodná veličina:***
- veličina, jejíž hodnota je jednoznačně určena výsledkem náhodného pokusu
- vlivem náhodných činitelů může nabýt různých hodnot, proto nelze její konkrétní hodnotu před provedením náhodného pokusu jednoznačně určit
- náhodnou veličinu pokládáme za danou, pokud známe všechny její možné hodnoty a pravděpodobnosti výskytu každé z nich
- symbolické značení – *X, Y, Z*, …
- příklady náhodných veličin: počet bodů, které padnou na hrací kostce, počet poruch určitého zařízení, doba čekání na obsluhu v určité prodejně, atd.

***Zákon rozdělení náhodné veličiny:*** pravidlo, které každé hodnotě nebo množině hodnot z každého intervalu přiřazuje pravděpodobnost, že náhodná veličina nabude této hodnoty nebo hodnoty z určitého intervalu. Je to pravděpodobnostní model empirické náhodné veličiny.

## 5.1 Popis rozdělení náhodné veličiny

### 5.1.1 Diskrétní náhodná veličina

***Distribuční funkce:*** pravděpodobnost, že NV nabude hodnoty menší nebo rovné *x*.

$$F(x) = P(X \le x) = \sum_{t \le x} P(t)$$

***Pravděpodobnostní funkce:*** pravděpodobnost, že NV nabude hodnoty rovné *x*.

$$P(x) = P(X = x)$$

$$\sum_{M} P(x) = 1,$$

kde $M$ je prostor hodnot NV X, tj. množina možných hodnot NV X.

### 5.1.2 Spojitá náhodná veličina

***Distribuční funkce***

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$$

***Hustota pravděpodobnosti***

$$f(x) = F'(x) = \frac{dF(x)}{dx}$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

## 5.2 Charakteristiky náhodných veličin

- číselné hodnoty, jejichž cílem je koncentrovat popis NV
- poskytují výstižný popis základních vlastností rozdělení NV.

***Podle vlastnosti rozdělení, kterou popisují, rozeznáváme:***

- charakteristiky polohy
- charakteristiky variability
- charakteristiky šikmosti
- charakteristiky špičatosti.

### 5.2.1 Charakteristiky polohy

**Střední hodnota** $\boldsymbol{E(X)}$

- tzv. očekávaná hodnota (z latinského expectatis)

***a) Diskrétní NV***

$$E(X) = \sum_{M} x \cdot P(x)$$

***b) Spojitá NV***

$$E(X) = \int_{M} x \cdot f(x) dx$$

**Modus** $\boldsymbol{\hat{x}}$

***a) Diskrétní NV***

- $\hat{x}$ je hodnota NV, která má největší pravděpodobnost výskytu
  $\hat{x}$ … max $P(x)$.

***b) Spojitá NV***

- $\hat{x}$ je bod, v němž je hustota pravděpodobnosti maximální, tj. lokální maximum hustoty pravděpodobnosti *f(x)*
  $\hat{x}$ … $f'(x) = 0$.

**Kvantily**

- používají se především kvantily *spojité* náhodné veličiny
- $x_p$ je hodnota, kterou hodnoty NV nepřekročí s pravděpodobností 100p %.

*Hodnota $x_p$ je 100p% kvantilem NV X, jestliže pro ni platí:*

$$F(x_p) = P(X \le x_p) = p\,.$$

### 5.2.2 Charakteristiky variability

**Rozptyl** $\boldsymbol{D(X)}$

***a) Diskrétní NV***

$$D(X) = \sum_{M} [x - E(X)]^2 \cdot P(x) = \sum_{M} x^2 \cdot P(x) - \left[ \sum_{M} x \cdot P(x) \right]^2$$

***b) Spojitá NV***

$$D(X) = \int_{M} [x - E(X)]^2 f(x) dx = \int_{M} x^2 \cdot f(x) dx - \left[ \int_{M} x \cdot f(x) dx \right]^2$$

**Směrodatná odchylka** $\boldsymbol{\sigma(X)}$

$$\boldsymbol{\sigma(X) = \sqrt{D(X)}}$$

# 6 PRAVDĚPODOBNOST – 3. ČÁST

***Rozdělení náhodné veličiny:***
- pravděpodobnostní model chování náhodné veličiny
- existuje celá řada rozdělení pro diskrétní i spojité náhodné veličiny.

## 6.1 Některá rozdělení diskrétních náhodných veličin

### 6.1.1 Alternativní rozdělení $A(\pi)$

- NV $X$ je počet nastoupení jevu $A$ při realizaci náhodného pokusu
- rozdělení nula-jedničkové náhodné veličiny
- *rozdělení má jeden parametr:*
  $\pi$ … pravděpodobnost nastoupení sledovaného jevu při náhodném pokusu.

***Pravděpodobnostní funkce:***

$$P(x) = \pi^x (1-\pi)^{1-x},\ x = 0, 1,\ 0 < \pi < 1,$$
$$= 0 \qquad \text{jinak.}$$

$$E(X) = \pi$$
$$D(X) = \pi(1-\pi)$$

### 6.1.2 Binomické rozdělení $Bi(n;\ \pi)$

- NV $X$ je počet výskytů náhodného jevu $A$ v $n$ nezávislých náhodných pokusech, je-li pravděpodobnost nastoupení jevu $A$ ve všech pokusech stejná ($\pi$)
- *rozdělení má dva parametry:*
  $n$ … počet nezávislých pokusů
  $\pi$ … pravděpodobnost nastoupení sledovaného jevu v jednom pokusu.

***Pravděpodobnostní funkce:***

$$P(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x},\ x = 0, 1, \ldots, n,\ 0 < \pi < 1,$$
$$= 0 \qquad \text{jinak.}$$

$$E(X) = n\pi$$
$$D(X) = n\pi(1-\pi)$$

*Např.:* NV $X$ je počet „šestek“, které padnou při deseti hodech kostkou.

### 6.1.3 Poissonovo rozdělení $Po(\lambda)$

- NV X je počet výskytů náhodného jevu $A$ v určitém časovém intervalu délky $t$ (tzn. za jednotku času), v jednotce plochy nebo objemu (v prostorové jednotce)
- *rozdělení má jeden parametr:*
  $\lambda$ … střední hodnota rozdělení.

***Pravděpodobnostní funkce:***

$$P(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}, \; x = 0, 1, 2, \ldots, \lambda > 0,$$
$$= 0 \qquad \text{jinak.}$$

$$E(X) = \lambda$$
$$D(X) = \lambda$$

*Např.:* NV X je počet poruch stroje za směnu, počet telefonních hovorů za hodinu, počet vad na 1 $m^2$ koberce.

***Aproximace Binomického rozdělení rozdělením Poissonovým:***

- podmínky pro aproximaci: počet pokusů musí být dostatečně velký (alespoň $n > 30$) a pravděpodobnost $\pi$ velmi malá (alespoň $\pi \leq 0{,}1$)
- při aproximaci udává *P(x)* přibližnou pravděpodobnost, že ve velkém počtu *n* nezávislých náhodných pokusů se sledovaný jev *A* vyskytne *x*-krát, jestliže pravděpodobnost výskytu jevu v jednom pokusu je velni malá.

*Např.:* NV *X* je počet vadných výrobků ve velké sérii, je-li pravděpodobnost výroby zmetku velmi malá.

### 6.1.4 Hypergeometrické rozdělení $\boldsymbol{Hyp(N; M; n)}$

- používá se v případě závislých pokusů, tzn. při výběru bez vracení
- NV X je počet vybraných prvků se sledovanou vlastností při závislých pokusech
- *rozdělení má tři parametry:*
  *N* … rozsah souboru, z něhož vybíráme
  *M* … počet prvků v základním souboru, které mají sledovanou vlastnost
  *n* … počet závislých pokusů (rozsah výběru).

***Pravděpodobnostní funkce:***

$$P(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \; x = max[0,\; M - N + n], \ldots, min[M, n],$$
$$= 0 \qquad \text{jinak.}$$

$$E(X) = n \cdot \frac{M}{N}$$

$$D(X) = n \cdot \frac{M}{N}\left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1}$$

*Použití:* např. při kontrole jakosti u malého počtu výrobků nebo v případě, kdy kontrola má ráz destrukční zkoušky (výrobek je zničen).

# 7 PRAVDĚPODOBNOST – 4. ČÁST

## 7.1 Některá rozdělení spojitých náhodných veličin

### 7.1.1 Exponenciální rozdělení $E(A;\delta)$

- NV $X$ je doba čekání do nastoupení sledovaného jevu, může-li tento jev nastat v kterémkoli okamžiku
- *rozdělení má jeden parametr:*
  $A$ … počáteční doba, během které sledovaný jev nastat nemůže.

***Hustota pravděpodobnosti:***

$$f(x) = \frac{1}{\delta} \cdot e^{-(x-A)/\delta}, \; x > A, \delta > 0, A \geq 0,$$
$$= 0 \qquad \text{jinak.}$$

***Distribuční funkce:***

$$F(x) = 1 - e^{-(x-A)/\delta}, \; x > A,$$
$$= 0 \qquad x \leq A.$$

$$E(X) = A + \delta$$
$$D(X) = \delta^2$$

*Např.:* NV $X$ je doba čekání zákazníka na obsluhu v prodejně, doba realizace dvou po sobě jdoucích telefonních hovorů, doba životnosti zařízení, u nichž dochází k poruše z náhodných příčin (nikoli v důsledku opotřebení).

*Použití:* V teorii spolehlivosti a životnosti, v teorii hromadné obsluhy (tzv. teorii front), v teorii obnovy.

### 7.1.2 Normální rozdělení $N(\mu;\, \sigma^2)$

- je vhodné všude tam, kde kolísání NV $X$ je způsobeno velkým počtem nepatrných a vzájemně nezávislých vlivů
- klasickým typem veličin, které se řídí tímto rozdělením, jsou *náhodné chyby*
- pomocí $N(\mu;\, \sigma^2)$ lze za jistých podmínek aproximovat řadu jiných rozdělení, a to jak spojitých, tak i nespojitých
- *rozdělení má dva parametry:*
  $\mu$ … střední hodnota,
  $\sigma^2$… rozptyl.

***Hustota pravděpodobnosti:***

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \; -\infty < x < \infty, -\infty < \mu < \infty, \; \sigma^2 > 0.$$

***Distribuční funkce:***

$$F(x)=\frac{1}{\sigma\cdot\sqrt{2\pi}}\cdot\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt\,,\ -\infty<x<\infty,-\infty<\mu<\infty,\sigma>0.$$

$$E(X)=\mu$$
$$D(X)=\sigma^2$$

- hustota pravděpodobnosti $f(x)$ je zvonovitá křivka, symetrická podle $x=\mu$ a její tvar závisí na parametru $\sigma^2$
- rozdělení $N(\mu;\ \sigma^2)$ je jednovrcholové, vrchol je v bodě $x=\mu$
- $\mu = modus = medián$.

***Normování NV X s normálním rozdělením:*** výpočet $F(x)$ normálního rozdělení je poměrně komplikovaný, proto se z důvodů usnadnění výpočtu transformuje náhodná veličina X, která má normální rozdělení s parametry $\mu$ a $\delta^2$ na *normovanou veličinu U*, která má *normované normální rozdělení*.

**Normované normální rozdělení $\boldsymbol{N(0;\ 1)}$**

- původní NV *X*, která má $\boldsymbol{N(\mu;\ \sigma^2)}$ normujeme, tzn. transformujeme na NV *U*, která má $\boldsymbol{N(0;\ 1)}$
- hodnoty $\Phi(u)$ a kvantilů $u_p$ lze tabelovat.

$$U=\frac{x-\mu}{\sigma},\ E(U)=0,\ D(U)=1.$$

***Vztah pro výpočet F(x):***

$$F(x)=\Phi\left(\frac{x-\mu}{\sigma}\right)=\Phi(u)\ .$$

***Hustota pravděpodobnosti:***

$$\varphi(u)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{u^2}{2}},\ -\infty<u<\infty.$$

***Distribuční funkce:***

$$\Phi(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u}e^{-\frac{t^2}{2}}dt\,,\ -\infty<u<\infty.$$

***Tabulky normovaného normálního rozdělení:***

Tabelovány jsou obvykle hodnoty $\Phi(u)$ jen pro $u\geq 0$ a kvantily $u_p$ jen pro $P\geq 0{,}5$, neboť vzhledem k symetrii $N(0;\ 1)$ podle bodu $u=0$ platí vztahy:

$\Phi(-u) = 1 - \Phi(u)$, $u_P = -u_{1-P}$,

$u_p = \frac{x_p - \mu}{\sigma} \quad \Rightarrow \quad x_p = \sigma u_p + \mu$.

## 7.2 Rozdělení některých funkcí náhodných veličin

- mají zvláštní význam pro řešení některých matematicko-statistických úloh
- stejné značení pro náhodné veličiny i jejich hodnoty
- používají se především kvantily těchto rozdělení, jejich hodnoty jsou tabelovány.

### 7.2.1 Rozdělení $\chi^2$ $\chi^2(\nu)$

- NV $\chi^2$ je součtem ν nezávislých NV $U$ s normovaným normálním rozdělením
- *rozdělení má jeden parametr:*
  ν … počet stupňů volnosti
- kvantily jsou tabelovány pro $v = 1, 2, \ldots, 30$ a pro vybrané pravděpodobnosti.

$$\chi^2 = \sum_{i=1}^{v} U_i^2 = U_1^2 + U_2^2 + \ldots + U_v^2$$

### 7.2.2 Rozdělení Studentovo (t) $t(\nu)$

- NV $t$ je podílem dvou nezávislých NV, a to NV $U$ s rozdělením $N(0;\ 1)$ a NV $\chi^2$ s rozdělením $\chi^2(\nu)$
- *rozdělení má jeden parametr:*
  $\nu$ … počet stupňů volnosti
- kvantily jsou tabelovány pro $v = 1, 2, \ldots, 30$ a pro vybrané pravděpodobnosti
- používá se především pro výběry malého rozsahu ($n < 30$)
- rozdělení je symetrické podle bodu $t = 0$, pro kvantily proto platí vztah $t_p = -t_{1-p}$.

$$t = \frac{U}{\sqrt{\frac{\chi^2}{\nu}}}$$

### 7.2.3 Rozdělení Fisherovo (Snedecorovo) $F(\nu_1;\ \nu_2)$

- NV $F$ je podílem dvou nezávislých NV, a to NV $\chi_1^2$ s rozdělením $\chi^2(\nu_1)$ a NV $\chi_2^2$ s rozdělením $\chi^2(\nu_2)$
- *rozdělení má dva parametry:*
  $\nu_1$… počet stupňů volnosti NV $\chi_1^2$ (v čitateli)
  $\nu_2$… počet stupňů volnosti NV $\chi_2^2$ (ve jmenovateli)

$$F = \frac{\frac{\chi_1^2}{\nu_1}}{\frac{\chi_2^2}{\nu_2}}$$

# Příklad 1

## Zadání příkladu:

U 32 pracovníků jisté firmy byl zjišťován počet dětí do patnácti let. Uspořádejte tyto hodnoty do tabulky rozdělení četností. Od každého typu četností jednu vyberte a interpretujte.

1, 0, 4, 2, 1, 1, 0, 0 , 3, 2, 0, 1, 3, 2, 2, 0, 1, 1, 2, 0, 0, 1, 2, 1, 0, 2, 0, 2, 0, 2 ,0, 2

## Vypracování příkladu:

Tabulka prostého rozdělení četností

| Počet dětí $x_i$ | $n_i$ | $p_i$ | Kumulativní četnost | |
|---|---|---|---|---|
| | | | absolutní | relativní |
| 0 | 11 | 0,34375 | 11 | 0,34375 |
| 1 | 8 | 0,25000 | 19 | 0359375 |
| 2 | 10 | 0,31250 | 29 | 0,90625 |
| 3 | 2 | 0,06250 | 31 | 0,96875 |
| 4 | 1 | 0,03125 | 32 | 1,00000 |
| Celkem | 32 | 1,00000 | x | x |

## Interpretace:

Celkem 8 pracovníků má jedno dítě do patnácti let.
Z celkového počtu pracovníků má 31,25 % dvě děti do patnácti let.
Celkem 19 pracovníků má jedno a méně dětí do patnácti let.
Z celkového počtu pracovníků má 90,625 % dvě a méně dětí do patnácti let.

## STATGRAPHICS:

Describe – Categorical Data – Tabulation

V nabídce Tables and Graphs je třeba zaškrtnout položku Frequency Table.

Frequency Table for Pocet_deti

| | | | *Relative* | *Cumulative* | *Cum. Rel.* |
|---|---|---|---|---|---|
| *Class* | *Value* | *Frequency* | *Frequency* | *Frequency* | *Frequency* |
| 1 | 0 | 11 | 0,3438 | 11 | 0,3438 |
| 2 | 1 | 8 | 0,2500 | 19 | 0,5938 |
| 3 | 2 | 10 | 0,3125 | 29 | 0,9063 |
| 4 | 3 | 2 | 0,0625 | 31 | 0,9688 |
| 5 | 4 | 1 | 0,0313 | 32 | 1,0000 |

## EXCEL:

Do jednoho sloupce vypíšeme pod sebe jednotlivé obměny proměnné, tedy 0, 1, 2, 3, 4.

Do vedlejšího sloupce vedle každé obměny vypočteme absolutní četnosti:

Vzorce – Další funkce – Statistická

Zvolíme funkci COUNTIF.

V panelu Argumenty funkce zadáme do jednotlivých řádků:

***Oblast:*** hodnoty proměnné

***Kritérium:*** postupně obměny proměnné, např. 0, 1, atd.

=COUNTIF(A1:A32;0) = 11 atd. postupně pro všechny obměny.

Dole vytvoříme součtový řádek:

Vzorce – Mat. a trig. – SUMA.

Do sloupce vedle absolutních četností vypočteme relativní četnosti:

Postupně zadáváme pro jednotlivé obměny:

= absolutní četnost/ rozsah souboru.

Dole vytvoříme součtový řádek:

Vzorce – Mat. a trig. – SUMA.

Do sloupce vedle relativních četností vypočteme kumulativní absolutní četnosti:

Postupně zadáváme pro jednotlivé obměny:

= absolutní četnost pro první obměnu + absolutní četnost pro druhou obměnu atd.

Dole vytvoříme součtový řádek:

Vzorce – Mat. a trig. – SUMA.

Do sloupce vedle relativních četností vypočteme kumulativní relativní četnosti:

Postupně zadáváme pro jednotlivé obměny:

= relativní četnost pro první obměnu + relativní četnost pro druhou obměnu atd.

Dole vytvoříme součtový řádek:

Vzorce – Mat. a trig. – SUMA.

# Příklad 2

## Zadání příkladu:

U třiceti pracovníků jisté firmy bylo zjišťováno, jaký je převládající způsob jejich dopravy do zaměstnání (A = automobil, B = autobus, V = vlak, P = pěší chůze). Odpovědi uspořádejte do tabulky rozdělení četností a vždy jednu četnost od každého typu interpretujte. Data znázorněte za pomoci vhodného typu grafu.

A, A, P, B, B, A, A, P, V, P, V, A, A, B, B, V, B, B, A, A, P, P, B, A, A, B, P, A, A, B

## Vypracování příkladu:

Tabulka prostého rozdělení četností

| Způsob dopravy | $n_i$ | $p_i$ |
|---|---|---|
| Automobil | 12 | 0,4 |
| Autobus | 9 | 0,3 |
| Pěší chůze | 6 | 0,2 |
| Vlak | 3 | 0,1 |
| Celkem | 30 | 1,0 |

## Interpretace:

Celkem 9 pracovníků uvedlo jako převládající způsob dopravy do zaměstnání autobus.
Z celkového počtu pracovníků uvedlo 20 % jako převládající způsob dopravy do zaměstnání pěší chůzi.

## STATGRAPHICS:

Describe – Categorical Data – Tabulation

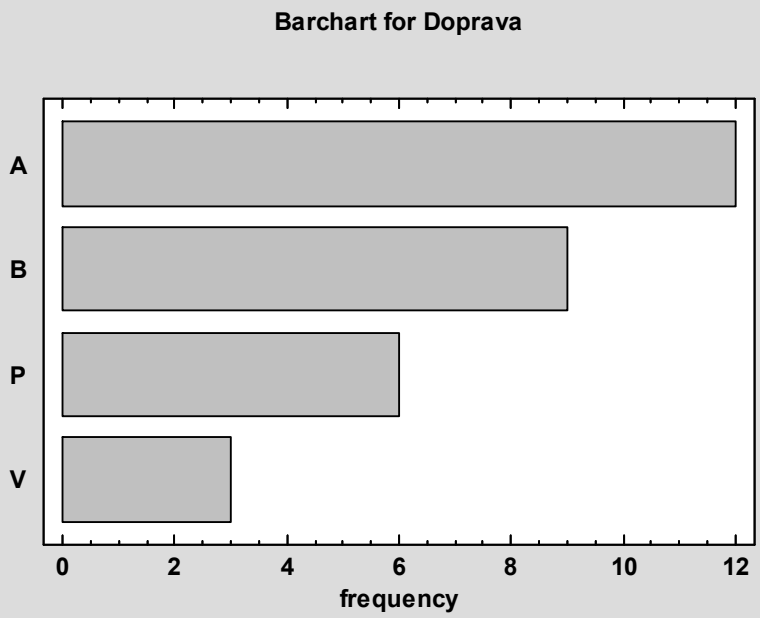
V nabídce Tables and Graphs je třeba zaškrtnout položku Frequency Table.

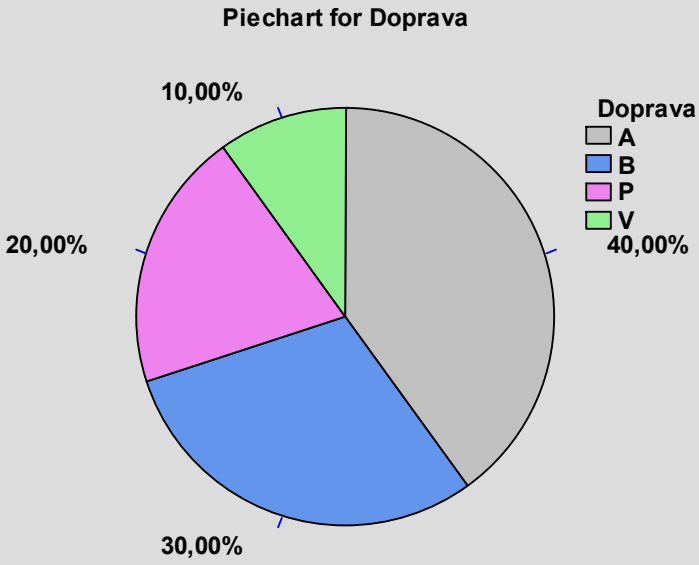
Frequency Table for Doprava

| | | | *Relative* | *Cumulative* | *Cum. Rel.* |
|---|---|---|---|---|---|
| *Class* | *Value* | *Frequency* | *Frequency* | *Frequency* | *Frequency* |
| 1 | A | 12 | 0,4000 | 12 | 0,4000 |
| 2 | B | 9 | 0,3000 | 21 | 0,7000 |
| 3 | P | 6 | 0,2000 | 27 | 0,9000 |
| 4 | V | 3 | 0,1000 | 30 | 1,0000 |

Vhodným typem grafu pro znázornění výše uvedené tabulky rozdělení četností je sloupcový graf (Barchart) nebo výsečový graf (Piechart).

V nabídce Tables and Graphs je třeba zaškrtnout položku Barchart resp. Piechart.


Barchart for Doprava
A
B
P
V
0
2
4
6
8
10
12
frequency


Piechart for Doprava
10,00%
20,00%
30,00%
40,00%
Doprava
A
B
P
V

## EXCEL:

Do jednoho sloupce vypíšeme pod sebe jednotlivé obměny proměnné, tedy A, B, P, V.

Do vedlejšího sloupce vedle každé obměny vypočteme absolutní četnosti:

Vzorce – Další funkce – Statistická

Zvolíme funkci COUNTIF.

V panelu Argumenty funkce zadáme do jednotlivých řádků:

***Oblast:*** hodnoty proměnné

***Kritérium:*** postupně obměny proměnné, např. A, B, atd.

=COUNTIF(A1:A30;"A") = 12 atd. postupně pro všechny obměny.

Dole vytvoříme součtový řádek:

Vzorce – Mat. a trig. – SUMA.

Do sloupce vedle absolutních četností vypočteme relativní četnosti:

Postupně zadáváme pro jednotlivé obměny:

= absolutní četnost/ rozsah souboru.

Dole vytvoříme součtový řádek:

Vzorce – Mat. a trig. – SUMA.

**Výsečový graf:**

Vložení – Graf – Výsečový

Jako data uvedeme četnosti jednotlivých obměn.

**Sloupcový graf:**

Vložení – Graf – Sloupcový

Jako data uvedeme četnosti jednotlivých obměn.

# Příklad 3

## Zadání příkladu:

Máme k dispozici následující hodnoty věku pracovníků jisté firmy:

45, 22, 28, 31, 39, 35, 44, 48, 52, 36, 27, 26, 35, 47, 58, 54, 47, 41, 33, 32, 55, 24, 22, 25, 38, 34, 45, 47, 31, 30, 33, 43, 58, 29, 28.

Uspořádejte údaje to tabulky intervalového rozdělení četností a od každého typu četností jednu vyberte a interpretujte. Vytvořené intervalové rozdělení četností znázorněte vhodným typem grafu.

## Vypracování příkladu:

Rozsah souboru $= n = 35$

Počet intervalů stanovíme pomocí Sturgesova pravidla: $k \approx 1 + 3{,}3 \log_{10} n = 1 + 3{,}3 \log_{10} 35 \doteq 6$

V souboru určíme základní potřebné charakteristiky:

$x_{\min} = 22\,;\; x_{\max} = 58\,;\; R = x_{\min} - x_{\max} = 36$

- Aby bylo intervalové rozdělení přehledné a hranice intervalů byly tvořeny celými čísly, zvolíme jako dolní mez prvního intervalu číslo 20 a jako horní mez posledního intervalu číslo 62. Rozpětí těchto čísel je 42, po rozdělení na 6 intervalů je délka intervalu 7.
- U každého intervalu je třeba stanovit jeho střed, který lze v případě potřeby používat k výpočtům.
- Pravidlo pro zařazování hodnot do intervalů: hodnoty na hranici intervalů jsou zařazovány do nižšího intervalu.

Tabulka intervalového rozdělení četností věku pracovníků

| Hranice intervalu v letech | Střed intervalu v letech | $n_i$ | $p_i$ | Kumulativní četnost | |
|---|---|---|---|---|---|
| | | | | absolutní | relativní |
| 20-27 | 23,5 | 6 | 0,1714 | 6 | 0,1714 |
| 27-34 | 30,5 | 10 | 0,2857 | 16 | 0,4571 |
| 34-41 | 37,5 | 6 | 0,1714 | 22 | 0,6286 |
| 41-48 | 44,5 | 8 | 0,2286 | 30 | 0,8571 |
| 48-55 | 51,5 | 3 | 0,0857 | 33 | 0,9429 |
| 55-62 | 58,5 | 2 | 0,0571 | 35 | 1,0000 |
| Celkem | x | 35 | 1,0000 | x | x |

## Interpretace:

Celkem 10 pracovníků je ve věku od 27 do 34 let.
Z celkového počtu pracovníků jich je 17,14 % ve věku od 34 do 41 let.
Celkem 30 pracovníků je ve věku od 20 do 48 let.
Z celkového počtu pracovníků je 94,29 % ve věku od 20 do 55 let.

Vhodným typem grafu pro znázornění intervalového rozdělení četností je ***histogram četností***.

## STATGRAPHICS:

Desribe – Numeric Data – One-Variable Analysis

V nabídce Tables and Graphs je třeba zaškrtnout položku Frequency Table.

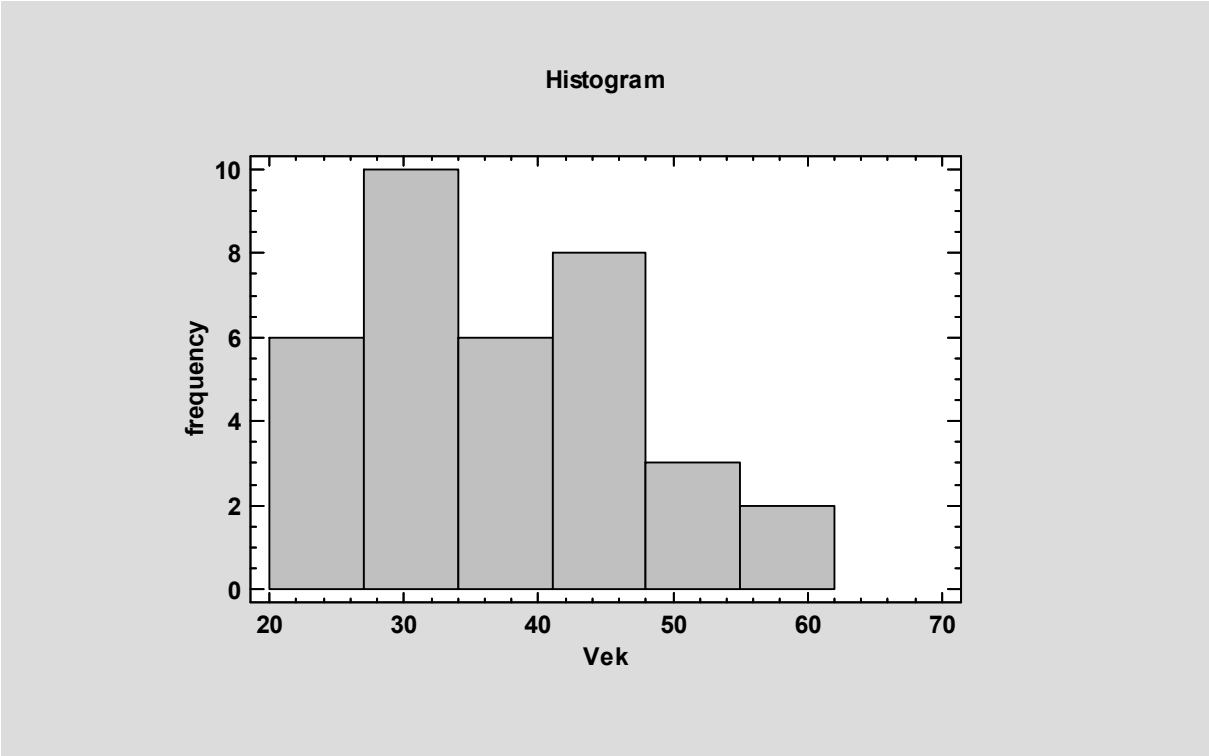
V nabídce Pane Options zadáme počet tříd 6, dolní mez 20, horní mez 62.

Frequency Tabulation for Vek

| | *Lower* | *Upper* | | | *Relative* | *Cumulative* | *Cum. Rel.* |
|---|---|---|---|---|---|---|---|
| *Class* | *Limit* | *Limit* | *Midpoint* | *Frequency* | *Frequency* | *Frequency* | *Frequency* |
| | at or below | 20 | | 0 | 0,0000 | 0 | 0,0000 |
| 1 | 20 | 27,0 | 23,5 | 6 | 0,1714 | 6 | 0,1714 |
| 2 | 27 | 34,0 | 30,5 | 10 | 0,2857 | 16 | 0,4571 |
| 3 | 34 | 41,0 | 37,5 | 6 | 0,1714 | 22 | 0,6286 |
| 4 | 41 | 48,0 | 44,5 | 8 | 0,2286 | 30 | 0,8571 |
| 5 | 48 | 55,0 | 51,5 | 3 | 0,0857 | 33 | 0,9429 |
| 6 | 55 | 62,0 | 58,5 | 2 | 0,0571 | 35 | 1,0000 |
| | above | 62 | | 0 | 0,0000 | 35 | 1,0000 |

Mean = 37,7714 Standard deviation = 10,5163

V nabídce Tables and Graphs je třeba zaškrtnout položku Frequency Histogram.




## Excel:

Do prvního sloupce zadáme data. Dále určíme počet intervalů, šířku intervalu a hranice intervalů (viz postup v ručním řešení příkladu). Horní hranice intervalů zadáme do vedlejšího sloupce. Vedle sloupce, ve kterém jsou horní hranice intervalů, podsvítíme vedlejší pole.

Automatické shrnutí – Další funkce – Vložit funkci – Statistické – Četnosti

V panelu Argumenty funkce zadáme do jednotlivých řádků:

***Data:*** hodnoty proměnné

***Hodnoty:*** horní hranice intervalů

Zadání argumentů funkce je nutno potvrdit stiskem kláves Ctrl+Shift+Enter.

Dole vytvoříme součtový řádek:

Vzorce – Mat. a trig. – SUMA.

Do sloupce vedle absolutních četností vypočteme relativní četnosti:

Postupně zadáváme pro jednotlivé obměny:

= absolutní četnost/ rozsah souboru.

Dole vytvoříme součtový řádek:

Vzorce – Mat. a trig. – SUMA.

Do sloupce vedle relativních četností vypočteme kumulativní absolutní četnosti:

Postupně zadáváme pro jednotlivé obměny:

= absolutní četnost pro první obměnu + absolutní četnost pro druhou obměnu atd.

Dole vytvoříme součtový řádek:

Vzorce – Mat. a trig. – SUMA.

Do sloupce vedle relativních četností vypočteme kumulativní relativní četnosti:

Postupně zadáváme pro jednotlivé obměny:

= relativní četnost pro první obměnu + relativní četnost pro druhou obměnu atd.

Dole vytvoříme součtový řádek:

Vzorce – Mat. a trig. – SUMA.

**Sloupcový graf:**

Vložení – Graf – Sloupcový

Jako data uvedeme četnosti jednotlivých intervalů.

# Příklad 1

## Zadání příkladu:

V rámci marketingového průzkumu bylo 500 respondentů dotázáno, jakou barvu by preferovali při koupi nového automobilu. Jejich odpovědi jsou uvedeny v tabulce.

| Barva automobilu | Počet respondentů |
|---|---|
| Bílá | 87 |
| Šedá | 22 |
| Černá | 69 |
| Červená | 113 |
| Zelená | 94 |
| Modrá | 69 |
| Žlutá | 46 |
| Celkem | 500 |

Změřte variabilitu odpovědí respondentů vhodnou mírou a její hodnotu interpretujte. Dále změřte vhodnou charakteristikou úroveň hodnot a výsledek interpretujte.

## Vypracování příkladu:

| Barva automobilu | $n_i$ | $n_i^2$ |
|---|---|---|
| Bílá | 87 | 7 569 |
| Šedá | 22 | 484 |
| Černá | 69 | 4761 |
| Červená | 113 | 12 769 |
| Zelená | 97 | 8 836 |
| Modrá | 69 | 4 761 |
| Žlutá | 46 | 2 116 |
| Celkem | 500 | 41 296 |

$$M = \frac{n^2 - \sum_{i=1}^{k} n_i^2}{n(n-1)} = \frac{500^2 - 41296}{500(500-1)} \doteq 0{,}8365$$

$\hat{x} = Červená$

## Interpretace:

Variabilita preferované barvy automobilu, měřená mírou mutability, je značně vysoká. Z celkového počtu všech možných dvojic respondentů uvedlo 83,65 % různou preferovanou barvu nového automobilu.

Typickou (modální, nejčetnější) hodnotou je odpověď „Červená“. Jinak řečeno, nejvíce respondentů by při nákupu nového automobilu preferovalo červenou barvu.

# Příklad 2

## Zadání příkladu:

V předvolebním průzkumu bylo zjišťováno, zda se respondenti zúčastní nejbližších voleb. Odpovědi respondentů jsou uvedeny v tabulce.

| Odpověď | Podíl respondentů v % |
|---|---|
| Ano | 64,8 |
| Ne | 21,7 |
| Nevím | 13,5 |

Změřte variabilitu odpovědí respondentů vhodnou mírou a její hodnotu interpretujte. Dále změřte vhodnou charakteristikou úroveň hodnot a výsledek interpretujte.

## Vypracování příkladu:

| Odpověď | Podíl respondentů v % | $p_i$ | $p_i^2$ |
|---|---|---|---|
| Ano | 64,8 | 0,648 | 0,419904 |
| Ne | 21,7 | 0,217 | 0,047089 |
| Nevím | 13,5 | 0,135 | 0,018225 |

$$NOMVAR = 1 - \sum_{i=1}^{k} p_i^2 = 1 - 0{,}485218 = 0{,}514782$$

$$\hat{x} = Ano$$

## Interpretace:

Variabilita odpovědí respondentů, měřená pomocí nominální variance, je středně vysoká.

Typickou (modální, nejčetnější) hodnotou je odpověď „Ano“. Jinak řečeno, největší podíl respondentů uvedl odpověď „Ano“.

# Příklad 3

## Zadání příkladu:

Máme k dispozici následující hodnoty věku pracovníků jisté firmy v dokončených letech:

45, 22, 28, 31, 39, 35, 44, 48, 52, 36, 27, 26, 35, 47, 58, 54, 47, 41, 33, 32, 55, 24, 22, 25, 38, 34, 45, 47, 31, 30, 33, 43, 58, 29, 28.

Stanovte všechny kvartily proměnné věk, prostřední kvartil interpretujte. Dále stanovte druhý tercil, třetí kvintil, sedmý decil a dvacátý druhý percentil a jejich hodnoty interpretujte.

## Vypracování příkladu:

Nejdříve setřídíme data do neklesající řady.

22, 22, 24, 25, 26, 27, 28, 28, 29, 30, 31, 31, 32, 33, 33, 34, 35, 35, 36, 38, 39, 41, 43, 44, 45, 45, 47, 47, 47, 48, 52, 54, 55, 58, 58.

- Pro jednotlivé druhy kvantilů nejprve stanovíme jejich pořadová čísla v neklesající řadě a pak k těmto pořadovým číslům přiřadíme příslušnou hodnotu.
- Pokud je pořadové číslo mezi dvěma čísly, která nejsou celá, je pořadovým číslem daného kvantilu celé číslo, které se nalézá mezi nimi.
- Pokud je pořadové číslo mezi dvěma celými čísly, stanovíme pořadové číslo daného kvantilu jako aritmetický průměr těchto dvou celých čísel.
- Pro stanovení mediánu lze použít zjednodušený postup.

***První kvartil*** $\tilde{x}_{25}\left(x_{0,25}\right)$

Pořadové číslo prvního kvartilu: $n \cdot p < m_p < n \cdot p + 1$

$$35 \cdot 0{,}25 < m_{0,25} < 35 \cdot 0{,}25 + 1$$

$$8{,}75 < m_{0,25} < 9{,}75$$

Pořadové číslo prvního kvartilu je 9, jde tedy o devátý člen v neklesající řadě hodnot věku.

$\tilde{x}_{25} = 29$

***Druhý kvartil (medián)*** $\tilde{x}_{50}\left(x_{0,5}, \tilde{x}\right)$

Pořadové číslo mediánu: $\tilde{x} = x_{\left(\frac{n+1}{2}\right)}$, protože rozsah souboru *n* je liché číslo

$$\tilde{x} = x_{\left(\frac{36}{2}\right)} = x_{18}$$

Pořadové číslo mediánu je 18, jde o prostřední člen v neklesající řadě hodnot věku.

$\tilde{x} = 35$

***Třetí kvartil*** $\tilde{x}_{75}\left(x_{0,75}\right)$

Pořadové číslo prvního kvartilu: $n \cdot p < m_p < n \cdot p + 1$

$$35 \cdot 0{,}75 < m_{0,75} < 35 \cdot 0{,}75 + 1$$

$$26{,}25 < m_{0,75} < 27{,}25$$

Pořadové číslo třetího kvartilu je 27, jde tedy o dvacátý sedmý člen v neklesající řadě hodnot věku.

$\tilde{x}_{75} = 47$

***Druhý tercil*** $\tilde{x}_{66,6667}\left(x_{0,666667}\right)$

Pořadové číslo prvního kvartilu: $n \cdot p < m_p < n \cdot p + 1$

$$35 \cdot 0{,}666667 < m_{0,666667} < 35 \cdot 0{,}666667 + 1$$

$$23{,}3 < m_{0,666667} < 24{,}3$$

Pořadové číslo první kvartilu je 24, jde tedy o dvacátý čtvrtý člen v neklesající řadě hodnot věku.

$\tilde{x}_{66,6667} = 44$

***Třetí kvintil*** $\tilde{x}_{60}\left(x_{0,6}\right)$

Pořadové číslo prvního kvartilu: $n \cdot p < m_p < n \cdot p + 1$

$$35 \cdot 0{,}6 < m_{0,6} < 35 \cdot 0{,}6 + 1$$

$$21 < m_{0,6} < 22$$

Pořadové číslo první kvartilu je 21,5, jde tedy o aritmetický průměr dvacátého prvního a dvacátého členu v neklesající řadě hodnot věku.

$$\tilde{x}_{0,6} = \frac{39 + 41}{2} = 40$$

***Sedmý decil*** $\tilde{x}_{70}\left(x_{0,7}\right)$

Pořadové číslo prvního kvartilu: $n \cdot p < m_p < n \cdot p + 1$

$$35 \cdot 0{,}7 < m_{0,7} < 35 \cdot 0{,}7 + 1$$

$$24{,}5 < m_{0,6} < 25{,}5$$

Pořadové číslo první kvartilu je 25, jde tedy o dvacátý pátý člen v neklesající řadě hodnot věku.

$\tilde{x}_{0,7} = 45$

***Dvacátý druhý percentil*** $\tilde{x}_{22}\left(x_{0,22}\right)$

Pořadové číslo prvního kvartilu: $n \cdot p < m_p < n \cdot p + 1$

$$35 \cdot 0{,}22 < m_{0,22} < 35 \cdot 0{,}22 + 1$$

$$7{,}7 < m_{0,22} < 8{,}7$$

Pořadové číslo první kvartilu je 8, jde tedy o osmý člen v neklesající řadě hodnot věku.

$\tilde{x}_{0,22} = 28$

## Interpretace:

$\tilde{x} = 35$ Padesát procent pracovníků je ve věku 35 nebo méně let.

$\tilde{x}_{66,6667} = 44$ 66,6667 % pracovníků je ve věku 44 nebo méně let.

$\tilde{x}_{0,6} = 40$ 60 % pracovníků je ve věku 40 nebo méně let.

$\tilde{x}_{0,7} = 45$ 70 % pracovníků je ve věku 45 nebo méně let.

$\tilde{x}_{0,22} = 28$ 22 % pracovníků je ve věku 28 nebo méně let.

## STATGRAPHICS:

Desribe – Numeric Data – One-Variable Analysis

V nabídce Tables and Graphs je třeba zaškrtnout položku Percentiles.

V proceduře Percentiles zvolíme nabídku Pane Options, ve které zadáme požadovaná procenta pro jednotlivé druhy kvantilů (např. 25 pro dolní kvartil atd.).

**Percentiles for Vek**

| | *Percentiles* |
|---|---|
| 25,0% | 29,0 |
| 50,0% | **35,0** |
| 75,0% | 47,0 |
| 66,6667% | **44,0** |
| 60,0% | **40,0** |
| 70,0% | **45,0** |
| 22,0% | **28,0** |
| 95,0% | 58,0 |
| 99,0% | 58,0 |

## EXCEL:

Vzorce – Další funkce – Statistická

Zvolíme funkci PERCENTIL.

V panelu Argumenty funkce zadáme do řádku Pole hodnoty A1:A32.

Do řádku K postupně zadáváme hodnoty jednotlivých kvantilů.

k = 0,25 $\tilde{x}_{25} = 29{,}5$

k = 0,5 $\tilde{x} = 35$

k = 0,75 $\tilde{x}_{75} = 46$

k = 0,666667 $\tilde{x}_{66,6667} = 43,666678$

k = 0,6 $\tilde{x}_{6} = 39,8$

k = 0,7 $\tilde{x}_{70} = 44,8$

k = 0,22 $\tilde{x}_{22} = 28,48$

Odlišnosti ve vypočtených hodnotách některých kvantilů ve STATGRAPHICSu a v EXCELu jsou dány rozdílnými algoritmy v těchto softwarech.

# Příklad 4

## Zadání příkladu:

U 32 pracovníků jisté firmy byl v rámci průzkumu zjištěn počet dětí do patnácti let:

1, 0, 4, 2, 1, 1, 0, 0 , 3, 2, 0, 1, 3, 2, 2, 0, 1, 1, 2, 0, 0, 1, 2, 1, 0, 2, 0, 2, 0, 2 ,0, 2

Stanovte modus a aritmetický průměr proměnné počet dětí do patnácti let a vypočtené hodnoty interpretujte. Dále změřte variabilitu proměnné směrodatnou odchylkou a variačním koeficientem, stanovené hodnoty rovněž interpretujte.

## Vypracování příkladu:

Tabulka pro výpočet

| Počet dětí $x_i$ | $n_i$ | $x_i n_i$ | $x_i^2 n_i$ |
|---|---|---|---|
| 0 | 11 | 0 | 0 |
| 1 | 8 | 8 | 8 |
| 2 | 10 | 20 | 40 |
| 3 | 2 | 6 | 18 |
| 4 | 1 | 4 | 16 |
| Celkem | 32 | 38 | 82 |

Pro výpočet z tabulky rozdělení četností používáme vážené tvary příslušných vzorců.

$$\hat{x} = 0$$

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i} = \frac{38}{32} = 1{,}1875$$

$$s_x^2 = \frac{\sum_{i=1}^{k} x_i^2 n_i}{\sum_{i=1}^{k} n_i} - \left( \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i} \right)^2 = \frac{82}{32} - \left( \frac{38}{32} \right)^2 \doteq 1{,}1523437$$

$$s_x = \sqrt{s_x^2} = \sqrt{1{,}1523437} \doteq 1{,}0735$$

$$V_x = \frac{s_x}{\bar{x}} = \frac{1{,}0735}{1{,}1875} \doteq 0{,}904$$

## Interpretace:

$\hat{x} = 0$ V daném souboru je nejvíce pracovníků bezdětných.

$\bar{x} = 1{,}1875$ Průměrný počet dětí do patnácti let na jednoho pracovníka je 1,1875.

$s_x = 1{,}0735$ V průměru se počet dětí jednotlivých pracovníků liší od svého aritmetického o 1,0735 v obou směrech.

$V_x \doteq 0{,}904$ Směrodatná odchylka se na průměru hodnot podílí 90,4 %. Jedná se o velmi vysokou relativní variabilitu.

## STATGRAPHICS:

Desribe – Numeric Data – One-Variable Analysis

V nabídce Tables and Graphs je třeba zaškrtnout položku Summary Statistics.

V proceduře Summary Statistics zvolíme nabídku Pane Options, ve které zaškrtneme požadované charakteristiky.

**Summary Statistics for Pocet_deti**

| | |
|---|---|
| Count | 32 |
| Average | **1,1875** |
| Mode | **0** |
| Variance | 1,18952 |
| Standard deviation | **1,09065** |

$\hat{x} = 0$

$\bar{x} = 1{,}1875$

Statgraphics počítá výběrový rozptyl a výběrovou směrodatnou odchylku, takže hodnotu směrodatné odchylky je třeba přepočítat tak, aby odpovídala klasickému tvaru:

$$s_x = s'_x \sqrt{\frac{n-1}{n}} = s_x = 1{,}09065 \cdot \sqrt{\frac{31}{32}} \doteq 1{,}0735$$

Variační koeficient dopočítáme podle vzorce:

$$V_x = \frac{s_x}{\bar{x}} = \frac{1{,}0735}{1{,}1875} \doteq 0{,}904$$

## EXCEL:

Vzorce – Další funkce – Statistická

Postupně zvolíme funkce, které chceme vypočítat.

V panelu Argumenty funkce vždy zadáme do řádku 1 hodnoty A1:A32.

MODE = 0

AVERAGEA = 1,1875

VAR = 1,15234375

SMODCH = 1,073472752

# Příklad 1

## Zadání příkladu:

Na lístcích jsou napsána čísla od 1 do 10, přičemž každé číslo je napsáno na různém počtu lístků. Číslo 1 je na jednom lístku, číslo 2 na dvou lístcích, číslo 3 na třech lístcích atd. Číslo deset je tedy na deseti lístcích. Všechny lístky jsou vloženy do osudí a pečlivě promíchány. Stanovte pravděpodobnost, že:

a) Při třech náhodných po sobě jdoucích výběrech bez vracení bude první vytažené číslo 1, druhé vytažené číslo 2 a třetí vytažené číslo 3.
b) Při třech náhodných po sobě jdoucích výběrech s vracením bude první vytažené číslo 1, druhé vytažené číslo 2 a třetí vytažené číslo 3.

## Vypracování příkladu:

Celkový počet lístků v osudí = $1+2+3+\ldots+10 = 55$

a) Sledujeme celkem tři jevy, a to A, B a C. Vzhledem k tomu, že se jedná o výběr bez vracení, tedy o závislé pokusy, je třeba při výpočtu stanovit podmíněné pravděpodobnosti.

Jev A je vytažení lístku s číslem 1. Pravděpodobnost tohoto jevu není podmíněná. Lístek s číslem 1 je v osudí pouze jeden.

**1. krok:** $P(A) = \frac{1}{55} \doteq 0{,}018182$

Jev B je vytažení lístku s číslem 2. Stanovíme podmíněnou pravděpodobnost nastoupení jevu B za podmínky, že v prvním kroku nastal jev A, tedy byl vytažen lístek s číslem 1. Lístky s číslem 2 jsou v osudí celkem dva.

**2. krok:** $P(A/B) = \frac{2}{54} = 0{,}037037$

Jev C je vytažení lístku s číslem 3. Stanovíme podmíněnou pravděpodobnost nastoupení jevu C za podmínky, že v prvním kroku nastal jev A a v druhém kroku jev B. Lístky s číslem 3 jsou v osudí celkem tři.

**3. krok:** $P(C/A\cap B) = \frac{3}{53} \doteq 0{,}056604$

Pravděpodobnost, že při výběru bez vracení bude první vytažené číslo 1, druhé vytažené číslo 2 a třetí vytažené číslo 3 vypočteme jako součin pravděpodobností, stanovených pro jednotlivé kroky.

$$P(A\cap B\cap C) = P(A)\cdot P(B/A)\cdot P(C/A\cap B) = \frac{1}{55}\cdot\frac{2}{54}\cdot\frac{3}{53} \doteq 0{,}000038$$

## Interpretace:

Pravděpodobnost, že při třech náhodných po sobě jdoucích výběrech bez vracení bude první vytažené číslo 1, druhé vytažené číslo 2 a třetí vytažené číslo 3 je 0,0038 %. Tato pravděpodobnost je velmi malá.

b) Sledujeme celkem tři jevy, a to A, B a C. Vzhledem k tomu, že se jedná o výběr s vracením, tedy o nezávislé pokusy, je třeba stanovit nepodmíněné pravděpodobnosti jednotlivých jevů.

Jev A je vytažení lístku s číslem 1. Pravděpodobnost tohoto jevu není podmíněná. Lístek s číslem 1 je v osudí pouze jeden.

**1. krok:** $P(A) = \frac{1}{55} \doteq 0{,}018182$

Jev B je vytažení lístku s číslem 2. Pravděpodobnost tohoto jevu není podmíněná. Lístky s číslem 2 jsou v osudí celkem dva.

**2. krok:** $P(B) = \frac{2}{55} \doteq 0{,}036364$

Jev C je vytažení lístku s číslem 3. Pravděpodobnost tohoto jevu není podmíněná. Lístky s číslem 3 jsou v osudí celkem tři.

**3. krok:** $P(C) = \frac{3}{55} \doteq 0{,}054546$

Pravděpodobnost, že při výběru s vracením bude první vytažené číslo 1, druhé vytažené číslo 2 a třetí vytažené číslo 3 vypočteme jako součin pravděpodobností, stanovených pro jednotlivé kroky.

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C) = \frac{1}{55} \cdot \frac{2}{55} \cdot \frac{3}{55} \doteq 0{,}000036$$

## Interpretace:

Pravděpodobnost, že při třech náhodných po sobě jdoucích výběrech s vracením bude první vytažené číslo 1, druhé vytažené číslo 2 a třetí vytažené číslo 3 je 0,0036 %. Tato pravděpodobnost je velmi malá.

# Příklad 2

## Zadání příkladu:

Dodávku Kinder vajíček s překvapením obdržel supermarket v krabicích po 200 kusech. Jedná se o vajíčka s figurkami afrických zvířat, přičemž výrobce deklaruje, že 40 % vajíček obsahuje figurku zebry, 30 % figurku antilopy, 20 % figurku slona a 10 % figurku lva. Stanovte pravděpodobnost, že při náhodném výběru tří vajíček z plné krabice bude:

a) ve všech třech vajíčkách lev
b) ve všech třech vajíčkách zebra.

## Vypracování příkladu:

a) Sledujeme nastoupení jevu A, kterým je vybrání tří vajíček s překvapením s figurkou lva.

$$P(A) = \frac{m}{n}$$

***Případy příznivé nastoupení jevu A (m):***

V krabici je celkem 200 vajíček, z toho jich 20 obsahuje figurku lva. Příznivé případy jsou všechny trojice vajíček, které obsahují figurku lva.

$$C_3(20) = \frac{n!}{k!(n-k)!} = \frac{20!}{3!17!} = 1140$$

***Případy možné (n):***

V krabici je celkem 200 vajíček, možné případy jsou všechny trojice vajíček, které lze z 200 vajíček vytvořit.

$$C_3(200) = \frac{n!}{k!(n-k)!} = \frac{200!}{3!197!} = 1313400$$

$$P(A) = \frac{m}{n} = \frac{1140}{1313400} \doteq 0{,}000868$$

## Interpretace:

Pravděpodobnost, že při náhodném výběru tří vajíček z plné krabice bude ve všech vajíčkách figurka lva, činí 0,0868 %. Tato pravděpodobnost je velmi malá.

b) Sledujeme nastoupení jevu A, kterým je vybrání tří vajíček s překvapením s figurkou zebry.

$$P(A) = \frac{m}{n}$$

***Případy příznivé nastoupení jevu A (m):***

V krabici je celkem 200 vajíček, z toho jich 80 obsahuje figurku lva. Příznivé případy jsou všechny trojice vajíček, které obsahují figurku zebry.

$$C_3(80) = \frac{n!}{k!(n-k)!} = \frac{80!}{3!77!} = 82160$$

***Případy možné (n):***

V krabici je celkem 200 vajíček, možné případy jsou všechny trojice vajíček, které lze z 200 vajíček vytvořit.

$$C_3(200) = \frac{n!}{k!(n-k)!} = \frac{200!}{3!197!} = 1313400$$

$$P(A) = \frac{m}{n} = \frac{82160}{1313400} \doteq 0{,}062555$$

## Interpretace:

Pravděpodobnost, že při náhodném výběru tří vajíček z plné krabice bude ve všech vajíčkách figurka zebry, činí 6,2555 %. Tato pravděpodobnost je malá.

# Příklad 1

## Zadání příkladu:

Náhodná veličina $X$ je popsána následující pravděpodobnostní funkcí:

$$P(x) = 0{,}5^x \cdot 0{,}5 \qquad x = 0,1,2,3$$
$$= 0{,}5^4 \qquad x = 4$$
$$= 0 \qquad jinak$$

a) Vyjádřete tuto pravděpodobnostní funkci tabulkou.
b) Stanovte modus, střední hodnotu a rozptyl této náhodné veličiny.

## Vypracování příkladu:

a) Tabulka obsahuje všechny hodnoty, kterých může daná náhodná veličina nabýt, a jejich pravděpodobnosti. Jednotlivé pravděpodobnosti vypočteme dosazením příslušných hodnot náhodné veličiny do pravděpodobnostní funkce.

$P(0) = 0{,}5^0 \cdot 0{,}5 = 0{,}5$

$P(1) = 0{,}5^1 \cdot 0{,}5 = 0{,}25$

$P(2) = 0{,}5^2 \cdot 0{,}5 = 0{,}125$

$P(3) = 0{,}5^3 \cdot 0{,}5 = 0{,}0625$

$P(4) = 0{,}5^4 = 0{,}0625$

***Tabulka pravděpodobnostní funkce***

| **x** | 0 | 1 | 2 | 3 | 4 | Celkem |
|---|---|---|---|---|---|---|
| **P(x)** | 0,5 | 0,25 | 0,125 | 0,0625 | 0,0625 | 1,0000 |

## Interpretace:

Pravděpodobnost, že náhodná veličina nabude hodnoty 0, je 50 %.

Pravděpodobnost, že náhodná veličina nabude hodnoty 1, je 25 %.

Pravděpodobnost, že náhodná veličina nabude hodnoty 2, je 12,5 %.

Pravděpodobnost, že náhodná veličina nabude hodnoty 3, je 6,25 %.

Pravděpodobnost, že náhodná veličina nabude hodnoty 4, je 6,25 %.

b) ***Modus náhodné veličiny***

$\hat{x} = 0$

***Střední hodnota náhodné veličiny*** $E(X)$

$$E(X) = \sum_{M} x \cdot P(x) = 0{,}9375$$

***Rozptyl náhodné veličiny*** $D(X)$

$$D(X) = \sum_{M} [x - E(X)]^2 \cdot P(x) = \sum_{M} x^2 \cdot P(x) - \left[\sum_{M} x \cdot P(x)\right]^2 = 2{,}5125 - 0{,}9375^2 = 1{,}633594$$

Tabulka s výpočty

| ***x*** | *P(x)* | *xP(x)* | *$x^2$P(x)* |
|---|---|---|---|
| 0 | 0,5 | 0 | 0 |
| 1 | 0,25 | 0,25 | 0,25 |
| 2 | 0,125 | 0,25 | 1,00 |
| 3 | 0,0625 | 0,1875 | 0,5625 |
| 4 | 0,0625 | 0,25 | 1,000 |
| Celkem | 1,0000 | 0,9375 | 2,5125 |

## Interpretace:

Modus náhodné veličiny *X*, tedy hodnota s nejvyšší pravděpodobností nastoupení, je 0.

Střední hodnota náhodné veličiny *X* je 0,9375.

Rozptyl náhodné veličiny *X* je 1,633594.

# Příklad 2

## Zadání příkladu:

Spojitá náhodná veličina $X$ je popsána následující hustotou pravděpodobnosti:

$$f(x) = 3x^2 \qquad 0 < x < 1$$
$$= 0 \qquad \textit{jinak}$$

Vypočtěte kvartily této náhodné veličiny.

## Vypracování příkladu:

Nejdříve je třeba stanovit distribuční funkci náhodné veličiny $X$.

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} t^2 dt = x^3$$

$$F(x) = x^3 \qquad 0 < x < 1$$
$$= 0 \qquad \textit{jinak}$$

Kvartily stanovíme na základě následujícího vztahu:

$$F(x_p) = P(X \le x_p) = p$$

***První kvartil*** $(x_{0,25})$***:***

$$F(x_{0,25}) = P(X \le x_{0,25}) = 0,25$$

$$x_{0,25}^3 = 0,25$$

$$x_{0,25} = \sqrt[3]{0,25} \doteq 0,63$$

***Druhý kvartil*** $(x_{0,5})$***:***

$$F(x_{0,5}) = P(X \le x_{0,5}) = 0,5$$

$$x_{0,5}^3 = 0,5$$

$$x_{0,5} = \sqrt[3]{0,5} \doteq 0,794$$

***Třetí kvartil*** $(x_{0,5})$***:***

$$F(x_{0,75}) = P(X \le x_{0,75}) = 0,75$$

$$x_{0,75}^3 = 0,75$$

$$x_{0,75} = \sqrt[3]{0,75} \doteq 0,909$$

## Interpretace:

První (dolní) kvartil náhodné veličiny $X$ je 0,63, druhý (prostřední) kvartil, tedy medián, je 0,794, a třetí (horní) kvartil je 0,909.

# Příklad 1

## Zadání příkladu:

Pravděpodobnost, že spotřeba elektrické energie ve všední den sledovaného období přesáhne stanovenou normu je 0,3. Stanovte pravděpodobnost, že během pěti všedních dnů sledovaného období:

a) nebude norma překročena
b) bude norma překročena maximálně dvakrát
c) bude norma překročena nejméně čtyřikrát.

## Vypracování příkladu:

Jedná se o nezávislé náhodné pokusy, přičemž známe pravděpodobnost nastoupení sledovaného jevu (překročení normy) v jednom pokusu (spotřeba elektrické energie ve všední den sledovaného období).

Vhodným modelem je binomické rozdělení $Bi(n;\pi)$:

- $n$… počet nezávislých pokusů
- NV $X$… spotřeba elektrické energie
- jev $A$… překročení normy; $P(A)=0{,}3$

Parametry rozdělení:

$n = 5$

$\pi = 0{,}3$

Pravděpodobnostní funkce:

$$P(x)=\binom{n}{x}\pi^{x}(1-\pi)^{n-x}$$

a) $P(X=0)=P(0)=\binom{5}{0}0{,}3^{0}(1-0{,}3)^{5-0}\doteq 0{,}1681$

b) $P(X\le 2)=P(0)+P(1)+P(2)=\binom{5}{0}0{,}3^{0}(1-0{,}3)^{5-0}+\binom{5}{1}0{,}3^{1}(1-0{,}3)^{5-1}+$

$+\binom{5}{2}0{,}3^{2}(1-0{,}3)^{5-2}\doteq 0{,}8369$

c) $P(X\ge 4)=P(4)+P(5)=\binom{5}{4}0{,}3^{4}(1-0{,}3)^{5-4}+\binom{5}{5}0{,}3^{5}(1-0{,}3)^{5-5}\doteq 0{,}0308$

## Interpretace:

a) Pravděpodobnost, že během pěti všedních dnů sledovaného období nebude překročena norma, je 16,81 %. Tato pravděpodobnost je poměrně malá.

b) Pravděpodobnost, že během pěti všedních dnů sledovaného období bude norma překročena maximálně dvakrát, je 83,69 %. Tato pravděpodobnost je značně velká.

c) Pravděpodobnost, že během pěti všedních dnů sledovaného období bude norma překročena nejméně čtyřikrát, je 3,08 %. Tato pravděpodobnost je velmi malá.

## STATGRAPHICS:

Desribe – Distribution Fitting – Probability Distributions

V panelu Probability Distributions zaškrtneme Binomial.

V panelu Binomial Options zadáme oba parametry rozdělení.

V nabídce Tables and Graphs je třeba zaškrtnout položku Cumulative Distributions.

V proceduře Cumulative Distributions zvolíme nabídku Pane Options, ve které zadáme požadované hodnoty NV, jejichž pravděpodobnosti budeme počítat, tedy 0, 2 a 4.

**Cumulative Distribution**
Distribution: Binomial

Lower Tail Area (<)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0 | 0,0 | | | | |
| 2 | **0,52822** | | | | |
| 4 | 0,96922 | | | | |

Probability Mass (=)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0 | **0,16807** | | | | |
| 2 | **0,3087** | | | | |
| 4 | **0,02835** | | | | |

Upper Tail Area (>)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0 | 0,83193 | | | | |
| 2 | 0,16308 | | | | |
| 4 | **0,0024295** | | | | |

a) $P(X=0)=0{,}16807$

b) $P(X\le 2)=P(X<2)+P(X=2)=0{,}52822+0{,}3087=0{,}83692$

c) $P(X\ge 4)=P(X>4)+P(X=4)=0{,}0024295+0{,}02835=0{,}0307795$

## EXCEL:

Vzorce – Další funkce – Statistická

Zvolíme funkci BINOMDIST.

V panelu Argumenty funkce zadáme do jednotlivých řádků:

***Úspěch:*** hodnotu NV $X$

***Pokusy:*** počet nezávislých náhodných pokusů $n$

***Prst_úspěchu:*** parametr $\pi$

***Počet:*** NEPRAVDA pro pravděpodobnostní funkci

a) BINOMDIST(3; 5; 0,3; NEPRAVDA)

$P(X = 0) = 0{,}16807$

b) BINOMDIST(2; 5; 0,3; PRAVDA)

$P(X \leq 2) = 0{,}83692$

c) BINOMDIST(3; 5; 0,3; PRAVDA)

$P(X \geq 4) = 1 - P(X \leq 3) = 0{,}0307795$

# Příklad 2

## Zadání příkladu:

Bylo zjištěno, že na 1 000 m tkaniny je v průměru 5 kazů. Pro expedici jsou připraveny stometrové balíky tkaniny. Stanovte pravděpodobnost, že náhodně vybraný balík:

a) bude bez kazu
b) bude mít méně než tři kazy
c) bude mít nejméně dva kazy.

## Vypracování příkladu:

Náhodná veličina $X$ je počet kazů na 100 m tkaniny. Vhodným modelem pro tuto NV je Poissonovo rozdělení $Po(\lambda)$.

Stanovení parametru $\lambda$:

Na 1000 m tkaniny je v průměru 5 kazů. Na 100 m tkaniny je tedy v průměru 0,5 kazu.

$\lambda = 0{,}5$

Pravděpodobnostní funkce:

$$P(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$$

a) $\lambda = 0{,}5\,;\ x = 0$

$$P(X = 0) = P(0) = e^{-0{,}5} \cdot \frac{0{,}5^0}{0!} \doteq 0{,}6065$$

b) $\lambda = 0{,}5\,;\ x = 3$

$$P(X < 3) = P(0) + P(1) + P(2) = e^{-0{,}5} \cdot \frac{0{,}5^0}{0!} + e^{-0{,}5} \cdot \frac{0{,}5^1}{1!} +$$

$$e^{-0{,}5} \cdot \frac{0{,}5^2}{2!} == 0{,}6065307 + 0{,}3032653 + 0{,}0758163 \doteq 0{,}9856$$

c) $\lambda = 0{,}5\,;\ x = 2$

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - [P(0) + P(1)] = 1 - (0{,}6065307 + 0{,}3032653) \doteq 0{,}0902$$

## Interpretace:

a) Pravděpodobnost, že náhodně vybraný balík bude vzkazu, je 60,65 %. Tato pravděpodobnost je středně velká.

b) Pravděpodobnost, že náhodně vybraný balík bude mít méně než tři kazy, je 98,56 %. Tato pravděpodobnost je značně velká.

c) Pravděpodobnost, že náhodně vybraný balík bude mít nejméně dva kazy, je 9,02 %. Tato pravděpodobnost je malá.

## STATGRAPHICS:

Desribe – Distribution Fitting – Probability Distributions

V panelu Probability Distributions zaškrtneme Poisson.

V panelu Poisson Options zadáme parametr rozdělení.

V nabídce Tables and Graphs je třeba zaškrtnout položku Cumulative Distributions.

V proceduře Cumulative Distributions zvolíme nabídku Pane Options, ve které zadáme požadované hodnoty NV, jejichž pravděpodobnosti budeme počítat, tedy 0, 2, a 3.

**Cumulative Distribution**
Distribution: Poisson

Lower Tail Area (<)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0 | 0,0 | | | | |
| 3 | **0,985612** | | | | |
| 2 | 0,909796 | | | | |

Probability Mass (=)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0 | **0,606531** | | | | |
| 3 | 0,0126361 | | | | |
| 2 | **0,0758164** | | | | |

Upper Tail Area (>)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0 | 0,393469 | | | | |
| 3 | 0,00175154 | | | | |
| 2 | **0,0143876** | | | | |

a) $P(X=0)=0{,}606531$

b) $P(X<3)=0{,}985612$

c) $P(X\geq 2)=P(X=2)+P(X>2)=0{,}0758164+0{,}0143876\doteq 0{,}0902$

## EXCEL:

Vzorce – Další funkce – Statistická

Zvolíme funkci POISSON.

V panelu Argumenty funkce zadáme do jednotlivých řádků:

***X:*** hodnotu NV $X$

***Střední:*** střední hodnota $\lambda$

***Součet:*** PRAVDA pro distribuční funkci

a) POISSON(0; 0,5; NEPRAVDA)
$P(X = 0) = 0{,}606531$

b) POISSON(2; 0,5; PRAVDA)
$P(X < 3) = P(X \leq 2) = 0{,}985612$

c) POISSON(1; 0,5; PRAVDA)
$P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0{,}909796 = 0{,}090204$

# Příklad 1

## Zadání příkladu:

Předpokládejme, že maximální hmotnost nákupu, kterou unese určitý typ plastikových nákupních tašek, je náhodná veličina s normálním rozdělením. Střední hodnota této náhodné veličiny je 5 kg a směrodatná odchylka 1 kg. Stanovte pravděpodobnost, že se nákupní taška protrhne při hmotnosti nákupu menší než 4,5 kg.

## Vypracování příkladu:

Náhodná veličina $X$… maximální hmotnost nákupu, kterou unese plastiková taška.

Normální rozdělení $N(\mu;\sigma^2)$ má dva parametry:

$\mu = E(X) = 5$

$\sigma^2 = D(X) = 1$

Při výpočtu je třeba provést převod na normovanou náhodnou veličinu:

$$U = \frac{x-\mu}{\sigma};\ F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \Phi(u)$$

$\mu = 5; \sigma = 1; x = 4{,}5$

$$P(X < 4{,}5) = P(X \le 4{,}5) = F(4{,}5) = \Phi\left(\frac{4{,}5-5}{1}\right) = \Phi(-0{,}5) = 1-\Phi(0{,}5) = 1-0{,}6915 = 0{,}3085$$

Hodnotu distribuční funkce $\Phi(0{,}5) = 0{,}3085$ vyhledáme ve statistických tabulkách.

## Interpretace:

Pravděpodobnost, že se nákupní taška protrhne při hmotnosti nákupu menší než 4,5 kg, je 30,85 %.

## STATGRAPHICS:

Desribe – Distribution Fitting – Probability Distributions

V panelu Probability Distributions zaškrtneme Normal.

V panelu Normal Options zadáme oba parametry rozdělení.

V nabídce Tables and Graphs je třeba zaškrtnout položku Cumulative Distributions.

V proceduře Cumulative Distributions zvolíme nabídku Pane Options, ve které zadáme požadovanou hodnotu NV, tedy 4,5.

**Cumulative Distribution**
Distribution: Normal

Lower Tail Area (<)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 4,5 | **0,308536** | | | | |

Probability Density

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 4,5 | 0,352065 | | | | |

Upper Tail Area (>)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 4,5 | 0,691464 | | | | |

$P(X < 4{,}5) = 0{,}308536$

## EXCEL:

Vzorce – Další funkce – Statistická

Zvolíme funkci NORMDIST.

V panelu Argumenty funkce zadáme do jednotlivých řádků:

***X:*** hodnotu NV $X$

***Střed_hodn:*** střední hodnota $\mu$

***SM_odch:***směrodatná odchylka $\sigma$

***Součet:*** PRAVDA pro distribuční funkci

NORMDIST(4,5; 5; 1; PRAVDA)

$P(X < 4{,}5) = 0{,}308537539$

# Příklad 2

## Zadání příkladu:

Doba čekání zákazníka na obsluhu v určitém typu prodejny je náhodná veličina, která se řídí exponenciálním rozdělením. Střední hodnota této náhodné veličiny je 50 sekund. Stanovte pravděpodobnost, že náhodně vybraný zákazník bude obsloužen:

a) nejdéle za 30 sekund
b) nejdříve za 20 sekund.

## Vypracování příkladu:

Náhodná veličina $X$… doba čekání zákazníka na obsluhu.

Exponenciální rozdělení $E(A;\delta)$ má dva parametry:

$E(X) = A + \delta = 50$

$A = 0$ (zákazník může být obsloužen ihned po příchodu)

$\delta = 50$

Distribuční funkce:

$F(x) = 1 - e^{-(x-A)/\delta}$

a) $A = 0; \delta = 50; x = 30$

$P(X \le 30) = F(30) = 1 - e^{-30/50} = 1 - 0{,}5488116 \doteq 0{,}4512$

b) $A = 0; \delta = 50; x = 20$

$P(X \ge 20) = 1 - F(20) = 1 - \left(1 - e^{-20/50}\right) = 1 - 0{,}32968 \doteq 0{,}6703$

## Interpretace:

a) Pravděpodobnost, že náhodně vybraný zákazník bude obsloužen nejdéle za 30 sekund, je 45,12 %.

b) Pravděpodobnost, že náhodně vybraný zákazník bude obsloužen nejdříve za 20 sekund, je 67,03 %.

## STATGRAPHICS:

Desribe – Distribution Fitting – Probability Distributions

V panelu Probability Distributions zaškrtneme Exponential.

V panelu Exponential Options zadáme parametr $\delta$, parametr $A$ je automaticky roven 0.

V nabídce Tables and Graphs je třeba zaškrtnout položku Cumulative Distributions.

V proceduře Cumulative Distributions zvolíme nabídku Pane Options, ve které zadáme požadované hodnoty NV, tedy 20 a 30.

**Cumulative Distribution**
Distribution: Exponential

Lower Tail Area (<)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 20 | 0,32968 | | | | |
| 30 | **0,451188** | | | | |

Probability Density

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 20 | 0,0134064 | | | | |
| 30 | 0,0109762 | | | | |

Upper Tail Area (>)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 20 | **0,67032** | | | | |
| 30 | 0,548812 | | | | |

a) $P(X \le 30) = P(X < 30) = 0{,}451188$

b) $P(X \ge 20) = P(X > 20) = 0{,}67032$

## EXCEL:

Vzorce – Další funkce – Statistická

Zvolíme funkci EXPONDIST.

V panelu Argumenty funkce zadáme do jednotlivých řádků:

***X:*** hodnotu NV $X$

***Lambda:*** převrácená hodnota parametru $\delta$

***Součet:*** PRAVDA pro distribuční funkci

NEPRAVDA pro hustotu pravděpodobnosti.

a) EXPONDIST(30; 0,02; PRAVDA)

$P(X \le 30) = P(X < 30) = 0{,}451188$

b) EXPONDIST(20; 0,02; PRAVDA)

$P(X \ge 20) = 1 - P(X \le 20) = 1 - 0{,}32968 = 0{,}67032$

# Příklad 1

## Zadání příkladu:

U stovky náhodně vybraných výrobků z produkce určitého podniku byla zjišťována spotřeba materiálu na jeden výrobek. Z výběrových dat byla vypočtena průměrná spotřeba materiálu 150 g a rozptyl spotřeby materiálu 16 $g^2$ .V jakých mezích lze se spolehlivostí 95 % očekávat průměrnou spotřebu materiálu na jeden výrobek? Spotřeba materiálu na jeden výrobek je náhodná veličina, která se řídí normálním rozdělením.

## Vypracování příkladu:

$n = 100\,;\, 1-\alpha = 0{,}95$

$\bar{x} = 150$

$s_x^2 = 16\,; s_x = 4$

Budeme konstruovat interval spolehlivosti pro střední hodnotu normálně rozdělené náhodné veličiny. Protože se jedná o velký výběr z normálního rozdělení s neznámým rozptylem $\sigma_x^2$, použijeme pro výpočet oboustranného intervalu spolehlivosti následující vzorec:

$$P\left(\bar{x} - u_{1-\frac{\alpha}{2}} \cdot \frac{s_x}{\sqrt{n}} < \mu < \bar{x} + u_{1-\frac{\alpha}{2}} \cdot \frac{s_x}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(150 - u_{0{,}025} \cdot \frac{4}{\sqrt{100}} < \mu < 150 + u_{0{,}975} \cdot \frac{4}{\sqrt{100}}\right) = 0{,}95$$

$$P\left(150 - 1{,}96 \cdot \frac{4}{\sqrt{100}} < \mu < 150 + 1{,}96 \cdot \frac{4}{\sqrt{100}}\right) = 0{,}95$$

$$P(149{,}216 < \mu < 150{,}784) = 0{,}95$$

Kvantily normovaného normálního rozdělení vyhledáme ve statistických tabulkách.

Se spolehlivostí 95 % je možno průměrnou spotřebu materiálu na jeden výrobek očekávat v intervalu 149,216 g až 150,784 g.

## STATGRAPHICS:

Desribe – Numeric Data – Hypothesis Tests

V panelu Hypothesis Tests zaškrtneme parametr Normal Mean.

V panelu Hypothesis Tests dále zadáme:

Sample Mean: 150

Sample Sigma: 4

Sample Size: 100

Po potvrzení se objeví panel Hypothesis Tests Options, ve kterém zadáme:

Alernative Hypothesis: Not Equal

Alpha: 5,0

**Hypothesis Tests**
Sample mean = 150,0
Sample standard deviation = 4,0
Sample size = 100

95,0% confidence interval for mean: 150,0 +/- 0,793688 **[149,206;150,794]**

Null Hypothesis: mean = 0,5
Alternative: not equal
Computed t statistic = 373,75
P-Value = 0,0
Reject the null hypothesis for alpha = 0,05.

$$P\left(\bar{x} - u_{1-\frac{\alpha}{2}} \cdot \frac{s_x}{\sqrt{n}} < \mu < \bar{x} + u_{1-\frac{\alpha}{2}} \cdot \frac{s_x}{\sqrt{n}}\right) = 1 - \alpha$$

$$P(149{,}206 < \mu < 150{,}794) = 0{,}95$$

Se spolehlivostí 95 % je možno průměrnou spotřebu materiálu na jeden výrobek očekávat v intervalu 149,206 g až 150,794 g.

*Pozn.:* Nepatrná odlišnost výsledků oproti ručnímu výpočtu je způsobena skutečností, že příslušné kvantily jsou ve STATGRAPHICSu stanoveny na více desetinných míst než v tabulkách.

## EXCEL:

Pro výpočet intervalu spolehlivosti z výběrových charakteristik neexistuje v Excelu žádná speciální procedura. Je třeba postupovat podle vzorce jako u ručního výpočtu. Příslušný kvantil lze stanovit následujícím způsobem:

Vzorce – Další funkce – Statistická

Zvolíme funkci NORMSINV.

V panelu Argumenty funkce zadáme řádku:

***Prst:*** pravděpodobnost pro hledaný kvantil (např. 0,975 atd.)

# Příklad 2

## Zadání příkladu:

V rámci rozsáhlého průzkumu v oblasti ekologie bylo mimo jiné zjišťováno, jaký podíl dospělé populace v jistém kraji třídí odpad. Tisíc náhodně vybraných dospělých osob bylo dotázáno, zda třídí odpad či nikoli. Z celkového počtu dotázaných uvedlo 386, že odpad třídí. Na základě výše uvedených údajů stanovte se spolehlivostí 95 % minimální podíl dospělých osob, které v daném kraji odpad třídí.

## Vypracování příkladu:

$n = 1\,000;\ 1-\alpha = 0{,}95$

$$p = \frac{386}{1\,000} = 0{,}386$$

Budeme konstruovat jednostranný interval spolehlivosti pro relativní četnost $\pi$. Protože potřebujeme stanovit dolní mez, jedná se o levostranný interval spolehlivosti. Použijeme následující vzorec:

$$P\left[\pi > p - u_{1-\alpha} \cdot \sqrt{\frac{p(1-p)}{n}}\right] = 1-\alpha$$

$$P\left[\pi > 0{,}386 - u_{0{,}95} \cdot \sqrt{\frac{0{,}386(1-0{,}386)}{1\,000}}\right] = 0{,}95$$

$$P\left(\pi > 0{,}386 - 1{,}645 \cdot \sqrt{\frac{0{,}386 \cdot 0{,}614}{1\,000}}\right) = 0{,}95$$

$$P(\pi > 0{,}3606753) = 0{,}95$$

Se spolehlivostí 95 % je minimální podíl dospělých osob, které v daném kraji třídí odpad, 36,07 %.

## STATGRAPHICS:

Desribe – Numeric Data – Hypothesis Tests

V panelu Hypothesis Tests zaškrtneme parametr Binomial Proportion.

V panelu Hypothesis Tests dále zadáme:

Sample Proportion: 0,386

Sample Size: 1 000

Po potvrzení se objeví panel Hypothesis Tests Options, ve kterém zadáme:

Alernative Hypothesis: Greater Than

Alpha: 5,0

**Hypothesis Tests**
Sample proportion = 0,386

Sample size = 1000

Approximate 95,0% lower confidence bound for p: **[0,360451]**

Null Hypothesis: proportion = 0,5
Alternative: greater than
P-Value = 1,0
Do not reject the null hypothesis for alpha = 0,05.

$$P\left[\pi > p - u_{1-\alpha} \cdot \sqrt{\frac{p(1-p)}{n}}\right] = 1 - \alpha$$

$$P(\pi > 0{,}360451) = 0{,}95$$

Se spolehlivostí 95 % je minimální podíl dospělých osob, které v daném kraji třídí odpad, 36,05 %.

*Pozn.:* Nepatrná odlišnost výsledků oproti ručnímu výpočtu je způsobena skutečností, že STATGRAPHICS používá pro výpočet upravený vzorec.

## EXCEL:

Pro výpočet intervalu spolehlivosti z výběrových charakteristik neexistuje v Excelu žádná speciální procedura. Je třeba postupovat podle vzorce jako u ručního výpočtu. Příslušný kvantil lze stanovit následujícím způsobem:

Vzorce – Další funkce – Statistická

Zvolíme funkci NORMSINV.

V panelu Argumenty funkce zadáme řádku:

***Prst:*** pravděpodobnost pro hledaný kvantil (např. 0,95 atd.)

# Příklad 3

## Zadání příkladu:

Výrobce balí třtinový cukr do balíčků po 500 gramech. Deklaruje přitom, že směrodatná odchylka hmotnosti cukru v balíčku nepřekročí 5 gramů. V rámci kontroly jakosti bylo náhodně vybráno 30 balíčků třtinového cukru od tohoto výrobce a z výsledků byly stanoveny následující výběrové charakteristiky: $\bar{x} = 499$ g, $s_x = 3{,}8$ g. Stanovte mez, o které lze s 95% spolehlivostí prohlásit, že ji směrodatná odchylka hmotnosti cukru v balíčku nepřekročí. Hmotnost cukru v balíčku je náhodná veličina s normálním rozdělením.

## Vypracování příkladu:

$n = 30;\ 1-\alpha = 0{,}95$

$s_x = 3{,}8$

Budeme konstruovat interval spolehlivosti pro rozptyl normálně rozdělené náhodné veličiny. Protože potřebujeme stanovit horní mez, jedná se o pravostranný interval spolehlivosti. Použijeme následující vzorec:

$$P\left[\sigma^2 < \frac{(n-1)s_x^2}{\chi_\alpha^2(n-1)}\right] = 1-\alpha$$

$$P\left[\sigma^2 < \frac{(30-1)3{,}8^2}{\chi_{0{,}05}^2(29)}\right] = 0{,}95$$

$$P\left(\sigma^2 < \frac{29\cdot 3{,}8^2}{17{,}7}\right) = 0{,}95$$

$$P\left(\sigma^2 < 23{,}658757\right) = 0{,}95$$

$$P\left(\sigma < 4{,}8640268\right) = 0{,}95$$

Kvantil rozdělení $\chi^2$ vyhledáme ve statistických tabulkách.

S 95% spolehlivostí lze prohlásit, že směrodatná odchylka hmotnosti cukru v balíčku nepřekročí 4,86 g.

## STATGRAPHICS:

Desribe – Numeric Data – Hypothesis Tests

V panelu Hypothesis Tests zaškrtneme parametr Normal Sigma.

V panelu Hypothesis Tests dále zadáme:

Sample Sigma: 3,8

Sample Size: 30

Po potvrzení se objeví panel Hypothesis Tests Options, ve kterém zadáme:

Alernative Hypothesis: Less Than

Alpha: 5,0

**Hypothesis Tests**
Sample standard deviation = 3,8
Sample size = 30

95,0% upper confidence bound for sigma: **[4,86288]**

Null Hypothesis: standard deviation = 0,5
Alternative: less than
Computed chi-square statistic = 1675,04
P-Value = 1,0
Do not reject the null hypothesis for alpha = 0,05.

$$P\left[\sigma^2 < \frac{(n-1)s_x^2}{\chi_\alpha^2(n-1)}\right] = 1-\alpha$$

$$P\left(\sigma^2 < 23{,}647602\right) = 0{,}95$$

$$P\left(\sigma < 4{,}86288\right) = 0{,}95$$

S 95% spolehlivostí lze prohlásit, že směrodatná odchylka hmotnosti cukru v balíčku nepřekročí 4,86 g.

*Pozn.:* Nepatrná odlišnost výsledků oproti ručnímu výpočtu je způsobena skutečností, že příslušný kvantil je ve STATGRAPHICSu stanoven na více desetinných míst než v tabulkách.

## EXCEL:

Pro výpočet intervalu spolehlivosti z výběrových charakteristik neexistuje v Excelu žádná speciální procedura. Je třeba postupovat podle vzorce jako u ručního výpočtu. Příslušný kvantil lze stanovit následujícím způsobem:

Vzorce – Další funkce – Statistická

Zvolíme funkci CHINV.

V panelu Argumenty funkce zadáme jednotlivých řádků:

***Prst:*** pravděpodobnost pro hledaný kvantil (např. 0,05 atd.)

***Volnost:*** počet stupňů volnosti (např. 29)

# Příklad 1

## Zadání příkladu:

Předpokládá se, že podíl kuřáků v dospělé populaci ČR je 35 %. V rámci statistického průzkumu byl proveden náhodný výběr 200 dospělých osob, z nichž 76 uvedlo, že jsou kuřáci. Na 5% hladině významnosti rozhodněte, zda na základě těchto výsledků lze konstatovat, že podíl kuřáků v dospělé populaci ČR přesahuje 35 %.

## Vypracování příkladu:

$n = 200\ ; \pi = 0{,}35 ; \alpha = 0{,}05$

$$p = \frac{76}{200} = 0{,}38$$

**1.** $H_0 : \pi = 0{,}35$

$H_1 : \pi > 0{,}35$

**2.** $$U = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0{,}38 - 0{,}35}{\sqrt{\frac{0{,}35 \cdot 0{,}65}{200}}} \doteq 0{,}89$$

**3.** $W \equiv \{U; U \geq u_{1-\alpha}\}$

$W \equiv \{U; U \geq u_{0{,}95}\}$

$W \equiv \{U; U \geq 1{,}645\}$

Kvantil normovaného normálního rozdělení vyhledáme ve statistických tabulkách.

**4.** Závěr testu:

Testové kritérium neleží v kritickém oboru, proto nezamítáme $H_0$ a nepřijímáme $H_1$. Na 5% hladině významnosti jsme neprokázali, že podíl kuřáků v dospělé populaci ČR přesahuje 35 %.

## STATGRAPHICS:

Desribe – Numeric Data – Hypothesis Tests

V panelu Hypothesis Tests zaškrtneme parametr Binomial Proportion.

V panelu Hypothesis Tests dále zadáme:

Null Hypothesis: 0,35

Sample Proportion: 0,38

Sample Size: 200

Po potvrzení se objeví panel Hypothesis Tests Options, ve kterém zadáme:

Alernative Hypothesis: Greater Than

**Hypothesis Tests**
Sample proportion = 0,38
Sample size = 200

Approximate 95,0% lower confidence bound for p: [0,322669]

Null Hypothesis: proportion = 0,35
Alternative: greater than
P-Value = **0,207428**
Do not reject the null hypothesis for alpha = 0,05.

**2.** $H_0 : \pi = 0{,}35$
$H_1 : \pi > 0{,}35$

**2.** $U = \dfrac{p - \pi_0}{\sqrt{\dfrac{\pi_0(1-\pi_0)}{n}}} = \dfrac{0{,}38 - 0{,}35}{\sqrt{\dfrac{0{,}35 \cdot 0{,}65}{200}}} \doteq 0{,}89$

**3.** $P - Value = 0{,}207428 > 0{,}05$

**4.** Závěr testu:

Nejnižší hladina významnosti pro zamítnutí $H_0$ je 0,207428, proto na 5% hladině významnosti nezamítáme $H_0$ a nepřijímáme $H_1$. Na 5% hladině významnosti jsme neprokázali, že podíl kuřáků v dospělé populaci ČR přesahuje než 35 %.

## EXCEL:

Pro provedení tohoto typu testu nemá Excel speciální proceduru, která by stanovila P –Value, jako je tomu ve STATGRAPHICSu. Je třeba postupovat jako u ručního výpočtu. Příslušný kvantil lze stanovit následujícím způsobem:

Vzorce – Další funkce – Statistická

Zvolíme funkci NORMSINV.

V panelu Argumenty funkce zadáme řádku:

***Prst:*** pravděpodobnost pro hledaný kvantil (např. 0,975 atd.)

# Příklad 2

## Zadání příkladu:

Předpokládá se, že průměrná pevnost proužků netkané textilie je 88 jednotek. U dvaceti náhodně vybraných proužků netkané textilie byla změřena její pevnost a byly spočteny výběrové charakteristiky. Výběrový průměr činil 85,1 jednotek, výběrová směrodatná odchylka byla 3,25 jednotek. Na 5% hladině významnosti testujte hypotézu, že průměrná pevnost proužků netkané textilie je 88 jednotek. Předpokládejte normalitu rozdělení náhodné veličiny, kterou je pevnost proužků netkané textilie.

## Vypracování příkladu:

$n = 20; \mu = 88; \alpha = 0{,}05$

$\bar{x} = 85{,}1$

$s_x = 3{,}25$

**1.** $H_0 : \mu = 88$
$H_1 : \mu \neq 88$

**2.** $t = \dfrac{\bar{x} - \mu_0}{\dfrac{s_x}{\sqrt{n}}} = \dfrac{85{,}1 - 88}{\dfrac{3{,}25}{\sqrt{20}}} \doteq -3{,}99$

**3.** $W \equiv \left\{t; t \leq t_{\frac{\alpha}{2}}(n-1) \cup t \geq t_{1-\frac{\alpha}{2}}(n-1)\right\}$

$W \equiv \left\{t; t \leq t_{0{,}025}(19) \cup t \geq t_{0{,}975}(19)\right\}$

$W \equiv \left\{t; t \leq -2{,}093 \cup t \geq 2{,}093\right\}$

Vzhledem k malému rozsahu výběru použijeme při konstrukci kritického oboru kvantily Studentova rozdělení, které vyhledáme ve statistických tabulkách.

**4.** Závěr testu:

Testové kritérium leží v kritickém oboru, proto zamítáme $H_0$ a přijímáme $H_1$. Na 5% hladině významnosti jsme prokázali, že průměrná pevnost proužků netkané textilie není 88 jednotek.

## STATGRAPHICS:

Desribe – Numeric Data – Hypothesis Tests

V panelu Hypothesis Tests zaškrtneme parametr Normal Mean.

V panelu Hypothesis Tests dále zadáme:

Null Hypothesis: 88

Sample Mean: 85,1

Sample Sigma: 3,25

Sample Size: 20

Po potvrzení se objeví panel Hypothesis Tests Options, ve kterém zadáme:

Alernative Hypothesis: Not Equal

**Hypothesis Tests**
Sample mean = 85,1
Sample standard deviation = 3,25
Sample size = 20

95,0% confidence interval for mean: 85,1 +/- 1,52105 [83,5789;86,6211]

Null Hypothesis: mean = 88,0
Alternative: not equal
Computed t statistic = **-3,99052**
P-Value = **0,000783276**
Reject the null hypothesis for alpha = 0,05.

**1.** $H_0 : \mu = 88$
$H_1 : \mu \neq 88$

**2.** $t = \dfrac{\bar{x} - \mu_0}{\dfrac{s_x}{\sqrt{n}}} = \dfrac{85{,}1 - 88}{\dfrac{3{,}25}{\sqrt{20}}} = -3{,}99052$

**3.** $P - value = 0{,}000783276 < 0{,}05$

**4.** Závěr testu:

Nejnižší hladina významnosti pro zamítnutí $H_0$ je 0,000783276, proto na 5% hladině významnosti zamítáme $H_0$ a přijímáme $H_1$. Na 5% hladině významnosti jsme prokázali, že průměrná pevnost proužků netkané textilie není 88 jednotek.

## EXCEL:

Pro provedení tohoto typu testu nemá Excel speciální proceduru, která by stanovila P –Value, jako je tomu ve STATGRAPHICSu. Je třeba postupovat jako u ručního výpočtu.

# Příklad 1

## Zadání příkladu:

Výsledky kontroly 100 náhodně vybraných DVD lze popsat rozdělením četností počtu vad viz tabulka.

| Počet vad | $n_i$ |
|---|---|
| 0 | 65 |
| 1 | 20 |
| 2 | 10 |
| 3+ | 5 |
| Celkem | 100 |

N 5% hladině významnosti rozhodněte, zda lze počet vad na DVD považovat za náhodnou veličinu, která se řídí Poissonovým rozdělením s parametrem $\lambda = 0{,}55$.

## Vypracování příkladu:

| Počet vad | $n_i$ Empirická četnost | $\pi_{0,i}$ Teoretická četnost | $n\pi_{0,i}$ Očekávaná četnost | $\frac{(n_i - n\pi_{0,i})^2}{n\pi_{0,i}}$ |
|---|---|---|---|---|
| 0 | 65 | 0,57695 | 57,695 | 0,92492 |
| 1 | 20 | 0,31732 | 31,732 | 4,33757 |
| 2 | 10 | 0,08726 | 8,726 | 0,18600 |
| 3 | 5 | 0,01847 | 1,847 | 5,38246 |
| Celkem | 100 | 1,00000 | 100,00 | 10,83095 |

Teoretická četnost = pravděpodobnostní funkce $Po(0{,}55)$.

Hodnoty pro parametr $\lambda = 0{,}55$ nelze nalézt v běžných statistických tabulkách, proto je třeba každou hodnotu vypočítat dosazením do pravděpodobnostní funkce Poissonova rozdělení.

$$P(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$$

$$\lambda = 0{,}55; x = 0 \quad P(0) = e^{-0{,}55} \cdot \frac{\lambda^0}{0!} \doteq 0{,}57695$$

$$\lambda = 0{,}55; x = 1 \quad P(1) = e^{-0{,}55} \cdot \frac{\lambda^1}{1!} \doteq 0{,}31732$$

$$\lambda = 0{,}55; x = 0 \quad P(2) = e^{-0{,}55} \cdot \frac{\lambda^2}{2!} \doteq 0{,}08726$$

$$\lambda = 0{,}55; x = 0 \quad P(3) = e^{-0{,}55} \cdot \frac{\lambda^3}{3!} \doteq 0{,}01847$$

**1.** $H_0 : vyhovuje \ Po(0{,}55)$
$H_1 : non \ H_0$

**2.** $G = \sum_{i=1}^{k} \frac{(n_i - n\pi_{0,i})^2}{n\pi_{0,i}} = 10{,}83095$

**3.** $W \equiv \{G; G \geq \chi^2_{1-\alpha}(k-1)\}$

$W \equiv \{G; G \geq \chi^2_{0,95}(3)\}$

$W \equiv \{G; G \geq 7{,}81\}$

Kvantil rozdělení $\chi^2$ vyhledáme ve statistických tabulkách.

**4.** Závěr testu:

Testové kritérium leží v kritickém oboru, proto zamítáme $H_0$ a přijímáme $H_1$. Na 5% hladině významnosti jsme prokázali, že počet vad na DVD nelze považovat za náhodnou veličinu, která se řídí Poissonovým rozdělením s parametrem $\lambda = 0{,}55$.

## STATGRAPHICS:

Desribe – Distribution Fitting – Fitting Uncensored Data

V panelu Distribution Fitting Options zaškrtneme parametr Poisson.

V panelu Tables and Graphs zaškrtneme Goodness-of-Fit Tests.

Pomocí nabídky Pane Options vybereme položku Chi-Square Test.

**Goodness-of-Fit Tests for Pocet_vad**
Chi-Square Test

| | *Lower* | *Upper* | *Observed* | *Expected* | |
|---|---|---|---|---|---|
| | *Limit* | *Limit* | *Frequency* | *Frequency* | *Chi-Square* |
| at or below | | 0,0 | 65 | 57,69 | 0,92 |
| | 1,0 | 1,0 | 20 | 31,73 | 4,34 |
| | 2,0 | | 15 | 10,57 | 1,85 |

Chi-Square = 7,11648 with 1 d.f. P-Value = **0,00763611**

**1.** $H_0 : vyhovuje\ \ Po(0{,}55)$

$H_1 : non\ \ H_0$

**2.** $G = \sum_{i=1}^{k} \frac{(n_i - n\pi_{0,i})^2}{n\pi_{0,i}} = 7{,}11648$

**3.** $P - Value = 0{,}00763611 < 0{,}05$

**4.** Závěr testu:

Nejnižší hladina významnosti pro zamítnutí $H_0$ je 0,00763611, proto na 5% hladině významnosti zamítáme $H_0$ a přijímáme $H_1$. Na 5% hladině významnosti jsme prokázali, že počet vad na DVD nelze považovat za náhodnou veličinu, která se řídí Poissonovým rozdělením s parametrem $\lambda = 0{,}55$.

## MS EXCEL:

Do jednoho sloupce zadáme empirické četnosti, do vedlejšího sloupce očekávané četnosti, vypočtené pro Poissonovo rozdělení s parametrem $\lambda = 0{,}55$. V tomto případě byly navíc sloučeny poslední dvě třídy z důvodu nedostatečného obsazení poslední třídy.

Vzorce – Další funkce – Statistická

Zvolíme funkci CHITEST.

V panelu Argumenty funkce zadáme do jednotlivých řádků:

***Aktuální:*** empirické četnosti

***Očekávané:*** teoretické (očekávané) četnosti

Výsledkem testu je hodnota $P - Value = 0{,}028494 < 0{,}05$. Na 5% hladině významnosti proto zamítáme $H_0$ a přijímáme $H_1$. Prokázali jsme, že počet vad na DVD nelze považovat za náhodnou veličinu, která se řídí Poissonovým rozdělením s parametrem $\lambda = 0{,}55$.

# *Test shody rozptylů dvou normálních rozdělení*

## 1. Formulace hypotéz

$$H_0 : \sigma_1^2 = \sigma_2^2$$

*a*) $H_1 : \sigma_1^2 \neq \sigma_2^2$
*b*) $H_1 : \sigma_1^2 > \sigma_2^2$
*c*) $H_1 : \sigma_1^2 < \sigma_2^2$

## 2. Volba testového kritéria

$$F = \frac{s_1'^2}{s_2'^2} \quad \approx F(n_1 - 1; n_2 - 1)$$

## 3. Stanovení kritického oboru

*a*) $W \equiv \left\{ F; F \leq F_{\frac{\alpha}{2}}(n_1 - 1; n_2 - 1) \cup F \geq F_{1-\frac{\alpha}{2}}(n_1 - 1; n_2 - 1) \right\}$
*b*) $W \equiv \{F; F \geq F_{1-\alpha}(n_1 - 1; n_2 - 1)\}$
*c*) $W \equiv \{F; F \leq F_{\alpha}(n_1 - 1; n_2 - 1)\}$

# Metody zkoumání závislostí

- zkoumání závislosti dvou event. více proměnných, měření síly této závislosti, atd.
- cílem je hlubší vniknutí do podstaty sledovaných jevů a procesů, přiblížení k tzv. příčinným souvislostem.

## Dvourozměrná tabulka rozdělení četností

- je elementární formou popisu závislosti
- rozlišujeme různé typy tabulek.

**Korelační tabulka:** obě proměnné jsou numerické.

**Kontingenční tabulka:** alespoň jedna proměnná je slovní.

**Asociační tabulka:** obě proměnné jsou alternativní.

**Čtyřpolní tabulka:** obě proměnné nabývají pouze dvou obměn.

***Dvourozměrná tabulka rozdělení četností***

| $y_j$ / $x_i$ | $y_1$ | $y_2$ | . . . | $y_s$ | Součty četností $n_{i\bullet}$ |
|---|---|---|---|---|---|
| $x_1$ | $n_{11}$ | $n_{12}$ | . . . | $n_{1s}$ | $n_{1\bullet}$ |
| $x_2$ | $n_{21}$ | $n_{22}$ | . . . | $n_{2s}$ | $n_{2\bullet}$ |
| . . . | . . . | . . . | . . . | . . . | . . . |
| $x_r$ | $n_{r1}$ | $n_{r2}$ | . . . | $n_{rs}$ | $n_{r\bullet}$ |
| Součty četností $n_{\bullet j}$ | $n_{\bullet 1}$ | $n_{\bullet 2}$ | . . . | $n_{\bullet s}$ | n |

***<u>Symbolika:</u>***

$n_{ij}$ ..... sdružené (simultánní) absolutní četnosti

$n_{i\bullet}$, $n_{\bullet j}$ ..... okrajové (marginální) absolutní četnosti

$p_{ij}$ ..... sdružené relativní četnosti

$p_{i\bullet}$, $p_{\bullet j}$ ..... marginální relativní četnosti

$$n_{i\bullet} = \sum_{j=1}^{s} n_{ij} \; ; \quad n_{\bullet j} = \sum_{i=1}^{r} n_{ij}$$

$$\sum_{i=1}^{r} n_{i\bullet} = \sum_{j=1}^{s} n_{\bullet j} = \sum_{i=1}^{r} \sum_{j=1}^{s} n_{ij} = n$$

$$p_{i\bullet} = \frac{n_{i\bullet}}{n} \; ; \quad p_{\bullet j} = \frac{n_{\bullet j}}{n} \; ; \quad p_{ij} = \frac{n_{ij}}{n}$$

## Podmíněné rozdělení četností:

Rozdělení četností jedné proměnné, které odpovídá určité obměně druhé proměnné (tj. za podmínky, že druhá proměnná nabyla určité obměny).

***Podmíněné relativní četnosti:*** $p_{j/i} = \frac{n_{ij}}{n_{i\bullet}} \quad ; \quad p_{i/j} = \frac{n_{ij}}{n_{\bullet j}}$

***Podmíněný průměr:*** $\bar{y}_i = \frac{\sum_{j=1}^{s} y_j n_{ij}}{n_{i\bullet}} = \frac{\sum_{j=1}^{s} y_{ij}}{n_{i\bullet}}$

***Podmíněný rozptyl:*** $s_{yi}^2 = \frac{\sum_{j=1}^{s} (y_j - \bar{y}_i)^2 n_{ij}}{n_{i\bullet}} = \frac{\sum_{j=1}^{s} (y_{ij} - \bar{y}_i)^2}{n_{i\bullet}}$

# Grafické znázornění dvourozměrného rozdělení četností

- je další formou popisu závislosti
- různé typy grafů

- ***čára podmíněných průměrů***
- ***čára podmíněných rozptylů***
- ***bodový graf (diagram)***.

# Analýza rozptylu

- jednofaktorová,
- vícefaktorová.

## *Jednofaktorová analýza rozptylu*

- Zkoumá, zda lze změny hodnot měřitelné proměnné *y* vysvětlovat faktorem *x*, který může být slovní i číselná proměnná.
- **Použití :** AR slouží k ověření významnosti rozdílu mezi výběrovými průměry většího počtu náhodných výběrů.

- **Předpoklady AR :**

  **I.** Náhodné výběry jsou nezávislé.
  **II.** Každý z výběrů má normální rozdělení s neznámou střední hodnotou $\mu_1, \mu_2, ..., \mu_k$ a s neznámým rozptylem $\sigma_1^2, \sigma_2^2, ..., \sigma_k^2$.
  **III.** Rozptyly všech skupin jsou stejné**,** tj. $\sigma_1^2 = \sigma_2^2 = ..... = \sigma_k^2$.
  **IV.** Počet pozorování musí být větší než počet skupin, tj. $n > k$.
  **V.** Sledovaný statistický znak *y* je číselný, faktor *x* může být proměnná číselná i slovní.

- **Hlavní myšlenka AR :** Celkový rozptyl hodnot proměnné *y* ( $s_y^2$ ) rozložíme na 2 části :

  1. *meziskupinový rozptyl* $s_{ym}^2$ *(rozptyl podmíněných průměrů)*

  - Odráží variabilitu mezi skupinami.
  - Jde o rozptyl, který se vztahuje k vlivu, podle něhož bylo provedeno třídění hodnot *y*.
  - Meziskupinová variabilita je vysvětlitelná daným faktorem *x*.

  2. *vnitroskupinový (reziduální) rozptyl* $s_{yv}^2$ *(průměr podmíněných rozptylů)*

  - Odráží variabilitu uvnitř skupin.
  - Kolísání je důsledkem závislosti *y* na jiných faktorech než na *x*.
  - Vnitroskupinová variabilita není vysvětlitelná daným faktorem.

- Platí tedy : $s_y^2 = s_{ym}^2 + s_{yv}^2$,

kde $s_y^2 = \dfrac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}\right)^2}{n} = \dfrac{S_y}{n}$, kde

$y_{ij}$ ..... jednotlivé hodnoty sledované proměnné,
$\bar{y}$ ..... celkový průměr,
n ..... rozsah výběru,
$S_y$ ..... celkový součet čtverců ( = součet čtverců odchylek hodnot od celkového průměru).

$$s_{ym}^2 = \frac{\sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2 n_i}{n} = \frac{S_{ym}}{n}$$ , kde

$\bar{y}_i$ ..... podmíněný průměr,

$S_{ym}$ .... meziskupinový součet čtverců ( = součet čtverců odchylek hodnot podmíněných průměrů od celkového průměru).

$$s_{yv}^2 = \frac{\sum_{i=1}^{k} s_i^2 n_i}{n} = \frac{S_{yv}}{n}$$, kde

$s_i^2$ ..... podmíněný rozptyl,

$S_{yv}$ .... vnitroskupinový (reziduální) součet čtverců.

$$s_i^2 = \frac{\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n_i} \qquad \bar{y} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}}{n} \qquad \bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$$

- Jestliže platí : $s_y^2 = s_{ym}^2 + s_{yv}^2$, platí také $S_y = S_{ym} + S_{yv}$ (Za účelem zjednodušení lze tudíž používat pouze čitatele vzorců, tzv. součty čtverců.).

- **Test hypotézy o nezávislosti :**

  1) $H_0 : \mu_1 = \mu_2 = ..... = \mu_k$
  (NEBO $H_0$ : *y* nezávisí na *x*)
  $H_1 : non\ H_0$

  2) **Testovým kritériem** je statistika :

  $$F = \frac{\frac{S_{ym}}{k-1}}{\frac{S_{yv}}{n-k}}$$ , která má při platnosti $H_0$ rozdělení F s $\nu_1 = k-1$ a $\nu_2 = n-k$ stupni volnosti.

  3) **Kritický obor:**
  $W \equiv \{F, F \geq F_{1-\alpha}(k-1, n-k)\}$.

- Měření síly závislosti proměnné *y* na faktoru *x*:

  ***Poměr determinace P²*** : $P^2 = \frac{S_{ym}}{S_y}$; $P^2 \in \langle 0,1 \rangle$

  ***Poměr korelace P***: $P = \sqrt{P^2}$ ; $P \in \langle 0,1 \rangle$

# χ2 – test o nezávislosti v kontingenční tabulce

= test nezávislosti kategoriálních znaků.

- **Kontingenční tabulka** – dvourozměrná tabulka, do které jsou uspořádány údaje o dvou slovních proměnných (příp. jedné proměnné slovní a druhé číselné).

*Kontingenční tabulka:*

| $b_j$ / $a_i$ | $b_1$ | $b_2$ | ..... | $b_s$ | Součty četností $n_{i\bullet}$ |
|---|---|---|---|---|---|
| $a_1$ | $n_{11}$ | $n_{12}$ | ..... | $n_{1s}$ | $n_{1\bullet}$ |
| $a_2$ | $n_{21}$ | $n_{22}$ | ..... | $n_{2s}$ | $n_{2\bullet}$ |
| . . . | . . . | . . . | . . . | . . . | . . . |
| $a_r$ | $n_{r1}$ | $n_{r2}$ | ..... | $n_{rs}$ | $n_{r\bullet}$ |
| Součty četností $n_{\bullet j}$ | $n_{\bullet 1}$ | $n_{\bullet 2}$ | ..... | $n_{\bullet s}$ | n |

- Četnosti $n_{ij}$ uvnitř tabulky se nazývají *sdružené (simultánní) četnosti*.
- Na okrajích tabulky najdeme *okrajové (marginální) četnosti* $n_{i\bullet}$ a $n_{\bullet j}$, kde tečkou je naznačeno sčítání.
- Platí, že:

$$n = \sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij} \qquad \text{a} \qquad n = \sum_{i=1}^{r} n_{i\bullet} = \sum_{j=1}^{s} n_{\bullet j} .$$

- Podstatou $\chi^2$ – testu nezávislosti je porovnání empirických četností s teoretickými četnostmi, které bychom očekávali v případě nezávislosti.
- Teoretické četnosti značíme $n'_{ij}$ a vypočítáme je podle :

$$n'_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n} .$$

- **Předpoklad testu:** $\chi^2$ – test nezávislosti lze použít tehdy, jsou-li všechna políčka kontingenční tabulky dostatečně obsazena, tj. platí-li $n'_{ij} \geq 5$. Pokud tato podmínka splněna není, musíme některé třídy sloučit nebo zvětšit rozsah výběru.

- **Test:**

1) $H_0$ : znaky *a* a *b* jsou nezávislé
   $H_1$ : znaky *a* a *b* jsou závislé (non $H_0$)

2) **Testové kritérium:**

$$G = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(n_{ij} - n'_{ij}\right)^2}{n'_{ij}}, \qquad G \in \langle 0, n \cdot h \rangle,$$ h ... min (r – 1), (s – 1).

G má při platnosti $H_0$ rozdělení $\chi^2$ s $\nu = (r-1)\cdot(s-1)$ stupni volnosti.

3) **Kritický obor:**

$$W \equiv \left\{G, G > \chi^2_{1-\alpha}\left[(r-1)(s-1)\right]\right\}$$

- Pokud jsme testem prokázali závislost, má smysl posuzovat její sílu – to lze pomocí kontingenčních koeficientů:

  ***Cramérův koeficient kontingence***

  $$C_C = \sqrt{\frac{G}{n \cdot h}} \qquad C_C \in \langle 0,1 \rangle$$

  ***Pearsonův koeficient kontingence***

  $$C_P = \sqrt{\frac{G}{G+n}} \qquad C_P \in \langle 0,1 )$$

- **Speciálním případem** kontingenční tabulky je tabulka ***ASOCIAČNÍ*** – čtyřpolní dvourozměrná tabulka, která obsahuje 2 slovní proměnné, které nabývají pouze 2 hodnot (alternativní znaky).

| $A_i$ \ $B_j$ | $B_1$ | $B_2$ | $n_{i\bullet}$ |
|---|---|---|---|
| $A_1$ | $n_{11}$ | $n_{12}$ | $n_{1\bullet}$ |
| $A_2$ | $n_{21}$ | $n_{22}$ | $n_{2\bullet}$ |
| $n_{\bullet j}$ | $n_{\bullet 1}$ | $n_{\bullet 2}$ | n |

1) $H_0$ : znaky *A* a *B* jsou nezávislé
   $H_1$ : znaky *A* a *B* jsou závislé (non $H_0$)

2) $$G = n \cdot \frac{\left(n_{11}n_{22} - n_{12}n_{21}\right)^2}{n_{\bullet 1}n_{\bullet 2}n_{1\bullet}n_{2\bullet}}$$

   G má při platnosti $H_0$ rozdělení $\chi^2(1)$.

3) $$W \equiv \left\{G, G > \chi^2_{1-\alpha}(1)\right\}$$

- K měření těsnosti závislosti jevů *A* a *B* se používá **koeficient asociace $r_{AB}$** :

  $$r_{AB} = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_{1\bullet}n_{2\bullet}n_{\bullet 1}n_{\bullet 2}}}, \qquad r_{AB} \in \langle -1,1 \rangle$$

# Regresní analýza

- zkoumání jednostranné závislosti proměnné *y* (závislá, vysvětlovaná) na proměnné *x* (nezávislá, vysvětlující), resp. na proměnných $x_1, x_2, \ldots, x_k$;
- nezávislá proměnná = příčina, závislá proměnná = důsledek;
- důležitý je přitom směr závislosti, tzn. která proměnná je závislá, a která nezávislá;
- závislost většinou modelujeme nějakou matematickou funkcí (tzv. regresní funkce).

***Postup regresní analýzy:***

1. Volba typu regresní funkce.
2. Odhad parametrů zvolené regresní funkce.
3. Testování hypotéz o parametrech regresní funkce.
4. Ověření vhodnosti zvoleného regresního modelu.

## Jednoduchá regresní analýza

## 1. Volba typu regresní funkce

- při volbě typu regresní funkce lze uplatnit různá kritéria;
- volba by se měla v prvé řadě opírat o určitou teorii, tzn. vyplývat z věcného rozboru vztahů proměnných;
- vždy se snažíme o jednoduchost modelu (ne příliš mnoho parametrů);
- je třeba změřit vhodnou charakteristikou přilnavost regresní funkce k datům.

## 2. Odhad parametrů regresní funkce

### Regresní modely

- matematické modely, které vyjadřují představu o průběhu závislosti proměnných;
- umožňují odhady neznámých hodnot závisle proměnné *y* ze známých hodnot nezávisle proměnné *x*, resp. nezávisle proměnných $x_1, x_2, \ldots, x_k$.

***Obecný tvar modelu:***

$$y_i = \eta_i + \varepsilon_i = \eta(x_i) + \varepsilon_i \ , \quad i = 1, 2, \ldots, n.$$

***Symbolika:*** $\eta_i$ ... deterministická složka

$\varepsilon_i$ ... náhodná (rušivá) složka

**Typy modelů:** **1. aditivní (součtový)** – jeho složky se skládají sčítáním, je nejčastější.
**2. multiplikativní (součinový)** – jeho složky se skládají násobením.

***Teoretická regresní funkce:*** $\eta = \eta(x)$

- existují různé typy regresních funkcí;
- nejčastější jsou lineární regresní funkce;
- linearita se může hodnotit jak z hlediska proměnných, tak z hlediska parametrů;
- každá regresní funkce má určitý počet parametrů (jejich počet je *p*).

***Parametry regresní funkce:***

- neznámé konstanty; symbolicky je značíme řeckými písmeny $(\beta_0, \beta_1, \ldots, \beta_k)$;
- jejich hodnoty lze odhadnout z výběrových dat;

- k jejich odhadu je třeba zvolit takovou metodu, aby odhady měly co nejlepší vlastnosti.

**1) Funkce lineární z hlediska parametrů**

| | |
|---|---|
| ***přímka*** | $\eta = \beta_0 + \beta_1 x$ |
| ***rovina*** | $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ |
| ***nadrovina*** | $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$ |
| ***parabola*** | $\eta = \beta_0 + \beta_1 x + \beta_2 x^2$ |
| ***hyperbola*** | $\eta = \beta_0 + \beta_1 x^{-1}$ |
| ***logaritmická funkce*** | $\eta = \beta_0 + \beta_1 \ln x$ |
| ***polynom*** | $\eta = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k$ |

**2) Funkce nelineární z hlediska parametrů**

| | |
|---|---|
| ***exponenciální funkce*** | $\eta = \beta_0 \beta_1^x$ |
| ***mocninná funkce*** | $\eta = \beta_0 x^{\beta_1}$ |
| ***Tőrnquistova křivka*** | $\eta = \dfrac{\beta_0 x}{x + \beta_1}$ |

# I. Jednoduchá lineární regrese

- regresní funkce je lineární z hlediska parametrů;
- má jednu vysvětlující proměnnou (regresor) *x*.

***Teoretická (hypotetická) regresní funkce:*** $\eta = \beta_0 + \beta_1 x$

- $\beta_0, \beta_1$ … *parametry*; $x$ … *regresor*
- nutno provést odhad teoretické regresní funkce, tzn. odhad neznámých parametrů $\beta_0, \beta_1$;
- nejlepší metodou odhadu parametrů lineární regresní funkce je **metoda nejmenších čtverců**;
- takový postup zaručí, že výběrová regresní funkce bude co nejlépe přiléhat k výběrovým hodnotám.

***Empirická (výběrová) regresní funkce:*** $\hat{\eta} = Y = b_0 + b_1 x$

$b_0, b_1$ … *odhady parametrů*; $b_0 = \hat{\beta}_0$; $b_1 = \hat{\beta}_1$

- odhad parametrů se provádí metodou nejmenších čtverců;
- když odhadneme parametry, získáme tzv. výběrovou regresní funkci.

## Metoda nejmenších čtverců

- lze ji použít pouze k odhadu parametrů funkcí lineárních v parametrech (v lineární regresi);
- princip: parametry odhadujeme tak, aby pro ně byl minimální součet čtverců reziduí.

$$y_i = \eta_i + \varepsilon_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

$$y_i = Y_i + \hat{\varepsilon}_i = b_0 + b_1 x_i + \hat{\varepsilon}_i$$

***Reziduum:*** $\hat{\varepsilon}_i = y_i - Y_i = y_i - b_0 - b_1 x_i = e_i$

$$S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left( y_i - b_0 - b_1 x_i \right)^2 \qquad \Rightarrow \quad \text{minimalizovat}$$

1. Stanovíme parciální derivace a položíme je rovny 0.
2. Vznikne soustava dvou rovnic (tzv. normální rovnice).
3. Vyřešíme je a získáme vzorce pro výpočet $b_0$ *a* $b_1$.

***Vzorce pro výpočet parametrů výběrové regresní přímky:***

$$b_1 = \frac{\dfrac{\sum_{i=1}^{n} x_i y_i}{n} - \dfrac{\sum_{i=1}^{n} x_i}{n} \cdot \dfrac{\sum_{i=1}^{n} y_i}{n}}{\dfrac{\sum_{i=1}^{n} x_i^2}{n} - \left( \dfrac{\sum_{i=1}^{n} x_i}{n} \right)^2} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{s_x^2} = \frac{s_{xy}}{s_x^2}$$

$$b_0 = \frac{\sum y_i \cdot \sum x_i^2 - \sum x_i \cdot \sum x_i \cdot y_i}{n \sum x_i^2 - \left( \sum x_i \right)^2} = \bar{y} - b_1 \cdot \bar{x}$$

$b_1$ … ***výběrový regresní koeficient (směrnice výběrové regresní přímky)***

- udává, jak velká průměrná změna proměnné *y* odpovídá zvýšení proměnné *x* o jednotku.

$s_{xy}$ … ***kovariance***

- symetrická míra, tzn. $s_{xy} = s_{yx}$.

## II. Nelineární regrese

- není-li regresní funkce lineární v parametrech, nelze její parametry odhadnout metodou nejmenších čtverců;
- pro odhad parametrů se používá řada různých metod;
- častá je metoda linearizující transformace (např. zlogaritmování);
- většinou následují další metody pro zlepšení vlastností odhadů;
- výpočetně značně náročné (využití statistických programů).

# 3. Testování hypotéz o parametrech regresní funkce

***t – testy***

- dílčí testy o nulových hodnotách jednotlivých regresních parametrů.

***1)*** $H_0 : \beta_j = 0, \quad j = 1, 2, \ldots, k$

$H_1 : non\ H_0$

**2)** ***Testové kritérium:***

$t = \frac{b_j}{s(b_j)}$ , $j = 1, 2, \ldots, k.$ Statistika ***t*** má při platnosti $H_0$ rozdělení t s (n-p) stupni volnosti.

**3)** ***Kritický obor:***

$$W \equiv \left\{ t;\ t \le t_{\frac{\alpha}{2}}(n-p) \cup t \ge t_{1-\frac{\alpha}{2}}(n-p) \right\}$$

***4) Závěr testu:***

Pokud leží hodnota testového kritéria v kritickém oboru, zamítáme $H_0$ a přijímáme $H_1$, jinak řečeno, test je statisticky významný. Testovaný parametr je statisticky významný, je v regresní funkci přínosný.

# 4. Ověření vhodnosti zvoleného regresního modelu

## ***Celkový F – test***

- testujeme vhodnost modelu jako celku;
- analýza rozptylu.

***1)*** $H_0 : \beta_0 = c,\ \beta_1, \beta_2, \ldots\ldots, \beta_k = 0$ (regresní funkce nemá žádný význam, tj. není vhodná)
$H_1 : non\ H_0$

**2)** ***Testové kritérium:***

$F = \frac{S_T / p-1}{S_R / n-p}$ ; Statistika **F** má při platnosti $H_0$ rozdělení F s (p – 1) a (n – p) stupni volnosti .

**Rozklad celkového součtu čtverců:** $S_y = S_T + S_R$

$S_y \ldots$ ***celkový součet čtverců***; charakterizuje celkovou variabilitu proměnné *y*.

$S_T \ldots$ ***teoretický součet čtverců***; část variability, kterou lze vysvětlit zvolenou regresní funkcí.

$S_R \ldots$ ***reziduální součet čtverců***; část variability, kterou nelze zvolenou regresní funkcí vysvětlit.

$$S_y = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$S_T = \sum_{i=1}^{n} (Y_i - \bar{y})^2 \ ; \quad S_R = \sum_{i=1}^{n} (y_i - Y_i)^2 \ .$$

**3)** ***Kritický obor:***

$$W \equiv \{F;\ F > F_{1-\alpha}(p-1;\ n-p)\}$$

***4) Závěr testu:***

Pokud leží hodnota testového kritéria v kritickém oboru, zamítáme $H_0$ a přijímáme $H_1$, jinak řečeno, test je statisticky významný. Model lze považovat za vhodný.

# Kritéria pro posouzení kvality regresní funkce

## *1. Index determinace*

- za vhodnější se považuje ta regresní funkce, u které je hodnota $I^2$ vyšší.

***Index determinace:*** $I^2 = \frac{S_T}{S_y}$; $I^2 \in \langle 0;1 \rangle$

- udává, jaký podíl variability proměnné *y* lze vysvětlit zvolenou regresní funkcí (lze udávat i v %).
- je zároveň mírou těsnosti závislosti proměnné *y* na proměnné *x*.
- obecná míra, nezávislá na typu regresní funkce (lze použít i pro měření nelineární závislosti).
- tato míra není symetrická.

*Pozn.:* Při srovnávání funkcí s rozdílným počtem parametrů musíme hodnotu $I^2$ upravit (penalizovat), neboť u funkce s vyšším počtem parametrů vychází hodnota $I^2$ automaticky vyšší. Existují různé formy penalizace, např.:

$$I^2_{adj} = 1 - (1 - I^2) \cdot \frac{n-1}{n-p} = 1 - \frac{(n-1)S_R}{(n-p)S_y} \ .$$

*Pozn.: adjusted = upravený.*

***Index korelace:*** $I = \pm\sqrt{I^2}$; $I \in \langle -1;1 \rangle$

## *2. Testové kritérium F*

- za vhodnější je považována ta funkce, u níž je hodnota F vyšší.

## *3. Reziduální součet čtverců a reziduální rozptyl*

***Reziduální součet čtverců:*** $S_R = \sum_{i=1}^{n} (y_i - Y_i)^2$

- za vhodnější považujeme funkci, která má reziduální součet čtverců nižší. Reziduální součet čtverců lze použít pouze tehdy, když srovnáváme funkce se stejným počtem parametrů.

***Reziduální rozptyl:*** $s_R^2 = \frac{S_R}{n-p}$

- za vhodnější považujeme funkci, která má reziduální rozptyl nižší. Reziduální rozptyl můžeme použít vždy, bez ohledu na to, kolik parametrů mají srovnávané regresní funkce.

## Vícenásobná lineární regrese

- zkoumáme závislost proměnné *y* na dvou či více vysvětlujících proměnných (regresorech) $x_1, x_2, \ldots, x_k$ ;
- volba typu regresní funkce je obtížná; vhodné použití statistických programů;
- nejčastěji proto volíme lineární regresní funkci.

***Teoretická vícenásobná lineární regresní funkce:***

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ \ldots \ + \beta_k x_k \ .$$

# Korelační analýza

- zabývá se především intenzitou vzájemného vztahu proměnných;
- je základní metodou měření síly lineární závislosti číselných proměnných;
- „correlatió“ = vzájemná souvislost (z lat.).

*! Z výpočetních a interpretačních hledisek se regresní a korelační analýza prolínají, nelze mezi nimi stanovit ostrou hranici.*

## Sdružené regresní přímky

$Y = a_{yx} + b_{yx}x$ popisuje závislost *y* na *x*

$X = a_{xy} + b_{xy}y$ popisuje závislost *x* na *y*

$a_{yx} = \bar{y} - b_{yx}\bar{x}$ $\qquad b_{yx} = \frac{s_{xy}}{s_x^2}$

$a_{xy} = \bar{x} - b_{xy}\bar{y}$ $\qquad b_{xy} = \frac{s_{xy}}{s_y^2}$

1. $b_{yx} = b_{xy} = 0 \Rightarrow$ *x* a *y* jsou korelačně nezávislé; sdružené regresní přímky svírají pravý úhel.
2. $b_{yx} = \frac{1}{b_{xy}} \Rightarrow$ *x* a *y* jsou úplně závislé; sdružené regresní přímky svírají nulový úhel (splývají).

## Míry těsnosti lineární závislosti

***Koeficient determinace:*** $r_{yx}^2 = r_{xy}^2 = b_{yx} \cdot b_{xy} = \frac{s_{xy}}{s_x^2} \cdot \frac{s_{xy}}{s_y^2} = \frac{s_{xy}^2}{s_x^2 \cdot s_y^2}$ **;** $\qquad r_{xy}^2 \in \langle 0;1 \rangle$

***Koeficient korelace:*** $r_{yx} = r_{xy} = \sqrt{r_{yx}^2} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\left(\overline{x^2} - \bar{x}^2\right) \cdot \left(\overline{y^2} - \bar{y}^2\right)}}$ **;** $\qquad r_{xy} \in \langle -1;1 \rangle$

- měří sílu lineární závislosti, nikoli závislosti obecně (lineární závislost = korelovanost).
- koeficient je symetrický.

***Interpretace:***

1. znaménko +/– udává směr závislosti:

   $r_{xy} > 0 \Rightarrow$ přímá závislost

   $r_{xy} < 0 \Rightarrow$ nepřímá závislost

2. $|r_{xy}|$ udává sílu závislosti:

$r_{xy} = 0 \quad \Rightarrow$ lineární nezávislost

$|r_{xy}| = 1 \quad \Rightarrow$ funkční (úplná) závislost

$|r_{xy}| \to 0 \quad \Rightarrow$ slabá lineární závislost

$|r_{xy}| \to 1 \quad \Rightarrow$ silná lineární závislost

## ***Test hypotézy o nulové hodnotě korelačního koeficientu***

**1)** $H_0 : \rho_{yx} = 0$ (lineární nezávislost *x* a *y*)

$H_1 :$ *non* $H_0$

**2)** ***Testové kritérium:***

$$t = \frac{r_{yx} \cdot \sqrt{n-2}}{\sqrt{1-r_{yx}^2}} \; ;$$ Statistika **t** má při platnosti $H_0$ rozdělení t s (n-2) stupni volnosti .

**3)** ***Kritický obor:***

$$W \equiv \left\{ t;\, t \le t_{\frac{\alpha}{2}}(n-2) \cup t \ge t_{1-\frac{\alpha}{2}}(n-2) \right\}$$

***4) Závěr testu:***

Pokud leží hodnota testového kritéria v kritickém oboru, zamítáme $H_0$ a přijímáme $H_1$, tzn. prokázali jsme hypotézu o lineární závislosti proměnných *x* a *y*.

## ***Pořadová korelace***

- Pokud chceme získat rychlou představu o síle závislosti mezi 2 kvantitativními znaky nebo určit závislost mezi pořadími znaků, nahradíme původní hodnoty $x_i$ a $y_i$ jejich pořadovými čísly $i_x$ a $i_y$ podle toho, která místa hodnoty zaujímají v uspořádané řadě.

***Spearmanův koeficient pořadové korelace:*** $r_s = 1 - \frac{6\sum_{i=1}^{n}(i_x - i_y)^2}{n(n^2-1)}$ ; $r_s \in \langle -1;1 \rangle$

- varianta korelačního koeficientu;
- měří sílu lineární závislosti dvou pořadí.

***Interpretace:*** stejná jako u korelačního koeficientu.

## ***Test hypotézy o nezávislosti pořadovou korelací***

**1)** $H_0 : \rho_S = 0$ (nezávislost pořadí)
$H_1 : non\ H_0$

**2)** ***Testové kritérium:***

$t = \frac{r_S \cdot \sqrt{n-2}}{\sqrt{1-r_S^2}}$ ; Statistika **t** má při platnosti $H_0$ rozdělení t s (n-2) stupni volnosti .

**3)** ***Kritický obor:***

$$W \equiv \left\{ t;\ t \le t_{\frac{\alpha}{2}}(n-2) \cup t \ge t_{1-\frac{\alpha}{2}}(n-2) \right\}$$

***4) Závěr testu:***

Pokud leží hodnota testového kritéria v kritickém oboru, zamítáme $H_0$ a přijímáme $H_1$, tzn. prokázali jsme hypotézu o lineární závislosti obou pořadí.

# *ČASOVÉ ŘADY*

- posloupnost chronologicky uspořádaných pozorování.
- posloupnosti věcně a prostorově srovnatelných pozorování, která jsou jednoznačně uspořádána z hlediska času.
- existují různé typy časových řad.

## A. Rozdělení ČŘ podle časového hlediska rozhodného pro zjišťování údajů:

**1. ČŘ intervalové:**

- řady intervalových ukazatelů (př. počet rozvodů za rok v ČR).
- intervalový ukazatel = ukazatel, jehož velikost závisí na délce intervalu, za který je sledován.
- lze vytvořit součty (resp. průměry), je nutná stejná délka intervalů.
- pokud jsou intervaly různě dlouhé, je třeba provést přepočet na jednotkový interval (tzv. očišťování od důsledků kalendářních variací).

**2. ČŘ okamžikové:**

- řady okamžikových (stavových) ukazatelů (př. počet obyvatel ČR k 31.12.).
- okamžikový ukazatel = ukazatel, jehož hodnoty se vztahují ke konkrétnímu časovému okamžiku.
- součet nedává reálný smysl, průměr nelze stanovit běžným způsobem.
- k charakterizování úrovně hodnot těchto ČŘ používáme *průměr chronologický*:

a) Pokud jsou vzdálenosti mezi časovými okamžiky stejné, použijeme ***prostý chronologický průměr***:

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + ..... + \frac{y_{n-1} + y_n}{2}}{n-1} = \frac{\frac{y_1}{2} + y_2 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1}$$

b) Pokud jsou mezi časovými okamžiky nestejné vzdálenosti, použijeme ***vážený chronologický průměr***:

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} \cdot d_1 + \frac{y_2 + y_3}{2} \cdot d_2 + ..... + \frac{y_{n-1} + y_n}{2} \cdot d_{n-1}}{d_1 + d_2 + ..... + d_{n-1}} = \frac{\sum_{i=1}^{n-1} \bar{y}_i d_i}{\sum_{i=1}^{n-1} d_i}$$

$d_i$ ..... vzdálenost mezi dvěma časovými okamžiky (= délka intervalu).

## B. Rozdělení ČŘ podle periodicity:

**1. ČŘ roční (dlouhodobé):**

- jejich periodicita je jeden rok a více.
- aplikace jiných postupů než u ČŘ krátkodobých.

**2. ČŘ krátkodobé:**

- jejich periodicita je kratší než jeden rok.
- ČŘ čtvrtletní, měsíční, týdenní, atd.

## C. Rozdělení ČŘ podle způsobu vyjádření ukazatelů:

**1. ČŘ naturálních ukazatelů:**

- hodnoty jsou vyjádřeny v naturálních jednotkách (př. ukazatele produkce).

**2. ČŘ peněžních ukazatelů:**

- hodnoty jsou vyjádřeny v peněžní formě.
- je nutno zajistit souměřitelnost hodnot v čase (změny cenové hladiny!).

## Základní charakteristiky časových řad

***Absolutní diference*** *(přírůstky)* – charakterizují absolutní změnu (přírůstek nebo úbytek) hodnoty ukazatele v časovém okamžiku *t* oproti období předcházejícímu (*t-1*).

**1. diference:** $\Delta_t^{(1)} = y_t - y_{t-1}\ , \quad t = 2, 3, \ldots, n$

**2. diference:** $\Delta_t^{(2)} = \Delta_t^{(1)} - \Delta_{t-1}^{(1)}, \quad t = 3, 4, \ldots, n$

**3. diference:** $\Delta_t^{(3)} = \Delta_t^{(2)} - \Delta_{t-1}^{(2)}, \quad t = 4, 5, \ldots, n$

atd.

***Průměrný absolutní přírůstek*** – aritmetický průměr jednotlivých prvních diferencí.

$$\overline{\Delta} = \frac{y_n - y_1}{n-1}$$

***Koeficient růstu*** – udává, kolikrát vzrostla hodnota ČŘ v časovém okamžiku **t** proti období předcházejícímu.

– $k_t \cdot 100 - 100$ udává o kolik procent vzrostla (event. klesla) hodnota ukazatele v časovém okamžiku *t* oproti období předcházejícímu.

$$k_t = \frac{y_t}{y_{t-1}}\ , \quad t = 2, 3, \ldots\ldots, n$$

***Průměrný koeficient růstu*** – geometrický průměr jednotlivých koeficientů růstu.

$$\overline{k} = \sqrt[n-1]{k_2 \cdot k_3 \cdot \ldots\ldots \cdot k_n} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

# Základní koncepce modelování časových řad

- cílem modelování ČŘ je jejich analýza a prognóza vývoje.
- snaha pochopit na základě abstrakce mechanismus chování ČŘ.

## 1. Jednorozměrný model

- nejjednodušší koncepce modelování ČŘ.
- reálná hodnota ČŘ ( $y_t$ ) je funkcí času ( $t$ ).

$$y_t = f(t;\varepsilon_t)\ ;\quad Y_t = f(t)\ ;\quad t = 1,2,\ldots,n$$

$$y_t = Y_t + \varepsilon_t$$

$$\varepsilon_t = y_t - Y_t$$

$t$ … časová proměnná

$y_t$ … reálná hodnota ukazatele v čase *t*

$Y_t$ … modelová (teoretická) hodnota ukazatele v čase *t*

$\varepsilon_t$ … nepravidelná (náhodná) složka (porucha) v čase *t*.

### *Klasický (formální) model:*

- zkoumá vliv časového faktoru na analyzovaný ukazatel.
- nezkoumá věcné příčiny dynamiky ČŘ.
- jeden z možných klasických přístupů je **dekompozice ČŘ**, tj. rozklad ČŘ na čtyři složky časového pohybu; popis jednotlivých forem pohybu.
- ČŘ může, ale nemusí všechny tyto složky obsahovat.

$T_t$ … ***trendová složka (trend)*** = dlouhodobá tendence ve vývoji hodnot analyzovaného ukazatele v čase.

$S_t$ … ***sezónní složka*** = pravidelně se opakující odchylka od trendu; periodicita maximálně 1 rok.

$C_t$ … ***cyklická složka*** = kolísání okolo trendu v důsledku dlouhodobého vývoje; periodicita delší než 1 rok.

$\varepsilon_t$ … ***náhodná složka*** = náhodné výkyvy, které nemají systematický charakter.

*Podle způsobu rozkladu rozlišujeme dva základní typy modelu:*

***aditivní model:*** $y_t = T_t + S_t + C_t + \varepsilon_t$

***multiplikativní model:*** $y_t = T_t \cdot S_t \cdot C_t \cdot \varepsilon_t$

## 2. Vícerozměrný model

- je založen na předpokladu, že vývoj analyzovaného ukazatele není ovlivněn pouze časovým faktorem, ale rovněž skupinou jiných, souvisejících ukazatelů.
- jedná se o tzv. *příčinné (faktorové) ukazatele*.
- pomocí těchto ukazatelů se snažíme vývoj analyzovaného ukazatele vysvětlit.

$$Y_t = f(t;x_1,x_2,\ldots,x_p)\ ;\quad t = 1,2,\ldots,n$$

kde $x_1, x_2, \ldots, x_n$ jsou příčinné (faktorové) ukazatele.

## *Trendová analýza*

- popis trendu (vyrovnání, vyhlazení) pomocí matematické (analytické, „hladké“) funkce = trendová funkce.
- tzv. modely s konstantními parametry.
- informace o charakteru hlavní tendence ve vývoji ukazatele.
- modelování vývoje trendu v budoucnu (prognóza = stanovení očekávaných hodnot ukazatele).

Předpoklad: $S_t = 0$ a $C_t = 0$ $\Rightarrow$ $Y_t = T_t$ ; $y_t = T_t + \varepsilon_t$.

### ***Lineární trend* :**

- nejběžněji používaný, předností je jednoduchost.

$T_t = \alpha_0 + \alpha_1 t \quad , \quad t = 1,2, ....., \ n$ .

- parametry $\alpha_0$ a $\alpha_1$ odhadneme pomocí metody nejmenších čtverců, čímž získáme výslednou rovnici odhadované trendové přímky:

$\widehat{T_t} = a_0 + a_1 t \quad , \quad t = 1,2, ....., \ n$ .

$a_0 = \bar{y} - a_1 \bar{t}$

$$a_1 = \frac{\sum_{t=1}^{n} t\, y_t - \bar{t} \sum_{t=1}^{n} y_t}{\sum_{t=1}^{n} t^2 - n\overline{t^2}} = \frac{n\sum_{t=1}^{n} t\, y_t - \sum_{t=1}^{n} t \sum_{t=1}^{n} y_t}{n\sum_{t=1}^{n} t^2 - \left(\sum_{t=1}^{n} t\right)^2} = \frac{\overline{t\, y_t} - \bar{t} \cdot \overline{y_t}}{\overline{t^2} - \bar{t}^2}$$

$a_0$ …… počáteční hodnota trendu pro $t = 0$

$a_1$ …… průměrný absolutní přírůstek (úbytek) hodnoty $\widehat{T_t}$ při zvýšení časové proměnné $t$ o jednotku.

***Parabolický (kvadratický) trend :*** $T_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$

***Exponenciální trend :*** $T_t = \alpha_0 \cdot \alpha_1^t$

***Modifikovaná (posunutá) exponenciála:*** $T_t = \kappa + \alpha_0 \cdot \alpha_1^{\ t}$

***Logistický trend:*** $T_t = \dfrac{\kappa}{1 + \alpha_0 \cdot \alpha_1^{\ t}}$

Existuje řada dalších typů trendových funkcí, v konkrétní situaci je třeba vždy zvolit takový, který nejlépe vystihuje empirická data.

### **Volba vhodného typu trendové funkce**

- věcná analýza zkoumaného jevu.
- posouzení empirické časové řady (ex ante).
- analýza grafu (pozor na subjektivitu).

## *Míry úspěšnosti zvolené trendové funkce:*

- měří kvalitu zvoleného modelu (funkce), je jich celá řada.

***Střední kvadratická (čtvercová) chyba odhadu:*** $M.S.E. = \frac{\sum_{t=1}^{n}\left(y_t - \widehat{T}_t\right)^2}{n}$

- nejčastější měřítko kvality modelu.
- přednost dáváme vždy tomu modelu, u něhož je hodnota M.S.E. nejnižší.
- prostřednictvím M.S.E. můžeme srovnávat jen funkce se stejným počtem parametrů.

***Statistika F:*** $F = \frac{\frac{S_T}{p-1}}{\frac{S_R}{n-p}} = \frac{\frac{\sum\left(\widehat{T}_t - \bar{y}\right)^2}{p-1}}{\frac{\sum\left(y_t - \widehat{T}_t\right)^2}{n-p}}$ , kde $\bar{y} = \frac{\sum \widehat{T}_t}{n}$

- za nejlepší považujeme model, pro který je hodnota statistiky F nejvyšší.

***Index determinace:*** $I^2 = \frac{S_T}{S_y} = \frac{\sum\left(\widehat{T}_t - \bar{y}\right)^2}{\sum\left(y_t - \bar{y}\right)^2}$

- vhodnější je ta funkce, která vede k vyšší hodnotě $I^2$.
- u funkcí s vyšším počtem parametrů vychází $I^2$ nadhodnocené; v takovém případě je vhodné použít $I^2_{adj}$.

# *Klouzavé průměry*

- *adaptivní přístup* k modelování trendu ČŘ.
- rozsah období, v jehož rámci bude ČŘ vyrovnána, je výrazně kratší než období, za něž máme ČŘ k dispozici.
- tzv. *modely s proměnlivými parametry*.
- posloupnost empirických pozorování nahradíme řadou průměrů z těchto pozorování vypočtených.
- každý z těchto průměrů reprezentuje určitou skupinu pozorování.
- při postupném výpočtu průměrů postupujeme (kloužeme) vždy o jedno pozorování kupředu, přičemž nejstarší (tj. první) pozorování z dané skupiny vypouštíme.

! V prvé řadě je třeba stanovit počet pozorování , z nichž vypočteme jednotlivé klouzavé průměry *(m)*.

***Klouzavá část období interpolace (m):*** časový interval délky $m = 2p + 1$, který se posunuje po časové ose vždy o jednotku.

*Volba délky klouzavé části období interpolace:*

- nelze stanovit exaktními statistickými postupy.
- stanovujeme především na základě věcné analýzy (heuristicky).
- přednost dáváme průměrům nižšího řádu.
- u neperiodických ČŘ se nejčastěji volí délka klouzavé části 3, 5, 7.
- u ČŘ s periodickou složkou je délka klouzavých průměrů rovna periodě sezónních nebo cyklických výkyvů.

*Identifikace jednotlivých klouzavých částí:*

- jednotlivé klouzavé části reprezentujeme jejich středními body (angl. *target*).
- je-li *m* liché číslo, pak střední bod klouzavé části je číslo celé *(t)*.
- je-li *m* sudé číslo, pak střední bod klouzavé části není celé číslo *(t + 0,5)*.

## ***Prosté klouzavé průměry***

*Předpoklad:* na klouzavých částech o rozsahu *m = 2p + 1* je definován lineární trend.

$$\bar{y}_t = \frac{1}{m}\sum_{i=-p}^{p} y_{t,i} = \frac{y_{t-p} + y_{t-p+1} + ..... + y_{t+p-1} + y_{t+p}}{m} \quad , \qquad t = p+1,\ p+2, ....., n-p\ .$$

- střední body klouzavých částí jsou celá čísla ( $t$ ).
- při tomto postupu zůstane *p* hodnot na začátku ČŘ a *p* hodnot na konci ČŘ nevyrovnáno.

## ***Vážené klouzavé průměry***

*Předpoklad:* na klouzavých částech o rozsahu *m = 2p + 1* je definován parabolický trend.

$$\bar{y}_t = \sum_{i=-p}^{p} W_i y_{t,i}\ , \quad t = p+1, p+2, ...., n-p\ ,$$

kde

$$W_i = \frac{3}{4m\left(m^2-4\right)}\left(3m^2 - 7 - 20i^2\right), \quad i = -p, \ldots, -1, 0, 1, \ldots, p.$$

Pro váhy platí : $\sum_{i=-p}^{p} W_i = 1$ a $W_i = W_{-i}$ , tj. váhy jsou symetrické.

## ***Centrované klouzavé průměry***

- jsou speciálním případem vážených klouzavých průměrů.
- používáme je v případě, že rozsah klouzavé části je číslo *sudé*.
- střední body klouzavých částí již nejsou celá čísla, proto nelze přímo přiřadit hodnoty klouzavých průměrů k empirickým pozorováním ČŘ.

*Postup výpočtu:*

- první vypočtený klouzavý průměr přiřadíme střednímu bodu *t*, který není celočíselný.
- další klouzavý průměr přiřadíme střednímu bodu *t+1*, který opět není celočíselný.
- celočíselný, tedy interpretovatelný, je bod *t+0,5*, který leží mezi body *t* a *t+1*.
- hodnotu klouzavého průměru, odpovídající bodu *t+0,5*, vypočteme jako aritmetický průměr dvou sousedních klouzavých průměrů.

*Např. čtyřčlenný centrovaný klouzavý průměr:*

$$\frac{1}{2}\left[\frac{1}{4}(y_1+y_2+y_3+y_4)+\frac{1}{4}(y_2+y_3+y_4+y_5)\right]=\frac{1}{8}(y_1+2y_2+2y_3+2y_4+y_5)$$

*Je možno použít vyjádření formou operátoru:* $\frac{1}{8}[1,2,2,2,1]$.

***Význam klouzavých průměrů:*** především při očišťování ČŘ od sezónních vlivů (viz další výklad).

## *Popis sezónní složky*

- sezónní složka = periodicky se opakující obousměrné odchylky hodnot ČŘ od trendu.
- délka periody je maximálně jeden rok.
- oscilace vznikají v důsledku přímých či nepřímých příčin, které se rok co rok pravidelně opakují v důsledku koloběhu Země kolem Slunce (klimatické vlivy, zprostředkované vlivy – společenské standardy a zvyklosti ve stereotypech chování lidí, např. dovolené, víkendy, prázdniny, Vánoce)
- nejprve je třeba zjistit, zda ČŘ reálně vykazuje sezónní výkyvy (např. věcným rozborem).

*Další postup:*

1. kvantifikace sezónních výkyvů.
2. očištění ČŘ, tj. vyloučení sezónní složky.

*Cíl sezónního očišťování:*

- odkrytí základní dynamiky vývoje zkoumaných jevů.
- umožnění bezprostředního srovnání vývoje v jednotlivých sezónách v rámci roku.

## 1. Model konstantní sezónnosti (aditivní model):

$$y_{ij} = T_{ij} + S_{ij} + \varepsilon_{ij}\,,\ i = 1,2,\ldots,m;\ j = 1,2,\ldots,r.$$

kde $i$ ..... označuje pořadí roku
$j$ ..... označuje dílčí období v rámci roku (sezóny).

***Kvantifikace sezónních výkyvů :***

1. **empirické sezónní rozdíly (odchylky) =** $y_{ij} - \hat{T}_{ij}$;
2. **průměrné sezónní rozdíly**;
3. **standardizované sezónní rozdíly (= sezónní faktory rozdílové)**.

*Standardizace (normování):* součet sezónních rozdílů v rámci roku musí být roven 0, tzn. v rámci roku se sezónní výkyvy kompenzují.

## 2. Model proporcionální sezónnosti (multiplikativní model):

$$y_{ij} = T_{ij} \cdot S_{ij} \cdot \varepsilon_{ij}\,,\ i = 1,2,\ldots,m;\ j = 1,2,\ldots,r.$$

***Kvantifikace sezónních výkyvů :***

1. **empirické sezónní indexy =** $\dfrac{y_{ij}}{\hat{T}_{ij}}$;
2. **průměrné sezónní indexy**;
3. **standardizované sezónní indexy (= sezónní faktory indexní)**.

*Standardizace (normování):* součet sezónních indexů v rámci roku musí být roven *r*, tzn. v rámci roku se sezónní výkyvy kompenzují.

## *Sezónní očišťování:*

1. vyrovnání ČŘ klouzavými průměry vhodného typu (možno i jiným způsobem, např. trendovou funkcí).
2. výpočet sezónních faktorů (rozdílových nebo indexních).
3. očištění údajů původní ČŘ:
   *a) model konstantní sezónnosti*: od hodnot původní ČŘ ***odečteme*** příslušný rozdílový sezónní faktor.
   *b) model proporcionální sezónnosti:* hodnoty původní ČŘ ***vydělíme*** příslušným indexním sezónním faktorem.

# *Individuální jednoduché indexy a rozdíly*

- slouží k bezprostřednímu srovnávání dvou hodnot téhož ukazatele, který není složen z dílčích částí;
- prostý podíl (rozdíl) hodnot ukazatele;
- výpočet přímo, není třeba shrnování údajů.

## *Index množství (objemu)*

- charakterizuje změnu hodnoty sledovaného extenzitního ukazatele (q resp. Q) v běžném období proti období základnímu.

$$i_q = \frac{q_1}{q_0}$$

Odpovídající ***rozdíl (diference)***: $\Delta_q = q_1 - q_0$.

$q_1$ ..... hodnota extenzitního ukazatele v situaci 1, tj. v běžném období;
$q_0$ ..... hodnota extenzitního ukazatele v situaci 0, tj. v základním období.

## *Index hodnoty*

$$i_Q = \frac{Q_1}{Q_0}$$

Odpovídající ***rozdíl (diference)***: $\Delta_Q = Q_1 - Q_0$.

## *Index úrovně*

- charakterizuje změnu hodnoty sledovaného intenzitního ukazatele (p) v běžném období proti období základnímu.

$$i_p = \frac{p_1}{p_0}$$

Odpovídající ***rozdíl (diference)***: $\Delta_p = p_1 - p_0$

! Mezi ukazateli platí deterministický vztah $Q = p \cdot q$ ; mezi indexy platí vztah: $i_Q = i_p \cdot i_q$.

## *Časové indexy a rozdíly*

- časové indexy individuální jednoduché bývají často sdružené do delších časových řad.

- relativně či absolutně srovnáváme dvě hodnoty shodně prostorově a věcně vymezeného ukazatele ve dvou časových obdobích.

***Základní období:*** je základem srovnání, označujeme indexem 0 $(p_0, q_0, Q_0)$.

***Běžné (sledované) období:*** označujeme indexem 1 $(p_1, q_1, Q_1)$, volíme vždy časově bližší období.

## *1. Řetězové indexy a rozdíly*

- charakterizují změny hodnot vzhledem k předcházejícímu období;
- indexy (rozdíly) s měnícím se základem.

*Řetězové indexy:* $i_{i/i-1} = \frac{u_i}{u_{i-1}} \quad ; \quad i = 1,2,3,\ldots,n$.

*Řetězové rozdíly:* $\Delta_{i/i-1} = u_i - u_{i-1} \quad ; \quad i = 1,2,3,\ldots,n$.

## *2. Bazické indexy a rozdíly*

- charakterizují změny hodnot vzhledem k určitému, pevně stanovenému období;
- indexy (rozdíly) se stálým základem;
- důležitá je volba základního období, zvolit nějakou „normální“ hodnotu (ne extrémní, atypickou).

*Bazické indexy:* $i_{i/B} = \frac{u_i}{u_B} \quad ; \quad i = 1,2,3,\ldots,n$.

*Bazické rozdíly:* $\Delta_{i/B} = u_i - u_B \quad ; \quad i = 1,2,3,\ldots,n$.

## Vztahy bazických a řetězových indexů a rozdílů

- umožňují přepočet jedněch na druhé;
- používáme je v případě, že nemáme k dispozici jednotlivé údaje, ale pouze řadu indexů.

**Přepočet řetězových indexů a rozdílů na bazické**

$i_{n/1} = i_{2/1} \cdot i_{3/2} \cdot \ldots \cdot i_{n/n-1}$ - řetězové indexy postupně násobíme.

$\Delta_{n/1} = \Delta_{2/1} + \Delta_{3/2} + \ldots + \Delta_{n/n-1}$ - řetězové rozdíly postupně přičítáme.

**Přepočet bazických indexů a rozdílů na řetězové**

$i_{i/i-1} = \frac{i_{i/B}}{i_{i-1/B}}$ - za sebou následující bazické indexy dělíme.

$\Delta_{i/i-1} = \Delta_{i/B} - \Delta_{i-1/B}$ - za sebou následující bazické rozdíly odčítáme.

## *Individuální složené indexy a rozdíly*

- slouží ke srovnávání hodnot stejnorodých ukazatelů, složených z dílčích částí;
- hodnota srovnávaného ukazatele je získána shrnutím hodnot za dílčí části celku.

**Shrnování hodnot za dílčí části celku:**

- u extenzitních ukazatelů (Q, q) shrnujeme prostým součtem;
- u intenzitních ukazatelů (p) shrnujeme průměrem.

### *Index množství (objemu)*

$$I_q = \frac{\sum q_1}{\sum q_0} = \frac{\sum i_q q_0}{\sum q_0} = \frac{\sum q_1}{\sum \frac{q_1}{i_q}}$$

Odpovídající ***rozdíl (diference)***: $\Delta_q = \sum q_1 - \sum q_0$ .

### *Index hodnoty*

$$I_Q = \frac{\sum Q_1}{\sum Q_0} = \frac{\sum i_Q Q_0}{\sum Q_0} = \frac{\sum Q_1}{\sum \frac{Q_1}{i_Q}} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$$

Odpovídající ***rozdíl (diference)***: $\Delta_Q = \sum Q_1 - \sum Q_0 = \sum p_1 q_1 - \sum p_0 q_0$ .

### *Index úrovně (tj. index proměnlivého složení)*

- je konstruován jako podíl dvou průměrů;
- průměrujeme obsahově stejnou, ale časově jinak vymezenou veličinu;
- vahami je struktura extenzitního ukazatele;
- udává změnu průměrné hodnoty intenzitního ukazatele způsobenou daným činitelem za předpokladu konstantní hodnoty druhého činitele.

$$I_{\bar{p}} = \frac{\bar{p}_1}{\bar{p}_0} = \frac{\frac{\sum Q_1}{\sum q_1}}{\frac{\sum Q_0}{\sum q_0}} = \frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_0 q_0}{\sum q_0}} = \frac{\frac{\sum Q_1}{\sum \frac{Q_1}{p_1}}}{\frac{\sum Q_0}{\sum \frac{Q_0}{p_0}}}$$

Odpovídající ***rozdíl (diference)***: $\Delta_{\bar{p}} = \bar{p}_1 - \bar{p}_0 = \frac{\sum p_1 q_1}{\sum q_1} - \frac{\sum p_0 q_0}{\sum q_0}$ .

### *Na velikost hodnoty $I_{\bar{p}}$ mají vliv dva činitele:*

1. změna dílčích hodnot intenzitního ukazatele, tj. hodnot v dílčích částech celku;
2. změna složení (struktury) hodnot extenzitního ukazatele, tj. změna vah.

*! Pro analýzu a kvantifikaci vlivu těchto činitelů* je třeba *provést rozklad* $I_{\bar{p}}$ *na dva indexy*. Nejčastěji je používána tzv. *metoda postupných změn*, která předpokládá, že ukazatele se v čase mění postupně (hypotetická situace).

## *Rozklad $I_{\bar{p}}$ metodou postupných změn*

A. $I_{\bar{p}} = I_{SS}(q_0) \cdot I_{STR}(p_1)$

nebo

B. $I_{\bar{p}} = I_{SS}(q_1) \cdot I_{STR}(p_0)$

! Oba typy rozkladu jsou významově rovnocenné, tzn. že neexistují objektivní důvody pro preferenci jednoho z nich. V praxi vždy pracujeme pouze s jedním.

## *Index stálého složení $I_{SS}$*

- charakterizuje vliv změny intenzitního ukazatele při stálém složení (v běžném či základním období) na změnu průměrné hodnoty intenzitního ukazatele;
- slouží ke zjištění vlivu samotných změn dílčích hodnot intenzitního ukazatele na změnu vyjádřenou $I_{\bar{p}}$;
- v indexu se mění pouze dílčí hodnoty intenzitního ukazatele a složení (struktura) vah je stálá;
- váhy lze fixovat na úrovni situace 0 nebo 1.

$$I_{SS}(q_0) = \frac{\frac{\sum p_1 q_0}{\sum q_0}}{\frac{\sum p_0 q_0}{\sum q_0}} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \qquad \Leftrightarrow \qquad \text{Váhy ze situace 0.}$$

$$I_{SS}(q_1) = \frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_0 q_1}{\sum q_1}} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \qquad \Leftrightarrow \qquad \text{Váhy ze situace 1.}$$

## *Index struktury $I_{STR}$*

- charakterizuje vliv změny struktury při stálé hodnotě intenzitního ukazatele (v běžném či základním období) na změnu průměrné hodnoty intenzitního ukazatele;
- slouží ke zjištění vlivu změn ve struktuře extenzitního ukazatele;
- v indexu se mění pouze struktura vah a dílčí hodnoty intenzitního ukazatele jsou stálé;
- dílčí hodnoty intenzitního ukazatele lze fixovat na úrovni situace 0 nebo 1.

$$I_{STR}(p_0) = \frac{\frac{\sum p_0 q_1}{\sum q_1}}{\frac{\sum p_0 q_0}{\sum q_0}} \quad \Leftrightarrow \quad \text{Hodnoty intenzitního ukazatele fixujeme na úrovni situace 0.}$$

$$I_{STR}(p_1) = \frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_1 q_0}{\sum q_0}} \quad \Leftrightarrow \quad \text{Hodnoty intenzitního ukazatele fixujeme na úrovni situace 1.}$$

# *Souhrnné indexy*

- slouží ke srovnávání hodnot nestejnorodých extenzitních a intenzitních ukazatelů;
- existuje řada různých druhů souhrnných indexů, většinou mají v názvu jméno svého autora.

## *Indexy úrovně (cenové)*

- slouží ke srovnávání hodnot nestejnorodých intenzitních ukazatelů (např. změny cen různých druhů výrobků);
- nejpoužívanější jsou *agregátní formy souhrnných indexů úrovně.*

***Agregátní formy souhrnných indexů úrovně:***

- jsou založeny na použití převodních koeficientů , tj. extenzitních ukazatelů, pomocí kterých jsou nestejnorodé intenzitní ukazatele, tj. ceny souboru výrobků, převáděny na stejnorodé extenzitní ukazatele Q (vynásobením těmito převodními koeficienty);
- měří v podstatě vliv změny intenzitních ukazatelů na změnu hodnoty extenzitního ukazatele Q za předpokladu, že extenzitní ukazatel q je ve srovnávaných obdobích konstantní.

***Loweův cenový index:***

$$_{Lo}I(p)=\frac{\sum p_1q_c}{\sum p_0q_c}=\frac{\sum\frac{p_1}{p_0}p_0q_c}{\sum p_0q_c}=\frac{\sum p_1q_c}{\sum\frac{p_1q_c}{\frac{p_1}{p_0}}}$$

- charakterizuje změnu cen (intenzitních ukazatelů *p*) v běžném období proti období základnímu nějakého konstantního (hypotetickému) souboru extenzitních ukazatelů *q* (nositelů dané intenzity).

Podle toho, z jakého období jsou zvoleny převodní koeficienty, rozlišujeme dva druhy indexů:

***Laspeyresův cenový index :***

$$_{L}I_{p}=\frac{\sum p_1q_0}{\sum p_0q_0}=\frac{\sum\frac{p_1}{p_0}p_0q_0}{\sum p_0q_0}=\frac{\sum i_pQ_0}{\sum Q_0}=\frac{\sum p_1q_0}{\sum\frac{p_1q_0}{\frac{p_1}{p_0}}}$$

- převodní koeficienty *q* jsou ze základního období;
- měří změnu hodnot intenzitních ukazatelů *p* v běžném období proti období základnímu souboru extenzitních ukazatelů *q* ze základního období.

***Paascheho cenový index :***

$$_{P}I_{p}=\frac{\sum p_1q_1}{\sum p_0q_1}=\frac{\sum\frac{p_1}{p_0}p_0q_1}{\sum p_0q_1}=\frac{\sum p_1q_1}{\sum\frac{p_1q_1}{\frac{p_1}{p_0}}}=\frac{\sum Q_1}{\sum\frac{Q_1}{i_p}}$$

- převodní koeficienty *q* jsou z běžného období;
- měří změnu hodnot intenzitních ukazatelů *p* v běžném období proti období základnímu souboru extenzitních ukazatelů *q* z běžného období.

! Výše uvedené dva indexy dávají při srovnání stejných souborů cen odlišné výsledky, přičemž nelze logicky odůvodnit upřednostnění jednoho z nich. Proto se někdy používá jejich prostý geometrický průměr.

***Fisherův cenový index :*** $_F I_p = \sqrt{_L I_p \cdot {}_P I_p}$

## *Indexy množství (objemu)*

- slouží ke srovnávání hodnot nestejnorodých extenzitních ukazatelů (např. výroby, prodeje, dovozu, spotřeby, nákupu apod. různých druhů výrobků).
- jsou založeny na převodu nestejnorodých extenzitních ukazatelů, jejichž srovnávání se provádí, na stejnorodé veličiny za pomoci převodních koeficientů (tzv. souměřitelů), kterými jsou intenzitní ukazatele *p* (vynásobením těmito souměřiteli) a na srovnávání relací hodnot těchto nestejnorodých extenzitních ukazatelů *q* pomocí relací hodnot stejnorodých extenzitních ukazatelů $Q = q \cdot p$ , které tímto součinem vzniknou.

***Loweův objemový index:***

$$_{Lo} I_q = \frac{\sum q_1 p_c}{\sum q_0 p_c} = \frac{\sum \frac{q_1}{q_0} q_0 p_c}{\sum q_0 p_c} = \frac{\sum q_1 p_c}{\sum \frac{q_1 p_c}{\frac{q_1}{q_0}}}$$

***Laspeyresův objemový index :***

$$_L I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{\sum \frac{q_1}{q_0} q_0 p_0}{\sum q_0 p_0} = \frac{\sum i_q Q_0}{\sum Q_0} = \frac{\sum q_1 p_0}{\sum \frac{q_1 p_0}{\frac{q_1}{q_0}}}$$

***Paascheho objemový index :***

$$_P I_q = \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{\sum \frac{q_1}{q_0} q_0 p_1}{\sum q_0 p_1} = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{\frac{q_1}{q_0}}} = \frac{\sum Q_1}{\sum \frac{Q_1}{i_q}}$$

***Fisherův objemový index :***

$$_F I_q = \sqrt{_L I_q \cdot {}_P I_q}$$

## *Souhrnný index hodnoty*

$$I_Q = \frac{\sum Q_1}{\sum Q_0} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$$

- vyjadřuje změnu hodnoty produkce, tj. jak změnu objemu, tak změnu cen;
- hodnotu lze vždy sčítat, takže souhrnný index hodnoty má stejný tvar jako individuální složený index extenzitního ukazatele Q.

## *Vztahy mezi souhrnnými indexy a rozdíly úrovně a souhrnnými indexy a rozdíly objemu*

***Index hodnoty extenzitního ukazatele Q rozložíme*** na součin Laspeyresova souhrnného indexu úrovně a Paascheho indexu objemového.

*Předpoklad:* nejdříve se mění ze základního období na běžné hodnota intenzitního ukazatele *p*, pak teprve dochází ke změně extenzitního ukazatele *q*.

$$I_Q = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum p_1 q_0} = {}_L I_p \cdot {}_P I_q$$

Rozklad příslušného **rozdílu** (diference) je vyjádřen vztahem :

$$\Delta_Q = \sum p_1 q_0 - \sum p_0 q_0 + \sum p_1 q_1 - \sum p_1 q_0 \,.$$

nebo

***Index hodnoty extenzitního ukazatele Q rozložíme*** na součin Paascheho souhrnného indexu úrovně a Laspeyresova indexu objemového.

*Předpoklad:* nejdříve se mění ze základního období na běžné hodnota extenzitního ukazatele *q*, pak teprve dochází ke změně intenzitního ukazatele *p*.

$$I_Q = \frac{\sum p_1 q_1}{\sum p_0 q_1} \cdot \frac{\sum p_0 q_1}{\sum p_0 q_0} = {}_P I_p \cdot {}_L I_q$$

Rozklad příslušného **rozdílu** (diference) je vyjádřen vztahem :

$$\Delta_Q = \sum p_1 q_1 - \sum p_0 q_1 + \sum p_0 q_1 - \sum p_0 q_0 \,.$$

# Charakteristiky úrovně časových řad – řešený příklad

Máme k dispozici údaje o počtu dopravních nehod v ČR v letech 2007-2018. Charakterizujte úroveň hodnot této časové řady.
Výsledek interpretujte!

| Rok | Počet nehod |
|---|---|
| 2007 | 182 736 |
| 2008 | 160 376 |
| 2009 | 74 815 |
| 2010 | 75 522 |
| 2011 | 75 137 |
| 2012 | 81 404 |
| 2013 | 84 398 |
| 2014 | 85 859 |
| 2015 | 93 067 |
| 2016 | 98 864 |
| 2017 | 103 821 |
| 2018 | 104 764 |

*Zdroj: Ročenka nehodovosti na pozemních komunikacích za rok 2018*

## *Řešení:*

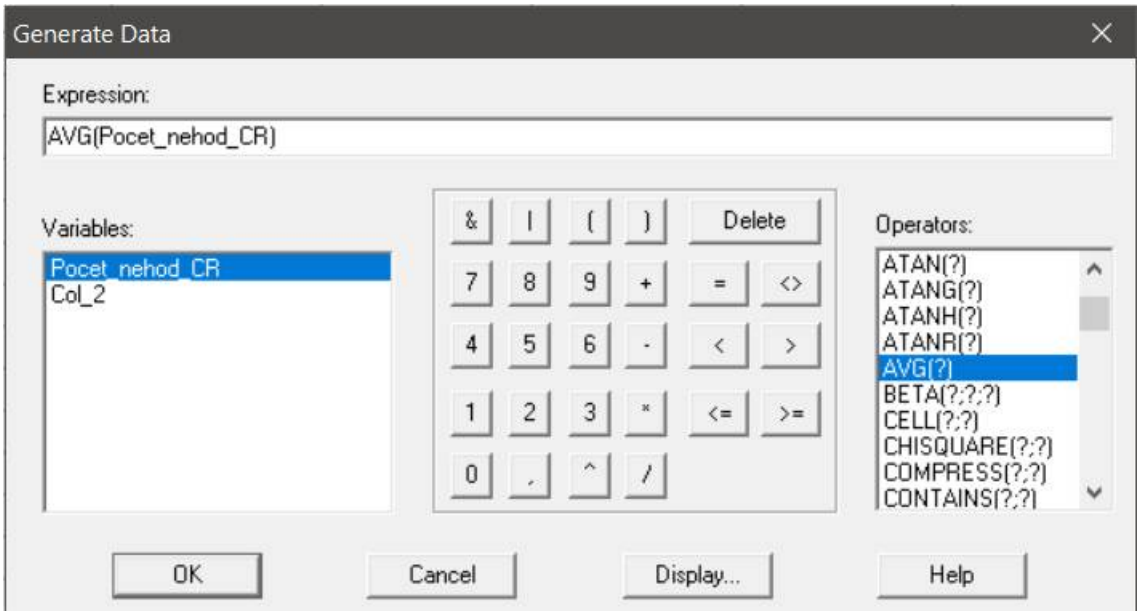
Počet nehod je ukazatel intervalový, z čehož vyplývá, že úroveň jeho hodnot je možné charakterizovat pomocí aritmetického průměru.

$$\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n} = \frac{1220763}{12} = 101730{,}25$$

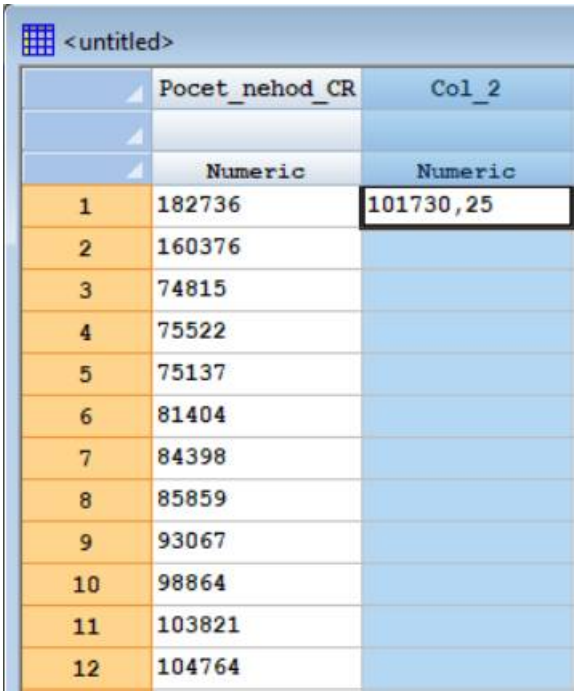
Průměrný počet nehod v ČR v letech 2007-2018 byl 101 730,25.

## Řešení v SGP:

Data o počtu nehod v ČR zadáme do prvního sloupce. Potom levým tlačítkem myši podsvítíme druhý sloupce a pravým tlačítkem vybereme nabídku *Generate Data*. Do řádku Expression vložíme dvojím kliknutím operátor AVG(?), který najdeme v levé části panelu v seznamu operátorů. Za otazník dosadíme dvojím kliknutím proměnnou Počet_nehod_CR z nabídky Variables vpravo – viz Obr. 1. Po stisknutí klávesy OK se v podsvíceném sloupci objeví hodnota aritmetického průměru, jak je vidět na Obr. 2.



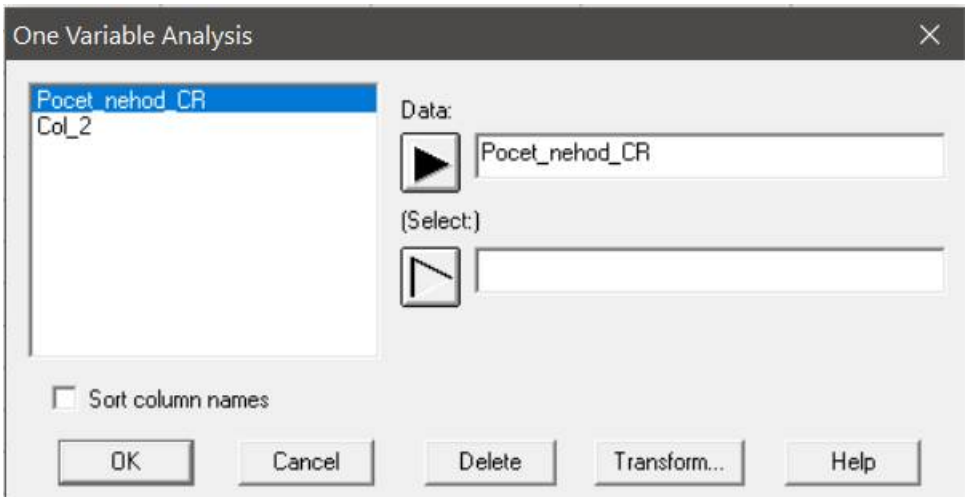

*Obrázek 1 – Použití operátoru AVG v Generate Data*




| | Pocet_nehod_CR | Col_2 |
|---|---|---|
| | Numeric | Numeric |
| 1 | 182736 | 101730,25 |
| 2 | 160376 | |
| 3 | 74815 | |
| 4 | 75522 | |
| 5 | 75137 | |
| 6 | 81404 | |
| 7 | 84398 | |
| 8 | 85859 | |
| 9 | 93067 | |
| 10 | 98864 | |
| 11 | 103821 | |
| 12 | 104764 | |

*Obrázek 2 – Vypočítaná hodnota aritmetického průměru*

Tento způsob výpočtu není jediný. Můžeme využít i nástrojů popisné statistiky v posloupnosti procedur Describe – Numeric Data – One-Variable Analysis … Do vstupního panelu zadáme do řádku Data hodnoty o počtu nehod v ČR a potvrdíme OK – viz Obr. 3. Nabídku Tables and Graphs můžeme jen potvrdit klávesou OK, protože výstup, který chceme, už je označený – zajímá nás *Summary Statistics*. V přehledu vypočítaných charakteristik najdeme aritmetický průměr pod názvem Average, jak ukazuje tabulka pod Obr. 3.




*Obrázek 3 – Vložení vstupních dat v proceduře One Variable Analysis*

**Summary Statistics for Pocet_nehod_CR**

| | |
|---|---|
| Count | 12 |
| Average | 101730, |
| Standard deviation | 34617,5 |
| Coeff. of variation | 34,0287% |
| Minimum | 74815,0 |
| Maximum | 182736, |
| Range | 107921, |
| Stnd. skewness | 2,44969 |
| Stnd. kurtosis | 1,54838 |

# Charakteristiky úrovně časových řad – řešený příklad

Máme k dispozici údaje o počtu dopravních nehod v ČR v letech 2007-2018. Charakterizujte úroveň hodnot této časové řady.
Výsledek interpretujte!

| Rok | Počet nehod |
|---|---|
| 2007 | 182 736 |
| 2008 | 160 376 |
| 2009 | 74 815 |
| 2010 | 75 522 |
| 2011 | 75 137 |
| 2012 | 81 404 |
| 2013 | 84 398 |
| 2014 | 85 859 |
| 2015 | 93 067 |
| 2016 | 98 864 |
| 2017 | 103 821 |
| 2018 | 104 764 |

*Zdroj: Ročenka nehodovosti na pozemních komunikacích za rok 2018*

## *Řešení:*

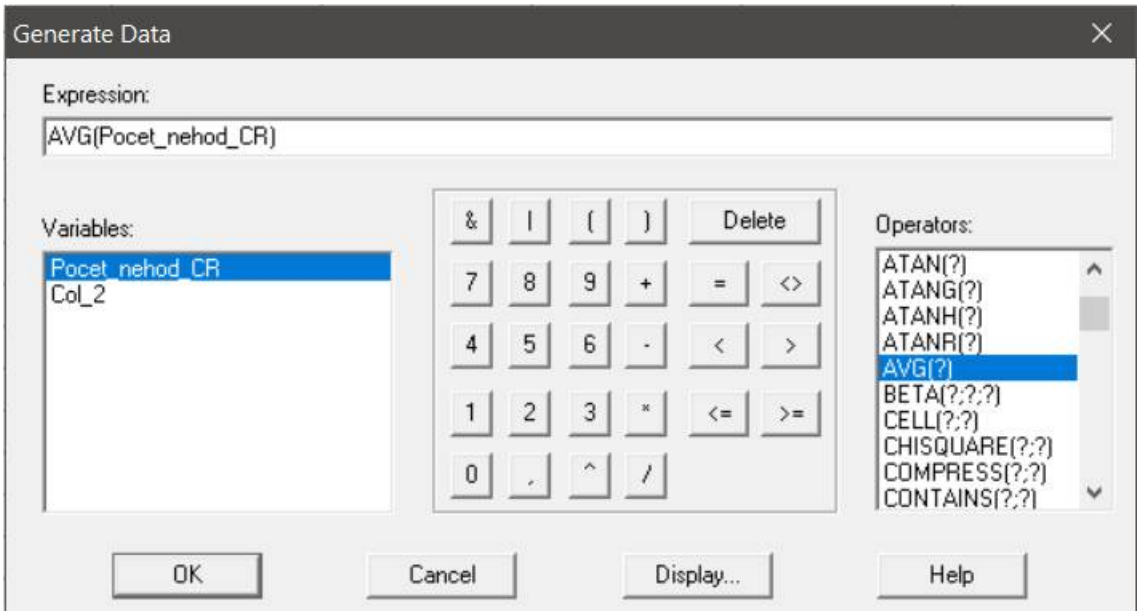
Počet nehod je ukazatel intervalový, z čehož vyplývá, že úroveň jeho hodnot je možné charakterizovat pomocí aritmetického průměru.

$$\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n} = \frac{1220763}{12} = 101730{,}25$$

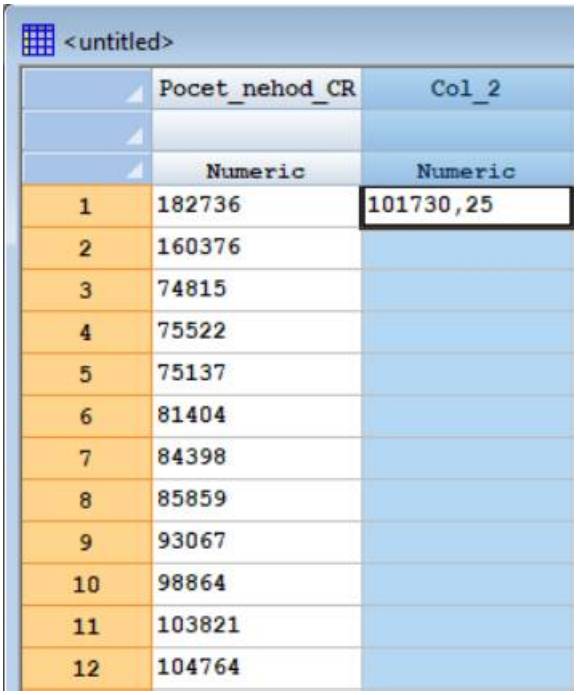
Průměrný počet nehod v ČR v letech 2007-2018 byl 101 730,25.

## Řešení v SGP:

Data o počtu nehod v ČR zadáme do prvního sloupce. Potom levým tlačítkem myši podsvítíme druhý sloupce a pravým tlačítkem vybereme nabídku *Generate Data*. Do řádku Expression vložíme dvojím kliknutím operátor AVG(?), který najdeme v levé části panelu v seznamu operátorů. Za otazník dosadíme dvojím kliknutím proměnnou Počet_nehod_CR z nabídky Variables vpravo – viz Obr. 1. Po stisknutí klávesy OK se v podsvíceném sloupci objeví hodnota aritmetického průměru, jak je vidět na Obr. 2.



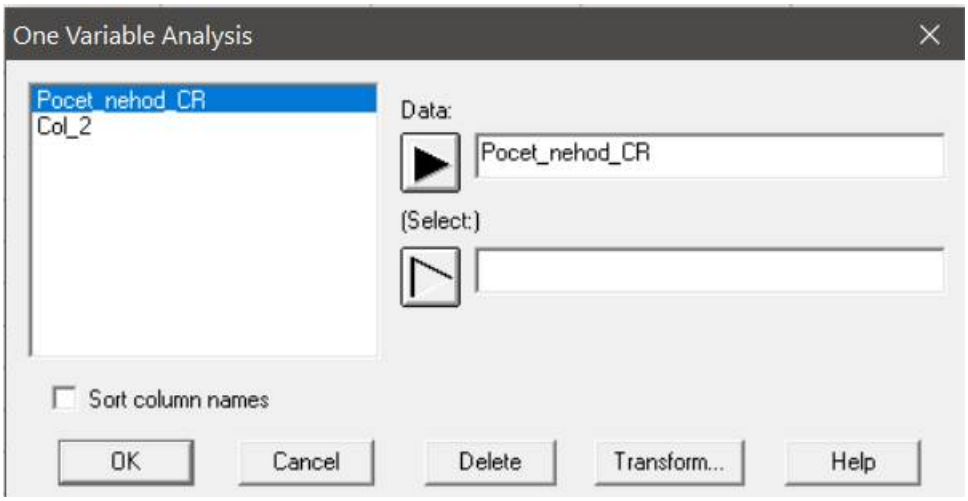

*Obrázek 1 – Použití operátoru AVG v Generate Data*




| | Pocet_nehod_CR | Col_2 |
|---|---|---|
| | Numeric | Numeric |
| 1 | 182736 | 101730,25 |
| 2 | 160376 | |
| 3 | 74815 | |
| 4 | 75522 | |
| 5 | 75137 | |
| 6 | 81404 | |
| 7 | 84398 | |
| 8 | 85859 | |
| 9 | 93067 | |
| 10 | 98864 | |
| 11 | 103821 | |
| 12 | 104764 | |

*Obrázek 2 – Vypočítaná hodnota aritmetického průměru*

Tento způsob výpočtu není jediný. Můžeme využít i nástrojů popisné statistiky v posloupnosti procedur Describe – Numeric Data – One-Variable Analysis … Do vstupního panelu zadáme do řádku Data hodnoty o počtu nehod v ČR a potvrdíme OK – viz Obr. 3. Nabídku Tables and Graphs můžeme jen potvrdit klávesou OK, protože výstup, který chceme, už je označený – zajímá nás *Summary Statistics*. V přehledu vypočítaných charakteristik najdeme aritmetický průměr pod názvem Average, jak ukazuje tabulka pod Obr. 3.




*Obrázek 3 – Vložení vstupních dat v proceduře One Variable Analysis*

**Summary Statistics for Pocet_nehod_CR**

| | |
|---|---|
| Count | 12 |
| Average | 101730, |
| Standard deviation | 34617,5 |
| Coeff. of variation | 34,0287% |
| Minimum | 74815,0 |
| Maximum | 182736, |
| Range | 107921, |
| Stnd. skewness | 2,44969 |
| Stnd. kurtosis | 1,54838 |

# Charakteristiky úrovně časových řad – řešený příklad 2

Máme k dispozici údaje o počtu nezaměstnaných v ČR k 31. 12. v letech 2008-2018. Charakterizujte úroveň hodnot této časové řady. Výsledek interpretujte!

| Rok | Počet nezaměstnaných k 31. 12. v tis. |
|---|---|
| 2008 | 229,8 |
| 2009 | 352,2 |
| 2010 | 383,7 |
| 2011 | 353,6 |
| 2012 | 366,9 |
| 2013 | 368,9 |
| 2014 | 323,6 |
| 2015 | 268,0 |
| 2016 | 211,4 |
| 2017 | 155,5 |
| 2018 | 121,6 |

*Zdroj: www.czso.cz*

## *Řešení:*

Počet nezaměstnaných je ukazatel okamžikový, z čehož vyplývá, že úroveň jeho hodnot je možné charakterizovat pomocí chronologického průměru. Aritmetický průměr zde není vhodný, protože součet hodnot počtu nezaměstnaných za roky 2008-2018 nedává smysl. V tomto případě je vhodné použít prostý tvar chronologického průměru, protože vzdálenosti mezi jednotlivými časovými okamžiky jsou stejné (vždy 1 rok).

$$\bar{y} = \frac{\frac{y_1}{2} + y_2 + \cdots + y_{n-1} + \frac{y_n}{2}}{n-1}$$

$$\bar{y} = \frac{\frac{229,8}{2} + 352,2 + 383,7 + 353,6 + 366,9 + 368,9 + 323,6 + 268 + 211,4 + 155,5 + \frac{121,6}{2}}{10} =$$

$$= \frac{2959,5}{10} = 295,95$$

Průměrný počet nezaměstnaných v ČR v letech 2008-2018 byl 295,5 tis. osob.

## Řešení v SGP:

SGP výpočet chronologického průměru neobsahuje.

# Charakteristiky úrovně časových řad – řešený příklad 2

Máme k dispozici údaje o počtu nezaměstnaných v ČR k 31. 12. v letech 2008-2018. Charakterizujte úroveň hodnot této časové řady. Výsledek interpretujte!

| Rok | Počet nezaměstnaných k 31. 12. v tis. |
|---|---|
| 2008 | 229,8 |
| 2009 | 352,2 |
| 2010 | 383,7 |
| 2011 | 353,6 |
| 2012 | 366,9 |
| 2013 | 368,9 |
| 2014 | 323,6 |
| 2015 | 268,0 |
| 2016 | 211,4 |
| 2017 | 155,5 |
| 2018 | 121,6 |

*Zdroj: www.czso.cz*

## *Řešení:*

Počet nezaměstnaných je ukazatel okamžikový, z čehož vyplývá, že úroveň jeho hodnot je možné charakterizovat pomocí chronologického průměru. Aritmetický průměr zde není vhodný, protože součet hodnot počtu nezaměstnaných za roky 2008-2018 nedává smysl. V tomto případě je vhodné použít prostý tvar chronologického průměru, protože vzdálenosti mezi jednotlivými časovými okamžiky jsou stejné (vždy 1 rok).

$$\bar{y} = \frac{\frac{y_1}{2} + y_2 + \cdots + y_{n-1} + \frac{y_n}{2}}{n-1}$$

$$\bar{y} = \frac{\frac{229,8}{2} + 352,2 + 383,7 + 353,6 + 366,9 + 368,9 + 323,6 + 268 + 211,4 + 155,5 + \frac{121,6}{2}}{10} =$$

$$= \frac{2959,5}{10} = 295,95$$

Průměrný počet nezaměstnaných v ČR v letech 2008-2018 byl 295,5 tis. osob.

## Řešení v SGP:

SGP výpočet chronologického průměru neobsahuje.

# Charakteristiky úrovně časových řad – řešený příklad 3

Máme k dispozici údaje o počtu obyvatel Liberce k 31. 12. v letech 2010-2019. Charakterizujte úroveň hodnot této časové řady. Výsledek interpretujte!

| **Rok** | **Počet obyvatel Liberce k 31. 12.** |
|---|---|
| 2010 | 101 625 |
| 2012 | 102 005 |
| 2015 | 102 562 |
| 2016 | 103 288 |
| 2018 | 103 979 |
| 2019 | 104 445 |

*Zdroj: www.obyvateleceska.cz*

## *Řešení:*

Počet obyvatel je ukazatel okamžikový, z čehož vyplývá, že úroveň jeho hodnot je možné charakterizovat pomocí chronologického průměru. Aritmetický průměr zde není vhodný, protože součet hodnot počtu obyvatel za roky 2010-2019 nedává smysl. V tomto případě je vhodné použít vážený tvar chronologického průměru, protože vzdálenosti mezi jednotlivými časovými okamžiky nejsou stejné. Do tabulky s daty si uděláme pomocný sloupec, kam zapíšeme hodnoty $d_i$ a další pomocné výpočty.

| **Rok** | **Počet obyvatel Liberce k 31. 12.** | **$d_i$** | $\overline{y}_i$ | $\overline{y}_i d_i$ |
|---|---|---|---|---|
| 2010 | 101 625 | | | |
| | | 2 | $\frac{101625 + 102005}{2} = 101815$ | 203630,0 |
| 2012 | 102 005 | | | |
| | | 3 | $\frac{102005 + 102562}{2} = 102283{,}5$ | 306850,5 |
| 2015 | 102 562 | | | |
| | | 1 | 102925,0 | 102925,0 |
| 2016 | 103 288 | | | |
| | | 2 | 103633,5 | 207267,0 |
| 2018 | 103 979 | | | |
| | | 1 | 104212,0 | 104212,0 |
| 2019 | 104 445 | | | |
| **Celkem** | **x** | **9** | **x** | **924884,5** |

$$\bar{y} = \frac{\sum_{i=1}^{n-1} \bar{y}_i d_i}{\sum_{i=1}^{n-1} d_i} = \frac{\frac{y_1 + y_2}{2} \cdot d_1 + \frac{y_2 + y_3}{2} \cdot d_2 + \cdots + \frac{y_{n-1} + y_n}{2} \cdot d_{n-1}}{d_1 + d_2 + \cdots + d_{n-1}} = \frac{924884{,}5}{9} = 102764{,}944$$

Průměrný počet obyvatel Liberce v letech 2010-2019 byl 102 764,944.

## Řešení v SGP:

SGP výpočet chronologického průměru neobsahuje.

# Charakteristiky úrovně časových řad – řešený příklad 3

Máme k dispozici údaje o počtu obyvatel Liberce k 31. 12. v letech 2010-2019. Charakterizujte úroveň hodnot této časové řady. Výsledek interpretujte!

| **Rok** | **Počet obyvatel Liberce k 31. 12.** |
|---|---|
| 2010 | 101 625 |
| 2012 | 102 005 |
| 2015 | 102 562 |
| 2016 | 103 288 |
| 2018 | 103 979 |
| 2019 | 104 445 |

*Zdroj: www.obyvateleceska.cz*

## *Řešení:*

Počet obyvatel je ukazatel okamžikový, z čehož vyplývá, že úroveň jeho hodnot je možné charakterizovat pomocí chronologického průměru. Aritmetický průměr zde není vhodný, protože součet hodnot počtu obyvatel za roky 2010-2019 nedává smysl. V tomto případě je vhodné použít vážený tvar chronologického průměru, protože vzdálenosti mezi jednotlivými časovými okamžiky nejsou stejné. Do tabulky s daty si uděláme pomocný sloupec, kam zapíšeme hodnoty $d_i$ a další pomocné výpočty.

| **Rok** | **Počet obyvatel Liberce k 31. 12.** | **$d_i$** | $\overline{y}_i$ | $\overline{y}_i d_i$ |
|---|---|---|---|---|
| 2010 | 101 625 | | | |
| | | 2 | $\frac{101625 + 102005}{2} = 101815$ | 203630,0 |
| 2012 | 102 005 | | | |
| | | 3 | $\frac{102005 + 102562}{2} = 102283{,}5$ | 306850,5 |
| 2015 | 102 562 | | | |
| | | 1 | 102925,0 | 102925,0 |
| 2016 | 103 288 | | | |
| | | 2 | 103633,5 | 207267,0 |
| 2018 | 103 979 | | | |
| | | 1 | 104212,0 | 104212,0 |
| 2019 | 104 445 | | | |
| **Celkem** | **x** | **9** | **x** | **924884,5** |

$$\bar{y} = \frac{\sum_{i=1}^{n-1} \bar{y}_i d_i}{\sum_{i=1}^{n-1} d_i} = \frac{\frac{y_1 + y_2}{2} \cdot d_1 + \frac{y_2 + y_3}{2} \cdot d_2 + \cdots + \frac{y_{n-1} + y_n}{2} \cdot d_{n-1}}{d_1 + d_2 + \cdots + d_{n-1}} = \frac{924884{,}5}{9} = 102764{,}944$$

Průměrný počet obyvatel Liberce v letech 2010-2019 byl 102 764,944.

## Řešení v SGP:

SGP výpočet chronologického průměru neobsahuje.

# Základní charakteristiky časových řad – řešený příklad

Máme k dispozici údaje o počtu dopravních nehod v ČR v letech 2007-2018. Charakterizujte vývoj počtu dopravních nehod pomocí:

a) prvních diferencí,
b) průměrného absolutního přírůstku,
c) koeficientů růstu,
d) průměrného koeficientu růstu.

Všechny výsledky interpretujte!
Doplňte též hodnoty druhých a třetích diferencí.

| Rok | Počet nehod |
|---|---|
| 2007 | 182 736 |
| 2008 | 160 376 |
| 2009 | 74 815 |
| 2010 | 75 522 |
| 2011 | 75 137 |
| 2012 | 81 404 |
| 2013 | 84 398 |
| 2014 | 85 859 |
| 2015 | 93 067 |
| 2016 | 98 864 |
| 2017 | 103 821 |
| 2018 | 104 764 |

*Zdroj: Ročenka nehodovosti na pozemních komunikacích za rok 2018*

## *Řešení:*

Všechny požadované charakteristiky časových řad je poměrně jednoduché spočítat i na kalkulačce, proto je zde uveden postup ručního výpočtu a následně i pomocí programu SGP.

a) První diference jsou konstruovány jako rozdíl hodnoty ukazatele v časovém okamžiku *t* a hodnoty ukazatele v okamžiku bezprostředně předcházejícím časovému okamžiku *t*, tedy *t-1*: $\Delta_t^{(1)} = y_t - y_{t-1}$. Hodnoty všech počítaných charakteristik budou doplněny postupně do následující tabulky.
   První řádek zůstane nevyplněn, protože hodnota *t* je rok 2007 a předcházející údaj není k dispozici. Tečka v řádku značí, že údaj existuje, ale nejsou k dispozici data k jeho výpočtu.
   Další hodnota je vypočtena jako $\Delta_2^{(1)} = y_2 - y_1 = 160376 - 182736 = -22360$.
   Tímto způsobem pokračujeme dále: $\Delta_3^{(1)} = y_3 - y_2 = 74815 - 160376 = -85561$.
   $\Delta_4^{(1)} = y_4 - y_3 = 75522 - 74815 = 707$
   $\Delta_5^{(1)} = y_5 - y_4 = 75137 - 75522 = -385$
   $\Delta_6^{(1)} = y_6 - y_5 = 81404 - 75137 = 6267$

$\Delta_7^{(1)} = y_7 - y_6 = 84398 - 81404 = 2994$
$\Delta_8^{(1)} = y_8 - y_7 = 85859 - 84398 = 1461$
$\Delta_9^{(1)} = y_9 - y_8 = 93067 - 85859 = 7208$
$\Delta_{10}^{(1)} = y_{10} - y_9 = 98864 - 93067 = 5797$
$\Delta_{11}^{(1)} = y_{11} - y_{10} = 103821 - 98864 = 4957$
$\Delta_{12}^{(1)} = y_{12} - y_{11} = 104764 - 103821 = 943$

| t | Rok | Počet nehod $y_t$ | $\Delta_t^{(1)}$ |
|---|---|---|---|
| 1 | 2007 | 182 736 | • |
| 2 | 2008 | 160 376 | -22 360 |
| 3 | 2009 | 74 815 | -85 561 |
| 4 | 2010 | 75 522 | 707 |
| 5 | 2011 | 75 137 | -385 |
| 6 | 2012 | 81 404 | 6 267 |
| 7 | 2013 | 84 398 | 2 994 |
| 8 | 2014 | 85 859 | 1 461 |
| 9 | 2015 | 93 067 | 7 208 |
| 10 | 2016 | 98 864 | 5 797 |
| 11 | 2017 | 103 821 | 4 957 |
| 12 | 2018 | 104 764 | 943 |
| **Celkem** | | | -77972 |

***Interpretace:***
Počet nehod v ČR se v roce 2008 snížil oproti roku 2007 o 22 360.
Počet nehod v ČR se v roce 2009 snížil oproti roku 2008 o 85 561.
Počet nehod v ČR se v roce 2010 zvýšil oproti roku 2009 o 707.
Počet nehod v ČR se v roce 2011 snížil oproti roku 2010 o 385.
Počet nehod v ČR se v roce 2012 zvýšil oproti roku 2011 o 6267.
atp.

b) Průměrný absolutní přírůstek je možné vyjádřit jako průměr prvních diferencí (tj. součet prvních diferencí dělených jejich počtem) nebo jako rozdíl poslední a první hodnoty časové řady dělený *n-1*. Oba způsoby výpočtu musejí dát pochopitelně stejný výsledek.

$$\bar{\Delta} = \frac{\sum_{t=2}^{n} \Delta_t^{(1)}}{n-1} = \frac{-77972}{11} = -7088{,}364$$

NEBO

$$\bar{\bar{\Delta}} = \frac{y_n - y_1}{n-1} = \frac{104764 - 182736}{11} = -7088{,}364$$

***Interpretace:***
Počet nehod v ČR se v letech 2007-2018 v průměru snížil o 7088,364 nehody.

c) Koeficient růstu je konstruován jako podíl hodnoty ukazatele v časovém okamžiku *t* a hodnoty ukazatele v časovém okamžiku *t-1*.

$$k_t = \frac{y_t}{y_{t-1}}$$

První řádek zůstane opět nevyplněn (jako u prvních diferencí), protože hodnota *t* je rok 2007 a předcházející údaj není k dispozici. Tečka v řádku značí, že údaj existuje, ale nejsou k dispozici data k jeho výpočtu.

$$k_2 = \frac{y_2}{y_1} = \frac{160376}{182736} = 0{,}8776$$
$$k_3 = \frac{y_3}{y_2} = \frac{74815}{160376} = 0{,}4665$$
$$k_4 = \frac{y_4}{y_3} = \frac{75522}{74815} = 1{,}0094$$
$$k_5 = \frac{y_5}{y_4} = \frac{75137}{75522} = 0{,}9949$$
$$k_6 = \frac{y_6}{y_5} = \frac{81404}{75137} = 1{,}0834$$
$$k_7 = \frac{y_7}{y_6} = \frac{84398}{81404} = 1{,}0368$$
$$k_8 = \frac{y_8}{y_7} = \frac{85859}{84398} = 1{,}0173$$
$$k_9 = \frac{y_9}{y_8} = \frac{93067}{85859} = 1{,}0840$$
$$k_{10} = \frac{y_{10}}{y_9} = \frac{98864}{93067} = 1{,}0623$$
$$k_{11} = \frac{y_{11}}{y_{10}} = \frac{103821}{98864} = 1{,}0501$$
$$k_{12} = \frac{y_{12}}{y_{11}} = \frac{104764}{103821} = 1{,}0091$$

| **t** | **Rok** | **Počet nehod $y_t$** | **$k_t$** |
|---|---|---|---|
| 1 | 2007 | 182 736 | • |
| 2 | 2008 | 160 376 | 0,8776 |
| 3 | 2009 | 74 815 | 0,4665 |
| 4 | 2010 | 75 522 | 1,0094 |
| 5 | 2011 | 75 137 | 0,9949 |
| 6 | 2012 | 81 404 | 1,0834 |
| 7 | 2013 | 84 398 | 1,0368 |
| 8 | 2014 | 85 859 | 1,0173 |
| 9 | 2015 | 93 067 | 1,0840 |
| 10 | 2016 | 98 864 | 1,0623 |
| 11 | 2017 | 103 821 | 1,0501 |
| 12 | 2018 | 104 764 | 1,0091 |

***Interpretace:***
Počet nehod v ČR se v roce 2008 snížil oproti roku 2007 o 12,24 % (0,8776*100-100).
Počet nehod v ČR se v roce 2009 snížil oproti roku 2008 o 53,35 %.
Počet nehod v ČR se v roce 2010 zvýšil oproti roku 2009 o 0,94 %.
Počet nehod v ČR se v roce 2011 snížil oproti roku 2010 o 0,51 %.
Počet nehod v ČR se v roce 2012 zvýšil oproti roku 2011 o 8,34 %.
atp.

d) Průměrný koeficient růstu je konstruován jako geometrický průměr koeficientů růstu. Lze ale také vypočítat jako *n-1*ní odmocnina z podílu poslední a první hodnoty časové řady.

$$\bar{k} = \sqrt[n-1]{k_2 \cdot k_3 \cdot \ldots \cdot k_n}$$

$$\bar{k} = \sqrt[11]{0{,}8776 \cdot 0{,}4665 \cdot 1{,}0094 \cdot \ldots \cdot 1{,}0091} = 0{,}9507$$

NEBO

$$\bar{k} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

$$\bar{k} = \sqrt[11]{\frac{104764}{182736}} = 0{,}9507$$

***Interpretace:***
Počet nehod v ČR se v letech 2007-2018 v průměru snížil o 4,93 %.

## Druhé a třetí diference

Druhé diference jsou definovány jako rozdíl sousedních prvních diferencí.
$\Delta_t^{(2)} = \Delta_t^{(1)} - \Delta_{t-1}^{(1)}$
První dva údaje není možné z dat, která máme k dispozici, vypočítat, proto opět píšeme do prvních dvou řádků puntík. Další hodnoty vypočítáme následovně:

$$\Delta_3^{(2)} = \Delta_3^{(1)} - \Delta_2^{(1)} = -85561 - (-22360) = -63201$$
$$\Delta_4^{(2)} = \Delta_4^{(1)} - \Delta_3^{(1)} = 707 - (-85561) = 86268$$
$$\Delta_5^{(2)} = \Delta_5^{(1)} - \Delta_4^{(1)} = -385 - 707 = -1092$$
$$\Delta_6^{(2)} = \Delta_6^{(1)} - \Delta_5^{(1)} = 6267 - (-385) = 6652$$
$$\Delta_7^{(2)} = \Delta_7^{(1)} - \Delta_6^{(1)} = 2994 - 6267 = -3273$$
$$\Delta_8^{(2)} = \Delta_8^{(1)} - \Delta_7^{(1)} = 1461 - 2994 = -1533$$
$$\Delta_9^{(2)} = \Delta_9^{(1)} - \Delta_8^{(1)} = 7208 - 1461 = 5747$$
$$\Delta_{10}^{(2)} = \Delta_{10}^{(1)} - \Delta_9^{(1)} = 5797 - 7208 = -1411$$
$$\Delta_{11}^{(2)} = \Delta_{11}^{(1)} - \Delta_{10}^{(1)} = 4957 - 5797 = -840$$
$$\Delta_{12}^{(2)} = \Delta_{12}^{(1)} - \Delta_{11}^{(1)} = 104764 - 103821 = -4014$$

| t | Rok | Počet nehod $y_t$ | $\Delta_t^{(1)}$ | $\Delta_t^{(2)}$ |
|---|---|---|---|---|
| 1 | 2007 | 182 736 | • | • |
| 2 | 2008 | 160 376 | -22 360 | • |
| 3 | 2009 | 74 815 | -85 561 | -63 201 |
| 4 | 2010 | 75 522 | 707 | 86 268 |
| 5 | 2011 | 75 137 | -385 | -1 092 |
| 6 | 2012 | 81 404 | 6 267 | 6 652 |
| 7 | 2013 | 84 398 | 2 994 | -3 273 |
| 8 | 2014 | 85 859 | 1 461 | -1 533 |
| 9 | 2015 | 93 067 | 7 208 | 5747 |
| 10 | 2016 | 98 864 | 5 797 | -1 411 |
| 11 | 2017 | 103 821 | 4 957 | -840 |
| 12 | 2018 | 104 764 | 943 | -4014 |

Třetí diference jsou definovány jako rozdíl sousedních druhých diferencí.
$\Delta_t^{(3)} = \Delta_t^{(2)} - \Delta_{t-1}^{(2)}$

První tři údaje není možné z dat, která máme k dispozici, vypočítat, proto opět píšeme do prvních tří řádků puntík. Další hodnoty už je možné spočítat.

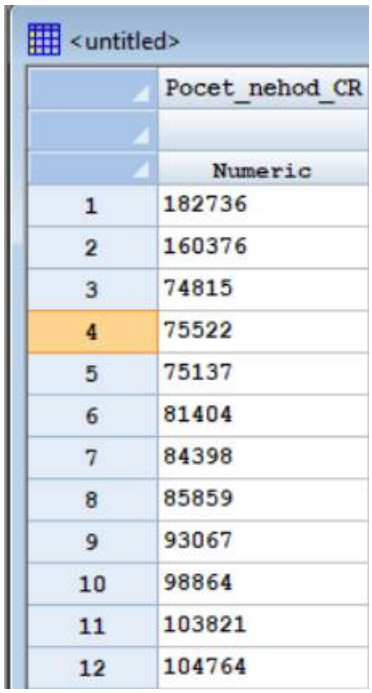
| t | Rok | Počet nehod $y_t$ | $\Delta_t^{(2)}$ | $\Delta_t^{(3)}$ |
|---|---|---|---|---|
| 1 | 2007 | 182 736 | • | • |
| 2 | 2008 | 160 376 | • | • |
| 3 | 2009 | 74 815 | -63 201 | • |
| 4 | 2010 | 75 522 | 86 268 | 149 469 |
| 5 | 2011 | 75 137 | -1 092 | -87 360 |
| 6 | 2012 | 81 404 | 6 652 | 7 744 |
| 7 | 2013 | 84 398 | -3 275 | -9 925 |
| 8 | 2014 | 85 859 | -1 533 | 1 740 |
| 9 | 2015 | 93 067 | 5747 | 7 280 |
| 10 | 2016 | 98 864 | -1 411 | -7 158 |
| 11 | 2017 | 103 821 | -840 | 571 |
| 12 | 2018 | 104 764 | -4014 | -3 174 |

$\Delta_4^{(3)} = \Delta_4^{(2)} - \Delta_3^{(2)} = 86268 - (-63201) = 149469$
$\Delta_5^{(3)} = \Delta_5^{(2)} - \Delta_4^{(2)} = -1092 - 86268 = -87360$
$\Delta_6^{(3)} = \Delta_6^{(2)} - \Delta_5^{(2)} = 6652 - (-1092) = 7744$
$\Delta_7^{(3)} = \Delta_7^{(2)} - \Delta_6^{(2)} = -3275 - 6652 = -9925$
$\Delta_8^{(3)} = \Delta_8^{(2)} - \Delta_7^{(2)} = -1533 - (-3275) = 1740$
$\Delta_9^{(3)} = \Delta_9^{(2)} - \Delta_8^{(2)} = 5747 - (-1533) = 7280$
$\Delta_{10}^{(3)} = \Delta_{10}^{(2)} - \Delta_9^{(2)} = -1411 - 5747 = -7158$
$\Delta_{11}^{(3)} = \Delta_{11}^{(2)} - \Delta_{10}^{(2)} = -840 - (-1411) = 571$
$\Delta_{12}^{(3)} = \Delta_{12}^{(2)} - \Delta_{11}^{(2)} = -4014 - (-840) = -3174$

Druhé a třetí diference se využívají při hledání vhodného modelu trendu (bude předmětem následujícího tématu).

## Řešení v SGP:

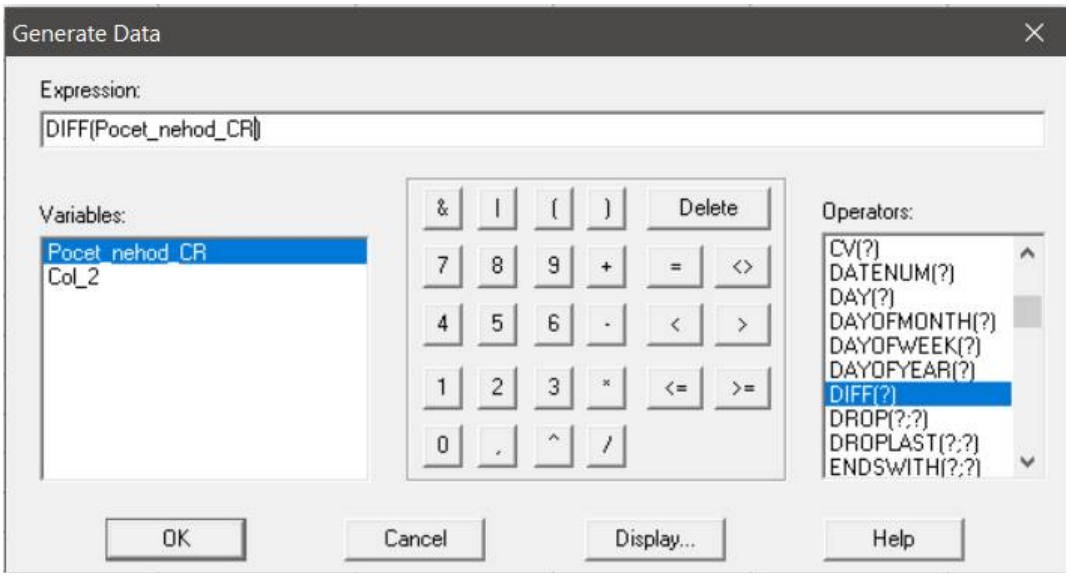
Prvním krokem při řešení příkladu v SGP je vložení vstupních dat. Údaje o počtu nehod v ČR vložíme jako jednu proměnnou, tj. do jednoho sloupce – viz Obr. 1.




| | Pocet_nehod_CR |
|---|---|
| | |
| | Numeric |
| 1 | 182736 |
| 2 | 160376 |
| 3 | 74815 |
| 4 | 75522 |
| 5 | 75137 |
| 6 | 81404 |
| 7 | 84398 |
| 8 | 85859 |
| 9 | 93067 |
| 10 | 98864 |
| 11 | 103821 |
| 12 | 104764 |

*Obrázek 1 – Vložení vstupních dat*

a) Pro výpočet prvních diferencí přeskočíme kurzorem na další sloupeček, levým tlačítkem ho podsvítíme a pravým tlačítkem myši zvolíme z nabídky procedur *Generate Data*. Ze sloupečku *Operators* v pravé části panelu vybereme operátor DIFF(?), který dvojím kliknutím dosadíme do řádku Expression, a za otazník zadáme proměnnou Počet_nehod_CR (dvojím kliknutím z tabulky Variables vlevo) – viz Obr. 2. Potvrdíme klávesou OK. Hodnoty prvních diferencí se objeví ve sloupci 2 – na Obr. 2 je už přejmenovaný na První_diference, aby bylo zřejmé, o jaké hodnoty se jedná.




*Obr. 2 – Zadání operátoru DIFF v Generate Data*

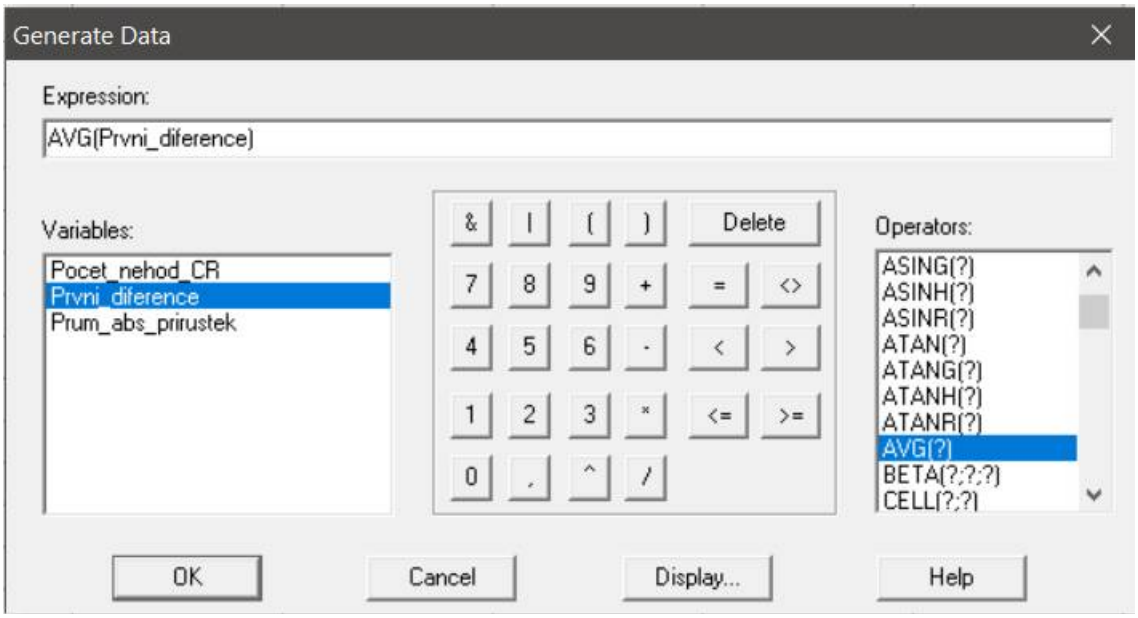
<untitled>

| | Pocet_nehod_CR | Prvni_diference |
|---|---|---|
| | Numeric | Numeric |
| 1 | 182736 | |
| 2 | 160376 | -22360 |
| 3 | 74815 | -85561 |
| 4 | 75522 | 707 |
| 5 | 75137 | -385 |
| 6 | 81404 | 6267 |
| 7 | 84398 | 2994 |
| 8 | 85859 | 1461 |
| 9 | 93067 | 7208 |
| 10 | 98864 | 5797 |
| 11 | 103821 | 4957 |
| 12 | 104764 | 943 |

*Obrázek 3 – Výsledek použití operátoru DIFF*

Interpretace vypočtených hodnot zde již není uvedena – je ji možné najít v první části řešeného příkladu, kde je ukázka ručního výpočtu.

b) Pro výpočet průměrného absolutního přírůstku využijeme rovněž operátoru v nabídce *Generate Data*. Víme, že průměrný absolutní přírůstek je aritmetický průměr prvních diferencí, proto použijeme operátor AVG. Třetí sloupec si nazveme Prum_abs_prirustek, podsvítíme ho a v nabídce *Generate Data* zadáme do řádku Expression operátor AVG(?). Za otazník zadáme proměnnou První_diference, jak ukazuje Obr. 4.




*Obrázek 4 – Použití operátoru AVG v Generate Data*

Po potvrzení klávesou OK, se objeví výsledná hodnota v podsvíceném sloupci – viz Obr. 5.

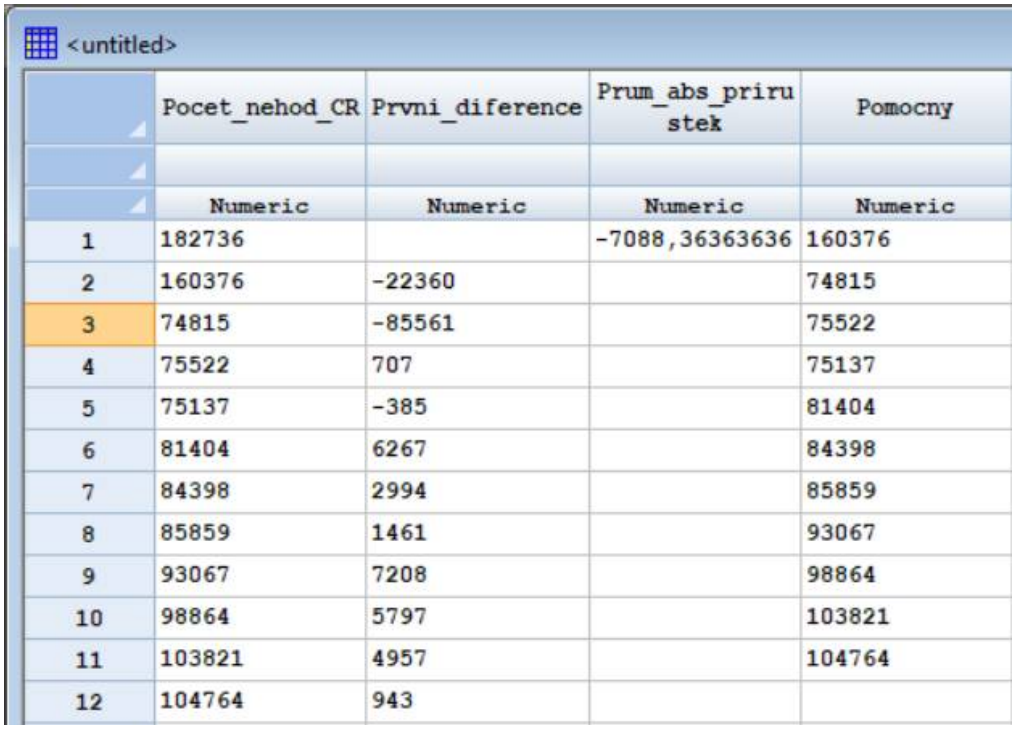
<untitled>

| | Pocet_nehod_CR | Prvni_diference | Prum_abs_prirustek |
|---|---|---|---|
| | Numeric | Numeric | Numeric |
| 1 | 182736 | | -7088,36363636 |
| 2 | 160376 | -22360 | |
| 3 | 74815 | -85561 | |
| 4 | 75522 | 707 | |
| 5 | 75137 | -385 | |
| 6 | 81404 | 6267 | |
| 7 | 84398 | 2994 | |
| 8 | 85859 | 1461 | |
| 9 | 93067 | 7208 | |
| 10 | 98864 | 5797 | |
| 11 | 103821 | 4957 | |
| 12 | 104764 | 943 | |

*Obrázek 5 – Vypočítaná hodnota průměrného absolutního přírůstku*

Interpretace vypočtené hodnoty viz ruční řešení.

c) Koeficienty růstu je možné v SGP vypočítat dvěma základními způsoby. První z nich opět využívá nabídku *Generate Data*. Nejprve je ale potřeba před výpočtem vytvořit pomocný sloupec s hodnotami proměnné Počet_nehod_CR bez první hodnoty. Viz Obr. 6.

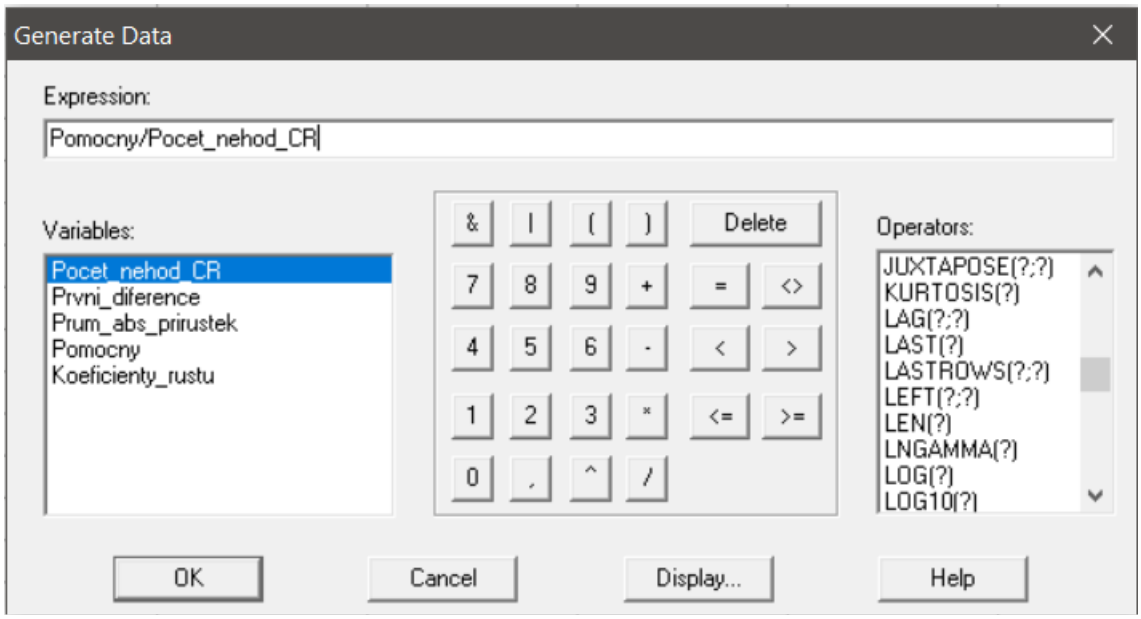


<untitled>

| | Pocet_nehod_CR | Prvni_diference | Prum_abs_prirustek | Pomocny |
|---|---|---|---|---|
| | | | | |
| | Numeric | Numeric | Numeric | Numeric |
| 1 | 182736 | | -7088,36363636 | 160376 |
| 2 | 160376 | -22360 | | 74815 |
| 3 | 74815 | -85561 | | 75522 |
| 4 | 75522 | 707 | | 75137 |
| 5 | 75137 | -385 | | 81404 |
| 6 | 81404 | 6267 | | 84398 |
| 7 | 84398 | 2994 | | 85859 |
| 8 | 85859 | 1461 | | 93067 |
| 9 | 93067 | 7208 | | 98864 |
| 10 | 98864 | 5797 | | 103821 |
| 11 | 103821 | 4957 | | 104764 |
| 12 | 104764 | 943 | | |

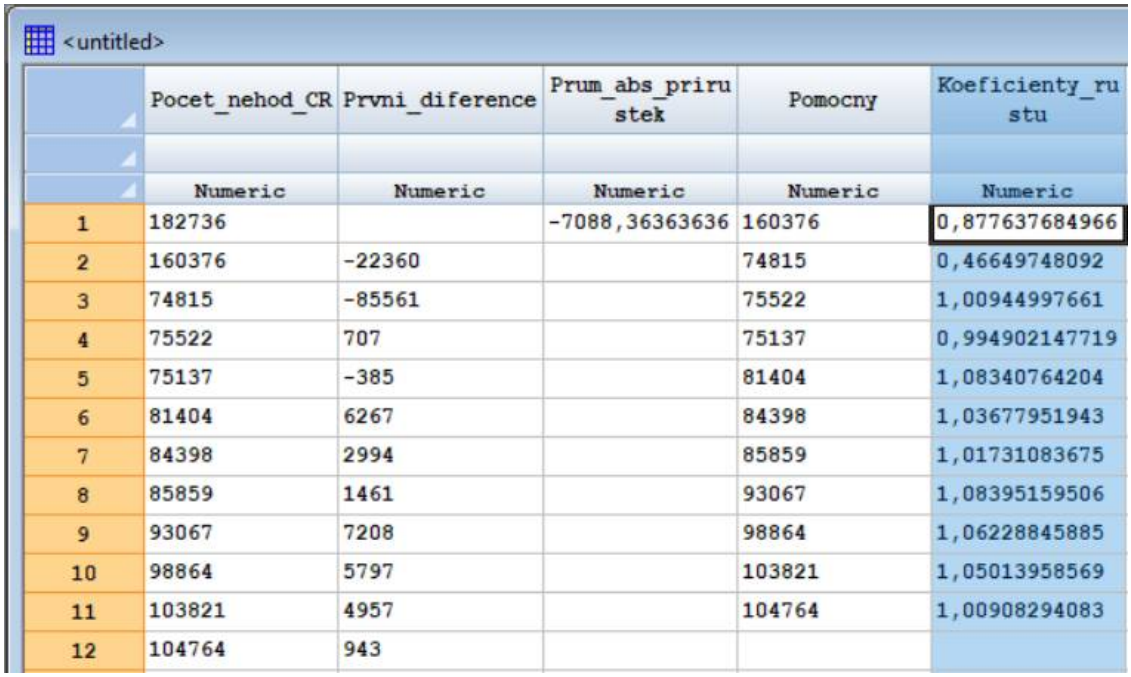
*Obrázek 6 – Vložení pomocné proměnné*

Kurzorem přeskočíme do dalšího volného sloupce, který podsvítíme, a v panelu *Generate Data* zadáme do řádku Expression vzorec „Pomocny/Počet_nehod_CR“, jak ukazuje Obr. 7.




*Obrázek 7 – Zadání vzorce pro výpočet koeficientů růstu v Generate Data*

Po stisknutí klávesy OK se hodnoty koeficientů růstu objeví v podsvíceném sloupci – viz Obr. 8 (už je zde přejmenován na Koeficienty_rustu).



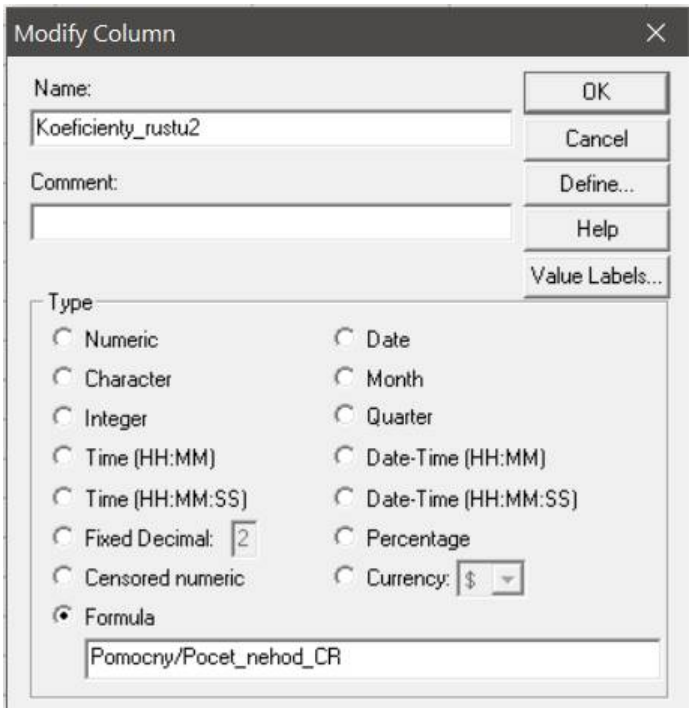
<untitled>

| | Pocet_nehod_CR | Prvni_diference | Prum_abs_prirustek | Pomocny | Koeficienty_rustu |
|---|---|---|---|---|---|
| | Numeric | Numeric | Numeric | Numeric | Numeric |
| 1 | 182736 | | -7088,36363636 | 160376 | 0,877637684966 |
| 2 | 160376 | -22360 | | 74815 | 0,46649748092 |
| 3 | 74815 | -85561 | | 75522 | 1,00944997661 |
| 4 | 75522 | 707 | | 75137 | 0,994902147719 |
| 5 | 75137 | -385 | | 81404 | 1,08340764204 |
| 6 | 81404 | 6267 | | 84398 | 1,03677951943 |
| 7 | 84398 | 2994 | | 85859 | 1,01731083675 |
| 8 | 85859 | 1461 | | 93067 | 1,08395159506 |
| 9 | 93067 | 7208 | | 98864 | 1,06228845885 |
| 10 | 98864 | 5797 | | 103821 | 1,05013958569 |
| 11 | 103821 | 4957 | | 104764 | 1,00908294083 |
| 12 | 104764 | 943 | | | |

*Obrázek 8 – Vypočítané hodnoty koeficientů růstu*

Interpretace vypočtených hodnot viz ruční řešení.

Druhý způsob výpočtu koeficientů růstu využívá nabídku *Modify Column*, kterou je možné vyvolat stejným způsobem jako *Generate Data*, tj. podsvítit další sloupec a pravým tlačítkem vyvolat příslušnou nabídku. V panelu *Modify Column* vyplníme nejprve název nové proměnné Koeficienty_rustu2 a pak zvolíme typ proměnné Formula. Do řádku Formula vepíšeme vzorec Pomocny/Počet_nehod_CR, jak je vidět na Obr. 9.



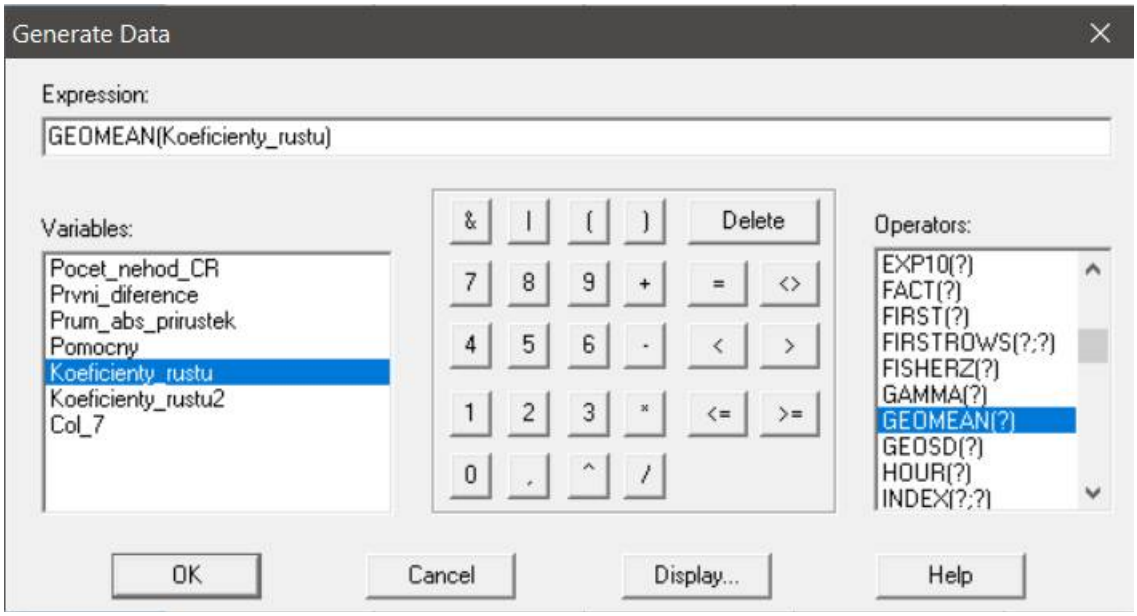

*Obrázek 9 – Zadání vzorce pro výpočet koeficientů růstu v Modify Column*

Po stisknutí klávesy OK se v podsvíceném sloupci objeví vypočítané hodnoty koeficientů růstu, jak ukazuje Obr. 10. Oproti předchozímu způsobu výpočtu jsou však šedé. To znamená, že pokud bychom např. udělali chybu ve vstupních datech, můžeme vstupní data opravit a daná oprava se promítne i do již vypočítaných hodnot.

| Koeficienty_rustu | Koeficienty_rustu2 |
|---|---|
| Numeric | Numeric |
| 0,877637684966 | 0,877637684966 |
| 0,46649748092 | 0,46649748092 |
| 1,00944997661 | 1,00944997661 |
| 0,994902147719 | 0,994902147719 |
| 1,08340764204 | 1,08340764204 |
| 1,03677951943 | 1,03677951943 |
| 1,01731083675 | 1,01731083675 |
| 1,08395159506 | 1,08395159506 |
| 1,06228845885 | 1,06228845885 |
| 1,05013958569 | 1,05013958569 |
| 1,00908294083 | 1,00908294083 |

*Obrázek 10 – Vypočítané hodnoty koeficientů růstu*

d) K výpočtu průměrného koeficientu růstu opět využijeme jeden z operátorů, protože víme, že je dán jako geometrický průměr koeficientů růstu. Podsvítíme nový sloupec a pravým tlačítkem myši opět zvolíme *Generate Data*, kde mezi operátory vybereme GEOMEAN (?). Vložíme tento operátor do řádku Expression a za otazník dosadíme proměnnou Koeficienty_rustu (nebo Koeficienty_rustu2) – viz Obr. 11. Potvrdíme OK.




*Obrázek 11 – Použití operátoru GEOMEAN v Generate Data*

V podsvíceném sloupci se objeví hodnota průměrného koeficientu růstu – viz Obr. 12.

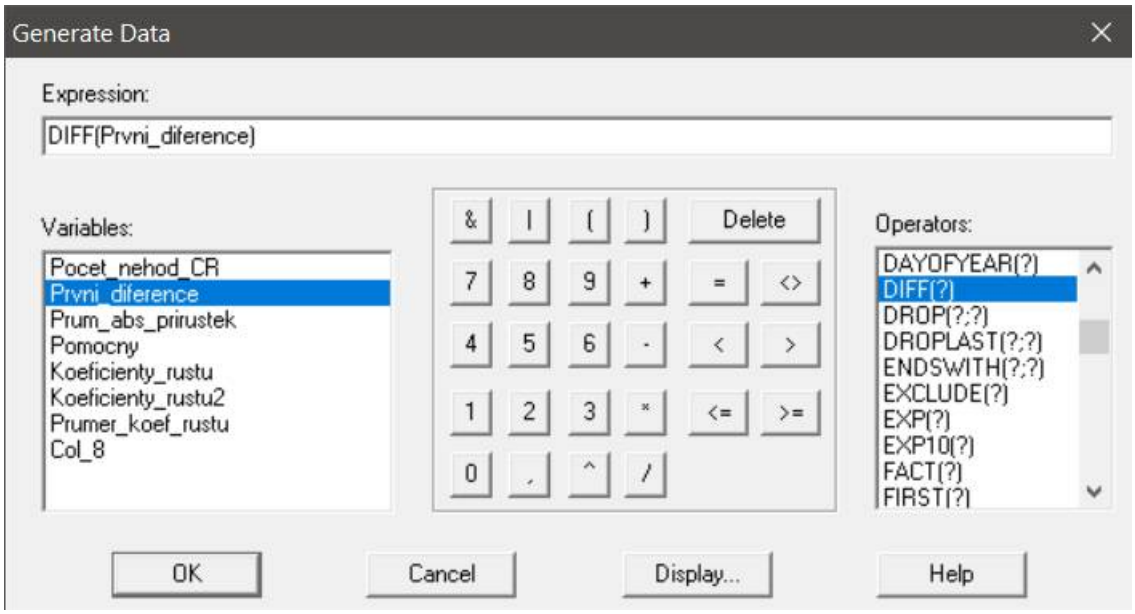
| Koeficienty_rustu | Koeficienty_rustu2 | Prumer_koef_rustu |
|---|---|---|
| Numeric | Numeric | Numeric |
| 0,877637684966 | 0,877637684966 | 0,950681995027 |
| 0,46649748092 | 0,46649748092 | |
| 1,00944997661 | 1,00944997661 | |
| 0,994902147719 | 0,994902147719 | |
| 1,08340764204 | 1,08340764204 | |
| 1,03677951943 | 1,03677951943 | |
| 1,01731083675 | 1,01731083675 | |
| 1,08395159506 | 1,08395159506 | |
| 1,06228845885 | 1,06228845885 | |
| 1,05013958569 | 1,05013958569 | |
| 1,00908294083 | 1,00908294083 | |

*Obrázek 12 – Vypočítaná hodnota průměrného koeficientu růstu*

Interpretace vypočtené hodnoty viz ruční výpočet.

## Druhé a třetí diference

Hodnoty druhých diferencí jsou definovány jako rozdíl sousedních prvních diferencí, proto k jejich výpočtu můžeme použít opět operátor DIFF. Podsvítíme tedy další sloupec, potvrdíme nabídku *Generate Data* a do řádku Expression zadáme operátor DIFF(?) a za otazník dosadíme proměnnou Prvni_diference – viz Obr. 13.




*Obrázek 13 – Použití operátoru DIFF v Generate Data*

Po potvrzení klávesou OK se v podsvíceném sloupci zobrazí vypočítané hodnoty druhých diferencí, jak je vidět na Obr. 14.

| Druhe_diference |
|---|
| Numeric |
| |
| |
| -63201 |
| 86268 |
| -1092 |
| 6652 |
| -3273 |
| -1533 |
| 5747 |
| -1411 |
| -840 |
| -4014 |

*Obrázek 14 – Vypočítané hodnoty druhých diferencí*

Hodnoty třetích diferencí získáme obdobným způsobem – jen za otazník k operátoru DIFF zadáme hodnoty druhých diferencí. Výsledné hodnoty můžeme vidět na Obr. 15.

| Treti_diference |
|---|
| Numeric |
| |
| |
| |
| 149469 |
| -87360 |
| 7744 |
| -9925 |
| 1740 |
| 7280 |
| -7158 |
| 571 |
| -3174 |

*Obrázek 15 – Vypočítané hodnoty třetích diferencí*

# Základní charakteristiky časových řad – řešený příklad

Máme k dispozici údaje o počtu dopravních nehod v ČR v letech 2007-2018. Charakterizujte vývoj počtu dopravních nehod pomocí:

a) prvních diferencí,
b) průměrného absolutního přírůstku,
c) koeficientů růstu,
d) průměrného koeficientu růstu.

Všechny výsledky interpretujte!
Doplňte též hodnoty druhých a třetích diferencí.

| Rok | Počet nehod |
|---|---|
| 2007 | 182 736 |
| 2008 | 160 376 |
| 2009 | 74 815 |
| 2010 | 75 522 |
| 2011 | 75 137 |
| 2012 | 81 404 |
| 2013 | 84 398 |
| 2014 | 85 859 |
| 2015 | 93 067 |
| 2016 | 98 864 |
| 2017 | 103 821 |
| 2018 | 104 764 |

*Zdroj: Ročenka nehodovosti na pozemních komunikacích za rok 2018*

## *Řešení:*

Všechny požadované charakteristiky časových řad je poměrně jednoduché spočítat i na kalkulačce, proto je zde uveden postup ručního výpočtu a následně i pomocí programu SGP.

a) První diference jsou konstruovány jako rozdíl hodnoty ukazatele v časovém okamžiku *t* a hodnoty ukazatele v okamžiku bezprostředně předcházejícím časovému okamžiku *t*, tedy *t-1*: $\Delta_t^{(1)} = y_t - y_{t-1}$. Hodnoty všech počítaných charakteristik budou doplněny postupně do následující tabulky.
   První řádek zůstane nevyplněn, protože hodnota *t* je rok 2007 a předcházející údaj není k dispozici. Tečka v řádku značí, že údaj existuje, ale nejsou k dispozici data k jeho výpočtu.
   Další hodnota je vypočtena jako $\Delta_2^{(1)} = y_2 - y_1 = 160376 - 182736 = -22360$.
   Tímto způsobem pokračujeme dále: $\Delta_3^{(1)} = y_3 - y_2 = 74815 - 160376 = -85561$.
   $\Delta_4^{(1)} = y_4 - y_3 = 75522 - 74815 = 707$
   $\Delta_5^{(1)} = y_5 - y_4 = 75137 - 75522 = -385$
   $\Delta_6^{(1)} = y_6 - y_5 = 81404 - 75137 = 6267$

$\Delta_7^{(1)} = y_7 - y_6 = 84398 - 81404 = 2994$
$\Delta_8^{(1)} = y_8 - y_7 = 85859 - 84398 = 1461$
$\Delta_9^{(1)} = y_9 - y_8 = 93067 - 85859 = 7208$
$\Delta_{10}^{(1)} = y_{10} - y_9 = 98864 - 93067 = 5797$
$\Delta_{11}^{(1)} = y_{11} - y_{10} = 103821 - 98864 = 4957$
$\Delta_{12}^{(1)} = y_{12} - y_{11} = 104764 - 103821 = 943$

| t | Rok | Počet nehod $y_t$ | $\Delta_t^{(1)}$ |
|---|---|---|---|
| 1 | 2007 | 182 736 | • |
| 2 | 2008 | 160 376 | -22 360 |
| 3 | 2009 | 74 815 | -85 561 |
| 4 | 2010 | 75 522 | 707 |
| 5 | 2011 | 75 137 | -385 |
| 6 | 2012 | 81 404 | 6 267 |
| 7 | 2013 | 84 398 | 2 994 |
| 8 | 2014 | 85 859 | 1 461 |
| 9 | 2015 | 93 067 | 7 208 |
| 10 | 2016 | 98 864 | 5 797 |
| 11 | 2017 | 103 821 | 4 957 |
| 12 | 2018 | 104 764 | 943 |
| **Celkem** | | | -77972 |

***Interpretace:***
Počet nehod v ČR se v roce 2008 snížil oproti roku 2007 o 22 360.
Počet nehod v ČR se v roce 2009 snížil oproti roku 2008 o 85 561.
Počet nehod v ČR se v roce 2010 zvýšil oproti roku 2009 o 707.
Počet nehod v ČR se v roce 2011 snížil oproti roku 2010 o 385.
Počet nehod v ČR se v roce 2012 zvýšil oproti roku 2011 o 6267.
atp.

b) Průměrný absolutní přírůstek je možné vyjádřit jako průměr prvních diferencí (tj. součet prvních diferencí dělených jejich počtem) nebo jako rozdíl poslední a první hodnoty časové řady dělený *n-1*. Oba způsoby výpočtu musejí dát pochopitelně stejný výsledek.

$$\bar{\Delta} = \frac{\sum_{t=2}^{n} \Delta_t^{(1)}}{n-1} = \frac{-77972}{11} = -7088{,}364$$

NEBO

$$\bar{\bar{\Delta}} = \frac{y_n - y_1}{n-1} = \frac{104764 - 182736}{11} = -7088{,}364$$

***Interpretace:***
Počet nehod v ČR se v letech 2007-2018 v průměru snížil o 7088,364 nehody.

c) Koeficient růstu je konstruován jako podíl hodnoty ukazatele v časovém okamžiku *t* a hodnoty ukazatele v časovém okamžiku *t-1*.

$$k_t = \frac{y_t}{y_{t-1}}$$

První řádek zůstane opět nevyplněn (jako u prvních diferencí), protože hodnota *t* je rok 2007 a předcházející údaj není k dispozici. Tečka v řádku značí, že údaj existuje, ale nejsou k dispozici data k jeho výpočtu.

$$k_2 = \frac{y_2}{y_1} = \frac{160376}{182736} = 0{,}8776$$
$$k_3 = \frac{y_3}{y_2} = \frac{74815}{160376} = 0{,}4665$$
$$k_4 = \frac{y_4}{y_3} = \frac{75522}{74815} = 1{,}0094$$
$$k_5 = \frac{y_5}{y_4} = \frac{75137}{75522} = 0{,}9949$$
$$k_6 = \frac{y_6}{y_5} = \frac{81404}{75137} = 1{,}0834$$
$$k_7 = \frac{y_7}{y_6} = \frac{84398}{81404} = 1{,}0368$$
$$k_8 = \frac{y_8}{y_7} = \frac{85859}{84398} = 1{,}0173$$
$$k_9 = \frac{y_9}{y_8} = \frac{93067}{85859} = 1{,}0840$$
$$k_{10} = \frac{y_{10}}{y_9} = \frac{98864}{93067} = 1{,}0623$$
$$k_{11} = \frac{y_{11}}{y_{10}} = \frac{103821}{98864} = 1{,}0501$$
$$k_{12} = \frac{y_{12}}{y_{11}} = \frac{104764}{103821} = 1{,}0091$$

| **t** | **Rok** | **Počet nehod $y_t$** | **$k_t$** |
|---|---|---|---|
| 1 | 2007 | 182 736 | • |
| 2 | 2008 | 160 376 | 0,8776 |
| 3 | 2009 | 74 815 | 0,4665 |
| 4 | 2010 | 75 522 | 1,0094 |
| 5 | 2011 | 75 137 | 0,9949 |
| 6 | 2012 | 81 404 | 1,0834 |
| 7 | 2013 | 84 398 | 1,0368 |
| 8 | 2014 | 85 859 | 1,0173 |
| 9 | 2015 | 93 067 | 1,0840 |
| 10 | 2016 | 98 864 | 1,0623 |
| 11 | 2017 | 103 821 | 1,0501 |
| 12 | 2018 | 104 764 | 1,0091 |

***Interpretace:***
Počet nehod v ČR se v roce 2008 snížil oproti roku 2007 o 12,24 % (0,8776*100-100).
Počet nehod v ČR se v roce 2009 snížil oproti roku 2008 o 53,35 %.
Počet nehod v ČR se v roce 2010 zvýšil oproti roku 2009 o 0,94 %.
Počet nehod v ČR se v roce 2011 snížil oproti roku 2010 o 0,51 %.
Počet nehod v ČR se v roce 2012 zvýšil oproti roku 2011 o 8,34 %.
atp.

d) Průměrný koeficient růstu je konstruován jako geometrický průměr koeficientů růstu. Lze ale také vypočítat jako *n-1*ní odmocnina z podílu poslední a první hodnoty časové řady.

$$\bar{k} = \sqrt[n-1]{k_2 \cdot k_3 \cdot \ldots \cdot k_n}$$

$$\bar{k} = \sqrt[11]{0{,}8776 \cdot 0{,}4665 \cdot 1{,}0094 \cdot \ldots \cdot 1{,}0091} = 0{,}9507$$

NEBO

$$\bar{k} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

$$\bar{k} = \sqrt[11]{\frac{104764}{182736}} = 0{,}9507$$

***Interpretace:***
Počet nehod v ČR se v letech 2007-2018 v průměru snížil o 4,93 %.

## Druhé a třetí diference

Druhé diference jsou definovány jako rozdíl sousedních prvních diferencí.
$\Delta_t^{(2)} = \Delta_t^{(1)} - \Delta_{t-1}^{(1)}$
První dva údaje není možné z dat, která máme k dispozici, vypočítat, proto opět píšeme do prvních dvou řádků puntík. Další hodnoty vypočítáme následovně:

$$\Delta_3^{(2)} = \Delta_3^{(1)} - \Delta_2^{(1)} = -85561 - (-22360) = -63201$$
$$\Delta_4^{(2)} = \Delta_4^{(1)} - \Delta_3^{(1)} = 707 - (-85561) = 86268$$
$$\Delta_5^{(2)} = \Delta_5^{(1)} - \Delta_4^{(1)} = -385 - 707 = -1092$$
$$\Delta_6^{(2)} = \Delta_6^{(1)} - \Delta_5^{(1)} = 6267 - (-385) = 6652$$
$$\Delta_7^{(2)} = \Delta_7^{(1)} - \Delta_6^{(1)} = 2994 - 6267 = -3273$$
$$\Delta_8^{(2)} = \Delta_8^{(1)} - \Delta_7^{(1)} = 1461 - 2994 = -1533$$
$$\Delta_9^{(2)} = \Delta_9^{(1)} - \Delta_8^{(1)} = 7208 - 1461 = 5747$$
$$\Delta_{10}^{(2)} = \Delta_{10}^{(1)} - \Delta_9^{(1)} = 5797 - 7208 = -1411$$
$$\Delta_{11}^{(2)} = \Delta_{11}^{(1)} - \Delta_{10}^{(1)} = 4957 - 5797 = -840$$
$$\Delta_{12}^{(2)} = \Delta_{12}^{(1)} - \Delta_{11}^{(1)} = 104764 - 103821 = -4014$$

| t | Rok | Počet nehod $y_t$ | $\Delta_t^{(1)}$ | $\Delta_t^{(2)}$ |
|---|---|---|---|---|
| 1 | 2007 | 182 736 | • | • |
| 2 | 2008 | 160 376 | -22 360 | • |
| 3 | 2009 | 74 815 | -85 561 | -63 201 |
| 4 | 2010 | 75 522 | 707 | 86 268 |
| 5 | 2011 | 75 137 | -385 | -1 092 |
| 6 | 2012 | 81 404 | 6 267 | 6 652 |
| 7 | 2013 | 84 398 | 2 994 | -3 273 |
| 8 | 2014 | 85 859 | 1 461 | -1 533 |
| 9 | 2015 | 93 067 | 7 208 | 5747 |
| 10 | 2016 | 98 864 | 5 797 | -1 411 |
| 11 | 2017 | 103 821 | 4 957 | -840 |
| 12 | 2018 | 104 764 | 943 | -4014 |

Třetí diference jsou definovány jako rozdíl sousedních druhých diferencí.
$\Delta_t^{(3)} = \Delta_t^{(2)} - \Delta_{t-1}^{(2)}$

První tři údaje není možné z dat, která máme k dispozici, vypočítat, proto opět píšeme do prvních tří řádků puntík. Další hodnoty už je možné spočítat.

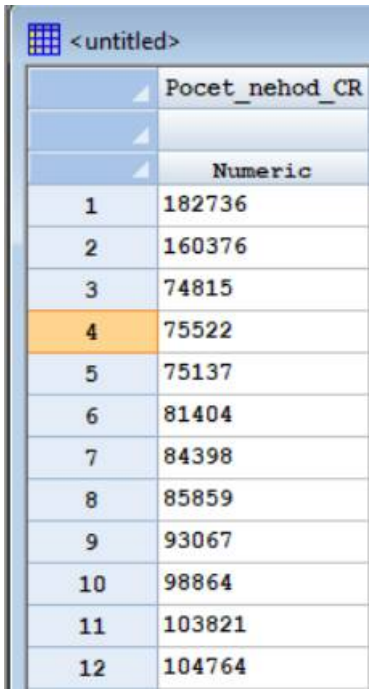
| t | Rok | Počet nehod $y_t$ | $\Delta_t^{(2)}$ | $\Delta_t^{(3)}$ |
|---|---|---|---|---|
| 1 | 2007 | 182 736 | • | • |
| 2 | 2008 | 160 376 | • | • |
| 3 | 2009 | 74 815 | -63 201 | • |
| 4 | 2010 | 75 522 | 86 268 | 149 469 |
| 5 | 2011 | 75 137 | -1 092 | -87 360 |
| 6 | 2012 | 81 404 | 6 652 | 7 744 |
| 7 | 2013 | 84 398 | -3 275 | -9 925 |
| 8 | 2014 | 85 859 | -1 533 | 1 740 |
| 9 | 2015 | 93 067 | 5747 | 7 280 |
| 10 | 2016 | 98 864 | -1 411 | -7 158 |
| 11 | 2017 | 103 821 | -840 | 571 |
| 12 | 2018 | 104 764 | -4014 | -3 174 |

$\Delta_4^{(3)} = \Delta_4^{(2)} - \Delta_3^{(2)} = 86268 - (-63201) = 149469$
$\Delta_5^{(3)} = \Delta_5^{(2)} - \Delta_4^{(2)} = -1092 - 86268 = -87360$
$\Delta_6^{(3)} = \Delta_6^{(2)} - \Delta_5^{(2)} = 6652 - (-1092) = 7744$
$\Delta_7^{(3)} = \Delta_7^{(2)} - \Delta_6^{(2)} = -3275 - 6652 = -9925$
$\Delta_8^{(3)} = \Delta_8^{(2)} - \Delta_7^{(2)} = -1533 - (-3275) = 1740$
$\Delta_9^{(3)} = \Delta_9^{(2)} - \Delta_8^{(2)} = 5747 - (-1533) = 7280$
$\Delta_{10}^{(3)} = \Delta_{10}^{(2)} - \Delta_9^{(2)} = -1411 - 5747 = -7158$
$\Delta_{11}^{(3)} = \Delta_{11}^{(2)} - \Delta_{10}^{(2)} = -840 - (-1411) = 571$
$\Delta_{12}^{(3)} = \Delta_{12}^{(2)} - \Delta_{11}^{(2)} = -4014 - (-840) = -3174$

Druhé a třetí diference se využívají při hledání vhodného modelu trendu (bude předmětem následujícího tématu).

## Řešení v SGP:

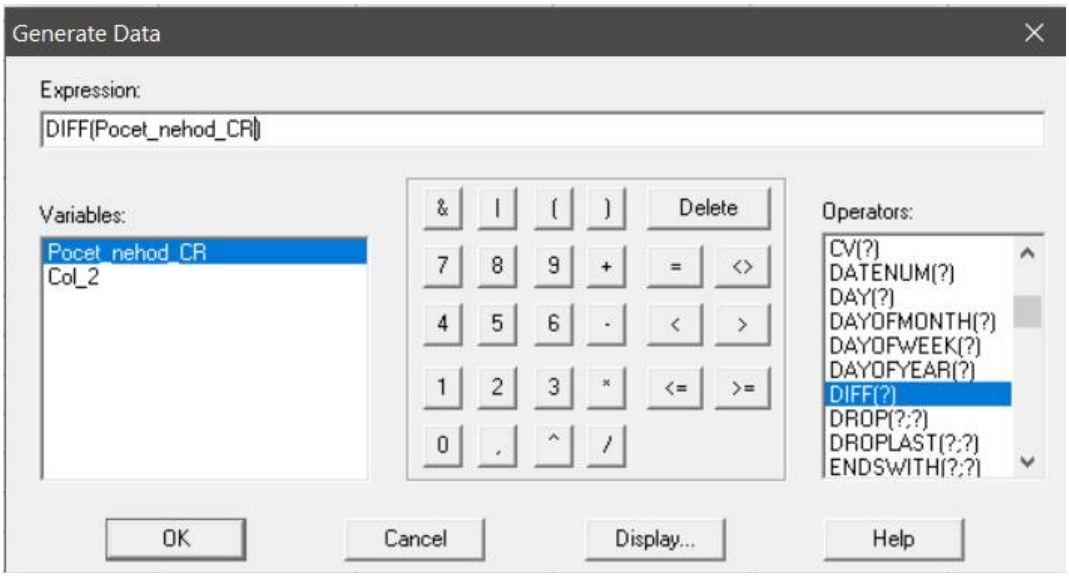
Prvním krokem při řešení příkladu v SGP je vložení vstupních dat. Údaje o počtu nehod v ČR vložíme jako jednu proměnnou, tj. do jednoho sloupce – viz Obr. 1.




| | Pocet_nehod_CR |
|---|---|
| | |
| | Numeric |
| 1 | 182736 |
| 2 | 160376 |
| 3 | 74815 |
| 4 | 75522 |
| 5 | 75137 |
| 6 | 81404 |
| 7 | 84398 |
| 8 | 85859 |
| 9 | 93067 |
| 10 | 98864 |
| 11 | 103821 |
| 12 | 104764 |

*Obrázek 1 – Vložení vstupních dat*

a) Pro výpočet prvních diferencí přeskočíme kurzorem na další sloupeček, levým tlačítkem ho podsvítíme a pravým tlačítkem myši zvolíme z nabídky procedur *Generate Data*. Ze sloupečku *Operators* v pravé části panelu vybereme operátor DIFF(?), který dvojím kliknutím dosadíme do řádku Expression, a za otazník zadáme proměnnou Počet_nehod_CR (dvojím kliknutím z tabulky Variables vlevo) – viz Obr. 2. Potvrdíme klávesou OK. Hodnoty prvních diferencí se objeví ve sloupci 2 – na Obr. 2 je už přejmenovaný na První_diference, aby bylo zřejmé, o jaké hodnoty se jedná.




*Obr. 2 – Zadání operátoru DIFF v Generate Data*

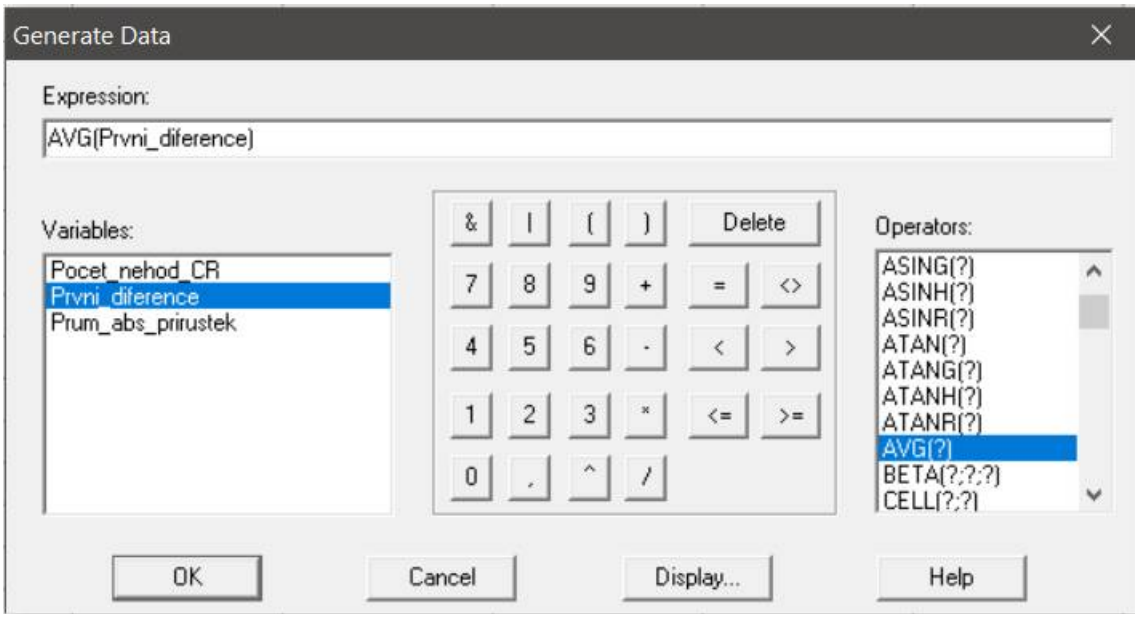
<untitled>

| | Pocet_nehod_CR | Prvni_diference |
|---|---|---|
| | Numeric | Numeric |
| 1 | 182736 | |
| 2 | 160376 | -22360 |
| 3 | 74815 | -85561 |
| 4 | 75522 | 707 |
| 5 | 75137 | -385 |
| 6 | 81404 | 6267 |
| 7 | 84398 | 2994 |
| 8 | 85859 | 1461 |
| 9 | 93067 | 7208 |
| 10 | 98864 | 5797 |
| 11 | 103821 | 4957 |
| 12 | 104764 | 943 |

*Obrázek 3 – Výsledek použití operátoru DIFF*

Interpretace vypočtených hodnot zde již není uvedena – je ji možné najít v první části řešeného příkladu, kde je ukázka ručního výpočtu.

b) Pro výpočet průměrného absolutního přírůstku využijeme rovněž operátoru v nabídce *Generate Data*. Víme, že průměrný absolutní přírůstek je aritmetický průměr prvních diferencí, proto použijeme operátor AVG. Třetí sloupec si nazveme Prum_abs_prirustek, podsvítíme ho a v nabídce *Generate Data* zadáme do řádku Expression operátor AVG(?). Za otazník zadáme proměnnou První_diference, jak ukazuje Obr. 4.




*Obrázek 4 – Použití operátoru AVG v Generate Data*

Po potvrzení klávesou OK, se objeví výsledná hodnota v podsvíceném sloupci – viz Obr. 5.

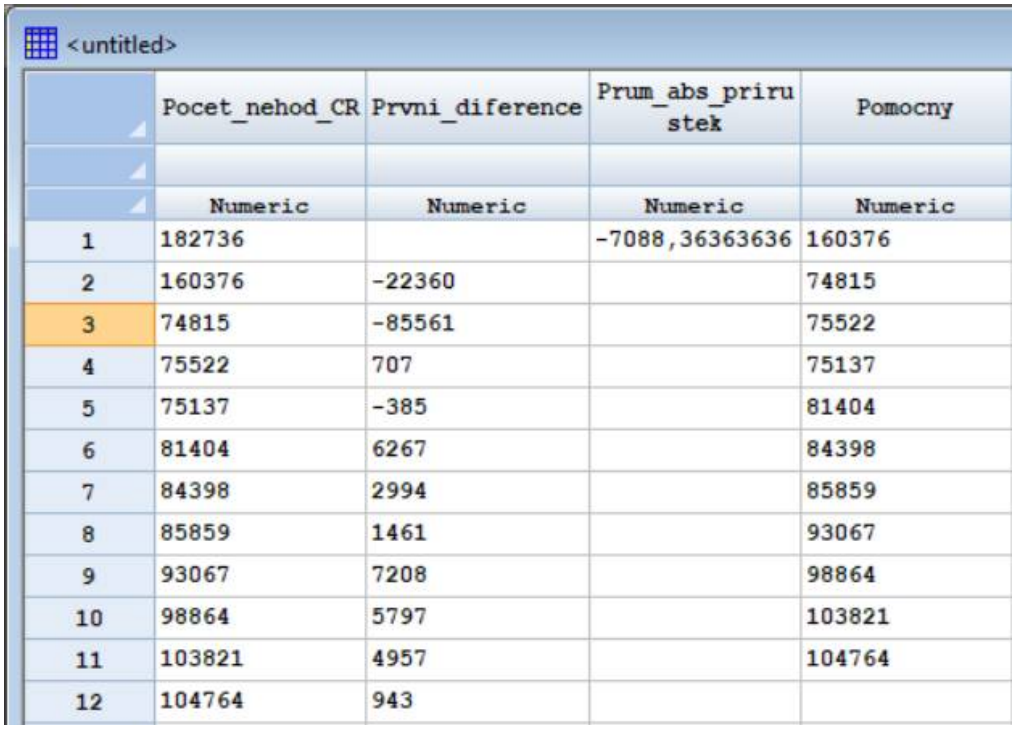
<untitled>

| | Pocet_nehod_CR | Prvni_diference | Prum_abs_prirustek |
|---|---|---|---|
| | Numeric | Numeric | Numeric |
| 1 | 182736 | | -7088,36363636 |
| 2 | 160376 | -22360 | |
| 3 | 74815 | -85561 | |
| 4 | 75522 | 707 | |
| 5 | 75137 | -385 | |
| 6 | 81404 | 6267 | |
| 7 | 84398 | 2994 | |
| 8 | 85859 | 1461 | |
| 9 | 93067 | 7208 | |
| 10 | 98864 | 5797 | |
| 11 | 103821 | 4957 | |
| 12 | 104764 | 943 | |

*Obrázek 5 – Vypočítaná hodnota průměrného absolutního přírůstku*

Interpretace vypočtené hodnoty viz ruční řešení.

c) Koeficienty růstu je možné v SGP vypočítat dvěma základními způsoby. První z nich opět využívá nabídku *Generate Data*. Nejprve je ale potřeba před výpočtem vytvořit pomocný sloupec s hodnotami proměnné Počet_nehod_CR bez první hodnoty. Viz Obr. 6.

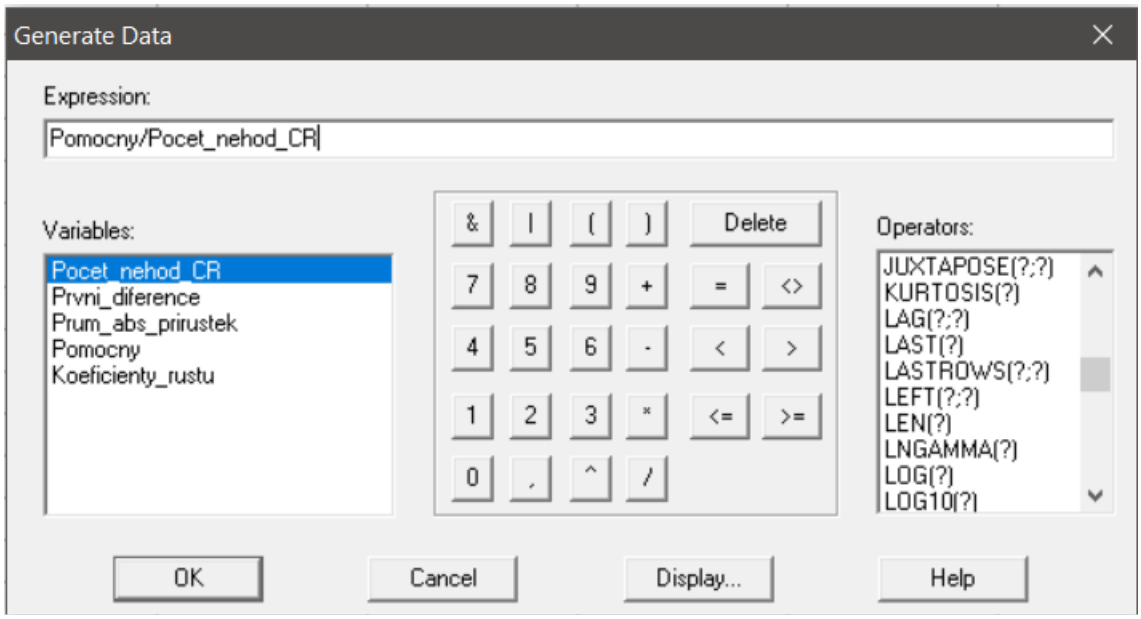


<untitled>

| | Pocet_nehod_CR | Prvni_diference | Prum_abs_prirustek | Pomocny |
|---|---|---|---|---|
| | | | | |
| | Numeric | Numeric | Numeric | Numeric |
| 1 | 182736 | | -7088,36363636 | 160376 |
| 2 | 160376 | -22360 | | 74815 |
| 3 | 74815 | -85561 | | 75522 |
| 4 | 75522 | 707 | | 75137 |
| 5 | 75137 | -385 | | 81404 |
| 6 | 81404 | 6267 | | 84398 |
| 7 | 84398 | 2994 | | 85859 |
| 8 | 85859 | 1461 | | 93067 |
| 9 | 93067 | 7208 | | 98864 |
| 10 | 98864 | 5797 | | 103821 |
| 11 | 103821 | 4957 | | 104764 |
| 12 | 104764 | 943 | | |

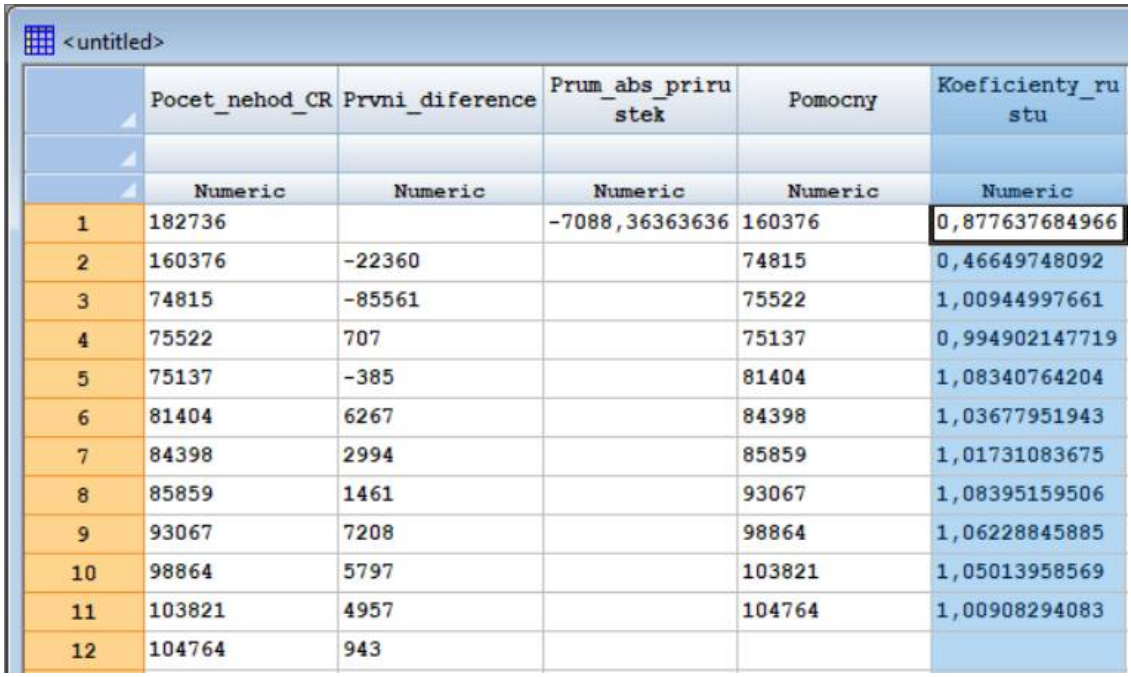
*Obrázek 6 – Vložení pomocné proměnné*

Kurzorem přeskočíme do dalšího volného sloupce, který podsvítíme, a v panelu *Generate Data* zadáme do řádku Expression vzorec „Pomocny/Počet_nehod_CR“, jak ukazuje Obr. 7.




*Obrázek 7 – Zadání vzorce pro výpočet koeficientů růstu v Generate Data*

Po stisknutí klávesy OK se hodnoty koeficientů růstu objeví v podsvíceném sloupci – viz Obr. 8 (už je zde přejmenován na Koeficienty_rustu).



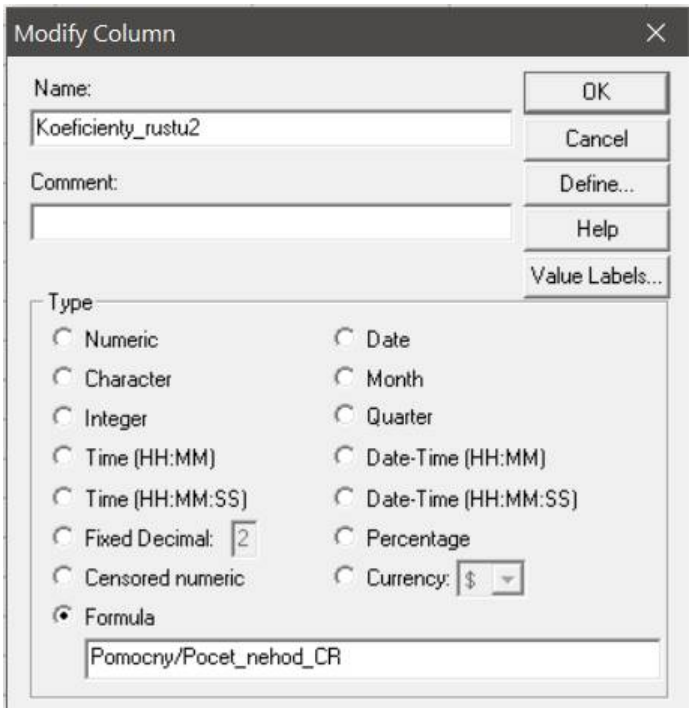
<untitled>

| | Pocet_nehod_CR | Prvni_diference | Prum_abs_prirustek | Pomocny | Koeficienty_rustu |
|---|---|---|---|---|---|
| | Numeric | Numeric | Numeric | Numeric | Numeric |
| 1 | 182736 | | -7088,36363636 | 160376 | 0,877637684966 |
| 2 | 160376 | -22360 | | 74815 | 0,46649748092 |
| 3 | 74815 | -85561 | | 75522 | 1,00944997661 |
| 4 | 75522 | 707 | | 75137 | 0,994902147719 |
| 5 | 75137 | -385 | | 81404 | 1,08340764204 |
| 6 | 81404 | 6267 | | 84398 | 1,03677951943 |
| 7 | 84398 | 2994 | | 85859 | 1,01731083675 |
| 8 | 85859 | 1461 | | 93067 | 1,08395159506 |
| 9 | 93067 | 7208 | | 98864 | 1,06228845885 |
| 10 | 98864 | 5797 | | 103821 | 1,05013958569 |
| 11 | 103821 | 4957 | | 104764 | 1,00908294083 |
| 12 | 104764 | 943 | | | |

*Obrázek 8 – Vypočítané hodnoty koeficientů růstu*

Interpretace vypočtených hodnot viz ruční řešení.

Druhý způsob výpočtu koeficientů růstu využívá nabídku *Modify Column*, kterou je možné vyvolat stejným způsobem jako *Generate Data*, tj. podsvítit další sloupec a pravým tlačítkem vyvolat příslušnou nabídku. V panelu *Modify Column* vyplníme nejprve název nové proměnné Koeficienty_rustu2 a pak zvolíme typ proměnné Formula. Do řádku Formula vepíšeme vzorec Pomocny/Počet_nehod_CR, jak je vidět na Obr. 9.



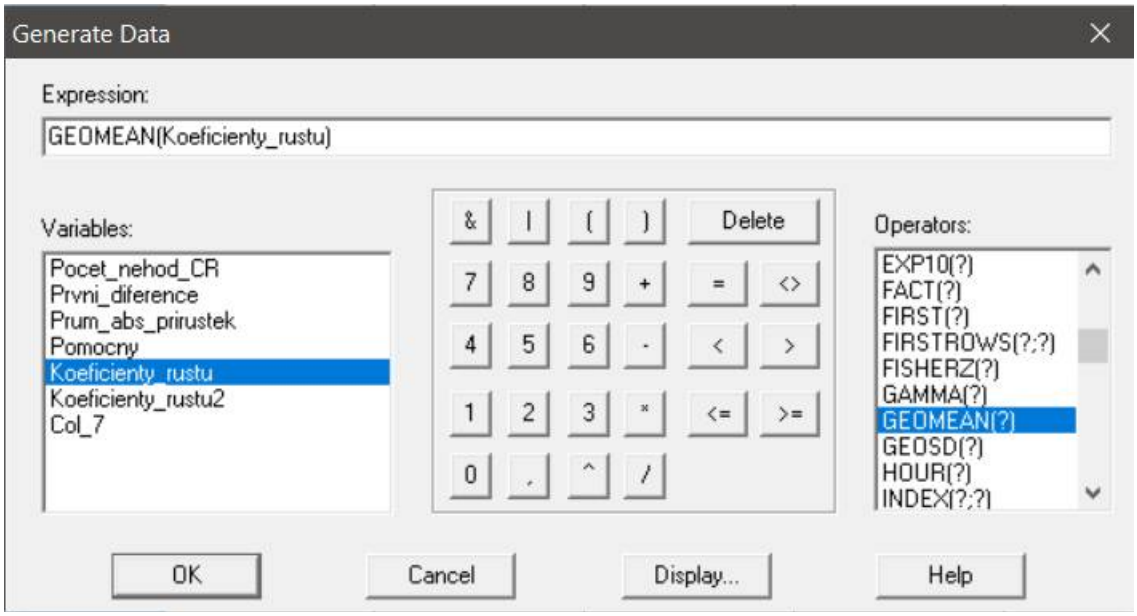

*Obrázek 9 – Zadání vzorce pro výpočet koeficientů růstu v Modify Column*

Po stisknutí klávesy OK se v podsvíceném sloupci objeví vypočítané hodnoty koeficientů růstu, jak ukazuje Obr. 10. Oproti předchozímu způsobu výpočtu jsou však šedé. To znamená, že pokud bychom např. udělali chybu ve vstupních datech, můžeme vstupní data opravit a daná oprava se promítne i do již vypočítaných hodnot.

| Koeficienty_rustu | Koeficienty_rustu2 |
|---|---|
| Numeric | Numeric |
| 0,877637684966 | 0,877637684966 |
| 0,46649748092 | 0,46649748092 |
| 1,00944997661 | 1,00944997661 |
| 0,994902147719 | 0,994902147719 |
| 1,08340764204 | 1,08340764204 |
| 1,03677951943 | 1,03677951943 |
| 1,01731083675 | 1,01731083675 |
| 1,08395159506 | 1,08395159506 |
| 1,06228845885 | 1,06228845885 |
| 1,05013958569 | 1,05013958569 |
| 1,00908294083 | 1,00908294083 |

*Obrázek 10 – Vypočítané hodnoty koeficientů růstu*

d) K výpočtu průměrného koeficientu růstu opět využijeme jeden z operátorů, protože víme, že je dán jako geometrický průměr koeficientů růstu. Podsvítíme nový sloupec a pravým tlačítkem myši opět zvolíme *Generate Data*, kde mezi operátory vybereme GEOMEAN (?). Vložíme tento operátor do řádku Expression a za otazník dosadíme proměnnou Koeficienty_rustu (nebo Koeficienty_rustu2) – viz Obr. 11. Potvrdíme OK.




*Obrázek 11 – Použití operátoru GEOMEAN v Generate Data*

V podsvíceném sloupci se objeví hodnota průměrného koeficientu růstu – viz Obr. 12.

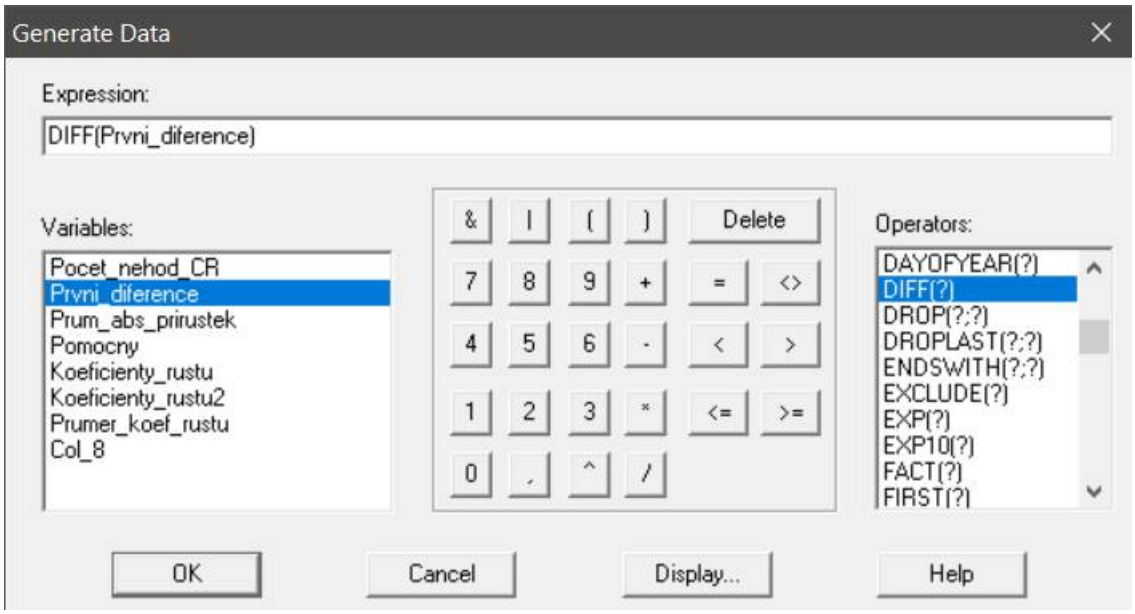
| Koeficienty_rustu | Koeficienty_rustu2 | Prumer_koef_rustu |
|---|---|---|
| Numeric | Numeric | Numeric |
| 0,877637684966 | 0,877637684966 | 0,950681995027 |
| 0,46649748092 | 0,46649748092 | |
| 1,00944997661 | 1,00944997661 | |
| 0,994902147719 | 0,994902147719 | |
| 1,08340764204 | 1,08340764204 | |
| 1,03677951943 | 1,03677951943 | |
| 1,01731083675 | 1,01731083675 | |
| 1,08395159506 | 1,08395159506 | |
| 1,06228845885 | 1,06228845885 | |
| 1,05013958569 | 1,05013958569 | |
| 1,00908294083 | 1,00908294083 | |

*Obrázek 12 – Vypočítaná hodnota průměrného koeficientu růstu*

Interpretace vypočtené hodnoty viz ruční výpočet.

## Druhé a třetí diference

Hodnoty druhých diferencí jsou definovány jako rozdíl sousedních prvních diferencí, proto k jejich výpočtu můžeme použít opět operátor DIFF. Podsvítíme tedy další sloupec, potvrdíme nabídku *Generate Data* a do řádku Expression zadáme operátor DIFF(?) a za otazník dosadíme proměnnou Prvni_diference – viz Obr. 13.




*Obrázek 13 – Použití operátoru DIFF v Generate Data*

Po potvrzení klávesou OK se v podsvíceném sloupci zobrazí vypočítané hodnoty druhých diferencí, jak je vidět na Obr. 14.

| Druhe_diference |
|---|
| Numeric |
| |
| |
| -63201 |
| 86268 |
| -1092 |
| 6652 |
| -3273 |
| -1533 |
| 5747 |
| -1411 |
| -840 |
| -4014 |

*Obrázek 14 – Vypočítané hodnoty druhých diferencí*

Hodnoty třetích diferencí získáme obdobným způsobem – jen za otazník k operátoru DIFF zadáme hodnoty druhých diferencí. Výsledné hodnoty můžeme vidět na Obr. 15.

| Treti_diference |
|---|
| Numeric |
| |
| |
| |
| 149469 |
| -87360 |
| 7744 |
| -9925 |
| 1740 |
| 7280 |
| -7158 |
| 571 |
| -3174 |

*Obrázek 15 – Vypočítané hodnoty třetích diferencí*

# Téma 11 – Příklad 2

## Zadání příkladu:

V následující tabulce jsou k dispozici údaje o cenách a prodeji 2 odrůdových vín z Vinařství Polívka v roce 2014 a 2015. Určete, jak se ve sledovaném období změnila cena, počet prodaných láhví a tržba za prodej odrůdy Veltlínské zelené, a jak za odrůdu Cabernet Moravia. Uvažujte relativní i absolutní vyjádření změn. Všechny výsledky interpretujte.

| **Odrůda** | **Cena v Kč za 0,75l** | | **Počet prodaných láhví (0,75l)** | |
|---|---|---|---|---|
| | **2014** | **2015** | **2014** | **2015** |
| **Veltlínské zelené** | 190 | 212 | 3500 | 3200 |
| **Cabernet Moravia** | 200 | 212 | 4100 | 4500 |

## Vypracování příkladu:

Vzhledem k tomu, že chceme určit změnu jednotlivých ukazatelů zvlášť pro každou odrůdu, je zřejmé, že srovnáváme hodnoty ukazatele ve dvou obdobích, kdy ukazatel není složen z dílčích částí. V takovém případě je pro relativní vyjádření změn jednotlivých ukazatelů vhodné použít individuální jednoduché indexy a pro absolutní vyjádření změn individuální jednoduché rozdíly.

Nejprve budeme charakterizovat změnu všech ukazatelů pro odrůdu **Veltlínské zelené**:

*a) Změna ceny:*

$$i_p = \frac{p_1}{p_0} = \frac{212}{190} = 1{,}116 \ \ (+11{,}6\ \%)$$

$$\Delta_p = p_1 - p_0 = 212 - 190 = 22 \ \ (+22\ \text{Kč})$$

*b) Změna počtu prodaných láhví:*

$$i_q = \frac{q_1}{q_0} = \frac{3200}{3500} = 0{,}914 \ \ (-8{,}6\ \%)$$

$$\Delta_q = q_1 - q_0 = 3200 - 3500 = -300 \ \ (-300\ \text{ks})$$

*c) Změna tržeb:*

Pro výpočet změny tržeb je potřeba provést pomocný výpočet, při kterém vycházíme z toho, že platí: $Q = p \cdot q$ .

| **Odrůda** | **Cena v Kč za 0,75l** | | **Počet prodaných láhví (0,75l)** | | **Tržby v Kč** | |
|---|---|---|---|---|---|---|
| | **2014** | **2015** | **2014** | **2015** | **2014** | **2015** |
| | $p_0$ | $p_1$ | $q_0$ | $q_1$ | $Q_0 = p_0 \cdot q_0$ | $Q_1 = p_1 \cdot q_1$ |
| **Veltlínské zelené** | 190 | 212 | 3 500 | 3 200 | 665 000 | 678 400 |
| **Cabernet Moravia** | 200 | 212 | 4 100 | 4 500 | | |

$$i_Q = \frac{Q_1}{Q_0} = \frac{678400}{665000} = 1{,}020\,(+2\ \%)$$

$$\Delta_Q = Q_1 - Q_0 = 678400 - 665000 = 13400\ \ (+13\ 400\ \text{Kč})$$

## Interpretace:

Cena láhve Veltlínského zeleného se v roce 2015 zvýšila oproti roku 2014 o 11,6 %, což představuje částku 22 Kč.

Počet prodaných láhví Veltlínského zeleného se v roce 2015 snížil oproti roku 2014 o 8,6 %, což představuje 300 ks láhví.

Tržba za prodej Veltlínského zeleného se v roce 2015 zvýšila oproti roku 2014 o 2 %, což představuje částku 13 400 Kč.

Nyní budeme charakterizovat změnu všech ukazatelů pro odrůdu **Cabernet Moravia**:

*a) Změna ceny:*

$$i_p = \frac{p_1}{p_0} = \frac{212}{200} = 1{,}060\ \ (+6\ \%)$$

$$\Delta_p = p_1 - p_0 = 212 - 200 = 12\ \ (+12\ \text{Kč})$$

*b) Změna počtu prodaných láhví:*

$$i_q = \frac{q_1}{q_0} = \frac{4500}{4100} = 1{,}098\ \ (+9{,}8\ \%)$$

$$\Delta_q = q_1 - q_0 = 4500 - 4100 = 400\ \ (+400\ \text{ks})$$

*c) Změna tržeb:*

Pro výpočet změny tržeb je opět potřeba provést pomocný výpočet, při kterém vycházíme z toho, že platí: $Q = p \cdot q$ .

| Odrůda | Cena v Kč za 0,75l | | Počet prodaných láhví (0,75l) | | Tržby v Kč | |
|---|---|---|---|---|---|---|
| | 2014 | 2015 | 2014 | 2015 | 2014 | 2015 |
| | $p_0$ | $p_1$ | $q_0$ | $q_1$ | $Q_0 = p_0 \cdot q_0$ | $Q_1 = p_1 \cdot q_1$ |
| **Veltlínské zelené** | 190 | 212 | 3 500 | 3 200 | 665 000 | 678 400 |
| **Cabernet Moravia** | 200 | 212 | 4 100 | 4 500 | 820 000 | 954 000 |

$$i_Q = \frac{Q_1}{Q_0} = \frac{954000}{820000} = 1{,}163\,(+16{,}3\ \%)$$

$$\Delta_Q = Q_1 - Q_0 = 954000 - 820000 = 134000\ \ (+134\ 000\ \text{Kč})$$

## Interpretace:

Cena láhve Cabernetu Sauvignon se v roce 2015 zvýšila oproti roku 2014 o 6 %, což představuje částku 12 Kč.

Počet prodaných láhví Cabernetu Sauvignon se v roce 2015 zvýšil oproti roku 2014 o 9,8 %, což představuje 400 ks láhví.

Tržba za prodej Cabernetu Sauvignon se v roce 2015 zvýšila oproti roku 2014 o 16,3 %, což představuje částku 134 000 Kč.

## Poznámka:

V programu STATGRAPHICS ani v MS Excel není žádná speciální procedura pro výpočet individuálních jednoduchých indexů a rozdílů.

# Téma 11 – Příklad 3

## Zadání příkladu:

Moravské vinařství prodává svá vína ve vlastní prodejně, přes sklad v Brně a v Praze. V následující tabulce jsou údaje o ceně ve všech 3 prodejních místech a počtu prodaných láhví (0,75 l) odrůdového vína Rulandské šedé v letech 2013 a 2014. Vypočítejte, jak se změnila průměrná cena vína, celkové prodané množství láhví a celkové tržby za prodej Rulandského šedého ve sledovaném období (v absolutním i relativním vyjádření). Všechny výsledky interpretujte.

| Prodejní místo | Cena v Kč za 0,75l | | Počet prodaných ks (0,75l) | |
|---|---|---|---|---|
| | 2013 | 2014 | 2013 | 2014 |
| Vlastní prodejna | 200 | 220 | 560 | 710 |
| Sklad Brno | 240 | 265 | 2 620 | 2 300 |
| Sklad Praha | 250 | 280 | 5 550 | 6 000 |

## Vypracování příkladu:

Vzhledem k tomu, že chceme určit změnu jednotlivých ukazatelů souhrnně za danou odrůdu, je zřejmé, že srovnáváme hodnoty ukazatele ve dvou obdobích, kdy ukazatel je složen z dílčích částí, a tyto dílčí části je potřeba nějakým způsobem shrnout. Zde pracujeme se stejnorodým ukazatelem, proto je pro relativní vyjádření změn jednotlivých ukazatelů vhodné použít individuální složené indexy a pro absolutní vyjádření změn individuální složené rozdíly.

*a) Změna průměrné ceny:*

Pro určení změny průměrné ceny je potřeba provést další doplňující výpočty:

| Prodejní místo | Cena v Kč za 0,75l | | Počet prodaných ks (0,75l) | | Tržby v Kč | |
|---|---|---|---|---|---|---|
| | 2013 | 2014 | 2013 | 2014 | 2013 | 2014 |
| | $p_0$ | $p_1$ | $q_0$ | $q_1$ | $Q_0 = p_0 \cdot q_0$ | $Q_1 = p_1 \cdot q_1$ |
| Vlastní prodejna | 200 | 220 | 560 | 710 | 112 000 | 156 200 |
| Sklad Brno | 240 | 265 | 2 620 | 2 300 | 628 800 | 609 500 |
| Sklad Praha | 250 | 280 | 5 550 | 6 000 | 1 387 500 | 1 680 000 |
| Celkem | x | x | 8730 | 9010 | 2 128 300 | 2 445 700 |

$$I_{\bar{p}} = \frac{\bar{p}_1}{\bar{p}_0} = \frac{\frac{\sum Q_1}{\sum q_1}}{\frac{\sum Q_0}{\sum q_0}} = \frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_0 q_0}{\sum q_0}} = \frac{\frac{2445700}{9010}}{\frac{2128300}{8730}} = 1{,}113 \ (+11{,}3\ \%)$$

$$\Delta_{\bar{p}} = \bar{p}_1 - \bar{p}_0 = \frac{\sum p_1 q_1}{\sum q_1} - \frac{\sum p_0 q_0}{\sum q_0} = \frac{2445700}{9010} - \frac{2128300}{8730} = 27{,}651 \ (+27{,}651\ \text{Kč})$$

*b) Změna počtu prodaných láhví:*

$$I_q = \frac{\sum q_1}{\sum q_0} = \frac{9010}{8730} = 1{,}032 \ \ (+3{,}2\ \%)$$

$$\Delta_q = \sum q_1 - \sum q_0 = 9010 - 8730 = 280 \ \ (+280\ \text{ks})$$

*c) Změna tržeb:*

$$I_Q = \frac{\sum Q_1}{\sum Q_0} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{2445700}{2128300} = 1{,}149 \ \ (+14{,}9\ \%)$$

$$\Delta_Q = \sum Q_1 - \sum Q_0 = \sum p_1 q_1 - \sum p_0 q_0 = 2445700 - 2128300 = 317400 \ \ (+317\ 400\ \text{Kč})$$

## Interpretace:

Průměrná cena láhve Rulandského šedého se v roce 2014 zvýšila oproti roku 2013 o 11,3 %, což představuje přibližně částku 27,70 Kč.

Celkový počet prodaných láhví Rulandského se v roce 2014 zvýšil oproti roku 2013 o 3,2 %, což představuje 280 ks láhví.

Celkové tržby za prodej Rulandského šedého se v roce 2014 zvýšily oproti roku 2013 o 14,9 %, což představuje částku 317 400 Kč.

## Poznámka:

V programu STATGRAPHICS ani v MS Excel není žádná speciální procedura pro výpočet individuálních složených indexů a rozdílů.

## Zadání příkladu – pokračování:

Dále určete, jak se na změně průměrné ceny láhve vína podílela změna cen v jednotlivých prodejních místech, a jak změna struktury prodeje. Výsledky interpretujte.

| **Prodejní místo** | **Cena v Kč za 0,75l** | | **Počet prodaných ks (0,75l)** | | **Tržby v Kč** | | $p_1 \cdot q_0$ |
|---|---|---|---|---|---|---|---|
| | **2013** | **2014** | **2013** | **2014** | **2013** | **2014** | |
| | $p_0$ | $p_1$ | $q_0$ | $q_1$ | $Q_0 = p_0 \cdot q_0$ | $Q_1 = p_1 \cdot q_1$ | |
| **Vlastní prodejna** | 200 | 220 | 560 | 710 | 112 000 | 156 200 | 123 200 |
| **Sklad Brno** | 240 | 265 | 2 620 | 2 300 | 628 800 | 609 500 | 694 300 |
| **Sklad Praha** | 250 | 280 | 5 550 | 6 000 | 1 387 500 | 1 680 000 | 1 554 000 |
| **Celkem** | **x** | **x** | **8730** | **9010** | **2 128 300** | **2 445 700** | **2 371 500** |

Pro zjištění vlivu dílčích činitelů na změnu průměrné ceny výrobku je potřeba provést rozklad indexu proměnlivého složení. Zde je použita k rozkladu *metoda postupných změn*.

a) Uvažujeme-li nejdříve změnu cen a potom změnu struktury prodeje, použijeme následující rozklad:

$I_{\bar{p}} = I_{SS}(q_0) \cdot I_{STR}(p_1)$ - relativní vyjádření vlivu činitelů, resp.

$\Delta_{\bar{p}} = \Delta_{SS}(q_0) + \Delta_{STR}(p_1)$ - absolutní vyjádření vlivu činitelů.

Je zřejmé, že budeme potřebovat další pomocný výpočet (v tabulce uveden modrou barvou).

$$I_{SS}(q_0) = \frac{\frac{\sum p_1 q_0}{\sum q_0}}{\frac{\sum p_0 q_0}{\sum q_0}} = \frac{\frac{2371500}{8730}}{\frac{2128300}{8730}} = 1{,}114 \ \ (+11{,}4\ \%)$$

$$I_{STR}(p_1) = \frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_1 q_0}{\sum q_0}} = \frac{\frac{2445700}{9010}}{\frac{2371500}{8730}} = 0{,}999 \ \ (-0{,}1\ \%)$$

$$I_{\bar{p}} = 1{,}114 \cdot 0{,}999 = 1{,}113$$

$$\Delta_{\bar{p}} = \frac{\sum p_1 q_0}{\sum q_0} - \frac{\sum p_0 q_0}{\sum q_0} + \frac{\sum p_1 q_1}{\sum q_1} - \frac{\sum p_1 q_0}{\sum q_0} = \frac{2371500}{8730} - \frac{2128300}{8730} + \frac{2445700}{9010} - \frac{2371500}{8730} =$$

$$= 27{,}858 + (-0{,}207) = 27{,}651$$

## Interpretace:

Vlivem změny cen v jednotlivých prodejních místech (při zachování struktury prodeje z r. 2013) došlo ke zvýšení průměrné ceny vína o 11,4 %, tj. o 27,858 Kč za láhev. Změna struktury prodaných vín (při cenové hladině r. 2014) způsobila pokles průměrné ceny o 0,1 %, tj. o 0,207 Kč.

b) Uvažujeme-li nejprve změnu struktury prodaných výrobků a potom změnu cen, využijeme tento tvar rozkladu:

$I_{\bar{p}} = I_{SS}(q_1) \cdot I_{STR}(p_0)$ - relativní vyjádření vlivu činitelů, resp.

$\Delta_{\bar{p}} = \Delta_{SS}(q_1) + \Delta_{STR}(p_0)$ - absolutní vyjádření vlivu činitelů.

Další pomocný výpočet je uveden v následující tabulce modrou barvou.

| Prodejní místo | Cena v Kč za 0,75l | | Počet prodaných ks (0,75l) | | Tržby v Kč | | $p_0 \cdot q_1$ |
|---|---|---|---|---|---|---|---|
| | 2013 | 2014 | 2013 | 2014 | 2013 | 2014 | |
| | $p_0$ | $p_1$ | $q_0$ | $q_1$ | $Q_0 = p_0 \cdot q_0$ | $Q_1 = p_1 \cdot q_1$ | |
| **Vlastní prodejna** | 200 | 220 | 560 | 710 | 112 000 | 156 200 | 142 000 |
| **Sklad Brno** | 240 | 265 | 2 620 | 2 300 | 628 800 | 609 500 | 552 000 |
| **Sklad Praha** | 250 | 280 | 5 550 | 6 000 | 1 387 500 | 1 680 000 | 1 500 000 |
| **Celkem** | **x** | **x** | **8730** | **9010** | **2 128 300** | **2 445 700** | **2 194 000** |

$$I_{SS}(q_1) = \frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_0 q_1}{\sum q_1}} = \frac{\frac{2445700}{9010}}{\frac{2194000}{9010}} = 1{,}114 \ (+11{,}4\ \%)$$

$$I_{STR}(p_0) = \frac{\frac{\sum p_0 q_1}{\sum q_1}}{\frac{\sum p_0 q_0}{\sum q_0}} = \frac{\frac{2194000}{9010}}{\frac{2128300}{8730}} = 0{,}999 \ (-0{,}1\ \%)$$

$$I_{\bar{p}} = 1{,}114 \cdot 0{,}999 = 1{,}113$$

$$\Delta_{\bar{p}} = \frac{\sum p_1 q_1}{\sum q_1} - \frac{\sum p_0 q_1}{\sum q_1} + \frac{\sum p_0 q_1}{\sum q_1} - \frac{\sum p_0 q_0}{\sum q_0} = \frac{2445700}{9010} - \frac{2194000}{9010} + \frac{2194000}{9010} - \frac{2128300}{8730} =$$

$$= 27{,}936 + (-0{,}284) = 27{,}652 \text{ (pozn.: drobný rozdíl pramení ze zaokrouhlování).}$$

## Interpretace:

Změna struktury prodaných výrobků při cenách r. 2013 vyvolala snížení průměrné ceny láhve vína o 0,1 %, tj. o 0,284 Kč. V důsledku změn cen v jednotlivých prodejních místech došlo ke zvýšení průměrné ceny láhve vína o 11,4 %, tj. o 27,936 Kč.

# Téma 11 – Příklad 3

## Zadání příkladu:

Moravské vinařství prodává svá vína ve vlastní prodejně, přes sklad v Brně a v Praze. V následující tabulce jsou údaje o ceně ve všech 3 prodejních místech a počtu prodaných láhví (0,75 l) odrůdového vína Rulandské šedé v letech 2013 a 2014. Vypočítejte, jak se změnila průměrná cena vína, celkové prodané množství láhví a celkové tržby za prodej Rulandského šedého ve sledovaném období (v absolutním i relativním vyjádření). Všechny výsledky interpretujte.

| Prodejní místo | Cena v Kč za 0,75l | | Počet prodaných ks (0,75l) | |
|---|---|---|---|---|
| | 2013 | 2014 | 2013 | 2014 |
| Vlastní prodejna | 200 | 220 | 560 | 710 |
| Sklad Brno | 240 | 265 | 2 620 | 2 300 |
| Sklad Praha | 250 | 280 | 5 550 | 6 000 |

## Vypracování příkladu:

Vzhledem k tomu, že chceme určit změnu jednotlivých ukazatelů souhrnně za danou odrůdu, je zřejmé, že srovnáváme hodnoty ukazatele ve dvou obdobích, kdy ukazatel je složen z dílčích částí, a tyto dílčí části je potřeba nějakým způsobem shrnout. Zde pracujeme se stejnorodým ukazatelem, proto je pro relativní vyjádření změn jednotlivých ukazatelů vhodné použít individuální složené indexy a pro absolutní vyjádření změn individuální složené rozdíly.

*a) Změna průměrné ceny:*

Pro určení změny průměrné ceny je potřeba provést další doplňující výpočty:

| Prodejní místo | Cena v Kč za 0,75l | | Počet prodaných ks (0,75l) | | Tržby v Kč | |
|---|---|---|---|---|---|---|
| | 2013 | 2014 | 2013 | 2014 | 2013 | 2014 |
| | $p_0$ | $p_1$ | $q_0$ | $q_1$ | $Q_0 = p_0 \cdot q_0$ | $Q_1 = p_1 \cdot q_1$ |
| Vlastní prodejna | 200 | 220 | 560 | 710 | 112 000 | 156 200 |
| Sklad Brno | 240 | 265 | 2 620 | 2 300 | 628 800 | 609 500 |
| Sklad Praha | 250 | 280 | 5 550 | 6 000 | 1 387 500 | 1 680 000 |
| Celkem | x | x | 8730 | 9010 | 2 128 300 | 2 445 700 |

$$I_{\bar{p}} = \frac{\bar{p}_1}{\bar{p}_0} = \frac{\frac{\sum Q_1}{\sum q_1}}{\frac{\sum Q_0}{\sum q_0}} = \frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_0 q_0}{\sum q_0}} = \frac{\frac{2445700}{9010}}{\frac{2128300}{8730}} = 1{,}113 \ (+11{,}3\ \%)$$

$$\Delta_{\bar{p}} = \bar{p}_1 - \bar{p}_0 = \frac{\sum p_1 q_1}{\sum q_1} - \frac{\sum p_0 q_0}{\sum q_0} = \frac{2445700}{9010} - \frac{2128300}{8730} = 27{,}651 \ (+27{,}651\ \text{Kč})$$

*b) Změna počtu prodaných láhví:*

$$I_q = \frac{\sum q_1}{\sum q_0} = \frac{9010}{8730} = 1{,}032 \ \ (+3{,}2\ \%)$$

$$\Delta_q = \sum q_1 - \sum q_0 = 9010 - 8730 = 280 \ \ (+280\ \text{ks})$$

*c) Změna tržeb:*

$$I_Q = \frac{\sum Q_1}{\sum Q_0} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{2445700}{2128300} = 1{,}149 \ \ (+14{,}9\ \%)$$

$$\Delta_Q = \sum Q_1 - \sum Q_0 = \sum p_1 q_1 - \sum p_0 q_0 = 2445700 - 2128300 = 317400 \ \ (+317\ 400\ \text{Kč})$$

## Interpretace:

Průměrná cena láhve Rulandského šedého se v roce 2014 zvýšila oproti roku 2013 o 11,3 %, což představuje přibližně částku 27,70 Kč.

Celkový počet prodaných láhví Rulandského se v roce 2014 zvýšil oproti roku 2013 o 3,2 %, což představuje 280 ks láhví.

Celkové tržby za prodej Rulandského šedého se v roce 2014 zvýšily oproti roku 2013 o 14,9 %, což představuje částku 317 400 Kč.

## Poznámka:

V programu STATGRAPHICS ani v MS Excel není žádná speciální procedura pro výpočet individuálních složených indexů a rozdílů.

## Zadání příkladu – pokračování:

Dále určete, jak se na změně průměrné ceny láhve vína podílela změna cen v jednotlivých prodejních místech, a jak změna struktury prodeje. Výsledky interpretujte.

| **Prodejní místo** | **Cena v Kč za 0,75l** | | **Počet prodaných ks (0,75l)** | | **Tržby v Kč** | | $p_1 \cdot q_0$ |
|---|---|---|---|---|---|---|---|
| | **2013** | **2014** | **2013** | **2014** | **2013** | **2014** | |
| | $p_0$ | $p_1$ | $q_0$ | $q_1$ | $Q_0 = p_0 \cdot q_0$ | $Q_1 = p_1 \cdot q_1$ | |
| **Vlastní prodejna** | 200 | 220 | 560 | 710 | 112 000 | 156 200 | 123 200 |
| **Sklad Brno** | 240 | 265 | 2 620 | 2 300 | 628 800 | 609 500 | 694 300 |
| **Sklad Praha** | 250 | 280 | 5 550 | 6 000 | 1 387 500 | 1 680 000 | 1 554 000 |
| **Celkem** | **x** | **x** | **8730** | **9010** | **2 128 300** | **2 445 700** | **2 371 500** |

Pro zjištění vlivu dílčích činitelů na změnu průměrné ceny výrobku je potřeba provést rozklad indexu proměnlivého složení. Zde je použita k rozkladu *metoda postupných změn*.

a) Uvažujeme-li nejdříve změnu cen a potom změnu struktury prodeje, použijeme následující rozklad:

$I_{\bar{p}} = I_{SS}(q_0) \cdot I_{STR}(p_1)$ - relativní vyjádření vlivu činitelů, resp.

$\Delta_{\bar{p}} = \Delta_{SS}(q_0) + \Delta_{STR}(p_1)$ - absolutní vyjádření vlivu činitelů.

Je zřejmé, že budeme potřebovat další pomocný výpočet (v tabulce uveden modrou barvou).

$$I_{SS}(q_0) = \frac{\frac{\sum p_1 q_0}{\sum q_0}}{\frac{\sum p_0 q_0}{\sum q_0}} = \frac{\frac{2371500}{8730}}{\frac{2128300}{8730}} = 1{,}114 \ \ (+11{,}4\ \%)$$

$$I_{STR}(p_1) = \frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_1 q_0}{\sum q_0}} = \frac{\frac{2445700}{9010}}{\frac{2371500}{8730}} = 0{,}999 \ \ (-0{,}1\ \%)$$

$$I_{\bar{p}} = 1{,}114 \cdot 0{,}999 = 1{,}113$$

$$\Delta_{\bar{p}} = \frac{\sum p_1 q_0}{\sum q_0} - \frac{\sum p_0 q_0}{\sum q_0} + \frac{\sum p_1 q_1}{\sum q_1} - \frac{\sum p_1 q_0}{\sum q_0} = \frac{2371500}{8730} - \frac{2128300}{8730} + \frac{2445700}{9010} - \frac{2371500}{8730} =$$

$$= 27{,}858 + (-0{,}207) = 27{,}651$$

## Interpretace:

Vlivem změny cen v jednotlivých prodejních místech (při zachování struktury prodeje z r. 2013) došlo ke zvýšení průměrné ceny vína o 11,4 %, tj. o 27,858 Kč za láhev. Změna struktury prodaných vín (při cenové hladině r. 2014) způsobila pokles průměrné ceny o 0,1 %, tj. o 0,207 Kč.

b) Uvažujeme-li nejprve změnu struktury prodaných výrobků a potom změnu cen, využijeme tento tvar rozkladu:

$I_{\bar{p}} = I_{SS}(q_1) \cdot I_{STR}(p_0)$ - relativní vyjádření vlivu činitelů, resp.

$\Delta_{\bar{p}} = \Delta_{SS}(q_1) + \Delta_{STR}(p_0)$ - absolutní vyjádření vlivu činitelů.

Další pomocný výpočet je uveden v následující tabulce modrou barvou.

| Prodejní místo | Cena v Kč za 0,75l | | Počet prodaných ks (0,75l) | | Tržby v Kč | | $p_0 \cdot q_1$ |
|---|---|---|---|---|---|---|---|
| | 2013 | 2014 | 2013 | 2014 | 2013 | 2014 | |
| | $p_0$ | $p_1$ | $q_0$ | $q_1$ | $Q_0 = p_0 \cdot q_0$ | $Q_1 = p_1 \cdot q_1$ | |
| **Vlastní prodejna** | 200 | 220 | 560 | 710 | 112 000 | 156 200 | 142 000 |
| **Sklad Brno** | 240 | 265 | 2 620 | 2 300 | 628 800 | 609 500 | 552 000 |
| **Sklad Praha** | 250 | 280 | 5 550 | 6 000 | 1 387 500 | 1 680 000 | 1 500 000 |
| **Celkem** | **x** | **x** | **8730** | **9010** | **2 128 300** | **2 445 700** | **2 194 000** |

$$I_{SS}(q_1) = \frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_0 q_1}{\sum q_1}} = \frac{\frac{2445700}{9010}}{\frac{2194000}{9010}} = 1{,}114 \ (+11{,}4\ \%)$$

$$I_{STR}(p_0) = \frac{\frac{\sum p_0 q_1}{\sum q_1}}{\frac{\sum p_0 q_0}{\sum q_0}} = \frac{\frac{2194000}{9010}}{\frac{2128300}{8730}} = 0{,}999 \ (-0{,}1\ \%)$$

$$I_{\bar{p}} = 1{,}114 \cdot 0{,}999 = 1{,}113$$

$$\Delta_{\bar{p}} = \frac{\sum p_1 q_1}{\sum q_1} - \frac{\sum p_0 q_1}{\sum q_1} + \frac{\sum p_0 q_1}{\sum q_1} - \frac{\sum p_0 q_0}{\sum q_0} = \frac{2445700}{9010} - \frac{2194000}{9010} + \frac{2194000}{9010} - \frac{2128300}{8730} =$$

$$= 27{,}936 + (-0{,}284) = 27{,}652 \text{ (pozn.: drobný rozdíl pramení ze zaokrouhlování).}$$

## Interpretace:

Změna struktury prodaných výrobků při cenách r. 2013 vyvolala snížení průměrné ceny láhve vína o 0,1 %, tj. o 0,284 Kč. V důsledku změn cen v jednotlivých prodejních místech došlo ke zvýšení průměrné ceny láhve vína o 11,4 %, tj. o 27,936 Kč.

# Téma 12 – Příklad 1

## Zadání příkladu:

V následující tabulce jsou k dispozici údaje o ceně a prodeji 3 odrůdových vín ve vinařství Polívka v roce 2014 a 2015. Určete, jak se souhrnně změnily ceny, množství prodaných výrobků a celkové tržby za prodej vín v roce 2015 oproti roku 2014. Všechny výsledky interpretujte.

| Odrůda | Cena v Kč za 0,75l | | Počet prodaných láhví (0,75l) | |
|---|---|---|---|---|
| | 2014 | 2015 | 2014 | 2015 |
| Veltlínské zelené | 190 | 212 | 3500 | 3200 |
| Cabernet Moravia | 200 | 212 | 4100 | 4500 |
| Müller Thurgau | 170 | 200 | 3900 | 3700 |

## Vypracování příkladu:

Vzhledem k tomu, že chceme určit změnu jednotlivých ukazatelů souhrnně, je potřeba zjistit, s jakým ukazatelem zde pracujeme. Je zřejmé, že jde o nestejnorodého ukazatele, jehož dílčí části nelze shrnovat ani pomocí součtu, ani pomocí průměru. Proto je pro vyjádření změn jednotlivých ukazatelů vhodné použít souhrnné indexy.

*a) Změna cen:*

K vyjádření souhrnné změny cen lze použít některý ze souhrnných cenových indexů. Zde bude uveden výpočet dvou z nich – Laspeyresova a Paascheho. Pomocné výpočty jsou uvedeny v následující tabulce.

| Odrůda | Cena v Kč za 0,75l | | Počet prodaných láhví (0,75l) | | Tržby v Kč | | $p_1 \cdot q_0$ | $p_0 \cdot q_1$ |
|---|---|---|---|---|---|---|---|---|
| | 2014 | 2015 | 2014 | 2015 | 2014 | 2015 | | |
| | $p_0$ | $p_1$ | $q_0$ | $q_1$ | $Q_0 = p_0 \cdot q_0$ | $Q_1 = p_1 \cdot q_1$ | | |
| Veltlínské zelené | 190 | 212 | 3500 | 3200 | 665 000 | 678 400 | 742 000 | 608 000 |
| Cabernet Moravia | 200 | 212 | 4100 | 4500 | 820 000 | 954 000 | 869 200 | 900 000 |
| Müller Thurgau | 170 | 200 | 3900 | 3700 | 663 000 | 740 000 | 780 000 | 629 000 |
| **Celkem** | **x** | **x** | **x** | **x** | **2 148 000** | **2 372 400** | **2 391 200** | **2 137 000** |

$$ {}_L I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{2391200}{2148000} = 1{,}113 \ \ (+11{,}3\ \%) $$

$$ {}_P I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{2372400}{2137000} = 1{,}110 \ \ (+11\ \%) $$

*b) Změna množství prodaných láhví:*

K vyjádření změny množství prodaných výrobků lze použít některý ze souhrnných objemových indexů. Zde bude uveden výpočet dvou z nich – Laspeyresova a Paascheho.

$$_L I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{2137000}{2148000} = 0{,}995 \text{ (-0,5 \%)}$$

$$_P I_q = \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{2372400}{2391200} = 0{,}992 \text{ (-0,8 \%)}$$

*c) Změna tržeb:*

$$I_Q = \frac{\sum Q_1}{\sum Q_0} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{2372400}{2148000} = 1{,}104 \text{ (+10,4 \%)}$$

## Interpretace:

Vezmeme-li v úvahu prodaný počet láhví jako v roce 2014, vzrostly ceny v roce 2015 o 11,3 % oproti roku 2014. (Laspeyresův cenový index)

Vezmeme-li v úvahu prodaný počet láhví jako v roce 2015, došlo k růstu cen v roce 2015 o 11,0 % oproti roku 2014. (Paascheho cenový index)

Bereme-li v úvahu ceny roku 2014, kleslo množství prodaných láhví o 0,5 % ve sledovaném období. (Laspeyresův objemový index)

Vezmeme-li v úvahu ceny roku 2015, snížilo se množství prodaných láhví o 0,8 % ve sledovaném období. (Paascheho objemový index)

Tržby za prodej vín vzrostly v roce 2015 oproti roku 2014 o 10,4 %.

## Poznámka:

V programu STATGRAPHICS ani v MS Excel není žádná speciální procedura pro výpočet souhrnných indexů a rozdílů.

## Zadání příkladu – pokračování:

Dále určete, jak změnu tržeb za prodej vín ovlivnila změna cen a jak změna počtu prodaných láhví. Výsledky interpretujte.

Pro zjištění vlivu dílčích činitelů na změnu tržeb je potřeba provést rozklad souhrnného indexu hodnoty. Zde je použita k rozkladu *metoda postupných změn*.

a) Uvažujeme-li nejdříve změnu cen a potom změnu prodaného množství při cenách roku 2015, použijeme následující rozklad:
$I_Q = {}_L I_p \cdot {}_P I_q = 1{,}113 \cdot 0{,}992$

## Interpretace:

Změna cen způsobila nárůst tržeb o 11,3 %, změna prodaného množství způsobila naopak pokles tržeb o 0,8 %.

b) Uvažujeme-li nejprve změnu prodaného množství a teprve potom změnu cen, využijeme tento tvar rozkladu:
$I_Q = {}_P I_p \cdot {}_L I_q = 1{,}110 \cdot 0{,}995$

## Interpretace:

Změna cen způsobila nárůst tržeb o 11 %, změna prodaného množství naopak pokles tržeb o 0,5 %.

# Téma 1 – Příklad 2

## Zadání příkladu

V následující tabulce je uvedeno 30 dvojic hodnot znaku x a y. Roztřiďte tyto hodnoty do tabulky dvourozměrného rozdělení četností a vypočítejte hodnoty podmíněných průměrů a podmíněných rozptylů proměnné y.

| Pořadí dvojice | $x_i$ | $y_j$ | Pořadí dvojice | $x_i$ | $y_j$ | Pořadí dvojice | $x_i$ | $y_j$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 11 | 2 | 3 | 21 | 3 | 4 |
| 2 | 1 | 2 | 12 | 2 | 1 | 22 | 4 | 4 |
| 3 | 1 | 1 | 13 | 3 | 3 | 23 | 3 | 5 |
| 4 | 1 | 3 | 14 | 3 | 2 | 24 | 3 | 3 |
| 5 | 2 | 1 | 15 | 3 | 2 | 25 | 4 | 3 |
| 6 | 2 | 4 | 16 | 3 | 1 | 26 | 4 | 4 |
| 7 | 2 | 2 | 17 | 3 | 5 | 27 | 1 | 4 |
| 8 | 2 | 2 | 18 | 2 | 3 | 28 | 2 | 5 |
| 9 | 1 | 4 | 19 | 1 | 1 | 29 | 4 | 5 |
| 10 | 1 | 4 | 20 | 2 | 1 | 30 | 4 | 1 |

## Vypracování příkladu

Korelační tabulka dvojrozměrně zobrazuje dvě číselné proměnné a jejich sdružené četnosti (absolutní). Nejprve zobrazíme unikátní hodnoty proměnných do řádků a sloupců – bývá zvykem používat řádky pro proměnnou *x* a sloupce pro proměnnou *y*. Vnitřní buňky tabulky je třeba zkonstruovat tak, aby jednotlivé hodnoty proměnných *x* a *y* měly v křížově odpovídající buňce absolutní četnost výskytu dat. Například pro x = 1 a y = 1 je četnost rovna třem, existují tři takové dvojice hodnot (v zadání s pořadovými čísly 1, 3, 19).

| $y_j$ / $x_i$ | 1 | 2 | **3** | **4** | 5 | Součty četností $n_{i.}$ | $\bar{y}_i$ | $s_i^2$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 1 | 1 | 3 | 0 | 8 | 2,5000 | 1,7500 |
| 2 | 3 | 2 | 2 | 1 | 1 | 9 | 2,4400 | 1,8242 |
| 3 | 1 | 2 | 2 | 1 | 2 | 8 | 3,1250 | 1,8594 |
| 4 | 1 | 0 | 1 | 2 | 1 | 5 | 3,4000 | 1,8400 |
| Součty četností $n_{.j}$ | 8 | 5 | 6 | 7 | 4 | **30** | x | x |

Jednou z kontrol správnosti je křížový součet okrajových četností, který musí dohromady představovat rozsah souboru[1].

Podmíněnou charakteristikou rozumíme určitou hodnotu deskriptivní statistiky pro proměnnou *y*, která platí za předpokladu určité hodnoty proměnné *x*. V našem případě budeme počítat podmíněný průměr pomocí vzorce:

$$\bar{y}_i = \frac{\sum_{j=1}^{l} y_j n_{ij}}{n_{i\cdot}} ;$$

kde $n_{ij}$ jsou příslušné sdružené absolutní četnosti a $n_{i\cdot}$ je okrajová absolutní četnost, která je součtem počtu hodnot *y* v případě určité hodnoty *x*.

Pro hodnotu x = 1 je podmíněný průměr roven hodnotě

$\bar{y}_i = \frac{1\cdot 3+2\cdot 1+3\cdot 1+4\cdot 3+5\cdot 0}{8} = 2{,}5$ . Takto postupujeme i pro další hodnoty x.

Podmíněný rozptyl bude vypočtený dle vzorce:

$$s_i^2 = \frac{\sum_{j=1}^{l} \left(y_j - \bar{y}_i\right)^2 n_{ij}}{n_{i\cdot}} .$$

Pro hodnotu x = 1 je podmíněný rozptyl roven hodnotě

$s_i^2 = \frac{(1-2{,}5)^2\cdot 3+(2-2{,}5)^2\cdot 1+(3-2{,}5)^2\cdot 1+(4-2{,}5)^2\cdot 3+(5-2{,}5)^2\cdot 0}{8} = 1{,}75$ . Takto postupujeme i pro další hodnoty x.

1V případě chybějících údajů je třeba postupovat podle některé ze známých metod jejich doplnění, nebo použít pouze kompletní dvojice (v SGP bývá označováno „Complete Cases Only“).

## Řešení v SGP

V programu Statgraphics stačí zadat všechny hodnoty dvojic do dvou samostatných sloupců – proměnných x a y. Je třeba pouze dbát na to, aby hodnoty párově správně odpovídaly. Výsledné vektory pak vypadají takto.

| | x | y |
|---|---|---|
| 1 | 1 | 1 |
| 2 | 1 | 2 |
| 3 | 1 | 1 |
| 4 | 1 | 3 |
| 5 | 2 | 1 |
| 6 | 2 | 4 |
| 7 | 2 | 2 |
| 8 | 2 | 2 |
| 9 | 1 | 4 |
| 10 | 1 | 4 |
| 11 | 2 | 3 |
| 12 | 2 | 1 |
| 13 | 3 | 3 |
| 14 | 3 | 2 |
| 15 | 3 | 2 |
| 16 | 3 | 1 |
| 17 | 3 | 5 |
| 18 | 2 | 3 |
| 19 | 1 | 1 |
| 20 | 2 | 1 |
| 21 | 3 | 4 |
| 22 | 4 | 4 |
| 23 | 3 | 5 |
| 24 | 3 | 3 |
| 25 | 4 | 3 |
| 26 | 4 | 4 |
| 27 | 1 | 4 |
| 28 | 2 | 5 |
| 29 | 4 | 5 |
| 30 | 4 | 1 |

Procedura v SGP: Describe – Categorical Data – Crosstabulation (Frequency Table)

Tvorbu dvourozměrné tabulky zařídí procedura Crosstabulation. Ve vstupním dialogu vybereme řádkovou proměnnou Row Variable (použijeme tradičně x) a sloupcovou proměnnou Column Variable (y). V okně Frequency Table můžeme přes doplňkový panel Pane Options zobrazit relativní četnosti vzhledem k řádku, sloupci, nebo celému souboru (Table Percentages). Dále jsou zde možnosti, které odkazují na chí-kvadrát test o nezávislosti

dvou kategoriálních proměnných (Expected Frequencies, Deviations, Chi-Square Values). Jejich použití by v tomto případě bylo ovšem chybné!

**Frequency Table for x by y**

| | 1 | 2 | 3 | 4 | 5 | Row Total |
|---|---|---|---|---|---|---|
| 1 | 3 | 1 | 1 | 3 | 0 | 8 |
| | 10,00% | 3,33% | 3,33% | 10,00% | 0,00% | 26,67% |
| 2 | 3 | 2 | 2 | 1 | 1 | 9 |
| | 10,00% | 6,67% | 6,67% | 3,33% | 3,33% | 30,00% |
| 3 | 1 | 2 | 2 | 1 | 2 | 8 |
| | 3,33% | 6,67% | 6,67% | 3,33% | 6,67% | 26,67% |
| 4 | 1 | 0 | 1 | 2 | 1 | 5 |
| | 3,33% | 0,00% | 3,33% | 6,67% | 3,33% | 16,67% |
| Column Total | 8 | 5 | 6 | 7 | 4 | 30 |
| | 26,67% | 16,67% | 20,00% | 23,33% | 13,33% | 100,00% |

Cell contents:
Observed frequency
Percentage of table

Procedura v SGP: Describe - Numeric Data – Subset Analysis
(Summary Statistics)

Při vstupním dialogu zadáme jako *Data* proměnnou, z jejíž hodnot budeme počítat výběrové charakteristiky (podmiňovanou), proměnnou *y*, do políčka *Codes* zadáme proměnnou podmiňující, tedy *x*.

V okně Summary Statistics vidíme jednotlivé podmíněné charakteristiky. Jejich zobrazení můžeme upravit v doplňkových možnostech Pane Options. Nezapomeňme, že jde o výběrové charakteristiky, takže v případě, že nás zajímají charakteristiky základní, musíme upravit výrazem (n-1/n). U charakteristik tvaru rozdělení je odlišnost výraznější, jelikož program Statgraphics užívá vzorců odlišných od klasických momentových měr.

**Summary Statistics**
Data variable: y

| *x* | *Count* | *Average* | *Variance* |
|---|---|---|---|
| 1 | 8 | 2,5 | 2,0 |
| 2 | 9 | 2,44444 | 2,02778 |
| 3 | 8 | 3,125 | 2,125 |
| 4 | 5 | 3,4 | 2,3 |
| Total | 30 | 2,8 | 2,02759 |

**The StatAdvisor**
This table shows sample statistics for the 4 levels of x.

## Interpratace

V případě, že proměnná x nabývá hodnoty jedna, je průměrná hodnota proměnné y rovna dvěma a půl. Rozptyl hodnot y je v tomto případě roven 1,75.

## Řešení v MS Excel

Dle verze MS Excel je nutné využít specifických statistických funkcí pro výpočty podmíněných charakteristik, případně využít odvozených funkcí (např. +IF)

**PRŮMĚR** – Průměr hodnot.

**VAR.P** – Rozptyl základního souboru (od Excel 2010).

**VAR.S –** Rozptyl výběru (od Excel 2010).

**VAR** – Rozptyl základního souboru.

**VAR.VÝBĚR** – Rozptyl výběru.

# Téma 2 – Příklad 1

## Zadání příkladu

Sledujeme výzkum, který se zabývá vlivem 4 typů učebních metod angličtiny na výsledky studentů, posuzované podle výsledků dosažených v závěrečném testu (v bodech). Údaje jsou v následující tabulce. Data byla získána náhodným výběrem, rozsah souboru je 20 studentů. Předpokládáme, že výkon studentů má normální rozdělení. Pomocí analýzy rozptylu posuďte závislost proměnné y (výsledky v bodech) na metodě výuky (x). Posuďte sílu případné závislosti.

| **Typ učební metody ($x_i$) i = 1,2,3,4** | **Počet bodů v závěrečném testu ($y_{ij}$) j = 1,2,3,4,5** | **Četnost ($n_i$)** | **Podmíněné průměry** $\acute{y}_i = \frac{\sum y_{ij}}{n_i}$ | **Podmíněné rozptyly** $s_i^2 = \frac{\sum (y_{ij} - \acute{y}_i)^2}{n_i}$ |
|---|---|---|---|---|
| 1. typ | 71 83 65 77 84 | 5 | 76 | 52 |
| 2. typ | 60 51 54 80 55 | 5 | 60 | 108,4 |
| 3. typ | 55 55 62 65 63 | 5 | 60 | 17,6 |
| 4. typ | 68 73 67 59 53 | 5 | 64 | 50,4 |
| **Suma** | | **20** | **X** | **X** |

## Vypracování příkladu

Testujeme hypotézu o neúčinnosti faktoru x na proměnnou y. Postup je obdobný jako u jiného statistického testování hypotéz a probíhá v několika krocích.

1. $H_0$: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (obecně *y* na *x* nezávisí) – rozdělení proměnné y mají na různých úrovních faktoru x stejné střední hodnoty μ
   $H_1$: non $H_0$
2. Testové kritérium
   $$F = \frac{\frac{S_{ym}}{k-1}}{\frac{S_{yv}}{n-k}} = \frac{\frac{860}{3}}{\frac{1142}{16}} = 4{,}016$$
3. $W \equiv (F, F \geq F_{1-\alpha}(k-1; n-k))$
   $W \equiv (F, F \geq F_{0,95}(3; 16))$
   $W \equiv (F, F \geq 3{,}2391)$

4. $4{,}0160 \geq 3{,}2391$
   Testové kritérium je prvkem kritického oboru. Nulovou hypotézu zamítáme, přijímáme hypotézu alternativní.

Na základě analýzy rozptylu tedy můžeme říci, že na 5% hladině významnosti **existuje statisticky významná závislost** výsledků studentů angličtiny (posuzovaných pomocí počtu bodů v závěrečném testu) na zvolené metodě výuky.

Sílu závislosti posoudíme pomocí poměru determinace. Jde o poměr meziskupinové variability na celkové variabilitě proměnné y. Výsledek je možné vyjádřit v procentech, jelikož jde o poměr části z celku, potenciálně tudíž vychází v intervalu od 0 do 1.

$$P^2 = \frac{S_{ym}}{S_y} = \frac{860}{2002} = 42{,}96\,\%$$

## Řešení v SGP

Důležité je správné zadání dat do datového listu. Každá proměnná musí mít formu datového vektoru. Závislou proměnnou jsou body studentů z testu (*Test*), faktorem je typ učební metody (*Ucebni_metoda*). Je třeba, aby závislá a nezávislá proměnná byly párově správně propojeny.

C:\Users\Jan\ownCloud\Škola\Výuka\Statistika\S

| | Ucebni_metoda | Test |
|---|---|---|
| | | body |
| 1 | 1 | 71 |
| 2 | 1 | 83 |
| 3 | 1 | 65 |
| 4 | 1 | 77 |
| 5 | 1 | 84 |
| 6 | 2 | 60 |
| 7 | 2 | 51 |
| 8 | 2 | 54 |
| 9 | 2 | 80 |
| 10 | 2 | 55 |
| 11 | 3 | 55 |
| 12 | 3 | 55 |
| 13 | 3 | 62 |
| 14 | 3 | 65 |
| 15 | 3 | 63 |
| 16 | 4 | 68 |
| 17 | 4 | 73 |
| 18 | 4 | 67 |
| 19 | 4 | 59 |
| 20 | 4 | 53 |
| 21 | | |
| 22 | | |
| 23 | | |

Procedura v SGP: Compare - Analysis of Variance - One-Way ANOVA (ANOVA Table, Variance check)

Při vstupním dialogu zadáme *Dependent Variable* jako proměnnou *Test*; *Factor* je *Ucebni_metoda*.

V programu Statgraphics můžeme ověřit jeden z předpokladů analýzy rozptylu – shodu skupinových rozptylů. Program nabízí čtyři různé testy (kromě nám známého Bartlettova testu jde o testy Leveneho, Cochrana a Hartleye). Výsledek testu vidíme ze stručné tabulky.

**Variance Check**

| | *Test* | *P-Value* |
|---|---|---|
| Bartlett's | 2,73239 | 0,434747 |

| *Comparison* | *Sigma1* | *Sigma2* | *F-Ratio* | *P-Value* |
|---|---|---|---|---|
| 1 / 2 | 8,06226 | 11,6404 | 0,479705 | 0,4943 |
| 1 / 3 | 8,06226 | 4,69042 | 2,95455 | 0,3190 |
| 1 / 4 | 8,06226 | 7,93725 | 1,03175 | 0,9766 |
| 2 / 3 | 11,6404 | 4,69042 | 6,15909 | 0,1062 |
| 2 / 4 | 11,6404 | 7,93725 | 2,15079 | 0,4765 |
| 3 / 4 | 4,69042 | 7,93725 | 0,349206 | 0,3326 |

Položka *Test* zobrazuje hodnotu testové statistiky. Hodnota *P-Value* ukazuje maximální možnou hladinu významnosti, na které nezamítáme nulovou hypotézu. V tomto případě tedy na 5% hladině významnosti nezamítáme nulovou hypotézu o shodě skupinových rozptylů. Předpoklad byl naplněn.

Doplňková tabulka zobrazuje porovnání a testy pro jednotlivé dvojice rozptylů (směrodatných odchylek).

Hlavní výstup je zobrazen v okně ANOVA Table.

**ANOVA Table for Test by Ucebni_metoda**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
|---|---|---|---|---|---|
| Between groups | 860,0 | 3 | 286,667 | 4,02 | 0,0262 |
| Within groups | 1142,0 | 16 | 71,375 | | |
| Total (Corr.) | 2002,0 | 19 | | | |

Sloupec Source zobrazuje rozklad variability na část meziskupinovou (Between groups) a vnitroskupinovou (Within groups). Příslušné sumy čtverců jsou ve sloupci Sum of Squares. Hodnota testové statistiky je označena jako F-Ratio. P-Value v tomto případě nasvědčuje tomu, že na 5% hladině významnosti zamítáme nulovou hypotézu a přijímáme hypotézu alternativní. Výpočet poměru determinace je třeba udělat jako podíl meziskupinové variability na celkové variabilitě (SGP nemá zvláštní funkcionalitu pro tento ukazatel).

Stejně jako v ostatních případech používá Statgraphics k vyhodnocení testu hypotézy ukazatel P-Value (v jiných programech např. Significance Level apod.), což je maximální možná hodnota hladiny významnosti, na které ještě nezamítáme nulovou hypotézu. Není tudíž nutné určovat kritický obor pro námi zvolenou hladinu významnosti.

## Interpretace

Na 5% hladině významnosti bylo prokázán statisticky významný vliv použité učební metody na výsledky studentů angličtiny posuzované výkonem v závěrečném testu. Výsledky studentů jsou vysvětleny faktorem učební metody asi z 42,96 % (42,96 % celkové variability proměnné y je vysvětleno pomocí faktoru x).

## Řešení v MS Excel

Dle verze programu se liší i možnosti zpracování analýzy rozptylu. Nejblíže výstupu ze Statgraphicsu je použití nástroje *Analýza dat*. Po kliknutí na toto tlačítko lze vybrat ANOVA: jednofaktorová. V následujícím dialogu určíte pole se vstupními daty. Lze je mít uspořádána ve sloupcích, nebo řádcích. Pozor ovšem, že každá kategorie musí mít samostatný sloupec/řádek (tím se liší zadávání od SGP). Lze také definovat, jestli vstup obsahuje hlavičky kategorií a určit hladinu významnosti pro testování. Nezapomeňte také určit pole pro výstup procedury (případně nový list).

Výstup je obdobný ANOVA Table ze Statgraphicsu. Variabilita je rozložena na meziskupinovou a vnitroskupinovou. Kromě P-Value je zobrazen i kvantil kritického oboru pro zadanou hladinu významnosti. V tabulce *Summary* jsou podmíněné charakteristiky pro jednotlivé kategorie (obdoba *Subset Analysis* v SGP).

# Asociační tabulka

## Zadání příkladu:

Zdravotní pojišťovna prováděla v jistém okresním městě průzkum očkování proti jisté nemoci. Náhodně bylo vybráno a dotázáno 60 jeho obyvatel, zdali očkování mají, nebo nemají. Výsledky jsou v tabulce 1. Otestujte hypotézu, zda pohlaví dotazované osoby a to, zdali má sledované očkování, jsou nezávislé veličiny. Hladina významnosti $\alpha = 5\ \%$

Tab. 1 - Průzkum očkování

| | Očkován | Neočkován | Součet $n_{i.}$ |
|---|---|---|---|
| Muž | 4 | 26 | 30 |
| Žena | 21 | 9 | 30 |
| Součet $n_{.j}$ | 25 | 35 | 60 |

## Vypracování příkladu

Pokud by sledované znaky měli zařadit podle typu proměnné, jedná se o nominální proměnné, které jsou typické tím, že nemají uspořádání a jsou diskrétní, jinak řečeno kategoriální. Označme sledované znaky písmeny A (pohlaví) a B (má/nemá očkování). Pro test nezávislosti dvou kategoriálních znaků, kdy oba nabývají pouze dvou hodnot, používáme tzv. *asociační tabulku* - viz tabulka 1. V řádcích jsou hodnoty jednoho znaku (v tomto případě pohlaví dotazované osoby), ve sloupcích hodnoty druhého znaku (stav očkování). Na průnicích řádků a sloupců odpovídajících všem čtyřem kombinacím těchto dvojic hodnot jsou absolutní četnosti výskytu těchto dvojic. V posledním sloupci jsou uvedeny řádkové součty, tj. celkový počet mužů a žen, v posledním řádku sloupcové součty - celkový počet očkovaných a neočkovaných osob. V pravém dolním rohu potom celkový počet dotazovaných osob, který se musí rovnat jak součtu řádkových, tak sloupcových součtů.

Jednotlivé četnosti označíme postupně $n_{11}, n_{12}, n_{12}, n_{22}$, kde první index je řádkový, druhý sloupcový. Jejich součet označíme $n$. Řádkové a sloupcové součty jsou již označeny v tabulce 1. Test nezávislosti provedeme standardně v pěti krocích:

1. **Formulovat hypotézu**
   $H_0$: Znaky A a B jsou nezávislé.
   $H_1$: Znaky A a B nejsou nezávislé.

2. **Zvolit testové kritérium (TK)**
   U asociační tabulky je TK následující:

$$\chi^2 = n\frac{(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}} \sim \chi^2(1)$$

   Uvedená veličina $\chi^2$ má asymptoticky rozdělení $\chi^2(1)$.

3. **Sestrojit kritický obor** $W$
   Kritický obor $W$ jsou ty hodnoty testového kritéria, které vedou k zamítnutí nulové hypotézy. V tomto případě platí:

$$W = \{\chi^2 : \chi^2 \geq \chi^2(1)_{1-\alpha}\}$$

   Konkrétně pro $\alpha = 5$ % je

$$W = \{\chi^2 : \chi^2 \geq \chi^2(1)_{0,95} = 3{,}84\}$$

4. **Spočítat hodnotu TK**

$$\chi^2 = 60.\frac{(4.9 - 21.26)^2}{30.30.25.35} = 19{,}817 \tag{1}$$

5. **Formulovat závěr**
   V tomto případě platí, že TK $\chi^2 \in W$, protože 19,817 $\geq$ 3,84. Takže nulovou hypotézu $H_0$ zamítáme a přijímáme $H_1$.

   **Závěr:** sledované znaky pohlaví a očkování nejsou nezávislé.

Pokud zamítneme nulovou hypotézu jako v tomto případě, je možné ještě změřit *sílu závislosti* sledovaných znaků pomocí **koeficientu asociace** $r_{AB} \in [-1, 1]$

$$r_{AB} = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_{1.}n_{2.}n_{.1}n_{.2}}} \tag{2}$$

Čím je $r_{AB}$ bližší jedné, tím jsou vyšší četnosti $n_{11}$ a $n_{22}$ (hlavní diagonála tabulky). Naopak čím je hodnota koeficientu asociace bližší -1, převládají četnosti $n_{12}$ a $n_{21}$. Síla závislosti roste s absolutní hodnotou $r_{AB}$. Zde je

$$r_{AB} = \frac{4.9 - 21.26}{\sqrt{30.30.25.35}} = \text{-}0{,}575 \tag{3}$$

## Program STATGRAPHICS

V programu STATGRAPHICS je výpočet snadný. Předpokládejme, že četnosti z asociační tabulky jsou zadány ve dvou proměnných: Ockovan (postupně 4, 21) a Neockovan (9, 26). Výchozím bodem je funkcionalita *Describe/Categorical Data/Contingency Tables (2-way)*.

Do pole Columns zadáme proměnné Ockovan a Neockovan a klikneme na OK. V dalším formuláři zaškrtneme tabulky *Analysis Summary, Frequency Table, Tests of Independence, Summary Statistics* a opět klikneme na OK. Ve vygenerované analýze v tabulce *Tests of Independence* je uvedena hodnota testového kritéria $\chi^2$ jako `Statistic`. Dále počet stupňů volnosti `Df` a hodnota `P-Value`. Pokud je P-Value nižší než hladina významnosti $\alpha$, nulovou hypotézu zamítáme. To je uvedeno i v textu pod tabulkou.

```
The StatAdvisor
This table shows the results of a hypothesis test run to determine
whether or not to reject the idea that the row and column classifications
are independent. Since the P-value is less than 0,05, we can reject the
hypothesis that rows and columns are independent at the 95,0% confidence
level. Therefore, the observed row for a particular case is related to
its column.
```

Uveďme nyní ještě postup, jak zjistit hodnoty kvantilů rozdělení $\chi^2(n-1)$ potřebné pro konstrukci kritického oboru. Ve STATGRAPHICS je najdeme ve formuláři *Describe/Distribution Fitting/Probability Distributions*, kde jsou uvedena všechna pravděpodobnostní rozdělení, která jsou v tomto programu k dispozici. V seznamu rozdělení vybereme Chi-Square a klikneme na OK. V dalším kroku je třeba zadat počet stupňů volnosti (D. F.), v tomto případě je to 1.

V dalším formuláři *Tables and Graphs* je třeba zaškrtnout tabulku *Inverse CDF*, ve které je možné zjistit hodnotu libovolného kvantilu příslušného rozdělení. Stačí najet na tabulku myší a kliknout pravým tlačítkem. Z menu zvolit *Pane Options* a do formuláře *Inverse CDF Options* zadat příslušnou hodnotu procent číslem z intervalu $(0, 1)$, zde konkrétně 0,95. Potom klikneme na OK a z tabulky odečteme hodnotu kvantilu, která je rovna 3,84.

STATGRAPHICS počítá i hodnotu koeficientu asociace $r_{AB}$ podle (2). Ta je v tabulce *Summary Statistics* jako `Pearson's R` a je rovna -0,5747, po zaokrouhlení -0,575, jak je uvedeno výše v (3).

## Interpretace výsledků

Na základě provedeného testu byla zamítnuta hypotéza nezávislosti pohlaví a očkování respondenta na dané hladině významnosti. Vypočtená hodnota TK (1) překračuje kritickou hodnotu 3,84, leží tedy v kritickém oboru $W$. Koeficient asociace $r_{AB}$ = -0,575. Sílu závislosti lze proto označit jako střední s tím, že dominují četnosti na vedlejší diagonále tabulky.

## Program MS Excel

Tabulkový kalkulátor k testování nezávislosti v asociační tabulce obecně použít lze, protože uvedené vzorce jsou velmi jednoduché, nicméně přesný postup výpočtů v programu MS Excel uvádět nebudeme. Kvantily rozdělení $\chi^2$ doporučujeme dohledat ve statistických tabulkách, které jsou nejsnáz dostupné na Internetu. Kvantilové funkce Excelu vzhledem ke špatným zkušenostem s nimi pro tyto účely nedoporučejeme.

# Téma 3 – Příklad 1

## Zadání příkladu

Na základě průzkumu provedeného u čtenářů časopisů A, B, C byla sestavena kontingenční tabulka. Rozhodněte, zda výběr časopisu a bydliště čtenáře jsou znaky závislé, neboli zda se struktura čtenářů z hlediska bydliště u časopisů A, B, C liší.

| | | Časopis | | |
|---|---|---|---|---|
| Bydliště | A | B | C | Celkem |
| Liberec | 75 | 75 | 50 | 200 |
| Jablonec | 40 | 70 | 40 | 150 |
| Česká Lípa | 35 | 5 | 10 | 50 |
| Celkem | 150 | 150 | 100 | 400 |

## Vypracování příkladu

Jelikož se jedná o dvě slovní proměnné, kdy závislá proměnná je nominální, bude k řešení využito chí-kvadrát testu o nezávislosti dvou kategoriálních proměnných. Tento test je oboustranný, zkoumá závislost proměnné *a* na proměnné *b* a naopak. Postup je obdobný jako u jiného statistického testování hypotéz a probíhá v několika krocích.

1. $H_0$: výběr časopisu je nezávislý na bydlišti respondenta (obecně *a* a *b* jsou nezávislé proměnné) – v tomto případě formuluje hypotézu jednostranně, jelikož nepředpokládáme, že opačný směr závislosti je smysluplný
   $H_1$: non $H_0$
2. Testové kritérium G (někdy označované jako chí-kvadrát) má tuto podobu:
   $$G = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(n_{ij} - n'_{ij}\right)^2}{n'_{ij}}$$
   Pro jeho určení je třeba vypočítat tzv. teoretické (hypotetické, očekávané) četnosti za pomoci vzorce:
   $$n'_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$$
   Např. pro četnost $n_{11}$ platí, že (150*200)/400 = 75. Teoretické četnosti jsou v následující tabulce vyznačeny modře a vyjadřují, jak by měla teoreticky kontingenční tabulka vypadat, kdyby mezi proměnnými byla nezávislost.

| | | Časopis | | |
|---|---|---|---|---|
| Bydliště | A | B | C | Celkem |
| Liberec | 75 75 | 75 75 | 50 50 | 200 |
| Jablonec | 40 56,25 | 70 56,25 | 40 37,5 | 150 |
| Česká Lípa | 35 18,75 | 5 18,75 | 10 12,5 | 50 |
| Celkem | 150 | 150 | 100 | 400 |

Testové kritériu má potom podobu:
0 + 0 + 0 + 4,694 + 3,361 + 0,167 + 14,083 + 10,083 + 0,5 = 32,889

3. $W \equiv \left\{ G,\ G > \chi^2_{1-\alpha}\left[(r-1)(s-1)\right] \right\}$

   W ≡ (G, G ≥ $X^2(4)$)
   W ≡ (G, G ≥ 9,488)

4. 32,889 ≥ 9,488
   Testové kritérium je prvkem kritického oboru. Nulovou hypotézu zamítáme, přijímáme hypotézu alternativní.

Jelikož byla závislost prokázána, pokusíme se určit i její těsnost. K tomu nám slouží koeficienty kontingence:
Pearsonův:

$$C_P = \sqrt{\frac{G}{G+n}}$$

$C_P = 0{,}276$

nebo
Cramérův:

$$C_C = \sqrt{\frac{G}{n \cdot h}}$$ , kde h … min (r-1);(s-1)

$C_C = 0{,}203$

Na 5% hladině významnosti jsme prokázali, že **existuje statisticky významná závislost** mezi četností časopisů a bydlištěm čtenářů.

## Řešení v SGP

V tomto případě je výjimečně možné v případě programu Statgraphics užití dat ve formě tabulky (sdružených) četností. Pro správnou funkci stačí zadat pouze sdružené četnosti,

nikoliv okrajové (součtové) hodnoty. Stejně tak proměnná *Bydliste* je v tomto případě pouze pro označení řádků v tabulce. Z výpočetního hlediska nemá vliv.

<untitled>

| | Bydliste | A | B | C |
|---|---|---|---|---|
| 1 | Liberec | 75 | 75 | 50 |
| 2 | Jablonec | 40 | 70 | 40 |
| 3 | Ceska Lipa | 35 | 5 | 10 |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |

Procedura v SGP: Describe - Categorical Data - Contingency Tables
(Frequency Table, Chi-Square Test, Summary Statistics)

Při vstupním dialogu zadáme jako *Colums A, B* a *C*, tedy sloupce kontingenční tabulky. Do políčka *Labels* je možné zadat *Bydliste,* aby kontingenční tabulka vypadala stejně, jako naše vstupní tabulka.

V okně Frequency Table je možné vidět kompletní kontingenční tabulku. Z doplňkových možností Pane Options si můžeme zvolit různé pohledy na relativní vyjádření četností nebo také hodnoty očekávaných četností a dílčích příspěvků k testovému kritériu.

**Frequency Table**

| | A | B | C | Row Total |
|---|---|---|---|---|
| Liberec | 75 | 75 | 50 | 200 |
| | 18,75% | 18,75% | 12,50% | 50,00% |
| | 37,50% | 37,50% | 25,00% | |
| | 50,00% | 50,00% | 50,00% | |
| | 75,00 | 75,00 | 50,00 | |
| | 0,00 | 0,00 | 0,00 | |
| Jablonec | 40 | 70 | 40 | 150 |
| | 10,00% | 17,50% | 10,00% | 37,50% |
| | 26,67% | 46,67% | 26,67% | |
| | 26,67% | 46,67% | 40,00% | |
| | 56,25 | 56,25 | 37,50 | |
| | 4,69 | 3,36 | 0,17 | |
| Ceska Lipa | 35 | 5 | 10 | 50 |
| | 8,75% | 1,25% | 2,50% | 12,50% |
| | 70,00% | 10,00% | 20,00% | |
| | 23,33% | 3,33% | 10,00% | |
| | 18,75 | 18,75 | 12,50 | |
| | 14,08 | 10,08 | 0,50 | |
| Column Total | 150 | 150 | 100 | 400 |
| | 37,50% | 37,50% | 25,00% | 100,00% |

Cell contents:
- Observed frequency
- Percentage of table
- Percentage of row
- Percentage of column
- Expected frequency
- Contribution to chi-square

Hlavní výstup je zobrazen v okně Test of Independence. Položka *Statistic* zobrazuje hodnotu testové statistiky. Hodnota *P-Value* ukazuje maximální možnou hladinu významnosti, na které nezamítáme nulovou hypotézu. V tomto případě tedy na 5% hladině významnosti zamítáme nulovou hypotézu o nezávislosti obou sledovaných znaků.

**Tests of Independence**

| *Test* | *Statistic* | *Df* | *P-Value* |
|---|---|---|---|
| Chi-Square | 32,889 | 4 | 0,0000 |

Stejně jako v ostatních případech používá Statgraphics k vyhodnocení testu hypotézy ukazatel P-Value (v jiných programech např. Significance Level apod.), což je maximální možná hodnota hladiny významnosti, na které ještě nezamítáme nulovou hypotézu. Není tudíž nutné určovat kritický obor pro námi zvolenou hladinu významnosti.

Hodnoty koeficientů kontingence nalezneme v okně *Summary Statistics*. Pearsovou koeficientu odpovídá údaj *Contingency Coeff., Cramer's V* je koeficient Cramerův.

**Summary Statistics**

| | | *With Rows* | *With Columns* |
|---|---|---|---|
| *Statistic* | *Symmetric* | *Dependent* | *Dependent* |
| Lambda | 0,0667 | 0,0000 | 0,1200 |
| Uncertainty Coeff. | 0,0420 | 0,0443 | 0,0399 |
| Somer's D | -0,0550 | -0,0524 | -0,0579 |
| Eta | | 0,0840 | 0,2074 |

| *Statistic* | *Value* | *P-Value* | *Df* |
|---|---|---|---|
| Contingency Coeff. | 0,2756 | | |
| Cramer's V | 0,2028 | | |
| Conditional Gamma | -0,0870 | | |
| Pearson's R | -0,0607 | 0,2256 | 398 |
| Kendall's Tau b | -0,0551 | 0,2251 | |
| Kendall's Tau c | -0,0516 | | |

## Interpretace

Na 5% hladině významnosti bylo prokázáno, že struktura čtenářů se z hlediska bydliště u časopisů A, B a C statisticky významně liší. Vzhledem k tomu, že považujeme za smysluplný pouze jednostranný vliv, je možné formulovat odpověď takto. Tato závislost je slabá. Pearsonův kontingenční koeficient má hodnotu 0,2756 nebo Cramerův 0,2028.

## Řešení v Excelu

Dle verze programu je možné použít funkce **CHITEST(aktuální;očekávané)** k provedení výše uvedeného testu. Aktuální četnosti je termín pro četnosti empirické. Výstupem funkce je hodnota P-Value. Výpočet očekávaných četností je ovšem nutné provést pomocí matematických operací (aplikací vzorců).

# Téma 5 – Příklad 1

## Zadání příkladu

Z korelační tabulky určete, zda existuje lineární závislost (korelace) proměnné y na proměnné x. U případné závislosti posuďte její sílu a směr.

| | $y_j$ | |
|---|---|---|
| $x_i$ | **6** | **9** |
| **2** | 2 | - |
| **5** | 4 | 1 |
| **9** | - | 3 |
| **11** | 1 | 2 |

## Vypracování příkladu

Korelační tabulka zobrazuje dvě číselné proměnné. Z obou proměnných spočítáme výběrový Pearsonův korelační koeficient a otestujeme jeho významnost v základním souboru. Pokud test neprokáže jeho významnost, předpokládáme, že neexistuje statisticky významný vztah mezi oběma proměnnými. Korelační analýza zkoumá závislost jako oboustrannou.

Korelační koeficient spočítáme dle vzorce:

$$r_{yx} = r_{xy} = \sqrt{r_{yx}^2} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\left(\overline{x^2} - \bar{x}^2\right) \cdot \left(\overline{y^2} - \bar{y}^2\right)}}$$

Výběrový korelační koeficient v tomto případě vychází 0,6313. Nyní otestujeme jeho významnost v základním souboru.

1. $H_0: \rho_{yx} = 0$

   $H_1: non\ H_0$

2. Testové kritérium

   $$t = \frac{r_{yx} \cdot \sqrt{n-2}}{\sqrt{1-r_{yx}^2}} = 6{,}395$$

3. $$W \equiv \left\{ t;\ t \le t_{\frac{\alpha}{2}}(n-2) \cup t \ge t_{1-\frac{\alpha}{2}}(n-2) \right\}$$

   $$W \equiv (t;\ t \le -2{,}201 \cup t \ge 2{,}201)$$

4. Testové kritérium je prvkem kritického oboru. Nulovou hypotézu zamítáme, přijímáme hypotézu alternativní.

Sílu a směr závislosti posoudíme pomocí vypočteného korelačního koeficientu. V našem případě jde o středně silnou přímou lineární závislost. Definiční obor Pearsonova korelačního koeficientu se pohybuje od -1 v případě funkční nepřímé závislosti až po 1 v případě funkční přímé závislosti.

## Řešení v SGP

Pro program Statgraphics je nejprve nutné vytvořit z korelační tabulky dva datové vektory odpovídající proměnným *x* a *y*. Pro jednu proměnnou lze využít operátoru REP a vytvořit vektor z okrajových absolutních četností, např. REP ($x_i$;$n_i$). Druhou proměnnou ovšem musím párově přiřadit tak, aby správně odpovídaly jednotlivé hodnoty x a y. Výsledné vektory pak vypadají takto.

<untitled>

| | x | y |
|---|---|---|
| 1 | 2 | 6 |
| 2 | 2 | 6 |
| 3 | 5 | 6 |
| 4 | 5 | 6 |
| 5 | 5 | 6 |
| 6 | 5 | 6 |
| 7 | 5 | 9 |
| 8 | 9 | 9 |
| 9 | 9 | 9 |
| 10 | 9 | 9 |
| 11 | 11 | 6 |
| 12 | 11 | 9 |
| 13 | 11 | 9 |
| 14 | | |

Procedura v SGP: Describe – Numeric Data – Multiple-Variable Analysis (Correlations)

alternativně: Describe – Multivariate Methods - Multiple-Variable Analysis (Correlations)

Při vstupním dialogu zadáme do položky *Data* jednotlivé proměnné (nezáleží na pořadí, jelikož jde o oboustrannou závislost). V druhém kroku (Analysis Options) se nás program zeptá, zda chceme do analýzy zahrnout pouze kompletní dvojice dat, nebo všechna data (dojde k aproximaci chybějících údajů). V našem případě jsou obě možnosti ekvivalentní.

V okně Correlations vidíme výsledek testu. První údaj sděluje hodnotu Pearsonova korelačního koeficientu, druhý údaj rozsah výběru a třetí údaj je hodnotou P-Value.

Stejně jako v ostatních případech používá Statgraphics k vyhodnocení testu hypotézy ukazatel P-Value (v jiných programech např. Significance Level apod.), což je maximální možná hodnota hladiny významnosti, na které ještě nezamítáme nulovou hypotézu. Není tudíž nutné určovat kritický obor pro námi zvolenou hladinu významnosti.

**Correlations**

| | x | y |
|---|---|---|
| x | | 0,6313 |
| | | (13) |
| | | 0,0207 |
| y | 0,6313 | |
| | (13) | |
| | 0,0207 | |

Correlation
(Sample Size)
P-Value

Hodnotu testové statistiky tato procedura nezobrazuje, je ovšem možné použít například výstup z regresní analýzy. V tomto případě tedy na 5% hladině významnosti zamítáme nulovou hypotézu o nevýznamnosti korelačního koeficientu v populaci.

## Interpretace

Na základě korelační analýzy můžeme říci, že na 5% hladině významnosti **existuje statisticky významná lineární závislost** proměnných x a y. Závislost je středně silná a přímá.

## Řešení v MS Excel

Pro spočítání Pearsonova korelačního koeficientu je možné použít funkci CORREL. Výslednou hodnotu je pak možné podrobit výše popsanému testu.

# Téma 5 – Příklad 2

## Zadání příkladu

Následující tabulka zobrazuje pořadí skupiny dívek ve dvou různých soutěžích krásy. Můžeme říci, že hodnocení obou porot jsou lineárně nezávislá? Posuďte charakter případné závislosti.

| Dívka | Umístění u poroty A | Umístění u poroty B |
|---|---|---|
| **Esmeralda** | 2. | 6. |
| **Kasandra** | 5. | 5. |
| **Rosaneta** | 6. | 2. |
| **Julie** | 7. | 1. |
| **Betty** | 8. | 3. |
| **Gertruda** | 4. | 4. |
| **Žofie** | 1. | 8. |
| **Manuela** | 3. | 7. |

## Vypracování příkladu

Z obou proměnných spočítáme v tomto případě Spearmanův korelační koeficient a otestujeme jeho významnost v základním souboru. Pokud test neprokáže jeho významnost, předpokládáme, že neexistuje statisticky významný vztah mezi oběma pořadími. Korelační analýza zkoumá závislost jako oboustrannou.

Korelační koeficient spočítáme dle vzorce:

$$r_s = 1 - \frac{6\sum_{i=1}^{n}\left(i_x - i_y\right)^2}{n\left(n^2-1\right)}$$

Výběrový korelační koeficient v tomto případě vychází -0,881. Nyní otestujeme jeho významnost v základním souboru.

1. $H_0: \rho_S = 0$
2. $H_1: non\ H_0$
3. Testové kritérium

$$t = \frac{r_S \cdot \sqrt{n-2}}{\sqrt{1-r_S^2}} = -4{,}561$$

4. $W \equiv \left\{ t;\ t \le t_{\frac{\alpha}{2}}(n-2) \cup t \ge t_{1-\frac{\alpha}{2}}(n-2) \right\}$
   $W \equiv (t;\ t \le -2{,}365 \cup t \ge 2{,}365)$
5. Testové kritérium je prvkem kritického oboru. Nulovou hypotézu zamítáme, přijímáme hypotézu alternativní.

Sílu a směr závislosti posoudíme pomocí vypočteného korelačního koeficientu. V našem případě jde o silnou nepřímou lineární závislost. Definiční obor Spearmonova korelačního koeficientu se pohybuje od -1 v případě funkční nepřímé závislosti až po 1 v případě funkční přímé závislosti.

## Řešení v SGP

Zadání do programu není v tomto případě vůbec problematické. Stačí přepsat obě proměnné do samostatných sloupců.

Procedura v SGP: Describe – Numeric Data – Multiple-Variable Analysis (Rank Correlations)

alternativně: Describe – Multivariate Methods - Multiple-Variable Analysis (Rank Correlations)

Při vstupním dialogu zadáme do položky *Data* jednotlivé proměnné (nezáleží na pořadí, jelikož jde o oboustrannou závislost). V druhém kroku (Analysis Options) se nás program zeptá, zda chceme do analýzy zahrnout pouze kompletní dvojice dat, nebo všechna data (dojde k aproximaci chybějících údajů). V našem případě jsou obě možnosti ekvivalentní.

V okně Rank Correlations vidíme výsledek testu. První údaj sděluje hodnotu Spearmanova korelačního koeficientu, druhý údaj rozsah výběru a třetí údaj je hodnotou P-Value.

Stejně jako v ostatních případech používá Statgraphics k vyhodnocení testu hypotézy ukazatel P-Value (v jiných programech např. Significance Level apod.), což je maximální možná hodnota hladiny významnosti, na které ještě nezamítáme nulovou hypotézu. Není tudíž nutné určovat kritický obor pro námi zvolenou hladinu významnosti.

Hodnotu testové statistiky tato procedura nezobrazuje. V tomto případě tedy na 5% hladině významnosti zamítáme nulovou hypotézu o nezávislosti Spearmanova korelačního koeficientu v základním souboru.

## Interpretace

Na základě korelační analýzy můžeme říci, že na 5% hladině významnosti **existuje statisticky významná lineární závislost** hodnocení obou porot. Závislost je silná a nepřímá. To znamená, že dívky, které se líbí první porotě, nejsou tak oblíbené u poroty druhé a naopak.

## Řešení v MS Excel

Pro spočítání Spearmanova korelačního koeficientu je možné použít funkci CORREL v případě, že upravíme proměnnou na pořadí. V některých verzích existuje funkce SCORREL, která počítá hodnotu Spearmanova korelačního koeficientu. Výslednou hodnotu je pak možné podrobit výše popsanému testu.

V závislosti na verzi programu je také možné použít doplňkový analytický nástroj a po kliknutí na Correlation – Spearman lze zadat proměnné a následně formulovat, zda se jedná o

jednostrannou či oboustrannou hypotézu. Po kliknutí na OK lze z výstupu vyčíst hodnotu P-Value.

# Základní charakteristiky časových řad

## Zadání příkladu:

Určete základní charakteristiky časové řady středního počtu obyvatelstva České republiky (ozn. $P(t)$) od roku 1989 do roku 2011. Hodnoty jsou uvedeny v tabulce 1. Dále vytvořte vhodný graf této veličiny.

Tab. 1 - Počet obyvatel ČR letech 1989-2011 v tisících - ozn. $P_t$

| Rok | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 |
|---|---|---|---|---|---|---|---|---|
| $P_t$ | 10362 | 10363 | 10309 | 10318 | 10331 | 10336 | 10331 | 10315 |
| Rok | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 |
| $P_t$ | 10304 | 10295 | 10283 | 10272 | 10224 | 10201 | 10202 | 10207 |
| Rok | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | |
| $P_t$ | 10234 | 10267 | 10313 | 10430 | 10491 | 10517 | 10497 | |

*Zdroj dat: Český statistický úřad (ČSÚ)*

## Vypracování příkladu

Ve statistice dělíme sledované veličiny podle různých hledisek. Například na slovní a číselné, diskrétní a spojité, alternativní a množné apod. Mimo jiné známe i dělení na prostorové a časové. U prostorových veličin není okamžik, ke kterému napozorovaná hodnota veličiny přísluší, důležitý. U časových veličin je ale každá hodnota pevně svázána s časovým údajem svého vzniku, takže soubor, se kterým pracujeme, je množinou uspořádaných dvojic $[t, X_t]$, kde čas $t = 1, \ldots, T$ a veličina $X$ má hodnoty $X_1, \ldots, X_T$. Hodnota času $t$ je u časových řad nezbytná pro jejich analýzu, neboť zde zkoumáme například trendovou nebo sezónní složku, neboli chování veličiny $X$ jako funkci času $t$.

### Program STATGRAPHICS Centurion XVII

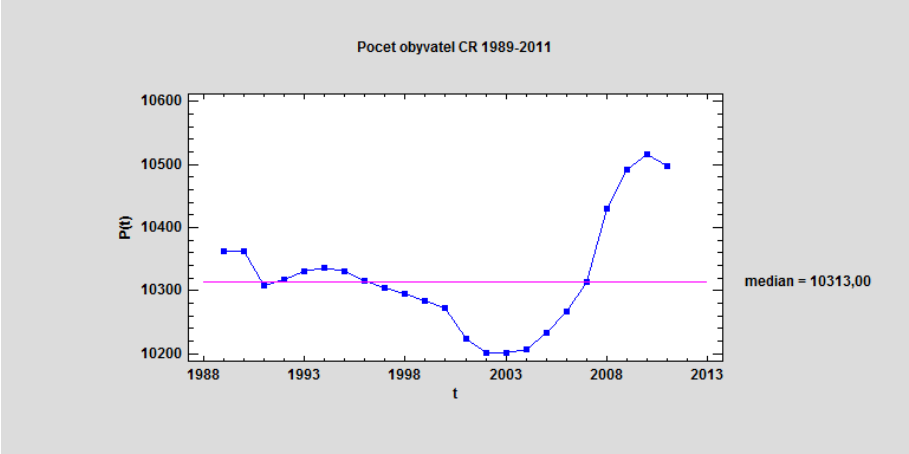
V programu STATGRAPHICS je pro grafy časových řad (ČŘ) vytvořena speciální funkcionalita. Najdeme ji v menu *Plot/Time Sequence Plots/Run Charts/Individuals*. Do povinného pole **Observations** je třeba zadat název proměnné, kde jsou hodnoty sledované veličiny $X_t$. Do pole **(Date/Time/Labels:)** je možné zadat odpovídající hodnoty času.

Vytvořený graf je možné upravovat přes menu, které se otevírá pravým tlačítkem myši, když je ukazatel myši na grafu, volba *Graphics Options*. V tomtéž menu je volba *Save Graph*, kterou lze graf vyexportovat do samostatného souboru. Grafické znázornění je ČŘ velmi důležité a slouží především k předběžné vizuální identifikaci trendů, periodicity a vůbec vývoje sledované veličiny v čase.

Nyní určíme základní číselné charakteristiky zkoumané řady. Těmi jsou popisné statistiky jako je **minimum, maximu, aritmetický průměr, medián, rozpětí a další**, ale také první diference, průměrný přírůstek a průměrný koeficient růstu.

Popisné statistiky lze získat přes *Describe/Numeric Data/One-variable Analysis*, kde do pole **Data** opět vložíme název sledované proměnné.

Obrázek 1: Počet obyvatel ČR v letech 1989-2011 v tis.




*Zdroj: Údaje ČSÚ, vlastní zpracování*

Výsledky:

$X_{min} = 10201$, $X_{MAX} = 10517$, rozpětí $R = 316$, $X_{0,25} = 10267$, medián $\tilde{X} = 10313$, $X_{0,75} = 10362$.

Průměr $\overline{X}$ **nelze** v tomto případě spočítat podle klasického vzorce $\overline{X} = \frac{\sum_{t=1}^{T} X_t}{T}$, protože se jedná o **okamžikovou časovou řadu**, kde součet hodnot nedává reálně příliš smysl. Použijeme proto vztah pro **prostý chronologický průměr**, neboť vzdálenosti mezi časovými okamžiky jsou stejné (za střední stav obyvatelstva v kalendářním roce je v ČR považován počet obyvatel daného území vždy o půlnoci z 30. 6. na 1. 7. sledovaného roku). Platí

$$\overline{X} = \frac{\frac{X_1}{2} + \sum_{t=2}^{T-1} X_t + \frac{X_T}{2}}{T-1}$$

Takže prostý chronologický průměr $\overline{X} = 10316{,}93$.

Nyní přikročme k výpočtu dalších charakteristik. **První diference** $\Delta_t^{(1)}$ zkoumané ČŘ vypočteme jako $\Delta_t^{(1)} = X_t - X_{t-1},\ t = 2, \ldots, T$. Jejich význam se projeví v trendové analýze ČŘ, která je dalším tématem přednášky. V programu STATGRAPHICS je pro výpočet prvních diferencí přímo vytvořena funkce `Diff(proměnná)`. Postup výpočtu je následující: označit myší volný sloupec datového listu kliknutím na záhlaví; kliknout na záhlaví sloupce pravým tlačítkem myši a v menu, které se otevře, vybrat položku *Generate Data*; do pole *Expression* napsat příkaz `Diff(proměnná)`, přičemž proměnná je ten sloupec datového listu, kde jsou hodnoty analyzované ČŘ.

Tab. 2 - První diference počtu obyvatel ČR letech 1989-2011 v tisících - ozn. $\Delta_t^{(1)}$

| Rok | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 |
|---|---|---|---|---|---|---|---|---|
| $\Delta_t^{(1)}$ | | 1 | -54 | 9 | 13 | 5 | -5 | -16 |
| Rok | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 |
| $\Delta_t^{(1)}$ | -11 | -9 | -12 | -11 | -48 | -23 | 1 | 5 |
| Rok | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | |
| $\Delta_t^{(1)}$ | 27 | 33 | 46 | 117 | 61 | 26 | -20 | |

*Zdroj dat: Český statistický úřad (ČSÚ) + vlastní zpracování*

**Průměrný přírůstek** ČŘ $\overline{\Delta}$ se spočítá jako jako aritmetický průměr jejích prvních diferencí. Po zjednodušení vzorce pro tento průměr vyjde vztah

$$\overline{\Delta} = \frac{X_T - X_1}{T - 1}$$

Zde konkrétně $\overline{\Delta} = 135/22 \doteq 6{,}13$.

Poslední charakteristikou je **průměrný koeficient růstu** $\overline{k}$. Je to geometrický průměr všech $T - 1$ koeficientů růstu $k_2, k_3, \ldots, k_{T-1}, k_T$, které se spočtou jako podíl hodnot ČŘ dvou sousedních období:

$$k_t = \frac{X_t}{X_{t-1}},\ t = 2, \ldots, T$$

Hodnota $k_t$ udává, kolikrát vzrostla hodnota ČŘ ve sledovaném období oproti období předchozímu. Geometrický průměr má potom tvar

$$\overline{k} = \sqrt[T-1]{k_2.k_3 \ldots k_{T-1}.k_T} = \sqrt[T-1]{\frac{X_T}{X_1}}$$

Zde je $\overline{k} = \sqrt[22]{\frac{10497}{10362}} \doteq 1{,}00059$. Pokud je hodnota $\overline{k} > 1$, potom řada v průměru meziročně roste. Pokud je menší než 1, jedná se o pokles. Percentuelní vyjádření získáme tak, že vypočteme výraz $100.(\overline{k} - 1)$. Pokud je kladný, jedná se opět o růst, v opačném případě o pokles. Pro sledovanou ČŘ je $100.(\overline{k} - 1) = 0{,}059\%$.

# Interpretace výsledků

Při interpretaci výsledků je nutné vedle sebe mít jak grafické znázornění, tak číselné charakteristiky. Závěry založené pouze na popisných statistikách mohou často vést k chybným závěrům, zejména v trendové analýze!

Z grafu je patrné, že počet obyvatel během prvních 13 let klesal. V roce 2011 dosáhl svého minima (10201), aby poté začal znovu stoupat (maximum 10517 v roce 2010). S tím souvisí i znaménko prvních diferencí, které je v prvním jmenovaném období vesměs záporné, ve zbytku naopak kladné. Rozpětí činí 316. Vzhledem k tomu, že průměr je větší než medián, je většina hodnot podprůměrná. Průměrný přírůstek činí 6,13. Lze proto říci, že od roku 1989 do roku 2011 se počet obyvatel každý rok v průměru zvýšil o 6130. Relativně vyjádřeno, každým rokem se počet obyvatel zvýšil o 0,59 ‰.

1) **Definujte pojem statistika.**
   - věda o sběru dat a zpracování hromadných údajů, zabývá se jevy, které mají hromadný charakter
   - hromadnost studována na statistických souborech
2) **Co je to popisná statistika?**
   - elementární metody sběru a zpracování informací
   - jednotkou je statistický soubor (osob, podniků, institucí, zvířat, zemí, atd.).
   - statistické soubory jsou tvořeny statistickými jednotkami, mají vlastnosti jednoznačně vymezeny.
3) **Co je matematická statistika a jak se dělí.**
   - moderní – zabývá se složitějšími metodami sběru a zpracování hromadných údajů; vytváří zvláštní druh matematických modelů – tzv. pravděpodobnostní modely – teorie pravděpodobnosti
4) **Typy statistických ukazatelů**
   - okamžikové, intervalové, primární, sekundární, extenzivní, intenzivní, stejnorodé, nestejnorodé
5) **Druhy statistických vlastností (2)**

6) **Statistické jednotky**
   - elementární jednotky stat. pozorování, jsou nositeli znaků
7) **Statistické znaky**
   - vlastnosti jednotek, která je předmětem zkoumání
   - kvalitativní – slovně vyjádřené – alternativní (2 obměny znaku); množné (více než 2 obměny)
   - kvantitativní – číselně vyjádřené, diskrétní (celočíselné), spojité (desetinné číslo, logaritmy)
8) **Statistický soubor**
   - množina jedinců, na kterých je prováděno statistické šetření
   - základní soubor – všechny jednotky s danou vlastností
   - výběrový soubor – vybrán ze základního, podmnožina je menší
9) **Rozdíl mezi ZS a VS**
   - VS je vlastně část ZS
   - VS je menší než ZS (úplné zjišťování, tvořen všemi jednotkami), VS(neúplné zjišťování)
10) **Základní etapy statistických prací**
    - statistické šetření (zjišťování) - získávání neznámých informací o znacích statistických jednotek , výsledkem statistického zjišťování jsou neuspořádané údaje
    - statistické zpracování
    - statistická analýza
11) **Co je statistické zjišťování?**
    - získávání neznámých informací o znacích statistických jednotek
    - výsledkem statistického zjišťování jsou neuspořádané údaje
    - ankety, dotazníky, experiment, výsledek vědeckého experimentu
    - pro zpřehlednění se data třídí
12) **Základní míry polohy rozdělení a k čemu slouží**
    - průměry: aritmetický; vážený ar. prům.; harmonický; geometrický; celkový ar. prům.; chronologický
    - ostatní střední hodnoty: medián, modus
    - měly by jedním číslem popsat střední úroveň hodnoty statistického znaku a umožnit jeho hlubší analýzy
    - reprezentují vhodnou střední hodnotu daného souboru kolem níž se soustřeďují hodnoty tohoto souboru
13) **Prosté x vážené charakteristiky polohy – rozdíl (3)**
    - prosté – u nesetříděných dat, máme-li relativní četnosti
    - vážené – u setříděných dat (tabulka rozdělení četností
14) **Průměr aritmetický**
    - součet všech hodnot znaků dělený počtem znaků
15) **Průměr geometrický**
    - n-tá odmocnina ze součinu znaků
16) **Jaké znáte míry založené na geometrickém průměru?**
    - všude tam, kde má smysl násobit hodnoty, např. průměrný koeficient růstu nebo Fisherův index
17) **V jakém oboru statistiky se můžeme setkat s geometrickým průměrem**
    - používá se u časových řad – průměrné tempo růstu (koeficient růstu)
18) **Průměr harmonický**
    - podíl počtu pozorování a sumy převrácených hodnot znaků
19) **Kdy a k čemu používáme harmonický průměr (3)**
    -v indexní analýze; průměr převrácených hodnot
20) **Průměr chronologický**
    - použití v okamžikové časové řadě
    - prostá forma – tam, kde délka mezi rozhodnými obdobími je stejná)
    - vážená forma – kde vahami jsou počty dní v měsíci,…

**21) Tempo a průměrný koeficient tempa růstu**
- počítání geometrického průměru

**22) Medián**
- x s vlnovkou - prostřední hodnota znaku v souboru uspořádaná podle velikosti
- lichý počet hodnot v souboru - střední hodnota
- sudý počet hodnot - průměr střední hodnoty

**23) Modus**
- hodnota, která se nejčastěji vyskytuje, hodnota znaku s největší četností

**24) Jak vypočítáte modus a medián spojité náhodné veličiny, znáte-li její distribuční funkci?**

**25) Uveďte situaci, kdy může medián popsat polohu statistického souboru lépe než průměr.**
- medián může popsat polohu statistického souboru lépe, pokud je nějaká hodnota hodně vychýlená, tzn., že se hodně liší od ostatních
- pak je průměr zkreslený a medián je lepší měrou polohy statistického souboru. př.: 4, 5, 5, 5, 5, 7, 48

**26) Pro která pravděpodobnostní rozdělení je jejich střední hodnota rovna mediánu a zároveň modu? Vysvětlete a uveďte příklady.**
Pro symetrická (Normální, studentovo)

**27) K čemu se používají podmíněné průměry**
– je to nejjednodušší způsob určení regresní závislosti (přímka podmíněných průměrů)- nelze však na jejich základě provádět odhady

**28) Co se stane s průměrem, rozptylem, směrodatnou odchylkou, mediánem a rozpětím statistického souboru, jestliže každá hodnota statistického souboru se:**
**a) zvětší dvakrát** - průměr a medián se zdvojnásobí, rozptyl se zvýší čtyřikrát; směrodatná odchylka a rozpětí statistického souboru se zvýší dvakrát
**b) zvětší o čtyři** – průměr a medián se zvětší o čtyři; rozptyl se nezmění; směrodatná odchylka a rozpětí statistického souboru se nezmění

**29) Rozdělení četnosti a co je intervalové rozdělení četnosti**
- rozdělení četností - u nespojitých znaků původně neuspořádané údaje roztřídit do rozdělení četností

**30) Jak se stanovuje interval relativní četnosti ZS u malých VS**
- výběr relativní četnosti se řídí binomickým rozdělením v případě výběru bez vracení se řídí hyperbolickým rozdělením
- výpočet vede ke složitým variacím, proto máme sestaveny tabulky a přímo odečítáme meze intervalu z tabulek

**31) Definice pojmu kumulativní četnost**
- absolutní a relativní
- vznikají postupným načítáním

**32) Druhy grafů**
- spojnicové, sloupcové(polygon, histogram), bodové, výsečové, speciální (kvartogram)

**33) Histogram**
- sloupcový graf - u intervalového rozdělení četností

**34) Které charakteristiky statistického souboru můžete přibližně zjistit z histogramu četnosti, aniž byste prováděli výpočet?**
- počet intervalů a jejich šířku, absolutní četnost intervalu a pokud jsou intervaly stejně dlouhé i modus

**35) Jaký graf používáme u jednorozměrných četností.**
- sloupcový

**36) Základní míry variability**
- absolutní: rozptyl, směrodatná odchylka, variační rozpětí, prům. odchylka
- relativní: variační koeficient, relativní průměrná odchylka

**37) Rozptyl**
- aritmetický průměr čtverců individuálních odchylek jednotlivých hodnot znaku od aritmetických průměrů
- nedostatek – jednotky jsou druhou mocninou původních jednotek

**38) Směrodatná odchylka v souboru výběrových průměrů**
- měří abs. Variabilitu
- je uvedena ve stejných měrných jednotkách jako zkoumaný stat. znak; s=odm.s na2
- prostá: S0=odm. z((sum(xi-x)na 2)/n)
- vážená: S0=odm. z((sum(xi-x)na 2*ni)/(sum.ni))
- informuje o proměnlivosti jednotlivých hodnot znaku kolem výběr. aritm. průměru

**39) Variační rozpětí**
- jednoduchá míra adaptability
- pouze odchylky mezi sebou - orientační

**40) Relativní ukazatele variability.**
- variační koeficient, relativní průměrná odchylka

**41) K čemu slouží variační koeficient? Jaká je jeho přednost?**
- variační koeficient je zákl. mírou relativní variability

- může se použít i tehdy pokud se znaky liší svou úrovní, což je výhoda
- počítá se jako podíl směrodatné odchylky a průměru

**42) Jak se změní variační koeficient, přičteme-li ke všem hodnotám souboru stejnou konstantu?**

Směrodatná odchylka v čitateli zůstane stejná a průměr ve jmenovateli se zvětší tuto konstantu => variační

**43) Lze vždy vypočítat variační koeficient souboru dat? Názor zdůvodněte.**

Ne. Variační koeficient se počítá jako podíl směr. odchylky a průměru => je-li např. průměr nulový, Variační koeficient vypočítat nelze.

**44) Kvantil**

- je hodnota, která rozděluje soubor hodnot na dvě části

**45) Kvartil**

- dělí soubor po 25%

**46) Rozdíl mezi charakteristikami šikmosti a špičatosti (3)**

- charakteristika šikmosti (nesouměrnosti)– ukazuje, jak soubor vypadá, stupeň koncentrace malých a velkých hodnot v souboru
- charakteristika špičatosti - ukazuje, jak jsou hodnoty nahloučeny kolem průměru

**47) Význam výběrového šetření v praxi (3)**

- pořizujeme výběrový soubor, aby nám poskytl informace o celém souboru
- hlavním nedostatkem je, že jsou zatíženy výběrovou chybou

**48) Výhody úplného zjišťování oproti neúplnému výběrovému zjišťování**

- úplné – při práci se základním souborem, nákladné, zdlouhavé, občas nemožné
- neúplné – při práci s výběrovým souborem, výběrový soubor musí být dobrým reprezentantem

**49) Vysvětlete pojmy oblastní a vícestupňový náhodný výběr**

- vícestupňový - výběr provádíme na více stupních (města – školy – fakulty – ročníky – studenti)
- oblastní - dvoustupňový výběr; v 1. stupni vybíráme oblast a ve 2. stupni vybíráme z oblasti jednotku

**50) Kvótní výběr - v čem spočívá**

- typ mechanického výběru při náhodném výběru

**51) Jaké znáte techniky pořízení náhodného výběru?**

- losování – opora výběru – výběr zastoupíme lístky
- tabulky náhodných čísel – generátor náhodných čísel
- mechanický výběr – systematické, každá n-tá jednotka v náhodně uspořádané posloupnosti speciální výběr

**52) Existuje rozdíl mezi stanovením intervalu u vracení a bez vracení?**

- s vracením – jednotku po výběru vracíme zpět
- bez vracení – rozsah ZS se zmenšuje, pravděpodobnost vybrání se zvětšuje
- u velkých souborů zbytečné zbytečné pracovat s vracením

**53) Jaký test k ověřování náhodnosti výběrového souboru? (3)**

**54) Metoda základního masivu**

- kdy se soubor skládá z několika velkých a mnoha malých jednotek
- zjišťování provádíme na velkých jednotkách

**55) Záměrný výběr**

- značná míra subjektivity toho, kdo vybírá
- vybere ty, o kterých si myslí, že dobře zastoupí soubor, ty blízké průměru, nelze vyvodit chyba

**56) Dělení (druhy) náhodného výběru**

- s vracením, bez vracení
- s nestejnou pravděpodobností vybrání
- prostý – se stejnými pravděpodobnostmi

**57) Náhodný jev**

- jev, který může nastat nebo nenastane v závislosti na náhodě a je výsledkem náhodného pokusu (charakterizuje výsledek náhodného pokusu kvalitativně)

**58) Náhodný pokus**

- realizace podmínek a vlivů, z nichž některé jsou známé a jiné náhodné

**59) Jev jistý, náhodný, nemožný**

- jev jistý - takový, který vždy nastane při každém provedení náhodného pokusu
- jev náhodný - jevy, které v závislosti na náhodě mohou, ale nemusí při uskutečňování daného komplexu podmínek nastat
- jev nemožný - náhodný jev, který nenastane při žádném provedení náhodného pokusu

**60) Klasické a statistické definice pravděpodobnosti**

- klasická - může li určitý pokus vykázat konečný počet n různých výsledků, které jsou stejně možné a jestliže m těchto výsledků má za následek nastoupení jevu A, kdežto zbylých n-m vylučuje: potom P(A)=m/n
- statistická - spojena s pojmem relativní četnosti; s rostoucím počtem pokusů se relativní četnost stabilizuje a přibližuje se k určitému konstrukčnímu číslu. P(A)= lim při n ku nekonečnu * M/N

**61) Matematická charakteristika pravděpodobnosti**

**62) Rozdíl mezi náhodnou veličinou a náhodným jevem (2)**

- náhodný jev – takový jev, který v závislosti na náhodě může, ale nemusí při uskutečňování daného komplexu podmínek nastat; charakterizuje výsledek náhodného pokusu kvalitativně (slovně)
- náhodná veličina – libovolná kvantitativní charakteristika náhodného pokusu; proměnná, která nabývá konkrétních hodnot, či hodnot z různých intervalů v závislosti na náhodě

**63) Zákon rozdělení náhodné veličiny**

- pravidlo, které každé hodnotě, nebo množině hodnot z každého intervalu přiřazuje pravděpodobnost, že náhodná veličina nebude této hodnoty, nebo hodnoty z tohoto intervalu
- tento zákon může být vyjádřen různou formou:

jako řada rozdělení pravděpodobností (grafem je polygon, diskrétní veličiny)
distribuční fce (univerzální zákon rozdělení, diskrétní i náhodné veličiny)
hustota pravděpodobnosti (spojité náhodné veličiny)

**64) Druhy rozdělení náhodných veličin**

- spojité (normální, exponenciální, chí-kvadratické,Studentovo t-rozdělení, F-rozdělení, rovnoměrné rozdělení)
- nespojité - diskrétní (Alternativní, Binomické, Poissonovo, Hypergeometrické, Geometrické)

**65) Charakterizujte normální rozdělení**

- náhodná veličina se řídí normálním rozdělením, je-li její střední hodnota μ a rozptyl ∂2
- grafem hustoty pravděpodobnosti je Gaussova křivka
- speciálním případem normálního rozdělení je normované normální rozdělení

**66) Binomické rozdělení je (možnosti) (2)**

- nejdůležitější typ rozdělení diskrétní náhodné veličiny
- rozdělením náhodné veličiny, která představuje počet výskytů jevu A při n nezávislých pokusech, přičemž pravděpodobnost výskytu jevu A je v každém pokusu konstantní

**67) Jaký je vztah binomického a Poissonova rozdělení?**

- má-li náhodná veličina X binomické rozdělení takové, že počet pokusů n je dostatečně veliké (nad 30), pravděpodobnost výskytu sledovaného jevu v jednom pokuse pod 0,1 a n konečné číslo, je možno toto rozdělení aproximovat Poissonovým rozdělením

**68)** Pravidlo tří sigma

- i když náhodná veličina X, která má normální rozdělení, může nabývat hodnot z intervalu od (- ∞,∞), je téměř nemožné, aby se pozorované hodnoty této veličiny odchylovaly od střední hodnoty o více než 3 sigma

**69) Co vyjadřuje zákon velkých čísel?**

- se zvyšováním počtu náhodných pokusů dochází k přibližování se empirické charakteristiky popisující výsledky těchto pokusů k charakteristice teoretické

**70) Co vyjadřuje centrální limitní věta?**

- vyjadřuje konvergenci pravděpodobnostních rozdělení k normálnímu rozdělení při dostatečně velkém rozsahu souboru.

**71) Co je normovaná náhodná veličina, jaké má charakteristiky a jaký má význam?**

- má normální rozdělení se střední hodnotou 0 a rozptylem 1
- význam je ve výpočtu distribuční funkce, která se z normálního rozdělení počítá obtížně

**72) V čem spočívá z pohledu teorie pravděpodobnosti průnik jevů A,B a sjednocení jevů A,B**

- průnik jevu A,B - spočívá v současné realizaci jak jevu A, tak jevu B
- sjednocení jevů A,B - spočívá v nastoupení alespoň jednoho z jevů a nebo B

**73) Je možné, aby existovaly 2 náhodné jevy, že pravděpodobnost jejich průniku je větší než pravděpodobnost jejich sjednocení?**

- Ne. Protože pravd. průniku může být max. rovná pravděp. sjednocení, když množiny splývají nebo je menší nebo množiny nemají průnik.

**74) Při výpočtu pravděpodobnosti projiti třemi zkouškami, z nichž každá má svou pravděpodobnost úspěchu používáme:**

a) sčítání
b) násobení
c) rozdíl

**75) Podmínky pro sčítání pravděpodobností a vzorec**

- jsou-li jevy A a B slučitelné, potom pravděpodobnosti jejich sjednocení se rovná součtu pravděpodobností jednotlivých jevů zmenšenému o pravděpodobnost jejich průniků
- v případě neslučitelných jevů je průnik těchto jevů jev nemožný

**76) Co jsou kvalitativní znaky, jak se dělí, příklady**

- kvalitativní znaky jsou znaky slovní, získané z anket, dotazníků
- dělíme na znaky alternativní -

**77) Jak ověříte nezávislost dvou kvalitativních znaků?**

Kontingenční tabulkou

**78) Jakým způsobem můžete zjistit, jestli existuje závislost mezi dvěma kvalitativními znaky a jakým v případě kvantitativních znaků?**

- závislost mezi 2 kvalitativními znaky ověřujeme kontingenční tabulkou a testem chi-kvadrát.
- závislost mezi 2 kvantitativními znaky měříme klasicky pomocí regresní a korelační analýzy, celkový F-test, Test t pro

jednotlivé parametry a koeficienty
- u 1 kvalitativního a 1 kvantitativního znaku se používá jednofaktorová analýza rozptylu

**79) Koeficient asociace slouží k:**
- vyjádření těsnosti 2 alternativních znaků

**80) Jak určíme nejvhodnější typ funkce při měření závislosti dvou kvantitativních znaků**
- zkušenost, logika, emp. metoda-korelační pole
- zkoušet = počítat - zpětně vybrat ten s nejvyšším korelačním charakterem

**81) Jaké jsou hlavní úlohy při měření závislosti 2 kvantitativních znaků**
- vystihnout průběh závislosti závisle proměnné na nezávisle proměnné, tak abychom mohli provádět odhady závisle proměnné na základě daných hodnot nezávisle proměnné
- změřit sílu závislosti, abychom mohli posoudit její sílu, intenzitu a abychom mohli zároveň posoudit přesnost odhadů z 1 bodu
- 1. úkol - regrese, 2. úkol – korelace

**82) Teoretický soubor výběrových průměrů**
- ze základního souboru vybereme všechny teoreticky možné VS, těch je nekonečně mnoho; v každém výběrovém souboru si vypočítáme výběrový průměr, všechny tyto průměry nám vytvoří teoretický soubor výběrových průměrů

**83) Statistická indukce**
- nejprve pořídíme výběrový soubor, na základě VS si spočítáme výběrové charakteristiky, na základě výběrových charakteristik odhadujeme charakteristiky ZS

**84) Jaké známe odhady (3)**
- bodový – jedno konkrétní číslo, které vybereme z VS, aby nám nahradilo ZS
- intervalový – stanovení intervalu, ve kterém ta neznámá charakteristika bude ležet, a určitou pravděpodobností

**85) Co je bodový odhad?**
- bodový – jedno konkrétní číslo, které vybereme z VS, aby nám nahradilo ZS

**86) Jaké znáte vlastnosti bodových odhadů a co vyjadřují, jaké jsou na ně kladeny požadavky?**
- nezkreslenost, nestrannost odhadu (střední hodnota výběrové statistiky = odhadované charakteristice)
- konzistence (odhad se s rostoucím rozsahem výběru blíží odhadované charakteristice základního souboru)
- vydatnost (co nejmenší rozptyl)
- postačujícnost (mimo ní neexistuje žádná jiná statistika poskytující další doplňující informace o odhadované charakteristice základního souboru)

**87) Proč musí být bodový odhad v základním souboru vydatný**
- můžeme použít více charakteristik odhadu, za nejvydatnější je ta, která má nejmenší rozptyl

**88) Jak spolu souvisí přesnost odhadu a spolehlivost odhadu?**
- čím širší interval spolehlivosti, tím ale menší přesnost
- spolehlivost = pravděpodobnost, se kterou bude odhadovaná charakteristika ležet v tom vymezeném intervalu; = maximální chyba, které se při odhadu s danou spolehlivostí můžeme dopustit

**89) Co znamená, že odhad je vychýlený? Znáte některé vychýlené odhady?**
- E(g) - G je tzv. zkreslení neboli vychýlení
- takový odhad vede k systematickému nadhodnocování či podhodnocování odhadované charakteristiky ZS

**90) Přípustná chyba u intervalového odhadu a k čemu jí používáme**
- chyba, které se při odhadu můžeme dopustit, aby hodnota padla do intervalu
- přesnost intervalového odhadu je charakterizována přípustnou chybou odhadu delta, která představuje polovinu délky intervalu spolehlivosti

**91) Přesnost odhadu: (2)**
- pravděpodobnost, s jakou se charakteristika nachází v intervalu
- vyneseme kritickou hodnotu příslušného rozdělení
- max chyba, které se při odhadu s danou spolehlivostí dopustíme vyjádřena hodnotou směrodatné odchylky souboru výběrových prům.
- ani jedna správně

**92) Jaké znáte metody pro získání odhadů parametrů regresních funkcí lineárních v parametrech? Napište princip metod.**
- požadavek kompenzace kladné a záporné odchylky empirických hodnot od hodnot vyrovnaných a metoda nejmenších čtverců (aby součet čtverců popsaných odchylek byl minimální)

**93) Jaké znáte metody pro získání odhadů parametrů regresních funkcí nelineárních v parametrech?**
- pro funkce nelineární v parametrech používáme linearizující transformaci
- pak použijeme metodu nejmenších čtverců, parciální derivace, dále dostaneme soustavu normálních rovnic a nakonec pomoci Cramerova pravidla (determinanty) vyjádříme $b_0$, $b_1$, ...

**94) U kterých z uvedených regresních funkcí lze k odhadu parametrů použít metodu nejmenších čtverců: přímka, parabola, exponenciála? Názor vysvětlete.**
- u přímky a paraboly, protože jsou lineární v parametrech na rozdíl od exponenciály, která není - tam je nutná lineární transformace.

**95) Pojmy: (3)**
a) alternativní hypotéza – popírá platnost nulové hypotézy

b) testovací kritérium – míra nesouhlasu výsledků pokusu s testovanou hypotézou (odpovídají-li data nulové hypotéze- testovací kritérium je rovno nule; čím více se výběrové hodnoty blíží k alternativní hypotéze, tím roste i testovací kritérium)

c) hladina významnosti – pravděpodobnost chyby 1. druhu; udává výši rizika, s jakým se $H_0$ zamítá, i když platí

**96) Alternativní hypotéza**

- popírá platnost nulové hypotézy
- přijímáme ji jestliže jsme nulovou hypotézu zamítli jako nesprávnou

**97) Vysvětlete pojem hladina významnosti a kritický obor.**

- hladina významnosti - pevná, předem zvolená pravděpodobnost chyby I. Druhu
- kritický obor - podprostor výběrového prostoru obsahující hodnoty svědčící ve prospěch alternativní hypotézy

**98) Vysvětlete pojem "síla testu" a "hladina významnosti".**

- síla testu - s jakou pravděpodobností se nedopustíme chyby II. druhu
- hladina významnosti - pevná, předem zvolená pravděpodobnost chyby I. druhu

**99) Co udává hladina významnosti alfa**

- udává výši rizika s jakým H0 zamítá, i když platí pravděpodobnost chyby 1. druhu; alfa=P(T je elem. K\H0)

**100) Testovací kriterium je**

- veličina vypočtená z výběrových hodnot, řídí se rozdělením(norm. studentovým), porovnáváme s krit. hodnotou

**101) Interval spolehlivosti a k čemu slouží**

- **i**nterval, ve kterém leží neznámé charakteristiky s určitou předem známou pravděpodobností. Prostřednictvím intervalu spol. posuzujeme přísnost odhadu. s rostoucí šířkou intervalu spol. klesá přesnost odhadu.

**102) Čím můžeme ovlivnit velikost intervalu spolehlivosti?**

- zvolenou hladinou spolehlivosti. Čím širší interval spolehlivosti, tím ale menší přesnost.

**103) Vysvětlete pojmy chyba I. druhu, chyba II. druhu a síla testu.**

a) Zamítnutí nulové hypotézy, ačkoliv ve skutečnosti platí

b) přijmutí nulové hypotézy, ačkoliv ve skutečnosti platí hypotéza alternativní

c) s jakou pravděpodobností se nedopustíme chyby II. druhu

**104) Jakým způsobem můžeme snížit pravděpodobnost chyby I. druhu a jak pravděpodobnost chyby II. druhu (při dané chybě I. druhu)?**

- chybu I. druhu snížíme tak, že místo 95% hladiny významnosti volíme 99%
- chybu II. druhu snížíme tím, že zvýšíme rozsah souboru

**105) Rozdíl mezi dvoustranným a jednostranným intervalem**

- dvoustranný - charakteristiky ZS jsou omezeny zdola i shora
- jednostranný - charakteristiky ZS jsou omezeny shora nebo zdola

**106) Obecný postup statistického testování**

- vybereme vhodný test (parametrický a neparametrický)
- formulace nulové a alternativní hypotézy
- volba hladiny významnosti
- volba testovacího kriteria
- určení kritického oboru
- výpočet hodnoty testovacího kriteria z výběrových hodnot
- rozhodnutí: jestliže vyp. hodnota t.k. padne do kritického oboru H0 se zamítá jinak se H0 nezamítá

**107) Rozdíl mezi parametrickými a neparametrickými testy**

- Parametrické - testy sloužící k ověřování hypotéz, které se týkají hodnot parametrů rozdělení - spolehlivé, vyžadují znalost ZS a hodnotu parametrů
- Neparametrické - místo původních hodnot pracujeme s pořadovými čísly - jednoduché, menší síla testu, nevyžadují znalost ZS

**108) Jaký závěr lze učinit, vede-li celkový F-test o regresní funkci k zamítnutí testované hypotézy (nejdříve uveďte, jak jsou při tomto testu formulovány testovaná a alternativní hypotéza).**

- testovaná = neexistence vlivu faktoru (o nezávislosti znaku Y na zkoumaném faktoru); Alternativní = existence
- závěr = podobnost skupinových průměrů na celý soubor

**109) Lze zvětšením rozsahu výběru ovlivnit šíři intervalu spolehlivosti odhadu střední hodnoty? Pokud ano, vysvětlete jak.**

- pokud máme větší rozsah souboru, vede to k užšímu intervalu spolehlivosti. Vzorek je více interpretativní.

**110) "Jestliže hypotézu □ = □0 zamítneme při jednostranném testu na určité hladině významnosti, pak jí zcela určitě zamítneme i při oboustranné alternativě při stejné hladině významnosti." Souhlasíte? Vysvětlete proč.**

- neplatí vždy <= dvoustranný test má při stejné hladině významnosti poloviční alfu a hodnota se do něj může, ale nemusí vejít

**111) Postup při použití testu minimální průkazové definice**

- u vyváženého modelu: dělíme diference mezi průměry ale neseřazujeme je podle velikosti; spočítáme hodnotu Dmin=g=odmocnina z (sr na 2/n); porovnáme diference s touto hodnotou; jestli je dif. větší než Dmin tak je rozdíl mezi průměrem statisticky významný

**112) Při testování spotřeby automobilu určité značky používáme:**

a) F-test
b) t-test
c) Bartlettův test

**113) Při testování více než dvou rozptylů se nepoužívá**

a) Hartleyův test
b) Bartlettův test
c) F-test

**114) Bartelův test**

- test rozptylu více souborů, H0 rozptyly se rovnají
- testuje, zda můžeme udělat analýzu rozptylů - podmínka - shodné rozptyly z normálního rozdělení
- pokud nesplňuje: a) transformace hodnot; b) doměření hodnot; vyloučení ext. hodnot

**115) Cochranův test**

- stejné rozsahy n

**116) Hartleyův test**

- řídí se f rozdělením

**117) Ferrarův – Glauberův test, k čemu se používá**

**118) Rozdílové testy**

**119) Pořadové testy**

**120) Rozptyl a jak jej definujeme**

- aritmetický průměr čtverců individuálních odchylek jednotlivých hodnot znaku x od aritmetických průměrů
- charakteristika variability
- informuje o proměnlivosti jednotlivých hodnot znaku kolem výběr. ar. prům.
- výběrový rozptyl s na 2 souboru: pozorované x1, x2....xn

**121) Jak provádíme rozklad celkového rozptylu? (4)**

- rozklad celkového rozptylu všech hodnot na dílčí rozptyly (vyjadřují podíl jednotlivých faktorů na celkovém rozptylu) a na rozptyl reziduální (zbytkový, vyjadřuje podíl faktorů působících náhodně )

**122) Kde prakticky využíváme rozklad rozptylu?**

- při počítání vnitro a meziskup. variability
- u regres. a korel. analýzy ke konstrukci indexů

**123) K čemu slouží analýza rozptylu a jaké jsou předpoklady jejího použití**

- zobecnění t-testů na více než 2 VS rozklad výběr. rozptylu na něk. částí, které jsou příslušné jednotlivým uvažovaným zdrojům variability, zkoumáme li vliv jednoho či více faktorů ne výsledky kvantitavního znaku
- analýza rozptylu je založena na předpokladu, že ze zjištěných hodnot lze provést odhad rozptylů - 2 způsoby: uvnitř tříd, mezi třídami
- typické použití v zemědělství - plemenictví

**124) Analýza rozptylu je (možnosti) (2)**

- je obecný postup, který patří do oblasti statistických metod – metodologický nástroj
- technika pro zakládání a vyhodnocování experimentu
- je více výběrový test, který testuje rozdíl mezi průměry
- zkoumá vliv jednoho i více faktorů na výsledný znak kvantitativní, každý faktor je sledován na několika úrovních (třídách), každá úroveň představuje 1 výběrový soubor; faktorem může být znak kvantitativní i kvalitativní
- pokud je úroveň každého faktoru pevně fixována – model analýzy rozptylu s pevnými efekty
- pokud jsou úrovně faktorů náhodně vybrány – model s náhodnými efekty

**125) Jaké míry jsou založeny na rozkladu rozptylu?**

- korelační koeficienty

**126) Jakou metodu byste použili k posouzení, jestli je prodej určitého zboží v průběhu dne rovnoměrně rozložen?**

Možná rozptyl - nebo řetězový index

***127)* Kde prakticky využíváme rozklad rozptylu?**

- při počítání vnitro- a meziskup. variability a u regres. a korel. analýzy ke konstrukci indexů.

**128) Metody podrobnějšího hodnocení výsledků analýzy rozptylu**

- metody mnohonásobných porovnání, s-metoda (Scheffeho metoda), t-metoda (metoda minimální důkazové diference), Cramerova metoda, Duncanova metoda

**129) Duncanova metoda**

**130) Postup při použití Duncanova testu**

- hodnotám přiřadíme pořadové číslo a spočítáme diferenci
- uspořádáme do tabulky, určíme krit. hodnotu diference a porovnáme diference s kritickými, jeli dif. větší než krit. pak je rozdíl statisticky významný

**131) Kramerova metoda a k čemu slouží**

- pomocí ní se provádí podrobnější hodnocení výsledku v analýze rozptylu u neváženého modelu
- rozdíl mezi průměry je statisticky významný, jestliže jejich diference je větší než výraz na pravé straně nerovnosti

**132) Scheffeho metoda**

- rozdíl mezi průměry je stat. významný, jestliže jejich diference je větší než vypočítaná hodnota

**133) Tukeyův test**

**134) Rozdíl mezi vyváženým a nevyváženým modelem analýzy rozptylu**

- vyvážený - stejný počet opakování v každém řádku
- nevyvážený - různý počet opakování v ...

**135) Rozdělení F při analýze rozptylu**

- F,alfa (m-1; n-m)
- rozdělení F pro f1=m-1 a f2=n-m stupňů volnosti

**136) K čemu se používá F-test**

- test o rozdílu 2 výběrových rozptylů

**137) Jaké testy předcházejí analýze rozptylu**

a) Bartletův
b) Cochranův
c) Hartleyův)

**138) Kdy používáme analýzu rozptylu + podmínky použitelnosti**

(1) používá se, když zkoumáme vliv jednoho či více faktorů na výsledný znak kvantitativní
(2) pomínky použitelnosti - normalita rozdělení ZS z kt. jsou pořízeny VS; statistická nezávislost náhodných chyb; existence shodných rozptylů ZS ze kt. jsou pořízeny VS

**139) Co je interakce u AR?**

**140) Jak se provádí a proč podrobnější vyhodnocení AR?**

**141) Modely AR dvojného třídění**

- zobecnění t-testů na více než 2 VS rozklad výběr. rozptylu na něk. částí, které jsou příslušné jednotlivým uvažovaným zdrojům variability, zkoumáme li vliv jednoho či více faktorů ne výsledky kvantitavního znaku
- analýza rozptylu je založena na předpokladu, že ze zjištěných hodnot lze provést odhad rozptylů - 2 způsoby: uvnitř tříd, mezi třídami
- typické použití v zemědělství - plemenictví

**142) Vypište neparametrické testy**

- testy dobré shody – $\chi 2$ test, Kalgor- Smirnovův test
- vlastní neparametrické testy – Wilcox- Whiteův test (parametrický t-test), znaménkový test (parametrický pár. t-test), Wilcoxonův test (parametrický t-test), Kruskal-Wallisův test (parametrická analýza rozptylu), Dixonův test, Friedmanův test, test iterací

**143) Výhody a nevýhody neparametrických testů (3)**

- v praxi se často setkáváme s výběry poměrně malých rozsahů, které pocházejí z výrazně nenormálních souborů, nebo ze souborů, o jejichž rozdělení nic nevíme
- nepředpokládáme specifikované rozdělení základního souboru, z něhož se získává náhodný výběr
- výhoda – nezávislost na tvaru rozdělení studovaných náhodných veličin, použitelnost pro studium jak znaků kvantitativních, tak znaků kvalitativních (obecnější použití); výpočetní jednoduchost
- nevýhody – menší síla (schopnost odhalit nesprávnost testované hypotézy)

**144) Které testy používáme jako neparametrickou obdobu párového t-testu a jak je provádíme**

- znaménkový: spočítáme diference mezi páry hodnot; počet kladných hodnot z+; počet záporných hodnot z-; menší z oboru je testovací kritérium z: z je větší než z,alfa (sečtou se počty ve skupinách a min. výsledná hodnota je testové kriterium, které se porovná s tabulkovou hodnotou)
- wilcoxonův: spočítáme diference; k nenulovým diferencím přiřadíme vzestupně pořadová čísla; pořadová čísla roztřídíme zpětně do dvou skupin podle znamének diferencí; v obou skupinách pořadová čísla sečteme a dostaneme W+ a W-; Wmin(W+,W-) menší z obou čísel je test. krit.; W je menší než W,alfa - H0 se zamítá;

**145) Který test používáme jako neparametrickou obdobu jednoduché analýzy rozptylu a jak se tento test provádí**

- Kruskal-Wallisův test: hodnoty sloučíme do jednoho výběru; seřadíme od největší do nejmenší; očíslujeme vzestupně; sečteme pořadová čísla pro každý řádek zvlášť; výpočet test. kritérium, porovnáme s kritickou hodnotou

**146) Postup při Wilcoxon-Whiteova testu**

- hodnoty sloučíme do jednoho výběru
- seřadíme od nejmenší do největší
- očíslujeme hodnoty vzestupně
- zpětně roztřídíme do VS
- sečteme pořadová čísla; T=min(Tx;Ty); T je menší než T,alfa - H0 se zamítá
- u souboru n větší než 20 normální aproximace odhadnem

**147) Test o shodě průměru základního souboru (2)**

- je určen k testování rozdílů mezi průměry u nezávislých výběrů
- používá se v případě párových výběrů
- používá se k testování rozdílu mezi výběrovým prům. a předpokládanou hodnotou prům. základního souboru
- určuje, zda prům. zákl. souboru je stat. významný
- ani jedna správně

**148) K čemu je Friedmanův test? (4)**

- máme více než 2 závislé soubory, tzn. k >2

**149) Uveďte aspoň 3 testy, u nichž použijete jako kritické hodnoty kvantity rozdělení chí-kvadrát. Uveďte, čeho se týkají a co říkají.**

- chi-kvadrát dobré shody
- test o parametru delta exponenciálního rozdělení
- test hypotézy o rozptylu v základním souboru

**150) Jaké znáte testy shody? K čemu slouží?**

- Chi-kvadrát test dobré shody, shoda 2 průměrů, rozptylů
- porovnávají dva různé soubory

**151) Uveďte alespoň dva případy, kdy použijete při testování hypotézy testované kriterium, které má rozdělení chí-kvadrát.**

- chi-kvadrát dobré shody, test o parametru delta expon. rozdělení a test hypotézy o rozptylu v zákl. souboru.

**152) Kdy používáme Fisherův test**

- když je n menší než 20 nebo n je prvkem (20,40) a jednina rel. čet. 15

**153) V kterých tabulkách se provádí $\chi 2$ test?**

**154) Jaké jsou základní předpoklady Kolmogorov-Smirnovova testu**

**155) Kdy používáme chí-kvadrát test?**

**156) Jak dělíme metody vícerozměrné stat. analýzy z hlediska klasifikace jednotek**

- shluková analýza
- diskriminační analýza
- faktorová analýza...

**157) Shluková analýza**

- vícerozměrná statistika
- souhrnný název pro řadu výpočetních postupů, jejichž cílem je rozklad daného souboru na několik relativně homogenních množin (shluků) a to tak, aby jednotky uvnitř jednotlivých shluků si byly co nejvíce podobné a jednotky patřící do jiných shluků co nejvíce nepodobné
- míra vzdálenosti - je jejím určením
- vytváří celky metody - nejbližšího souseda, nejvzdálenějšího souseda, průměrná vzdálenost - centrální, mediánová, warolová(graf)

**158) Diskriminační analýza**

- řeší problematiku vícerozměrné klasifikace
- předmětem je nalezení statisticky nejvhodnějšího způsobu rozlišení mezi 2 či více soubory statistických jednotek
- klasická úloha diskriminační analýzy spočívá v tom, že jsou předem známy 2 či více skupin jednotek a o každé víme do které skupiny patří
- na základě naměřených údajů se pro každou skupinu vypočítá diskriminační fce sloužící k dodatečnému zařazování nových jednotek
- do určité skupiny je zařazena ta jednotka, pro níž je pravděp. příslušnost této jednotky ke skupině největší
- slouží k dodatečnému zařazení znaků do předem stanovených skupin

**159) Faktorová analýza**

- umožňuje objasnit strukturu pozorovaných závislostí, redukuje počet výchozích proměnných pomocí hypotetických faktorů při minimální ztrátě informací a odhaduje skryté vztahy mezi proměnnými
- použití při zpracování dotazníků

**160) Kovarianční analýza**

- umožňuje rozčlenění variability proměnné y na části, které jsou přiřaditelné jednak kvalitativním a jednak kvantitativním vlivům

**161) Analýza hlavních komponent**

- podstatou je transformace souboru napozorovaných proměnných do nových proměnných, tzv. hlavních komponent, které jsou vzájemně nezávislé a jsou seřazeny dle velikosti svého příspěvku ke vysvětlení celkového rozptylu napozorovaných proměnných
- tato metoda citlivá na delta jednotek větší, než hodnoty normují
- použití: odhalení skrytých vztahů mezi proměnnými
- založena na bezezbytkovém vysvětlení celkového rozptylu proměnných

**162) Kanonická korelační analýza**

- vychází z logického členění proměnných do 2 skupin (výchozí + cílové veličiny) - každá skupina je charakterizována 1 soubornou veličinou = kanonickou proměnnou = lineární kombinace původních proměnných dané skupiny

**163) Jak se provádí odhad pro celou regresní přímku?**

**164) Rozdíl mezi jednoduchým a vícenásobným modelem regrese a korelace**

- počet nezávisle proměnných
- v modelu jednoduché stat. závislosti je předpokládáno, že změny závisle proměnných jsou vyvolány změnami jediné nezávislé proměnné, ostatní vlivy jsou považovány za náhodné
- teorie vícenásobné regrese a korelace je zobecněním teorie jednoduché regrese a korelace

**165) Korelace x regrese (2)**

- korelace –těsnost, síla, míra závislosti mezi kvantitativními statistickými znaky
- regrese – popis průběhu závislosti mezi kvantitativními znaky pomocí regresního modelu ( tímto modelem je regresní funkce

**166) Základní podmínka metody nejmenších čtverců**

- $\sum d_i^2=0$ - vede k jednomu závěru
- $\sum d_i=min$ - můžeme spočítat parametry

**167) Metoda nejmenších čtverců a k čemu se používá**

- matematická metoda s jejíž pomocí můžeme vypočítat parametry regresních fcí(s výjimkou fcí, které nemůžeme převést na aditivní tvar)

**168) Pás spolehlivosti regresní přímky – vysvětlit (3)**

**169) Jaká je regresní funkce nelineární v parametrech, jak parametry určujeme (3)**

**170) Co se dá zjistit z regresní přímky?**

**171) Co vyjadřuje regresní index a koeficient?**

**172) Test regresního koeficientu.**

**173) Podle jakých kritérií se vybírá nejvhodnější funkce pro vícenásobnou nelineární regresi?**

**174) Postup při metodě stupňovité regrese (4)**

**175) Napište rovnici regresní funkce a korelační koeficient, když máš zadáno x s pruhem (celkový průměr), y s pruhem (celkový průměr), s x, s y a s xy**

**176) Lze testovat významnost regresní funkce a určit její interval spolehlivosti? Pokud ano, napište stručný postup.**

**177) Určení závislosti u nelineární regrese**

**178) Vícenásobná regresní funkce a jednoduchá regresní funkce – rozdíl**

**179) Regrese na hlavních komponentách**

- vícenásobný regresní model předpokládá nezávislost nezávislé proměnné
- tento předpoklad bývá v praxi často nahrazen a proto se doporučuje do vícenásobní regrese nahradit nezávislé proměnou hlavními komponentami

**180) Regresní koeficient u lineární regrese**

- byx - vyjadřuje o kolik se změní závisle proměnná y, jestliže se nezávisle proměnná x změní o jednotku
- použití k odhadům změny

**181) Aditivita vícenásobného regresního modelu**

- funkci pro vícenásobnou regresi získáme jako součet jednoduchých regres. fcí

**182) Jakým způsobem určujeme parametry regresních funkcí? (3)**

- metoda nejmenších čtverců – obecná soustava normálních rovnic

**183) Beta koeficienty a k čemu se používají**

- normované přepočtení regresní koeficienty
- k určení podílu jedn. nezávisle proměnné u regres. odhadu závisle proměnné.(vícenásobná lineární regrese)

**184) Jak postupujeme při určení nejvhodnějšího typu regresní funkce**

- na základě zkušenosti, logické posouzení, empirické posouzení (korel. pole), výpočet mnoha funkcí-výběr nejlepší

**185) Jak provádíme výběr nejvhodnější nelineární regresní funkce? (3)**

- podle toho, která nelineární regresní funkce má nejvyšší hodnotu indexu korelace

**186) Co jsou to sdružené regresní přímky**

- hodnoty korelačních koef. převádí v hodnoty z a vždy se řídí alespoň přibližně normálním rozdělením

**187) Typy nelineárních regresních fcí**

- polynom, parabola 2. st., exponenciála, mocninná funkce, hyperbola, růstová funkce

**188) Interakce ve vícenásobné regresní analýze (2)**

- podíl nezávisle proměnných na regresním odhadu závisle proměnné
- posuzuje vhodnost regresní funkce
- představuje vzájemné působení proměnných (nejde to slovo přečíst)
- vyjmenuj vztahy mezi koeficienty
- ani jedna možnost

**189) Korelační pole**

- empirická metoda pro nalezení vhodné regresní funkce, vyjádření párové závislosti
- vyjadřuje vzdálenost mezi teoretickou a skutečnou hodnotou u vybrané funkce

**190) Korelační tabulka**

- kombinační tabulka, která vyjadřuje průběh závislosti 2 proměnných (kvalitativní znaky)
- na úhlopříčce je pevná závislost
- čím jsou více soustředěné kolem úhlopříčky, tím je silnější závislost a naopak
- těsnost závislostí v ní měříme korelačním poměrem

**191) Korelační tabulka**

- kombinační tabulka. která vyjadřuje průběh závislosti 2 proměnných(kvalit znaky)
- těsnost závislostí v ní měříme korelačním poměrem

**192) K čemu slouží Spearmanův koeficient korelace pořadových čísel**

- charakterizuje těsnost jednoduché závislosti kvantitativních znaků
- nabývá hodnot 0-1 původní hodnoty nahradíme pořadovými čísly je to u korel.
- jedná se o neparametrickou charakteristiku
- měří těsnost jakékoli statistické závislosti, která je monotónní
- poskytuje rychlou a dostatečně přesnou informaci o těsnosti sledované závislosti

**193) Totální koeficient korelace a determinace**

- totální koeficient korelace - vyjadřuje těsnost závislosti Y na všech nezávislých prom. X
- totální koeficient determinace - druhá mocnina korelace; udává z kolika % je nez. proměnná Y ovlivněna všemi uvažovanými nezávisle proměnnými X

**194) Koeficient vícenásobné totální korelace**

-měří těsnost závislosti závisle proměnných na všech uvažovaných nezávisle proměnných

**195) Koeficient parciální korelace**

-měří těsnost závisle proměnných na jedné z uvažovaných nezávisle proměnných

**196) Korelační index je (možnosti) (2)**

- měříme jím těsnost závislosti u nelineárních funkcí
- je v intervalu od 0 do 1, čím více se blíží 1 tím je závislost silnější, čím blíže 0 tím je závislost slabší

**197) Jak ověřujeme statistickou významnost korelačního indexu a korelačního koeficientu? (3)**

- korelační koeficient – podle t-Studentova rozdělení, tabulky kritických hodnot korelačního koeficientu , neprovádíme test, pouze porovnáváme hodnotu korel. koeficientu s tabulkovou , je-li vypočtená > tabulková, můžeme rovnou korel. koeficient považovat za statisticky významný
- korelační index - netestujeme

**198) Co měří dílčí korelační koeficient a co vícenásobný korelační koeficient?**

- dílčí korelační koeficient vyjadřuje těsnost lineárního vztahu mezi dvěma proměnnými při vyloučení vlivu jedné nebo více dalších proměnných
- vícenásobný koeficient korelace charakterizuje těsnost závislosti jedné proměnné na lineární kombinaci jiných proměnných

**199) Koeficient determinace - $r^2$; vyjádření v %**

- udává z kolika % je závisle proměnná ovlivněna uvažovanou nezáv. Proměnnou (u lineární regrese a korelace)

**200) Totální koeficient determinace**

- vyjadřujeme v % a udává nám z kolika % je závislá proměnná ovlivněna uvažovanými nezávisle proměnnými

**201) K čemu lze prakticky využít determinační index? O čem na jeho základě můžeme rozhodnout?**

- index determinace vynásobený 100 udává relativně v procentech tu část rozptylu závisle proměnné z, kterou se podařilo vysvětlit použitou regresní funkcí
- čím bližší je jedné, tím je daná závislost silnější.

**202) Může být vícenásobný korelační koeficient menší než některý z jednoduchých korelačních koeficientů?**

- nemůže <= vícenásobný je složen z jednoduchých

**203) Jakou metodu byste použili k posouzení, jestli je prodej určitého zboží v průběhu dne rovnoměrně rozložen?**

- rozptyl nebo řetězový index

**204) Posuďte, která varianta vztahů mezi koeficienty je možná:**

**a) byx=-5 , bxy=-0,2**

**b) ryx=-0,72 , bxy=-2,1**

**c) byx=0,5 , bxy=2,1**

**d) byx=-0,1 , bxy=0,7**

Možná je varianta A - Nesmí mít opačná znaménka a jejich součin musí být menší nebo roven jedné.

**205) Je možné, aby se korelační koeficient rovnal korelačnímu poměru? Kdy?**

- ano. Když je daná závislost naprosto shodná se svou regresní přímkou.

**206) Na jakém principu je založena konstrukce měr těsnosti závislosti?**

- na principu rozkladu rozptylů

**207) U kterých z uvedených regresních funkcí lze k odhadu parametrů použít metodu nejmenších čtverců: přímka, parabola, exponenciála ? Názor vysvětlete.**

- u přímky a paraboly, protože jsou lineární v parametrech na rozdíl od exponenciály, která není - tam je nutná lineární transformace.

**208) Jak určíme ve vícenásobné lineární regresní funkci podíl jednotlivých nezávisle proměnných v regresním odhadu závisle proměnné**

– **β koeficienty regresní fce** (dílčí regresní koeficienty na které jsou přepočítány nezávislé odchylky)- o kolik směrodatných odchylek se změní závisle proměnná jestliže se nezávisle proměnná změní o 1 odchylku

**209) Co je multikolinearita a jak ji zjistíš? (5)**

- je to závislost mezi vysvětlujícími proměnnými
- informace o ní čerpáme z matice korelačních koeficientů a jejího determinantu
- jestliže jsou párově nekorelované, multikolinearita neexistuje a D=1, pokud D=0, potom mluvíme o úplné multikolinearita
- většinou je škodlivá od hodnoty 0,75
- jestliže multikolinearita existuje, tak je některá proměnná x v modelu zbytečná

**210) Co je zdánlivá korelace?**

- zdánlivá korelace spočívá v tom, že je někdy možné pozorovat silnou závislost mezi proměnnými i v případě, kdy mezi proměnnými ve skutečnosti závislost buď skoro, nebo vůbec neexistuje.

**211) Doprovodná regrese**

- u analýzy určujeme, zda kromě regrese na kvantitativním znaku existuje i regrese na kvalitativním znaku - f-test

**212) Čím měříme těsnost závislosti**

- u lineárních závislostí - koef korelace, determinace
- u nelineárních závislostí - index korelace, determinace

**213) Kanonická korelační analýza**

- vycházejí z logického členění proměnných do dvou skupin
- každá skupina proměnných je charakterizována 1 souhrnnou veličinou, tzv. kanonickou proměnnou = lineární kombinace pův. prom. dané skup.

**214) Jak zjistíme zda je rozdíl korelačních koeficientů statisticky významný**

- tabulky - pomocí testu o korelačním koeficientu
- tr=ryx/(odm(1-r na 2xy)/(n-2))
- tr - má studentovo rozdělení pro n-2 stupňů volnosti
- t větší než t alfa - statisticky významný

**215) z-transformace, kdy ji používáme, její význam**

- hodnoty r1 převádíme na z1 a r2 na z2 u testu o rozdílu 2 korelačních koeficientů
- metoda pro intervalový odhad korelačního koeficientu, jestliže n<100
- při intervalovém odhadu korelačního koef. v případě že n je menší než 100 není možná normální aproximace

**216) Fischerova Z-transformace**

- speciální postup, kdy hodnoty korelačního koeficientu převádíme(pomocí tabulky) na hodnotu Z(r1=Z1, r2=Z2)
- testovací kriterium - u=(Z1/Z2)/(odm(1/(n1-3))+(1/(n2-3)))
- u je větší než u alfa ... H0 se zamítá

**217) Koeficient korelace – co vyjadřuje, jaké známe (4)**

**218) Jak určíme parametry mocninné funkce**

**219) Kdy se používá korelační poměr, jakých nabývá hodnot**

**220) Rozdíl mezi dílčím a totálním korelačním koeficientem**

**221) Rozdíl mezi volnou a pevnou závislostí**

**222) Index korelace – co to je, co vyjadřuje, jak je vymezen, kdy se používá**

**223) Jakých hodnot nabývá korel. index a jak ho interpretujeme; co vyjadřuje index determinace**

- korelační index nabývá hodnot<0;1>, čím více se blíží 1 tím je závislost silnější, čím blíže 0 tím je závislost slabší
- index determinace - I na 2xy - vyjádřen v % udává z kolika % je závisle proměnná ovlivněna uvažovanou nezávisle proměnnou

**224) Uveďte vztah mezi korelačním koeficientem a regresními koeficienty – ryx = √byx.bxy**

**225) Čím měříme těsnost závislosti**

– korelačním koeficientem (lineární regrese), korelačním indexem (nelineární regrese)

**226) Jak se vyjadřuje doprovodná regrese u kovariance?**

**227) Je možné, aby se korelační koeficient rovnal korelačnímu poměru? Kdy?**

Ano. Když je daná závislost naprosto shodná se svou regresní přímkou.

**228) Množné znaky - kontingenční tabulka**

- alespoň 1 znak většinou více než 2, měříme pouze test závislosti; Pearsenův koeficient kontingence vyzívá chí-test

**229) K čemu slouží normovaný koeficient kontingence**

- kontingenční tabulky mají různý počet řádků a sloupců - používáme ho pro porovnávání

**230) Koeficient asociace slouží k:**

- vyjádření těsnosti 2 alternativních znaků

**231) Čím určujeme těsnost závislosti v asociační tabulce**

- koeficient asociace <-1;1>, obdoba koeficientu korelace

**232) Asociační tabulka**

- vyjadřuje vztah 2 alternativních znaků
- asociační přímka, těsnost, koeficient asociace (Yuelův koeficient asociace) v tabulce prvky a a b b
- používají se zde upravené chí-testy

**233) Jaké funkce popisují průběh závislosti v asociační tabulce**

- lineární fce

**234) Jak měříme těsnost závislosti v kontingenční tabulce**

- Pearssonův koef kontingence - C (0,1)
- čupronův koef kont. - K (0,1)
- normovaný koef kontingence - Cn (0,1)
- vyjádřit průběh závislosti v kontingenční tabulce neumíme

**235) Jak ověřujeme závislost znaků v asociační tabulce**

- chí-kvadrát test, fischerův test

**236) Jaké funkce lze použít k popisu průběhu závislosti v asociační tabulce**

- pouze lineární asoc. přímku

**237) Rozdíl mezi korelační a kontingenční tabulkou?**

**238) Kontingenční tabulka, nakreslit obecné schéma**

**239) Fischer-pravděpodobnostní test tabulek**

**240) K čemu slouží normovaný koeficient kontingence**

- kontingenční tabulky mají různý počet řádků a sloupců - používáme ho pro porovnávání
- aby byl dostačující - postačující odhad

**241) Jaké testy se používají v asociační tabulce**

- upravený chý-kvadrát test; fisherův test
- jakými způsoby lze vyjádřit trend časové řady
- graficky, mechanicky (klouzavé prům.), analyticky(trend. fce.)

**242) Charakteristické rysy časových řad**

- uspořádaná řada údajů, které odlišují v čase
- rysy – trend (vývojové tendence v čr-vzestupný, sestupný, stacionární), kolísání (odchylky od rovnoměrného vývoje-periodické(cyklické, sezónní), náhodné(okamžikové, intervalové))

**243) Elementární charakteristiky čas. Řad**

- slouží k rychlé informaci o charakteru a chování ukazatele v čas. řadě

**244) Harmonická analýza, podstata, k čemu se používá a kde**

- soubor postupů používaných při analýze vlnitých pohybů u přírodních jevů

- v některých ČŘ se přesně pravidelně odchylují skutečné hodnoty od celkové tendence
- graf připomíná spojité periodické funkce goniometrického typu

**245) Co je autokorelace? Kde se s ní setkáváme a jak ji měříme?**

- je korelace mezi sousedními odchylkami od trendu
- ověřujeme ji pomocí Durbin- Watsonova testu
- závislost mezi 2 po sobě jdoucími členy v čř
- u takové čř nelze provést měření těsnosti závislosti
- zjišťuje se pomocí koeficient autokorelace (nízké hodnoty - neexistuje autokorelace)

**246) Jak se měří průměr v okamžikové a intervalové ČŘ?**

**247)** Analýza časových řad

- trend x periodická složka x náhodná složka

**248) Kdy používáme klouzavé průměry**

- při mechynickém popisu trendu - očišťuje čř od náhodného a period. kolísání

**249) K čemu používáme klouzavé průměry?**

- k vyrovnávání ČŘ
- nevyrovnáme ji celou najednou, ale po částech

**250) Co si představujete pod pojmem "dekompozice časové řady"?**

- rozklad čas. řady na 4 složky: trend, sezónní kolísání, cyklické kolísání a náhodná složka
- u aditivního modelu to sčítáme, u multiplikativního násobíme

**251) Jaké jsou předpoklady o náhodné složce časové řady? Proč ověřujeme jejich splnění a co to znamená, když se přesvědčíme, že nejsou splněny?**

- kompenzace v rámci časové řady, homoskedasticita (jejich variabilita se v čase nemění), jsou vzájemně nekorelované
- nesplnění = nepodařilo se v časové řadě eliminovat systematickou složku beze zbytku

**252) Co je extrapolace časové řady a jaký má význam?**

- dělání prognóz do budoucnosti
- nejčastější způsob použití extrapolačních kritérií je založen na simulaci spočívá v tom, že z analýzy řady oddělíme určitou část pozorování a na vhodný T funkce usuzujeme podle toho, jak dobře extrapoluje tato pozorování.

**253) Pomocí jakých měr posoudíte přesnost extrapolace?**

- Index korelace a determinace, u lin. záv. pomocí korel. koeficientu. Nebo pokud nemůžu model popsat žádnou závislosí tak poměrem determinace.

**254) Jak vyberete vhodnou metodu pro extrapolaci časové řady?**

- nakreslím graf a posoudím průběh, potom analyzuji růstové charakteristiky a diference

**255) Z údajů časových řad dvou ukazatelů y a x byl vypočten korelační koeficient ryx = 0,8. Jaké závěry z této hodnoty můžeme učinit?**

- mezi y a x existuje středně silná lineární závislost

**256) Z údajů časových řad ukazatele y a ukazatele x byl vypočten korelační koeficient ryx = 0,6. O čem vypovídá?**

Středně silná závislost

**257) Časová řada čtvrtletních údajů je rostoucí a sezónní indexy rostou úměrně trendu. Jak byste posoudili sezónnost takové časové řady?**

- jedná se o proporcionální sezónnost

**258) Co chápeme pod pojmem paralelismus časových řad a co způsobuje?**

Souběh časových řad: Systematické složky časových řad, zejm. celkový vývoj tendence, mohou mít velmi podobný průběh v čase, což může vést k pozorování silné závislosti mezi proměnnými, jež ve skutečnosti vůbec závislé nejsou. Proto je třeba zjistit, zda existuje nějaký vztah mezi náhodnými složkami, a pak teprve pak usuzovat na souv, mezi ČŘ.

**259) Popište základní princip exponenciálního vyrovnávání časové řady.**

Pozorování blízká současnosti jsou pro odhad parametrů důležitější, než pozorování vzdálenější, starší, a měla by jim proto být přisuzována větší váha.

**260) Jaké znáte metody pro získání odhadů parametrů trendových funkcí nelineárních v parametrech?**

- linearizovat model vhodnou transformací, pak požadavek kompenzace kladné a záporné odchylky empirických hodnot od hodnot vyrovnaných a metoda nejmenších čtverců (aby součet čtverců popsaných odchylek byl minimální)

**261) Náhodné kolísání časové řady, jak ho měříme**

- výsledek náhodných jevů, které nelze předurčit
- pomocí abs. prům. odch. a rel. prům. odch.

**262) Jak se určí průměr časové řady okamžikové a intervalové**

- okamžiková: y s pruhem=((y1/2+y2+...+yn-1*yn/2)/(n-1))
- intervalová - počítá se jako prostý ar. průměr

**263) Jakými zp. lze vyjádřit trend časové řady**

- graficky, mechanicky(klouz. prům.), analyticky(trend. fce.)

**264) Jak počítáme v čas. řadě koef. růstu a prům. koef. růstu**

- koef. růstu - Ki=(yi/(yi-1))*100
- prům. koef. růstu - k=ntá odm. z (k1*k2*...*kn)

**265) Sezónní index je**

- poměr skutečných a teoretických hodnot

**266) Extrapolace**

- jedná se o statistické prognózování kdy pomocí trendové fce a sezónních indexů je možno odhadnout budoucí vývoj daného ukazatele

**267) Interpolace**

- přibližné určení chybějící hodnoty sledováného ukazatele v ČR za předpokladu, že známe sousední hodnoty
- provádí se 2 způsoby: prostřednictvím 2 sousedních údajů; prostřednictvím všech hodnot v ČŘ

**268) Popište základní princip exponenciálního vyrovnávání časové řady.**

- Pozorování blízká současnosti jsou pro odhad parametrů důležitější, než pozorování vzdálenější, starší, a měla by jim proto být přisuzována větší váha.

**269) Co jsou odvozené časové řady a aspoň dva typy**

(z druhotných ukazatelů, např. HDP/na obyvatele)

**270) Čím vyjadřujeme vliv sezónního kolísání**

**271) Co je stacionární časová řada**

**272) Odvozené časové řady**

**273) Postup při korelaci časové řady.**

**274) Co jsou bazické a co řetězové indexy a jaký je mezi nimi vztah?**

- jsou to indexy, které se týkají časové posloupnosti
- bazický index má základ v zákl. období, řetězový vždy v předchozím období
- bazické se dají konstruovat pomocí násobení řetězových a řetězové zase dělením bazických

**275) Jaký index (indexy) použijete k posouzení změny cen vyráběné produkce u podniku, který vyrábí 7 různých výrobků?**

- Souhrnný index - různé druhy výrobků nemůžeme sčítat - Laspeyesův, Paascheho, Fischerův...

**276) Která znáte rozdělení spojitých náhodných veličin? Uveďte oblasti jejich použití.**

Normální, Logaritmicko-normální, Exponenciální, Chí-kvadrát, Studentovo, Rozdělení F

**277) Jaká je hlavní myšlenka rozkladu indexů metodou postupných změn?**

Index hodnot analyzovaného ukazatele rozkládáme na dílčí analytické indexy, které se skládají součinem.

**278) Může nastat situace, že index ceny hovězího masa stoupne, i když ve všech prodejnách cena poklesla? Zdůvodněte svůj názor.**

Z ekonomického hlediska ano. Asice v případě, že by indexy cen ostatního zboží a služeb stouply více.

**279) Jak budete postupovat, chcete-li zjistit, jak se na změně průměrné dovozní ceny určitého výrobku podílely změny cen z jednotlivých dovozních zemí a jak změna struktury dovozu podle dovozních zemí?**

Indexem proměnlivého složení

**280) Charakterizujte situaci, kdy použijete složený individuální cenový index a kdy souhrnný cenový index.**

Složený index: Chceme-li zjistit, jak se na změně průměrné ceny určitého výrobku podílely změny jednotlivých cen, a jak změna struktury Souhrnný index: nemůžeme-li sčítat různé druhy výrobků - Laspeyesův, Paascheho, Fischerův...

**281) Podle jakých hledisek dělíme indexy**

- věcné, časové, prostorové

**282) Sezónní index**

- podíl pův. hodnot a hodnot očištěných od sez. vlivů

**283) Souhrnné indexy úrovně - (2)**

- Indexy stejnorodý intenzity ukazatele; nestejnorodý extenzity ukazatele; srovnávající nest. prod.; nest. intenzity ukazatele; ani jedna

**284) Výsledný ukazatel intenzity je stejnorodý pokud:**

- **o**ba intenzivní ukazatele jsou stejnorodé; oba extenzivní ukazatele jsou stejnorodé; oba nestejnorodé; jeden extenziv. uk. je stejnorodý, druhý nestejnorodý

**285) Stejnorodé a nestejnorodé ukazatele**

- stejnorodé - lze li srovnat ext. ukaz. součtem v přir. měrných jedn.
- nestejnorodé - nelze je mezi sebou spočítat

**286) Indexy**

- podíl dvou hodnot téhož ukazatele; individuální a souhrnný; věcný, prostorový, časový
- individuální - indexy stejnorodých ukazatelů
- souhrnné - indexy nestejnorodých ukazatelů
- obě tyto skupiny se dělí na indexy množství a indexy úrovně

**287) Typy stat. ukazatelů**

- okamžikové, intervalové, primární, sekundární, extenzivní, intenzivní, stejnorodé, nestejnorodé

**288) Rozklad indexu hodnoty.**

**289) Rozdíl mezi souhrnným indexem množství a indexem úrovně.**

**290) Fischerův index – k čemu se používá.**

**291) Montgomeryho index – k čemu (3)**

**292) Stejnorodé a nestejnorodé ukazatele**

- stejnorodé - lze li srovnat ext. ukaz. součtem v přir. měrných jedn.
- nestejnorodé - nelze je mezi sebou spočítat

**293) K čemu Guttmanův koeficient prediktability (3)**

**294) Co je to vydatnost?**

**295) K čemu využíváme Theilův koeficient?**

**296) Co vyjadřuje Bayesův vzorec?**

- v případě, že jsou známy nejen nepodmíněné pravděpodobnosti P(Bi), ale i podmíněné pravděp. P(A/Bi) a je známo, že výsledkem je nastoupení jevu A, lze P(Bi) spočítat podle vzorce: P(Bi) = P(Bi)*P(A/Bi) / sum P(Bi)*P(A/Bi)

**297) Zemědělský závod**

- souhrn pozemků větších než 0,1 ha na kterých se hospodaří
- souhrn pozemků menší než 0,1 ha když:

**298) Jak odhadujeme v zemědělství ha výnosy**

- subjektivní odhad - celé území se rozdělí na obvody a v každém se odhaduje prům. výnos
- objektvní metoda - výběrové šetření, přímé měření; sledují se ha výnosy - klouz. prům.

**299) Kultura**

- pozemek s trvale určeným způsobem využívání půd(orná půda, vinice), plodina - brambory rané...

**300) Rozdíl mezi sklizňovou plochou a sklizenou plochou.**

**301) Statistická síť pro zemědělství.**

**302) Co je šetření Agrocenzus a kdy se toto šetření naposled uskutečnilo**

**303) Systém rodinných účtů**

**304) Vyjmenovat sčítací testy**

**305) Agrocenz**

- vychází z prahových hodnot; výběrová šetření; vysoké náklady

**306) Jaká měření provádí statistika v zemědělské rostlinné výrobě**

# Téma 11 – Příklad 1

## Zadání příkladu:

Máme k dispozici časovou řadu celkových dodávek motorové nafty v tis. tun v ČR v letech 2008 – 2012. Charakterizujte vývoj celkových dodávek nafty pomocí řetězových indexů a bazických indexů se základem v roce 2009. Alespoň jeden z indexů vždy interpretujte.

| **Rok** | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|
| **Dodávky nafty v tis. t** | 3458 | 3154 | 3310 | 3067 | 3180 |

## Vypracování příkladu:

Výpočet řetězových indexů se provádí podle: $i_{i/i-1} = \frac{u_i}{u_{i-1}};\ i = 1,2,3,\ldots,n$.

$u_i$ ..... hodnota ukazatele v i-tém období,

$u_{i-1}$ ..... hodnota ukazatele v období *i-1*.

Hodnotu prvního řetězového indexu nelze určit, protože bychom k výpočtu museli mít hodnotu ukazatele za rok 2007. Do Tab. 1 uděláme do pole k roku 2008 puntík, který nám dává informaci, že hodnota existuje, ale nemůžeme ji z poskytnutých dat vypočítat. Druhou hodnotu vypočítáme jako: $i_{2009/2008} = \frac{u_{2009}}{u_{2008}} = \frac{3154}{3458} = 0{,}912$. Další bude dána jako: $i_{2010/2009} = \frac{u_{2010}}{u_{2009}} = \frac{3310}{3154} = 1{,}049$. Obdobným způsobem spočítáme zbývající dva řetězové indexy.

Výpočet bazických indexů se provádí podle: $i_{i/B} = \frac{u_i}{u_B};\ i = 1,2,3,\ldots,n$.

$u_B$ ..... hodnota ukazatele v bazickém období.

V poli Tab. 1 u roku bazického zapíšeme hodnotu 1 (tj. 100 %). S bazickým rokem pak poměřujeme všechny ostatní hodnoty. Postupně tedy počítáme:

$$i_{2008/2009} = \frac{u_{2008}}{u_{2009}} = \frac{3458}{3154} = 1{,}096$$

$$i_{2010/2009} = \frac{u_{2010}}{u_{2009}} = \frac{3310}{3154} = 1{,}049$$

$$i_{2011/2009} = \frac{u_{2011}}{u_{2009}} = \frac{3067}{3154} = 0{,}972 \text{ atd.}$$

***Tabulka 1 – Hodnoty řetězových a bazických indexů***

| **Rok** | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|
| **Dodávky nafty v tis. t** | 3458 | 3154 | 3310 | 3067 | 3180 |
| **Řetězové indexy** | • | 0,912 | 1,049 | 0,927 | 1,037 |
| **Bazické indexy (1 = 2009)** | 1,096 | 1 | 1,049 | 0,972 | 1,008 |

## Interpretace:

*Řetězové indexy:*

V roce 2009 poklesly dodávky motorové nafty v ČR oproti roku 2008 o 8,8 %.

V roce 2010 vzrostly dodávky motorové nafty v ČR oproti roku 2009 o 4,9 %, tj. vzrostly 1,049 krát.

*Bazické indexy:*

V roce 2008 byly dodávky motorové nafty v ČR oproti roku 2009 vyšší o 9,6 %.

V roce 2011 poklesly dodávky motorové nafty v ČR oproti roku 2009 o 2,8 %.

## Zadání příkladu – pokračování:

Doplňte dále charakteristiku vývoje celkových dodávek nafty ve sledovaném období v absolutním vyjádření, tj. pomocí řetězových rozdílů a bazických rozdílů se základem v roce 2009.

Výpočet řetězových rozdílů se provádí podle: $\Delta_{i/i-1} = u_i - u_{i-1};\ i = 1,2,3,\ldots,n$.

První hodnotu řetězového rozdílu nelze z poskytnutých dat zjistit, proto do Tab. 2 uděláme k roku 2008 opět puntík. Další hodnotu už zjistit lze, a to jako:

$\Delta_{2009/2008} = u_{2009} - u_{2008} = 3154 - 3458 = -304.$

Další hodnotu řetězového rozdílu zjistíme obdobným způsobem:

$\Delta_{2010/2009} = u_{2010} - u_{2009} = 3310 - 3154 = 156.$

Podobně postupujeme i u výpočtu zbývajících dvou hodnot.

Výpočet bazických rozdílů se provádí podle: $\Delta_{i/B} = u_i - u_B;\ i = 1,2,3,\ldots,n$.

Nejprve zapíšeme hodnotu 0 do pole u roku bazického, tj. 2009, v Tab. 2. S tímto rokem pak všechny ostatní srovnáme. Nejprve tedy:

$\Delta_{2008/2009} = u_{2008} - u_{2009} = 3458 - 3154 = 304$

Potom spočítáme: $\Delta_{2010/2009} = u_{2010} - u_{2009} = 3310 - 3154 = 156.$

Další hodnota bazického rozdílu bude: $\Delta_{2011/2009} = u_{2011} - u_{2009} = 3067 - 3154 = -87.$

Obdobným způsobem zjistíme i poslední hodnotu bazického rozdílu.

***Tabulka 2 – Hodnoty řetězových a bazických rozdílů***

| **Rok** | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|
| **Dodávky nafty v tis. t** | 3458 | 3154 | 3310 | 3067 | 3180 |
| **Řetězové rozdíly** | • | -304 | 156 | -243 | 113 |
| **Bazické rozdíly (0 = 2009)** | 304 | 0 | 156 | -87 | 26 |

## Interpretace:

*Řetězové rozdíly:*

V roce 2009 poklesly dodávky motorové nafty v ČR oproti roku 2008 o 304 tis. t.

V roce 2010 vzrostly dodávky motorové nafty v ČR oproti roku 2009 o 156 tis. t.

*Bazické rozdíly:*

V roce 2008 byly dodávky motorové nafty v ČR oproti roku 2009 vyšší o 304 tis. t.

V roce 2011 poklesly dodávky motorové nafty v ČR oproti roku 2009 o 87 tis. t.

## Poznámka:

V programu STATGRAPHICS ani v MS Excel není žádná speciální procedura pro výpočet řetězových a bazických indexů a diferencí.

**STA2 - zk (60 min)** .....

jméno a příjmení

# Varianta S1

## Teoretické otázky:

**1)** Je pravdivé tvrzení, že 70 %-ní kvantil normovaného normálního rozdělení je záporné číslo? Vysvětlete svůj názor.
**2)** Jakou metodu byste použili k posouzení, jestli je prodej určitého zboží v průběhu dne rovnoměrně rozložen?
**3)** Co je multikolinearita? Kdy začíná být škodlivá?
**4)** Charakterizujte individuální jednoduché indexy. Uveďte příklad jejich použití.
**5)** Podle čeho se rozhodnete, jakou funkci zvolit k vyrovnání časové řady a jak ověříte, zda vaše volba byla správná?

## Příklady:

**1)** Náhodná veličina má rozdělení s hustotou pravděpodobnosti

$f(x) = cx^2$ pro $0 < x < 2$,
$= 0$ pro ostatní x.

Stanovte:
a) konstantu $c$,
b) rozptyl X,
c) distribuční funkci F (x). 3 BODY

**2)** Máme odhadnout podíl pracovníků, kteří mají zájem o zvyšování kvalifikace. Předpokládáme, že tento podíl nepřesáhne 20 %. Jak velký máme zvolit rozsah výběru, žádáme-li, aby při 95 %-ní spolehlivosti nepřekročila přípustná chyba 2 %? 1 BOD

**3)** O produkci podniku, který vyrábí dva druhy výrobků, máte v následující tabulce tyto údaje za rok 2000 a 2001:

| Výrobek | Hodnota výroby v tis. Kč | | Individuální jednoduchý index množství roku 2001/2000 |
|---|---|---|---|
| | 2000 | 2001 | |
| pastelky | 60 | 88 | 1,6 |
| fixy | 72 | 57,6 | 0,9 |

Určete, jak se celkově změnila úroveň prodejních cen ve firmě, jak se změnilo množství vyrobeného a prodaného zboží ve sledovaném období a jak se změnila hodnota výroby. Dále zhodnoťte, jak se na změně hodnoty výroby podílel faktor cen a jak faktor prodaného množství výrobků. 3 BODY

Celkový počet bodů: 12
Minimální počet bodů pro úspěšné složení zkoušky: 7,25
Minimální počet bodů z teorie: 1,75
Minimální počet bodů z příkladů: 2,5
Každá teoretická otázka je za 1 bod.

# Teoretické otázky - řešení:

**1)** Není to pravda. Normované normální rozdělení je symetrické podle nuly, proto medián je roven právě hodnotě 0. Tím pádem bude 70% kvantil tohoto rozdělení kladný, protože je větší než medián. Pozn.: Lze ověřit ve statistických tabulkách, které jsou ke zkoušce povolené, a na základě toho odvodit jednoduše vysvětlení.
**2)** Lze použít chí-kvadrát test dobré shody nebo Kolmogorovův-Smirnovův test.
**3)** Jedná se o lineární závislost mezi vysvětlujícími proměnnými ve vícerozměrném regresním modelu. Škodlivá začíná být tehdy, když párový korelační koeficient mezi dvojicí některých vysvětlujících proměnných je větší než 0,8 (někdy se uvádí i 0,75).
**4)** Slouží ke srovnávání dvou hodnot téhož ukazatele, který není složen z dílčích částí. Příklad: srovnání prodaného množství jednoho produktu u jedné firmy v aktuálním roce s množstvím téhož produktu (téže firmy) v roce předchozím.
**5)** Rozhodnutí – podle bodového diagramu vyberu funkci nebo funkce, které vypadají, že nejlépe přiléhají skutečným hodnotám. Ověřím pomocí celkového F-testu o vhodnosti vybrané trendové funkce a pomocí individuálních t-testů, kterými testuji přínos parametrů zvolené trendové funkce.

# Příklady – řešení:

**1)** a) Aby funkce byla hustotou pravděpodobnosti, musí platit obecně:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Konkrétně zde:

$$\int_{0}^{2} cx^2 dx = 1$$

Vhodné bude nejprve vyřešit integrál a pak výsledek položit rovno 1.

$$c\int_{0}^{2} x^2 dx = c\left[\frac{x^3}{3}\right]_0^2 = c \cdot \frac{8}{3}$$

Nyní výsledek položím rovno 1:

$$\frac{8}{3}c = 1$$
$$8c = 3$$
$$c = \frac{3}{8}$$

Pro konstantu 3/8 je tato funkce hustotou pravděpodobnosti, tj.:

$f(x) = \frac{3}{8}x^2$ pro 0 < x < 2

= 0 pro ost. x

b) Rozptyl spojité náhodné veličiny lze vypočítat obecně jako:

$$D(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - \left[\int_{-\infty}^{\infty} xf(x)dx\right]^2$$

$$D(X) = \int_0^2 x^2 \frac{3}{8}x^2 dx - \left[\int_0^2 x \frac{3}{8}x^2 dx\right]^2 = \frac{3}{8}\int_0^2 x^4\, dx - \left[\frac{3}{8}\int_0^2 x^3\, dx\right]^2 = \frac{3}{8}\left[\frac{x^5}{5}\right]_0^2 - \left[\frac{3}{8}\left[\frac{x^4}{4}\right]_0^2\right]^2 =$$
$$= \frac{3}{8}\cdot\frac{32}{5} - \left[\frac{3}{8}\cdot\frac{16}{4}\right]^2 = \frac{12}{5} - \frac{9}{4} = \frac{3}{20} = 0{,}15$$

Rozptyl této náhodné veličiny je 0,15.

c) Vztah distribuční funkce a hustoty pravděpodobnosti je obecně popsán vztahem:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Zde konkrétně:

$$F(x) = \int_0^x \frac{3}{8}t^2\, dt = \frac{3}{8}\int_0^x t^2\, dt = \frac{3}{8}\left[\frac{t^3}{3}\right]_0^x = \frac{1}{8}x^3$$

Distribuční funkce této náhodné velečiny je:

$$F(x) = \begin{cases} = 0 & \text{pro } x \leq 0 \\ \dfrac{x^3}{8} & pro\ 0 < x < 2 \\ = 1 & \text{pro } x \geq 2 \end{cases}$$

**2)** Výstupem, který hledáme, je minimální rozsah výběru pro dané šetření. Známé skutečnosti:

p = 0,2 (odhadovaný podíl zájemců o zvyšování kvalifikace)
1-α = 0,95 (požadovaná spolehlivost odhadu)
Δ = 0,02 (přípustná chyba odhadu)

Odhadujeme relativní četnost, proto vycházíme ze vzorce přípustné chyby odhadu relativní četnosti, jehož úpravou dostaneme:

$$n \geq \frac{u_{1-\frac{\alpha}{2}}^2 \cdot p(1-p)}{\Delta^2}$$

Ještě je potřeba zjistit v tabulkách či statistickém programu hodnotu kvantilu normovaného normálního rozdělení $u_{1-\frac{\alpha}{2}}$. Jde o kvantil $u_{0,975} = 1{,}96$.

$$n \geq \frac{1{,}96^2 \cdot 0{,}2 \cdot 0{,}8}{0{,}02^2}$$
$$n \geq 1537$$

Výběr by mělo tvořit alespoň 1537 pracovníků.

**3)** Popis známých údajů:

| **Výrobek** | **Hodnota výroby v tis. Kč** | | **Individuální jednoduchý index množství roku 2001/2000 ($i_q = q_1/q_0$)** |
|---|---|---|---|
| | **2000 ($Q_0$)** | **2001 ($Q_1$)** | |
| pastelky | 60 | 88 | 1,6 |
| fixy | 72 | 57,6 | 0,9 |

Máme dva různé produkty, jejichž množství ani cenu tudíž nemůžeme shrnovat, pracujeme tedy s nestejnorodým ukazatelem. Pro charakterizování změn úrovně prodejních cen, hodnoty výroby a množství vyrobených výrobků je vhodné použít souhrnné indexy.

- Změna vyrobeného množství – můžeme použít Laspeyresův nebo Paascheho objemový index, ovšem v upraveném tvaru:

  $${}_P I_q = \frac{\sum Q_1}{\sum \frac{Q_1}{i_q}}$$

  $${}_P I_q = \frac{\sum Q_1}{\sum \frac{Q_1}{i_q}} = \frac{145{,}6}{119} = 1{,}224$$

  Ve sledovaném období došlo k nárůstu vyrobeného množství výrobků o 22,4 %, bereme-li v úvahu ceny běžného období (rok 2001).

  NEBO

  $${}_L I_q = \frac{\sum i_q \cdot Q_0}{\sum Q_0} = \frac{160{,}8}{132} = 1{,}218$$

  Ve sledovaném období došlo k nárůstu vyrobeného množství výrobků o 21,8 %, bereme-li v úvahu ceny základního období (rok 2000).

- Změna hodnoty výroby – použijeme souhrnný hodnotový index:

  $$I_Q = \frac{\sum Q_1}{\sum Q_0} = \frac{145{,}6}{132} = 1{,}103$$

  Hodnota výroby ve sledovaném období vzrostla o 10,3 %.

- Při hodnocení změny cenové úrovně využijeme toho, co víme o rozkladu souhrnného indexu hodnoty:

  $$I_Q = {}_P I_p \cdot {}_L I_q$$

  $$1{,}103 = {}_P I_p \cdot 1{,}218$$
  $${}_P I_p = 0{,}906$$

  Ve sledovaném období došlo k poklesu cen výrobků o 9,4 %, bereme-li v úvahu vyrobené množství výrobků jako v běžném období, zde tedy v roce 2001.

  NEBO

  $$I_Q = {}_L I_p \cdot {}_P I_q$$
  $$1{,}103 = {}_L I_p \cdot 1{,}224$$
  $${}_L I_p = 0{,}901$$

  Ve sledovaném období došlo k poklesu cen výrobků o 9,9 %, bereme-li v úvahu vyrobené množství výrobků jako v základním období, zde tedy v roce 2000.

- Pro zhodnocení, jak se na změně hodnoty výroby podílel faktor cen a jak faktor prodaného množství výrobků využijeme předchozí výpočty, tj. rozklad souhrnného hodnotového indexu metodou postupných změn:

  $$1{,}103 = 0{,}906 \cdot 1{,}218$$

  Změna cen způsobila pokles hodnoty výroby o 9,4 %, změna vyrobeného množství výrobků způsobila nárůst hodnoty výroby o 21,8 %.

  Pozn.: Lze komentovat i druhou podobu rozkladu.

**STA2 – zk (60 min)** .....

jméno a příjmení

# Varianta S2

## Teoretické otázky:

**1)** Když zamítnete hypotézu $\mu = \mu_0$ při jednostranném testu, zamítnete ji při stejné hladině významnosti také při oboustranném testu? Proč?
**2)** Jaký je vztah mezi binomickým, hypergeometrickým a Poissonovým rozdělením?
**3)** K čemu slouží variační koeficient? Jaká je jeho přednost?
**4)** Co měří dílčí korelační koeficient a co vícenásobný korelační koeficient?
**5)** Co si představujete pod pojmem „dekompozice časové řady“?

## Příklady:

**1)** Určitá linka městské autobusové dopravy jezdí v době dopravní špičky v centru města průměrnou rychlostí 8 km/h. Vedení dopravního podniku uvažovalo o tom, zda by změna trasy dané linky vedla ke zvýšení průměrné rychlosti. Nová trasa proto byla projeta v deseti náhodně vybraných dnech v době dopravní špičky a byly zjištěny tyto průměrné rychlosti (v km/h):

8,5 9,5 7,8 8,2 9,0 7,5 8,2 7,8 9,0 8,5

Rozhodněte na 5 %-ní hladině významnosti, zda změna trasy vede ke zvýšení průměrné rychlosti, přičemž předpokládejte normální rozdělení rychlosti autobusu v centru v době dopravní špičky. 2 BODY

**2)** Při silniční kontrole byly u náhodně vybraných 200 vozidel zjišťovány závady na osvětlení a pneumatikách. Posuďte na hladině významnosti 1 %, zda existuje závislost mezi závadami na pneumatikách a osvětlení, případně změřte sílu a směr závislosti vhodnou charakteristikou.

2 BODY

| **Závady na pneumatikách** | **Závady na osvětlení** | |
|---|---|---|
| | ANO | NE |
| ANO | 32 | 12 |
| NE | 16 | 140 |

**3)** V následující tabulce máte údaje o vývozu osobních automobilů do určité země v letech 1990 – 1997. Posuďte, jakou trendovou funkci je vhodné použít k popisu trendu tohoto ukazatele (omezte svůj výběr na přímku, parabolu a exponenciálu). Své rozhodnutí zdůvodněte. Odhadněte vývoz automobilů (v tis. Kč) na rok 2000 na základě funkce, kterou jste vybrali jako nejvýstižnější. Uvažujte $\alpha = 0{,}05$. 3 BODY

| **Rok** | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 |
|---|---|---|---|---|---|---|---|---|
| **Vývoz v tis. Kč** | 60 | 58 | 55 | 53 | 51 | 48 | 46 | 44 |

Celkový počet bodů: 12
Minimální počet bodů pro úspěšné složení zkoušky: 7,25
Minimální počet bodů z teorie: 1,75
Minimální počet bodů z příkladů: 2,5
Každá teoretická otázka je za 1 bod.

# Teoretické otázky – řešení:

**1)** Tato hypotéza by zamítnuta být nemusela, a to z důvodu „nepřekrývání" se kritického oboru pro jednostranný a oboustranný test. Např. když $H_1: \mu \neq \mu_0$, potom může být kritický obor při zvolené hladině významnosti 0,05 následující: $W \equiv \left\{U; U \leq u_{\frac{\alpha}{2}} \cup U \geq u_{1-\frac{\alpha}{2}}\right\}$ a konkrétně
$W \equiv \{U; U \leq u_{0,025} \cup U \geq u_{0,975}\}$, tj. $W \equiv \{U; U \leq -1,96 \cup U \geq 1,96\}$.
U jednostranného testu, kdy je např. $H_1: \mu > \mu_0$, je kritický obor $W \equiv \{U; U \geq u_{1-\alpha}\}$, konkrétně tedy:
$W \equiv \{U; U \geq u_{0,95}\}$, tj. $W \equiv \{U; U \geq 1,645\}$.
Pokud hodnota testového kritéria bude např. U = 1,75, pak u oboustranného testu nevedla k zamítnutí $H_0$, ale u jednostranného ano.

**2)** Všechno to jsou rozdělení nespojitých náhodných veličin. Hypergeometrické rozdělení můžeme za určitých podmínek aproximovat buď Binomickým, nebo Poissonovým rozdělením. Binomické rozdělení lze za určitých podmínek aproximovat Poissonovým rozdělením.

**3)** Slouží k měření relativní variability číselných proměnných. Jeho výhodou je, že jde o bezrozměrné číslo, můžeme tak porovnávat variabilitu souborů, kde hodnoty jsou v různých měrných jednotkách.

**4)** Dílčí korelační koeficient měří sílu lineární závislosti závisle proměnné y na proměnné např. $x_1$ za předpokladu, že všechny ostatní vysvětlující proměnné (uvedené za tečkou) jsou konstantní. Vícenásobný korelační koeficient měří těsnost lineární závislosti závisle proměnné y na všech vysvětlujících proměnných dohromady.

**5)** Dekompozicí časové řady rozumíme rozklad časové řady na složky, které charakterizují různé druhy pohybů v čase. Rozlišujeme složku trendovou, sezónní, cyklickou a náhodnou.

# Příklady – řešení:

**1)** $H_0: \mu = 8$
$H_1: \mu > 8$

$$t = \frac{\bar{x} - \mu}{\frac{s'_x}{\sqrt{n}}}$$

$W \equiv \{t; t \geq t_{1-\alpha}(n-1)\}$
$W \equiv \{t; t \geq t_{0,95}(9)\}$
$W \equiv \{t; t \geq 1,833\}$

Z dat vypočítáme výběrové charakteristiky:
$\bar{x} = 8,4$
$s'_x = 0,629$

$$t = \frac{\bar{x} - \mu}{\frac{s'_x}{\sqrt{n}}} = \frac{8,4 - 8}{\frac{0,629}{\sqrt{10}}} = 2,011$$

$t \in W$, proto zamítáme $H_0$ a přijímáme $H_1$.

Na hladině významnosti 5 % jsme prokázali, že změna trasy autobusu MHD vede ke zvýšení průměrné rychlosti.

Poznámka: Lze řešit i ve statistickém programu. Rozhodnutí provádíme podle tzv. significance level, někde uváděné také jako P-Value. Zde je P- Value = 0,037591. Protože P-Value < α, zamítáme $H_0$ a přijímáme $H_1$.

**2)** Jde o kategoriální (nominální) znaky, proto je vhodnou metodou zkoumání závislostí mezi nimi test nezávislosti kategoriálních znaků. Zde obě proměnné nabývají pouze 2 obměn, jsou to tedy alternativní proměnné. Lze použít zjednodušenou verzi testu nezávislosti kategoriálních znaků.

$H_0$: Závady na pneumatikách a na osvětlení na sobě nezávisí
$H_1$: non $H_0$

$$G = n \cdot \frac{(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}}$$

$$W \equiv \{G; G \geq \chi^2_{1-\alpha}[(r-1)(s-1)]\}$$
$$W \equiv \{G; G \geq \chi^2_{0,99}(1)\}$$
$$W \equiv \{G; G \geq 6{,}639\}$$

$$G = 200 \cdot \frac{(32 \cdot 140 - 16 \cdot 12)^2}{48 \cdot 152 \cdot 44 \cdot 156} = 73{,}43$$

$G \in W$, tj. zamítáme $H_0$, přijímáme $H_1$.

Na hladině významnosti 1 % jsme prokázali, že mezi závadami na pneumatikách a na osvětlení existuje závislost.

Pozn.: Pokud řešíme ve statistickém programu, posuzujeme výsledek testu podle tzv. P-Value. Zde je P-Value = 0,0000, což je méně než α, proto zamítáme $H_0$ a přijímáme $H_1$.

Těsnost závislosti změříme pomocí koeficientu asociace:

$$r_{AB} = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}}}$$

$$r_{AB} = \frac{32 \cdot 140 - 16 \cdot 12}{\sqrt{48 \cdot 152 \cdot 44 \cdot 156}} = 0{,}606$$

Závislost mezi závadami je středně vysoká a přímá, tj. závady na pneumatikách se vyskytují častěji, když se zároveň vyskytnou závady na osvětlení.

**3)** Doporučeno řešit ve statistickém programu.

**Přímka**

## Simple Regression - vyvoz vs. t

Dependent variable: vyvoz
Independent variable: t
Linear model: Y = a + b*X

**Coefficients**

| | *Least Squares* | *Standard* | *T* | |
|---|---|---|---|---|
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| Intercept | 62,3214 | 0,232829 | 267,671 | 0,0000 |
| Slope | -2,32143 | 0,0461069 | -50,3488 | 0,0000 |

**Analysis of Variance**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
|---|---|---|---|---|---|
| Model | 226,339 | 1 | 226,339 | 2535,00 | 0,0000 |
| Residual | 0,535714 | 6 | 0,0892857 | | |
| Total (Corr.) | 226,875 | 7 | | | |

Correlation Coefficient = -0,998819
R-squared = 99,7639 percent
R-squared (adjusted for d.f.) = 99,7245 percent

*Rovnice*: $\hat{T}_t = 62{,}3214 - 2{,}3214t$

*Celkový F-test* (posouzení vhodnosti přímky k vyrovnání daných hodnot):

$H_0: \alpha_0 = c, \alpha_1 = 0$ (přímka není vhodná k vyrovnání daných hodnot)
$H_1$: non $H_0$

P-Value = 0,0000 < α, tj. zamítáme $H_0$, přijímáme $H_1$.
Na hladině významnosti 5 % jsme prokázali, že přímka je vhodným modelem k vyrovnání daných hodnot.

*Individuální t-testy o nulové hodnotě parametrů trendové funkce* (posouzení vhodnosti jednotlivých parametrů):

$H_0: \alpha_0 = 0$ (parametr $\alpha_0$ není pro funkci přínosný)
$H_1$: non $H_0$

P-Value = 0,0000 < α, tj. zamítáme $H_0$, přijímáme $H_1$.

Na hladině významnosti 5 % jsme prokázali, že parametr $\alpha_0$ je statisticky významný.

$H_0: \alpha_1 = 0$ (parametr $\alpha_1$ není pro funkci přínosný)
$H_1$: non $H_0$

P-Value = 0,0000 < α, tj. zamítáme $H_0$, přijímáme $H_1$.

Na hladině významnosti 5 % jsme prokázali, že parametr $\alpha_1$ je statisticky významný.

Závěr: Vzhledem k tomu, že celkový F-test je významný a oba t-testy také, je možné přímku použít k vyrovnání uvedených hodnot.

**Parabola**

## Polynomial Regression - vyvoz versus t

Dependent variable: vyvoz
Independent variable: t
Order of polynomial = 2

| | | *Standard* | *T* | |
|---|---|---|---|---|
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| CONSTANT | 62,5893 | 0,433234 | 144,47 | 0,0000 |
| t | -2,48214 | 0,220881 | -11,2375 | 0,0001 |
| t^2 | 0,0178571 | 0,0239579 | 0,745356 | 0,4896 |

Analysis of Variance

| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
|---|---|---|---|---|---|
| Model | 226,393 | 2 | 113,196 | 1173,89 | 0,0000 |
| Residual | 0,482143 | 5 | 0,0964286 | | |
| Total (Corr.) | 226,875 | 7 | | | |

R-squared = 99,7875 percent
R-squared (adjusted for d.f.) = 99,7025 percent

*Rovnice*: $\hat{T}_t = 62{,}5893 - 2{,}48214t + 0{,}0178571t^2$

*Celkový F-test* (posouzení vhodnosti paraboly k vyrovnání daných hodnot):

$H_0: \alpha_0 = c, \alpha_1 = \alpha_2 = 0$ (parabola není vhodná k vyrovnání daných hodnot)
$H_1$: non $H_0$

P-Value = 0,0000 < α, tj. zamítáme $H_0$, přijímáme $H_1$.
Na hladině významnosti 5 % jsme prokázali, že parabola je vhodným modelem k vyrovnání daných hodnot.

*Individuální t-testy o nulové hodnotě parametrů trendové funkce* (posouzení vhodnosti jednotlivých parametrů):

$H_0: \alpha_0 = 0$ (parametr $\alpha_0$ není pro funkci přínosný)
$H_1$: non $H_0$

P-Value = 0,0000 < α, tj. zamítáme $H_0$, přijímáme $H_1$.

Na hladině významnosti 5 % jsme prokázali, že parametr $\alpha_0$ je statisticky významný.

$H_0: \alpha_1 = 0$ (parametr $\alpha_1$ není pro funkci přínosný)
$H_1$: non $H_0$

P-Value = 0,0001 < α, tj. zamítáme $H_0$, přijímáme $H_1$.

Na hladině významnosti 5 % jsme prokázali, že parametr $\alpha_1$ je statisticky významný.

$H_0: \alpha_2 = 0$ (parametr $\alpha_2$ není pro funkci přínosný)
$H_1$: non $H_0$

P-Value = 0,4896 > α, tj. nezamítáme $H_0$, nepřijímáme $H_1$.

Na hladině významnosti 5 % jsme neprokázali, že parametr $\alpha_2$ je statisticky významný. To naznačuje, že počet parametrů je příliš velký a stačily by pravděpodobně dva.

Závěr: Vzhledem k tomu, že celkový F-test je významný a většina t-testů také, je možné parabolu použít k vyrovnání uvedených hodnot.

**Exponenciála**

## Simple Regression - vyvoz vs. t

Dependent variable: vyvoz
Independent variable: t
Exponential model: Y = exp(a + b*X)

**Coefficients**

| | *Least Squares* | *Standard* | *T* | |
|---|---|---|---|---|
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| Intercept | 4,14587 | 0,00497813 | 832,817 | 0,0000 |
| Slope | -0,0449638 | 0,000985816 | -45,6107 | 0,0000 |

NOTE: intercept = ln(a)

**Analysis of Variance**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
|---|---|---|---|---|---|
| Model | 0,084913 | 1 | 0,084913 | 2080,34 | 0,0000 |
| Residual | 0,000244902 | 6 | 0,000040817 | | |
| Total (Corr.) | 0,0851579 | 7 | | | |

Correlation Coefficient = -0,998561
R-squared = 99,7124 percent
R-squared (adjusted for d.f.) = 99,6645 percent

*Rovnice*: $\hat{T}_t = e^{4{,}14587 - 0{,}0449638t}$

*Celkový F-test* (posouzení vhodnosti exponenciály k vyrovnání daných hodnot):

$H_0: \alpha_0 = c, \alpha_1 = 0$ (exponenciála není vhodná k vyrovnání daných hodnot)
$H_1$: non $H_0$

P-Value = 0,0000 < α, tj. zamítáme $H_0$, přijímáme $H_1$.
Na hladině významnosti 5 % jsme prokázali, že exponenciála je vhodným modelem k vyrovnání daných hodnot.

*Individuální t-testy o nulové hodnotě parametrů trendové funkce* (posouzení vhodnosti jednotlivých parametrů):

$H_0: \alpha_0 = 0$ (parametr $\alpha_0$ není pro funkci přínosný)
$H_1$: non $H_0$

P-Value = 0,0000 < α, tj. zamítáme $H_0$, přijímáme $H_1$.

Na hladině významnosti 5 % jsme prokázali, že parametr $\alpha_0$ je statisticky významný.

$H_0: \alpha_1 = 0$ (parametr $\alpha_1$ není pro funkci přínosný)
$H_1$: non $H_0$

P-Value = 0,0000 < α, tj. zamítáme $H_0$, přijímáme $H_1$.

Na hladině významnosti 5 % jsme prokázali, že parametr $\alpha_1$ je statisticky významný.

Závěr: Vzhledem k tomu, že celkový F-test je významný a oba t-testy také, je možné exponenciálu použít k vyrovnání uvedených hodnot.

**Z výše uvedeného vyplývá, že všechny uvedené funkce jsou použitelné k vyrovnání daných hodnot. Otázkou je, která z těch funkcí je nejvýstižnější. K řešení použijeme vybraná matematicko-statistická kritéria:**

| **Kritérium** | **Přímka** | **Parabola** | **Exponenciála** |
|---|---|---|---|
| F | 2535 | 1173,89 | 2080,335 |
| $s_R^2$ | 0,08929 | 0,09643 | 0,00004 |
| $I_{adj}^2$ | 0,997 | 0,997 | 0,997 |

Podle hodnoty testového kritéria F se jako nejlepší jeví trendová přímka.
Podle hodnoty reziduálního rozptylu bychom jako nejvýstižnější funkci vybrali trendovou exponenciálu. Podle hodnoty upraveného indexu determinace bychom se mohli přiklonit k jakékoli z těchto tří funkcí. Záleží na tom, které kritérium si vybereme. Tu funkci pak zvolíme. Zde bychom pravděpodobně vybrali přímku, protože má nejvyšší hodnotu F a je zároveň ze všech tří uvedených funkcí nejjednodušší.

Předpověď vývozu automobilů na rok 2000 na základě trendové přímky:
$\hat{T}_t = 62{,}3214 - 2{,}3214t$
$\hat{T}_{11} = 62{,}3214 - 2{,}3214 \cdot 11$
$\hat{T}_{11} = 36{,}786$

Za předpokladu, že se trend vývoje nezmění, lze v roce 2000 očekávat vývoz automobilů v hodnotě 36,786 tis. Kč.

**STA2 – zk (60 min)** .....

jméno a příjmení

# Varianta S3

## Teoretické otázky:

**1)** K čemu slouží index determinace? Jakých hodnot může nabýt?
**2)** U kterých z uvedených funkcí nelze k odhadu parametrů použít metodu nejmenších čtverců: hyperbola, mocninná funkce, parabola, exponenciála? Názor vysvětlete!
**3)** Pro která pravděpodobnostní rozdělení je jejich střední hodnota rovna mediánu a zároveň modu? Vysvětlete a uveďte alespoň 2 příklady!
**4)** Jak budete postupovat, jestliže chcete zjistit, jak se na změně průměrné prodejní ceny určité komodity podílely změny cen u jednotlivých prodejců a jak změna struktury prodeje podle jednotlivých prodejců?
**5)** Vysvětlete pojem kvantil. Jaké skupiny kvantilů znáte? Blíže je charakterizujte!

## Příklady:

**1)** V následující tabulce jsou údaje o vývozu ekologických solárních zařízení v mld. Kč stálých cen:

| **Rok** | **1998** | **1999** | **2000** | **2001** | **2002** | **2003** |
|---|---|---|---|---|---|---|
| **Vývoz v mld. Kč** | 495 | 537,7 | 620,3 | 651,8 | 716,2 | 713,5 |

Charakterizujte vývoj vývozu solárních zařízení pomocí řetězových a bazických indexů (1998 = 1) a vypočítejte průměrný koeficient růstu a průměrný absolutní přírůstek vývozu za období 1998 – 2003. Výsledky interpretujte (u bazických a řetězových indexů stačí interpretace vždy 1 z nich)! 3 BODY

**2)** Vypočítejte celkový rozptyl, máme – li k dispozici následující údaje:

| **Skupina** | $\bar{x}_i$ | $n_i$ | $s_i$ |
|---|---|---|---|
| 1 | 5 | 10 | 2,0 |
| 2 | 4 | 15 | 0,8 |
| 3 | 2 | 25 | 0,5 |
| **Celkem** | x | 50 | x |

2 BODY

**3)** Mezi zákazníky Domu knih bylo náhodně vybráno 400 mužů a žen za účelem ověření hypotézy o nezávislosti preferovaného literárního žánru a pohlaví. Byly získány tyto údaje:

| **Pohlaví zákazníka** | **Literární žánr** | | | | | **Součet** |
|---|---|---|---|---|---|---|
| | sci – fi | milostné romány | vzdělávací literatura | válečné romány | historické romány | |
| Muž | 29 | 9 | 25 | 40 | 37 | 140 |
| Žena | 14 | 60 | 102 | 28 | 56 | 260 |
| **Součet** | 43 | 69 | 127 | 68 | 93 | 400 |

Na hladině významnosti $\alpha = 0{,}01$ ověřte hypotézu o nezávislosti literárního žánru a pohlaví zákazníka. Případně změřte vhodnou charakteristikou sílu závislosti. 2 BODY

Celkový počet bodů: 12
Minimální počet bodů pro úspěšné složení zkoušky: 7,25
Minimální počet bodů z teorie: 1,75
Minimální počet bodů z příkladů: 2,5
Každá teoretická otázka je za 1 bod.

# Teoretické otázky – řešení:

**1)** Slouží k měření těsnosti závislosti číselných proměnných, jejichž vztah je popsán pomocí nějaké matematické funkce. Zároveň je jedním z kritérií pro posouzení kvality vybrané regresní funkce. Nabývá hodnot z intervalu <0, 1>.

**2)** Není to možné u mocninné funkce a exponenciály, protože to jsou funkce nelineární v parametrech. Aplikace metody nejmenších čtverců u takových funkcí může vést k soustavě nelineárních rovnic, které nemusí mít jednoznačné řešení, nebo u nich nedokážeme parametry vyjádřit pomocí vhodných výpočtových vzorců.

**3)** Pro symetrická rozdělení. Př. normální, normované normální a Studentovo rozdělení.

**4)** Index proměnlivého složení, jímž charakterizujeme změnu průměrné ceny dané komodity, rozložíme např. pomocí metody postupných změn na dva dílčí indexy – index stálého složení a index struktury. Index stálého složení vyjadřuje, jak se na změně průměrné ceny podílely změny cen u jednotlivých prodejců, a index struktury vyjadřuje, jak se na změně průměrné ceny podílela změna struktury prodeje.

**5)** Kvantil je hodnota, která rozděluje uspořádaný statistický soubor na určitý počet stejně obsazených částí. Př. kvartily – tři kvantily ($\tilde{x}_{25}, \tilde{x}, \tilde{x}_{75}$), které rozdělují neklesající řadu hodnot na 4 stejně četné části. Decily – 9 kvantilů ($\tilde{x}_{10}, \tilde{x}_{20}, \tilde{x}_{30}, \tilde{x}_{40}, \tilde{x}, \tilde{x}_{60}, \tilde{x}_{70}, \tilde{x}_{80}, \tilde{x}_{90}$), které rozdělují uspořádaný statistický soubor na 10 stejně četných částí. Atd.

# Příklady – řešení:

**1)**

| **Rok** | **1998** | **1999** | **2000** | **2001** | **2002** | **2003** |
|---|---|---|---|---|---|---|
| **Vývoz v mld. Kč** | 495 | 537,7 | 620,3 | 651,8 | 716,2 | 713,5 |
| **Řetěz. index** | • | 1,0863 | 1,1536 | 1,0508 | 1,0988 | 0,9962 |
| **Baz. index (1 = 1998)** | 1 | 1,0863 | 1,2531 | 1,3168 | 1,4469 | 1,4414 |

Řetězové indexy jsou obecně definovány jako podíl hodnoty ukazatele v běžném a základním období, tj. pro rok 1998 nelze hodnotu řetězového indexu zjistit z poskytnutých dat, protože rok 1997, se kterým bychom měli srovnávat, není k dispozici. Pro rok 1999 počítáme podle:

$i_{1999/1998} = \frac{u_{1999}}{u_{1998}} = \frac{537,7}{495} = 1,0863$

Dále pak: $i_{2000/1999} = \frac{u_{2000}}{u_{1999}} = \frac{620,3}{537,7} = 1,1536$ atd.

*Interpretace:* V roce 2000 vzrostl vývoz solárních zařízení oproti roku 1999 o 15,36 %.

Bazické indexy jsou definované jako podíl hodnoty ukazatele v běžném a bazickém období. Zde je bazickým obdobím rok 1998, proto je u něj hodnota 1. Další index vypočítáme jako:

$$i_{1999/1998} = \frac{u_{1999}}{u_{1998}} = \frac{537,7}{495} = 1,0863$$

Dále pak:

$$i_{2000/1998} = \frac{u_{2000}}{u_{1998}} = \frac{620,3}{495} = 1,2531$$

$i_{2001/1998} = \frac{u_{2001}}{u_{1998}} = \frac{651,8}{495} = 1,3168$ atd.

*Interpretace:* V roce 2001 vzrostl vývoz solárních zařízení oproti roku 1998 o 31,68 %.

Průměrný absolutní přírůstek je možné vypočítat podle:

$$\bar{\Delta}=\frac{y_n-y_1}{n-1}$$

$$\bar{\Delta}=\frac{y_n-y_1}{n-1}=\frac{713{,}5-495}{6-1}=43{,}7$$

*Interpretace:* Vývoz solárních zařízení v letech 1998-2003 v průměru vzrostl o 43,7 mld. Kč.

Průměrný koeficient růstu je možné vypočítat podle:

$$\bar{k}=\sqrt[n-1]{\frac{y_n}{y_1}}$$

$$\bar{k}=\sqrt[n-1]{\frac{y_n}{y_1}}=\sqrt[5]{\frac{713{,}5}{495}}=1{,}076$$

*Interpretace:* Vývoz solárních zařízení v letech 1998-2003 v průměru vzrostl o 7,6 % (NEBO vzrostl 1,076krát).

**2)** Vzhledem k tomu, že máme soubor rozdělený do dílčích částí, ve kterých jsou nám známy dílčí charakteristiky, použijeme k řešení větu o rozkladu rozptylu.

| **Skupina** | $\bar{x}_i$ | $n_i$ | $s_i$ | $\bar{x}_i n_i$ | $\bar{x}_i^2 n_i$ | $s_i^2$ | $s_i^2 n_i$ |
|---|---|---|---|---|---|---|---|
| 1 | 5 | 10 | 2,0 | 50 | 250 | 4,00 | 40,00 |
| 2 | 4 | 15 | 0,8 | 60 | 240 | 0,64 | 9,60 |
| 3 | 2 | 25 | 0,5 | 50 | 100 | 0,25 | 6,25 |
| **Celkem** | x | 50 | x | 160 | 590 | x | 55,85 |

Pomocné výpočty jsou uvedeny v tabulce modrou barvou.

$$s_x^2=\overline{s^2}+s_{\bar{x}_i}^2$$

$$s_{\bar{x}_i}^2=\frac{\sum_{i=1}^k \bar{x}_i^2 n_i}{\sum_{i=1}^k n_i}-\left(\frac{\sum_{i=1}^k \bar{x}_i n_i}{\sum_{i=1}^k n_i}\right)^2$$

$$\overline{s_i^2}=\frac{\sum_{i=1}^k s_i^2 n_i}{\sum_{i=1}^k n_i}$$

$$s_{\bar{x}_i}^2=\frac{\sum_{i=1}^k \bar{x}_i^2 n_i}{\sum_{i=1}^k n_i}-\left(\frac{\sum_{i=1}^k \bar{x}_i n_i}{\sum_{i=1}^k n_i}\right)^2=\frac{590}{50}-\left(\frac{160}{50}\right)^2=1{,}56$$

$$\overline{s_i^2}=\frac{\sum_{i=1}^k s_i^2 n_i}{\sum_{i=1}^k n_i}=\frac{55{,}85}{50}=1{,}117$$

$$s_x^2=1{,}56+1{,}117=2{,}677$$

Celkový rozptyl má hodnotu 2,677.

**3)** Vhodnou metodou pro zkoumání závislosti preferovaného literárního žánru na pohlaví čtenáře je test nezávislosti kategoriálních znaků, neboť obě proměnné jsou nominální. Je doporučeno řešit prostřednictvím statistického programu. Prvním úkolem je ověřit předpoklady pro použití dané metody, tj. v každém políčku kontingenční tabulky musí být teoretické četnosti $n'_{ij} \geq 5$. V programu STATGRAPHICS Centurion 18 získáme tabulku teoretických četností:

**Frequency Table**

| | sci-fi | cervena_knih | vzdel_lit | valec_romany | hist_romany | Row Total |
|---|---|---|---|---|---|---|
| muz | 29 | 9 | 25 | 40 | 37 | 140 |
| | 15,05 | 24,15 | 44,45 | 23,80 | 32,55 | 35,00% |
| zena | 14 | 60 | 102 | 28 | 56 | 260 |
| | 27,95 | 44,85 | 82,55 | 44,20 | 60,45 | 65,00% |
| Column Total | 43 | 69 | 127 | 68 | 93 | 400 |
| | 10,75% | 17,25% | 31,75% | 17,00% | 23,25% | 100,00% |

Cell contents:
Observed frequency
Expected frequency

Teoretické četnosti jsou zvýrazněny modrou barvou. Vidíme, že všechny hodnoty převyšují hodnotu 5, proto je předpoklad pro použití testu nezávislosti kategoriálních dat splněn.

$H_0$: preferovaný literární žánr nezávisí na pohlaví zákazníka
$H_1$: non $H_0$

$$G = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(n_{ij} - n'_{ij}\right)^2}{n'_{ij}}$$

$W \equiv \{G; G > \chi^2_{1-\alpha}[(r-1)(s-1)]\}$
$W \equiv \{G; G > \chi^2_{99}(4)\}$
$W \equiv \{G; G > 13{,}277\}$

G = 65,508

G $\in W \Rightarrow$ zamítáme $H_0$, přijímáme $H_1$.

Na hladině významnosti 1 % jsme prokázali, že preferovaný literární žánr závisí na pohlaví zákazníka.

Pozn.: Pokud budeme celý příklad řešit v programu, závěr testu provedeme na základě P-Value:
**Tests of Independence**

| *Test* | *Statistic* | *Df* | *P-Value* |
|---|---|---|---|
| Chi-Square | 65,508 | 4 | 0,0000 |

P-Value = 0,0 < α, proto zamítáme $H_0$, přijímáme $H_1$.

Vzhledem k tomu, že závislost mezi proměnnými byla prokázána, má smysl změřit její sílu.
K tomu využijeme např. Cramérův koeficient kontingence:

$$C_C = \sqrt{\frac{G}{n \cdot h}} = \sqrt{\frac{65{,}508}{400 \cdot 1}} = 0{,}405$$

Závislost mezi preferovaným literárním žánrem a pohlavím zákazníka je spíše menší.

**STA2 – zk (60 min)** .....

jméno a příjmení

# Varianta S4

## Teoretické otázky:

**1)** K čemu slouží histogram rozdělení četností?
**2)** Stručně popište princip rozkladu souhrnného indexu hodnoty metodou postupných změn. Jaký má tento rozklad význam?
**3)** Časová řada měsíčních údajů je rostoucí a sezónní výkyvy rostou úměrně trendu. Jaký byste použili model časové řady a jak byste posoudili sezónnost takové časové řady?
**4)** Jakým způsobem lze snížit pravděpodobnost chyby I. druhu a jak pravděpodobnost chyby II. druhu (při dané chybě I. druhu)?
**5)** Co to znamená, když mezi 2 alternativními proměnnými existuje úplná záporná asociace? Vysvětlete!

## Příklady:

**1)** Mezi studenty vysoké školy proběhla anketa, která měla za cíl zjistit, z jakého důvodu se rozhodli na této škole studovat. Jejich odpovědi jsou zaznamenány v následující tabulce. Vypočítejte obě charakteristiky variability, které znáte. Interpretujte jejich hodnoty. Potom jejich hodnoty porovnejte – uveďte, proč se liší, a které byste dali při výpočtu přednost a proč. 2 BODY

| **Důvod studia** $x_i$ | **Počet studentů** $n_i$ |
|---|---|
| další sebevzdělávání | 10 |
| odklad vojny | 2 |
| prodloužení mládí | 8 |
| perspektiva při hledání zaměstnání | 9 |
| ostatní | 3 |
| **Součet** | 32 |

**2)** V následující tabulce jsou uvedeny údaje o počtu dětí v domácnosti a výdajích na školní potřeby v tis. Kč za rok. Pomocí analýzy rozptylu posuďte na hladině významnosti 5 %, zda existuje závislost výdajů na školní potřeby na počtu dětí v domácnosti. Případně změřte sílu závislosti vhodnou charakteristikou (Není třeba ověřovat podmínky pro použití analýzy rozptylu). 2 BODY

| **Počet dětí v domácnosti** $x_i$ | **Výdaje na školní potřeby v tis. Kč za rok** $y_{ij}$ | $n_i$ |
|---|---|---|
| 1 | 12 13,5 7 9,3 8,1 10,2 9 9,8 | 8 |
| 2 | 18 19,5 24 20,1 22,7 17,5 14,8 16,2 18,4 21,5 | 10 |
| 3 | 24,3 28,7 27,1 32 30,4 23 26,5 | 7 |
| 4 | 35 42,5 38,6 39,2 46 | 5 |
| **Součet** | | 30 |

**3)** V níže uvedené tabulce jsou údaje o prodeji 15-litrových plastových odpadkových košů v několika pobočkách určité firmy v lednu a srpnu 2003.

a) Určete, jak se změnila tržba, počet prodaných kusů a průměrná cena v srpnu oproti lednu.
b) Určete vliv změny cen v jednotlivých pobočkách a vliv změny struktury prodaných výrobků na změnu průměrné ceny.

3 BODY

| Pobočka | Cena za jednotku v Kč | | Tržby v Kč | |
|---|---|---|---|---|
| | leden | srpen | leden | srpen |
| Brno | 120 | 145 | 26 400 | 24 070 |
| Jihlava | 109 | 120 | 20 056 | 21 600 |
| Plzeň | 112 | 131 | 21 952 | 26 462 |
| Liberec | 105 | 126 | 22 260 | 28 980 |

Celkový počet bodů: 12
Minimální počet bodů pro úspěšné složení zkoušky: 7,25
Minimální počet bodů z teorie: 1,75
Minimální počet bodů z příkladů: 2,5
Každá teoretická otázka je za 1 bod.

# Teoretické otázky – řešení:

**1)** Slouží ke grafickému znázornění intervalového rozdělení četností.
**2)** Rozklad $I_Q$ může mít dvojí podobu:

$I_Q = {}_P I_p \cdot {}_L I_q$

NEBO

$I_Q = {}_L I_p \cdot {}_P I_q$

Význam: Je možné určit, jak se na změně hodnoty podílel faktor cen a jak faktor množství.

**3)** Použili bychom model proporcionální sezónnosti a velikost sezónních výkyvů bychom vyčíslili prostřednictví indexních sezónních faktorů.
**4)** Snížit pravděpodobnost chyby I. druhu můžeme sami – tím, že si koeficient α zvolíme menší, např. z 0,05 snížím na 0,01. Musíme ale přitom vzít v úvahu, že se při této aktivitě zároveň zvýší pravděpodobnost chyby II. druhu. Pravděpodobnost chyby II. druhu (při dané chybě I. druhu) snížíme pomocí zvětšení rozsahu výběru.
**5)** Znamená to, že pokud nastává jeden z jevů, druhý jev nenastává. Je to v případě, kdy koeficient asociace je roven -1. Použití v analýze zkoumání závislostí (alternativních znaků).

# Příklady – řešení:

**1)** Tabulka rozdělení četností obsahuje údaje o nominální proměnné, proto budeme počítat míru mutability a nominální varianci.

| **Důvod studia** $x_i$ | **Počet studentů** $n_i$ | $n_i^2$ | $p_i$ | $p_i^2$ |
|---|---|---|---|---|
| další sebevzdělávání | 10 | 100 | 0,3125 | 0,0977 |
| odklad vojny | 2 | 4 | 0,0625 | 0,0039 |
| prodloužení mládí | 8 | 64 | 0,2500 | 0,0625 |
| perspektiva při hledání zaměstnání | 9 | 81 | 0,2813 | 0,0791 |
| ostatní | 3 | 9 | 0,0938 | 0,0088 |
| **Součet** | 32 | 258 | 1,0000 | 0,252 |

Míru mutability vypočítáme podle:

$$M = \frac{n^2 - \sum_{i=1}^{k} n_i^2}{n(n-1)} = \frac{32^2 - 258}{32 \cdot 31} = \frac{766}{992} = 0{,}772$$

Pozn.: Pomocný výpočet v tabulce výše v modré barvě.

77,2 % dvojic studentů se odlišuje z pohledu důvodu studia na VŠ. Variabilita důvodu studia je poměrně vysoká.

Nominální varianci vypočítáme podle:

$$NOMVAR = 1 - \sum_{i=1}^{k} p_i^2$$

$$NOMVAR = 1 - 0{,}252 = 0{,}748$$

Pozn.: Pomocný výpočet v tabulce výše v hnědé barvě.

Variabilita důvodu studia je poměrně vysoká.

K charakterizování variability důvodu studia je zde vhodnější použít míru mutability, protože NOMVAR je charakteristika, která skutečnou variabilitu hodnot nominální proměnné podhodnocuje.

**2)** Doporučeno řešit ve statistickém programu.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

(NEBO $H_0$: Výdaje na školní potřeby nejsou závislé na počtu dětí v domácnosti)

$H_1$: non $H_0$

Výstup z programu STATGRAPHICS Centurion 18:

**ANOVA Table for vydaje_skol_potr by Pocet_deti**

| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
|---|---|---|---|---|---|
| Between groups | 3125,7 | 3 | 1041,9 | 114,66 | 0,0000 |
| Within groups | 236,266 | 26 | 9,08716 | | |
| Total (Corr.) | 3361,97 | 29 | | | |

F = 114,66

P-Value = 0,0000

P-Value < α, proto zamítáme $H_0$ a přijímáme $H_1$.

Na hladině významnosti 5 % jsme prokázali, že výše výdajů na školní potřeby závisí na počtu dětí v domácnosti.

Sílu závislosti lze změřit pomocí poměru determinace, tj.

$$P^2 = \frac{S_{ym}}{S_y}$$

$$P^2 = \frac{S_{ym}}{S_y} = \frac{3125{,}7}{3361{,}97} = 0{,}930$$

Pozn.: $S_{ym}$ …. Hodnotu najdeme v tabulce výše (modrá barva)

$S_y$ ….. Hodnotu najdeme v tabulce výše (růžová barva)

93 % z celkové variability výdajů na školní potřeby můžeme vysvětlit pomocí počtu dětí v domácnosti, tj. závislost výše výdajů na školní potřeby na počtu dětí v domácnosti je velmi silná.

**3)** Pracujeme se stejnorodým ukazatelem, jehož dílčí části lze shrnovat buď pomocí součtu (množství, tržby) nebo pomocí průměru (cena), proto k charakterizování změn indikátorů použijeme individuální složené indexy.

| Pobočka | Cena za jednotku v Kč | | Tržby v Kč | | Počet prodaných kusů | | $p_0q_1$ | $p_1q_0$ |
|---|---|---|---|---|---|---|---|---|
| | leden | srpen | leden | srpen | leden | srpen | | |
| **Brno** | 120 | 145 | 26 400 | 24 070 | 220 | 166 | 19920 | 31900 |
| **Jihlava** | 109 | 120 | 20 056 | 21 600 | 184 | 180 | 19620 | 22080 |
| **Plzeň** | 112 | 131 | 21 952 | 26 462 | 196 | 202 | 22624 | 25676 |
| **Liberec** | 105 | 126 | 22 260 | 28 980 | 212 | 230 | 24150 | 26712 |
| **Celkem** | x | x | 90668 | 101112 | 812 | 778 | 86314 | 106368 |

Pozn.: Pomocné výpočty jsou v tabulce uvedeny modře.

a)

$$I_Q = \frac{\sum Q_1}{\sum Q_0} = \frac{101112}{90668} = 1{,}115$$

$$\Delta_Q = \sum Q_1 - \sum Q_0 = 101112 - 90668 = 10444$$

Tržba se v srpnu zvýšila oproti lednu o 11,5%, tedy o 10 444 Kč.

$$I_q = \frac{\sum q_1}{\sum q_0} = \frac{778}{812} = 0{,}958$$

$$\Delta_q = \sum q_1 - \sum q_0 = 778 - 812 = -34$$

Počet prodaných kusů košů se snížil v srpnu oproti lednu o 4,2%, tedy o 34 ks.

$$I_{\bar{p}} = \frac{\bar{p}_1}{\bar{p}_0} = \frac{\frac{\sum Q_1}{\sum q_1}}{\frac{\sum Q_0}{\sum q_0}} = \frac{\frac{101112}{778}}{\frac{90668}{812}} = \frac{129{,}964}{111{,}66} = 1{,}164$$

$$\Delta_{\bar{p}} = \bar{p}_1 - \bar{p}_0 = \frac{\sum p_1 q_1}{\sum q_1} - \frac{\sum p_0 q_0}{\sum q_0} = 129{,}964 - 111{,}66 = 18{,}3$$

Průměrná cena košů se zvýšila v srpnu oproti lednu o 16,4%, tedy o 18,30 Kč.

b)

$$I_{\bar{p}} = I_{SS}(q_1) \cdot I_{STR}(p_0)$$

$$I_{SS}(q_1) = \frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_0 q_1}{\sum q_1}} = \frac{\frac{101112}{778}}{\frac{86314}{778}} = \frac{129{,}964}{110{,}943} = 1{,}171$$

$$I_{STR}(p_0) = \frac{\frac{\sum p_0 q_1}{\sum q_1}}{\frac{\sum p_0 q_0}{\sum q_0}} = \frac{\frac{86314}{778}}{\frac{90668}{812}} = \frac{110{,}943}{111{,}66} = 0{,}994$$

$$I_{\bar{p}} = I_{SS}(q_1) \cdot I_{STR}(p_0) = 1{,}171 \cdot 0{,}994 = 1{,}164$$

$$\Delta_{\bar{p}} = (129{,}964 - 110{,}943) + (110{,}943 - 111{,}66) = 19{,}021 + (-0{,}717) = 18{,}3$$

V důsledku změn cen na jednotlivých pobočkách (při zachování struktury prodeje z běžného období) vzrostla průměrná cena o 17,1%, tj. o 19 Kč a změna struktury prodaného zboží (při cenové hladině základního období) vyvolala pokles průměrné ceny o 0,6%, tj. o 0,7 Kč.

NEBO

$$I_{\bar{p}} = I_{SS}(q_0) \cdot I_{STR}(p_1)$$

$$I_{SS}(q_0) = \frac{\frac{\sum p_1 q_0}{\sum q_0}}{\frac{\sum p_0 q_0}{\sum q_0}} = \frac{\frac{106368}{812}}{\frac{90668}{812}} = \frac{130{,}995}{111{,}66} = 1{,}173$$

$$I_{STR}(p_1) = \frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_1 q_0}{\sum q_0}} = \frac{\frac{101112}{778}}{\frac{106368}{812}} = \frac{129{,}964}{130{,}995} = 0{,}992$$

$$I_{\bar{p}} = I_{SS}(q_0) \cdot I_{STR}(p_1) = 1{,}173 \cdot 0{,}992 = 1{,}164$$

$$\Delta_{\bar{p}} = (130{,}995 - 111{,}66) + (129{,}964 - 130{,}995) = 19{,}335 + (-1{,}031) = 18{,}3$$

V důsledku změn cen na jednotlivých pobočkách (při zachování struktury prodeje ze základního období) vzrostla průměrná cena o 17,3%, tj. o 19,3 Kč a změna struktury prodaného zboží (při cenové hladině běžného období) vyvolala pokles průměrné ceny o 0,8%, tj. o 1 Kč.

# Základní charakteristiky časových řad – řešený příklad

Máme k dispozici údaje o počtu dopravních nehod v ČR v letech 2007-2018. Charakterizujte vývoj počtu dopravních nehod pomocí:

a) prvních diferencí,
b) průměrného absolutního přírůstku,
c) koeficientů růstu,
d) průměrného koeficientu růstu.

Všechny výsledky interpretujte!
Doplňte též hodnoty druhých a třetích diferencí.

| Rok | Počet nehod |
|---|---|
| 2007 | 182 736 |
| 2008 | 160 376 |
| 2009 | 74 815 |
| 2010 | 75 522 |
| 2011 | 75 137 |
| 2012 | 81 404 |
| 2013 | 84 398 |
| 2014 | 85 859 |
| 2015 | 93 067 |
| 2016 | 98 864 |
| 2017 | 103 821 |
| 2018 | 104 764 |

*Zdroj: Ročenka nehodovosti na pozemních komunikacích za rok 2018*

## *Řešení:*

Všechny požadované charakteristiky časových řad je poměrně jednoduché spočítat i na kalkulačce, proto je zde uveden postup ručního výpočtu a následně i pomocí programu SGP.

a) První diference jsou konstruovány jako rozdíl hodnoty ukazatele v časovém okamžiku *t* a hodnoty ukazatele v okamžiku bezprostředně předcházejícím časovému okamžiku *t*, tedy *t-1*: $\Delta_t^{(1)} = y_t - y_{t-1}$. Hodnoty všech počítaných charakteristik budou doplněny postupně do následující tabulky.
   První řádek zůstane nevyplněn, protože hodnota *t* je rok 2007 a předcházející údaj není k dispozici. Tečka v řádku značí, že údaj existuje, ale nejsou k dispozici data k jeho výpočtu.
   Další hodnota je vypočtena jako $\Delta_2^{(1)} = y_2 - y_1 = 160376 - 182736 = -22360$.
   Tímto způsobem pokračujeme dále: $\Delta_3^{(1)} = y_3 - y_2 = 74815 - 160376 = -85561$.
   $\Delta_4^{(1)} = y_4 - y_3 = 75522 - 74815 = 707$
   $\Delta_5^{(1)} = y_5 - y_4 = 75137 - 75522 = -385$
   $\Delta_6^{(1)} = y_6 - y_5 = 81404 - 75137 = 6267$

$\Delta_7^{(1)} = y_7 - y_6 = 84398 - 81404 = 2994$
$\Delta_8^{(1)} = y_8 - y_7 = 85859 - 84398 = 1461$
$\Delta_9^{(1)} = y_9 - y_8 = 93067 - 85859 = 7208$
$\Delta_{10}^{(1)} = y_{10} - y_9 = 98864 - 93067 = 5797$
$\Delta_{11}^{(1)} = y_{11} - y_{10} = 103821 - 98864 = 4957$
$\Delta_{12}^{(1)} = y_{12} - y_{11} = 104764 - 103821 = 943$

| t | Rok | Počet nehod $y_t$ | $\Delta_t^{(1)}$ |
|---|---|---|---|
| 1 | 2007 | 182 736 | • |
| 2 | 2008 | 160 376 | -22 360 |
| 3 | 2009 | 74 815 | -85 561 |
| 4 | 2010 | 75 522 | 707 |
| 5 | 2011 | 75 137 | -385 |
| 6 | 2012 | 81 404 | 6 267 |
| 7 | 2013 | 84 398 | 2 994 |
| 8 | 2014 | 85 859 | 1 461 |
| 9 | 2015 | 93 067 | 7 208 |
| 10 | 2016 | 98 864 | 5 797 |
| 11 | 2017 | 103 821 | 4 957 |
| 12 | 2018 | 104 764 | 943 |
| **Celkem** | | | -77972 |

***Interpretace:***
Počet nehod v ČR se v roce 2008 snížil oproti roku 2007 o 22 360.
Počet nehod v ČR se v roce 2009 snížil oproti roku 2008 o 85 561.
Počet nehod v ČR se v roce 2010 zvýšil oproti roku 2009 o 707.
Počet nehod v ČR se v roce 2011 snížil oproti roku 2010 o 385.
Počet nehod v ČR se v roce 2012 zvýšil oproti roku 2011 o 6267.
atp.

b) Průměrný absolutní přírůstek je možné vyjádřit jako průměr prvních diferencí (tj. součet prvních diferencí dělených jejich počtem) nebo jako rozdíl poslední a první hodnoty časové řady dělený *n-1*. Oba způsoby výpočtu musejí dát pochopitelně stejný výsledek.

$$\bar{\Delta} = \frac{\sum_{t=2}^{n} \Delta_t^{(1)}}{n-1} = \frac{-77972}{11} = -7088{,}364$$

NEBO

$$\bar{\bar{\Delta}} = \frac{y_n - y_1}{n-1} = \frac{104764 - 182736}{11} = -7088{,}364$$

***Interpretace:***
Počet nehod v ČR se v letech 2007-2018 v průměru snížil o 7088,364 nehody.

c) Koeficient růstu je konstruován jako podíl hodnoty ukazatele v časovém okamžiku *t* a hodnoty ukazatele v časovém okamžiku *t-1*.

$$k_t = \frac{y_t}{y_{t-1}}$$

První řádek zůstane opět nevyplněn (jako u prvních diferencí), protože hodnota *t* je rok 2007 a předcházející údaj není k dispozici. Tečka v řádku značí, že údaj existuje, ale nejsou k dispozici data k jeho výpočtu.

$$k_2 = \frac{y_2}{y_1} = \frac{160376}{182736} = 0{,}8776$$
$$k_3 = \frac{y_3}{y_2} = \frac{74815}{160376} = 0{,}4665$$
$$k_4 = \frac{y_4}{y_3} = \frac{75522}{74815} = 1{,}0094$$
$$k_5 = \frac{y_5}{y_4} = \frac{75137}{75522} = 0{,}9949$$
$$k_6 = \frac{y_6}{y_5} = \frac{81404}{75137} = 1{,}0834$$
$$k_7 = \frac{y_7}{y_6} = \frac{84398}{81404} = 1{,}0368$$
$$k_8 = \frac{y_8}{y_7} = \frac{85859}{84398} = 1{,}0173$$
$$k_9 = \frac{y_9}{y_8} = \frac{93067}{85859} = 1{,}0840$$
$$k_{10} = \frac{y_{10}}{y_9} = \frac{98864}{93067} = 1{,}0623$$
$$k_{11} = \frac{y_{11}}{y_{10}} = \frac{103821}{98864} = 1{,}0501$$
$$k_{12} = \frac{y_{12}}{y_{11}} = \frac{104764}{103821} = 1{,}0091$$

| **t** | **Rok** | **Počet nehod $y_t$** | **$k_t$** |
|---|---|---|---|
| 1 | 2007 | 182 736 | • |
| 2 | 2008 | 160 376 | 0,8776 |
| 3 | 2009 | 74 815 | 0,4665 |
| 4 | 2010 | 75 522 | 1,0094 |
| 5 | 2011 | 75 137 | 0,9949 |
| 6 | 2012 | 81 404 | 1,0834 |
| 7 | 2013 | 84 398 | 1,0368 |
| 8 | 2014 | 85 859 | 1,0173 |
| 9 | 2015 | 93 067 | 1,0840 |
| 10 | 2016 | 98 864 | 1,0623 |
| 11 | 2017 | 103 821 | 1,0501 |
| 12 | 2018 | 104 764 | 1,0091 |

***Interpretace:***
Počet nehod v ČR se v roce 2008 snížil oproti roku 2007 o 12,24 % (0,8776*100-100).
Počet nehod v ČR se v roce 2009 snížil oproti roku 2008 o 53,35 %.
Počet nehod v ČR se v roce 2010 zvýšil oproti roku 2009 o 0,94 %.
Počet nehod v ČR se v roce 2011 snížil oproti roku 2010 o 0,51 %.
Počet nehod v ČR se v roce 2012 zvýšil oproti roku 2011 o 8,34 %.
atp.

d) Průměrný koeficient růstu je konstruován jako geometrický průměr koeficientů růstu. Lze ale také vypočítat jako *n-1*ní odmocnina z podílu poslední a první hodnoty časové řady.

$$\bar{k} = \sqrt[n-1]{k_2 \cdot k_3 \cdot \ldots \cdot k_n}$$

$$\bar{k} = \sqrt[11]{0{,}8776 \cdot 0{,}4665 \cdot 1{,}0094 \cdot \ldots \cdot 1{,}0091} = 0{,}9507$$

NEBO

$$\bar{k} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

$$\bar{k} = \sqrt[11]{\frac{104764}{182736}} = 0{,}9507$$

***Interpretace:***
Počet nehod v ČR se v letech 2007-2018 v průměru snížil o 4,93 %.

## Druhé a třetí diference

Druhé diference jsou definovány jako rozdíl sousedních prvních diferencí.
$\Delta_t^{(2)} = \Delta_t^{(1)} - \Delta_{t-1}^{(1)}$
První dva údaje není možné z dat, která máme k dispozici, vypočítat, proto opět píšeme do prvních dvou řádků puntík. Další hodnoty vypočítáme následovně:

$$\Delta_3^{(2)} = \Delta_3^{(1)} - \Delta_2^{(1)} = -85561 - (-22360) = -63201$$
$$\Delta_4^{(2)} = \Delta_4^{(1)} - \Delta_3^{(1)} = 707 - (-85561) = 86268$$
$$\Delta_5^{(2)} = \Delta_5^{(1)} - \Delta_4^{(1)} = -385 - 707 = -1092$$
$$\Delta_6^{(2)} = \Delta_6^{(1)} - \Delta_5^{(1)} = 6267 - (-385) = 6652$$
$$\Delta_7^{(2)} = \Delta_7^{(1)} - \Delta_6^{(1)} = 2994 - 6267 = -3273$$
$$\Delta_8^{(2)} = \Delta_8^{(1)} - \Delta_7^{(1)} = 1461 - 2994 = -1533$$
$$\Delta_9^{(2)} = \Delta_9^{(1)} - \Delta_8^{(1)} = 7208 - 1461 = 5747$$
$$\Delta_{10}^{(2)} = \Delta_{10}^{(1)} - \Delta_9^{(1)} = 5797 - 7208 = -1411$$
$$\Delta_{11}^{(2)} = \Delta_{11}^{(1)} - \Delta_{10}^{(1)} = 4957 - 5797 = -840$$
$$\Delta_{12}^{(2)} = \Delta_{12}^{(1)} - \Delta_{11}^{(1)} = 104764 - 103821 = -4014$$

| t | Rok | Počet nehod $y_t$ | $\Delta_t^{(1)}$ | $\Delta_t^{(2)}$ |
|---|---|---|---|---|
| 1 | 2007 | 182 736 | • | • |
| 2 | 2008 | 160 376 | -22 360 | • |
| 3 | 2009 | 74 815 | -85 561 | -63 201 |
| 4 | 2010 | 75 522 | 707 | 86 268 |
| 5 | 2011 | 75 137 | -385 | -1 092 |
| 6 | 2012 | 81 404 | 6 267 | 6 652 |
| 7 | 2013 | 84 398 | 2 994 | -3 273 |
| 8 | 2014 | 85 859 | 1 461 | -1 533 |
| 9 | 2015 | 93 067 | 7 208 | 5747 |
| 10 | 2016 | 98 864 | 5 797 | -1 411 |
| 11 | 2017 | 103 821 | 4 957 | -840 |
| 12 | 2018 | 104 764 | 943 | -4014 |

Třetí diference jsou definovány jako rozdíl sousedních druhých diferencí.
$\Delta_t^{(3)} = \Delta_t^{(2)} - \Delta_{t-1}^{(2)}$

První tři údaje není možné z dat, která máme k dispozici, vypočítat, proto opět píšeme do prvních tří řádků puntík. Další hodnoty už je možné spočítat.

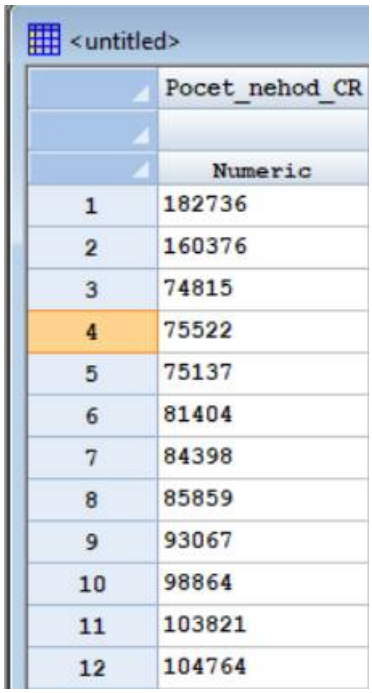
| t | Rok | Počet nehod $y_t$ | $\Delta_t^{(2)}$ | $\Delta_t^{(3)}$ |
|---|---|---|---|---|
| 1 | 2007 | 182 736 | • | • |
| 2 | 2008 | 160 376 | • | • |
| 3 | 2009 | 74 815 | -63 201 | • |
| 4 | 2010 | 75 522 | 86 268 | 149 469 |
| 5 | 2011 | 75 137 | -1 092 | -87 360 |
| 6 | 2012 | 81 404 | 6 652 | 7 744 |
| 7 | 2013 | 84 398 | -3 275 | -9 925 |
| 8 | 2014 | 85 859 | -1 533 | 1 740 |
| 9 | 2015 | 93 067 | 5747 | 7 280 |
| 10 | 2016 | 98 864 | -1 411 | -7 158 |
| 11 | 2017 | 103 821 | -840 | 571 |
| 12 | 2018 | 104 764 | -4014 | -3 174 |

$\Delta_4^{(3)} = \Delta_4^{(2)} - \Delta_3^{(2)} = 86268 - (-63201) = 149469$
$\Delta_5^{(3)} = \Delta_5^{(2)} - \Delta_4^{(2)} = -1092 - 86268 = -87360$
$\Delta_6^{(3)} = \Delta_6^{(2)} - \Delta_5^{(2)} = 6652 - (-1092) = 7744$
$\Delta_7^{(3)} = \Delta_7^{(2)} - \Delta_6^{(2)} = -3275 - 6652 = -9925$
$\Delta_8^{(3)} = \Delta_8^{(2)} - \Delta_7^{(2)} = -1533 - (-3275) = 1740$
$\Delta_9^{(3)} = \Delta_9^{(2)} - \Delta_8^{(2)} = 5747 - (-1533) = 7280$
$\Delta_{10}^{(3)} = \Delta_{10}^{(2)} - \Delta_9^{(2)} = -1411 - 5747 = -7158$
$\Delta_{11}^{(3)} = \Delta_{11}^{(2)} - \Delta_{10}^{(2)} = -840 - (-1411) = 571$
$\Delta_{12}^{(3)} = \Delta_{12}^{(2)} - \Delta_{11}^{(2)} = -4014 - (-840) = -3174$

Druhé a třetí diference se využívají při hledání vhodného modelu trendu (bude předmětem následujícího tématu).

## Řešení v SGP:

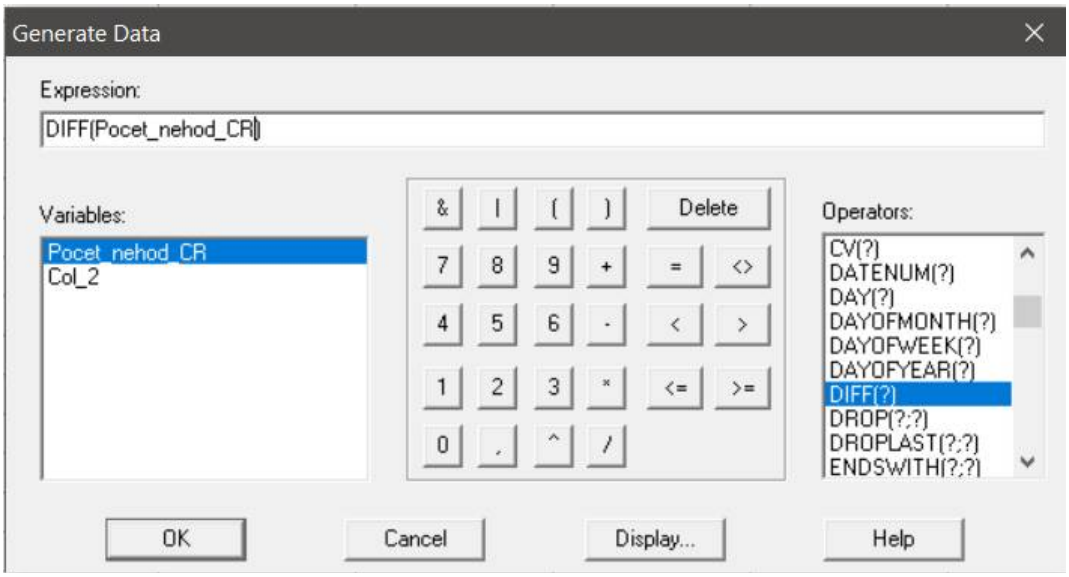
Prvním krokem při řešení příkladu v SGP je vložení vstupních dat. Údaje o počtu nehod v ČR vložíme jako jednu proměnnou, tj. do jednoho sloupce – viz Obr. 1.




| | Pocet_nehod_CR |
|---|---|
| | |
| | Numeric |
| 1 | 182736 |
| 2 | 160376 |
| 3 | 74815 |
| 4 | 75522 |
| 5 | 75137 |
| 6 | 81404 |
| 7 | 84398 |
| 8 | 85859 |
| 9 | 93067 |
| 10 | 98864 |
| 11 | 103821 |
| 12 | 104764 |

*Obrázek 1 – Vložení vstupních dat*

a) Pro výpočet prvních diferencí přeskočíme kurzorem na další sloupeček, levým tlačítkem ho podsvítíme a pravým tlačítkem myši zvolíme z nabídky procedur *Generate Data*. Ze sloupečku *Operators* v pravé části panelu vybereme operátor DIFF(?), který dvojím kliknutím dosadíme do řádku Expression, a za otazník zadáme proměnnou Počet_nehod_CR (dvojím kliknutím z tabulky Variables vlevo) – viz Obr. 2. Potvrdíme klávesou OK. Hodnoty prvních diferencí se objeví ve sloupci 2 – na Obr. 2 je už přejmenovaný na První_diference, aby bylo zřejmé, o jaké hodnoty se jedná.




*Obr. 2 – Zadání operátoru DIFF v Generate Data*

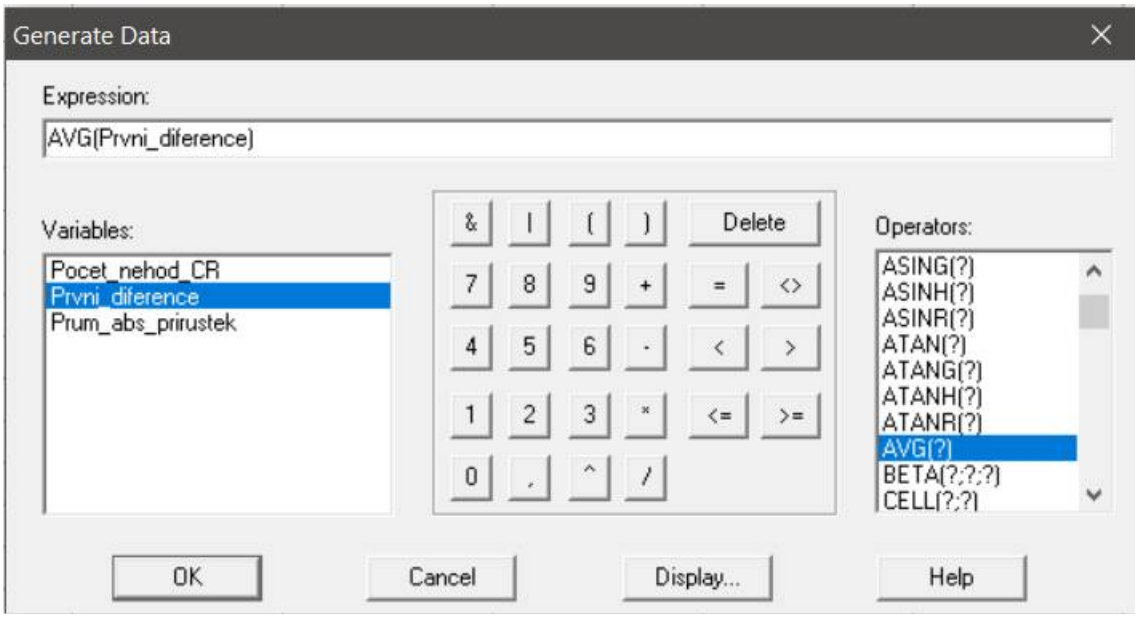
<untitled>

| | Pocet_nehod_CR | Prvni_diference |
|---|---|---|
| | Numeric | Numeric |
| 1 | 182736 | |
| 2 | 160376 | -22360 |
| 3 | 74815 | -85561 |
| 4 | 75522 | 707 |
| 5 | 75137 | -385 |
| 6 | 81404 | 6267 |
| 7 | 84398 | 2994 |
| 8 | 85859 | 1461 |
| 9 | 93067 | 7208 |
| 10 | 98864 | 5797 |
| 11 | 103821 | 4957 |
| 12 | 104764 | 943 |

*Obrázek 3 – Výsledek použití operátoru DIFF*

Interpretace vypočtených hodnot zde již není uvedena – je ji možné najít v první části řešeného příkladu, kde je ukázka ručního výpočtu.

b) Pro výpočet průměrného absolutního přírůstku využijeme rovněž operátoru v nabídce *Generate Data*. Víme, že průměrný absolutní přírůstek je aritmetický průměr prvních diferencí, proto použijeme operátor AVG. Třetí sloupec si nazveme Prum_abs_prirustek, podsvítíme ho a v nabídce *Generate Data* zadáme do řádku Expression operátor AVG(?). Za otazník zadáme proměnnou První_diference, jak ukazuje Obr. 4.




*Obrázek 4 – Použití operátoru AVG v Generate Data*

Po potvrzení klávesou OK, se objeví výsledná hodnota v podsvíceném sloupci – viz Obr. 5.

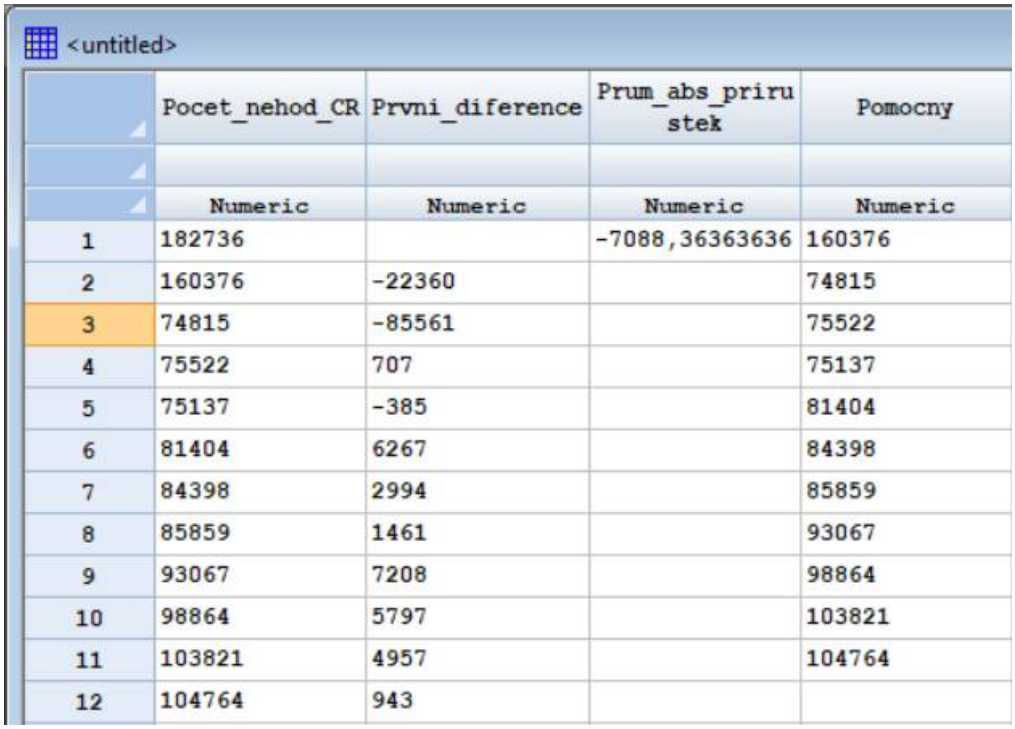
<untitled>

| | Pocet_nehod_CR | Prvni_diference | Prum_abs_prirustek |
|---|---|---|---|
| | Numeric | Numeric | Numeric |
| 1 | 182736 | | -7088,36363636 |
| 2 | 160376 | -22360 | |
| 3 | 74815 | -85561 | |
| 4 | 75522 | 707 | |
| 5 | 75137 | -385 | |
| 6 | 81404 | 6267 | |
| 7 | 84398 | 2994 | |
| 8 | 85859 | 1461 | |
| 9 | 93067 | 7208 | |
| 10 | 98864 | 5797 | |
| 11 | 103821 | 4957 | |
| 12 | 104764 | 943 | |

*Obrázek 5 – Vypočítaná hodnota průměrného absolutního přírůstku*

Interpretace vypočtené hodnoty viz ruční řešení.

c) Koeficienty růstu je možné v SGP vypočítat dvěma základními způsoby. První z nich opět využívá nabídku *Generate Data*. Nejprve je ale potřeba před výpočtem vytvořit pomocný sloupec s hodnotami proměnné Počet_nehod_CR bez první hodnoty. Viz Obr. 6.

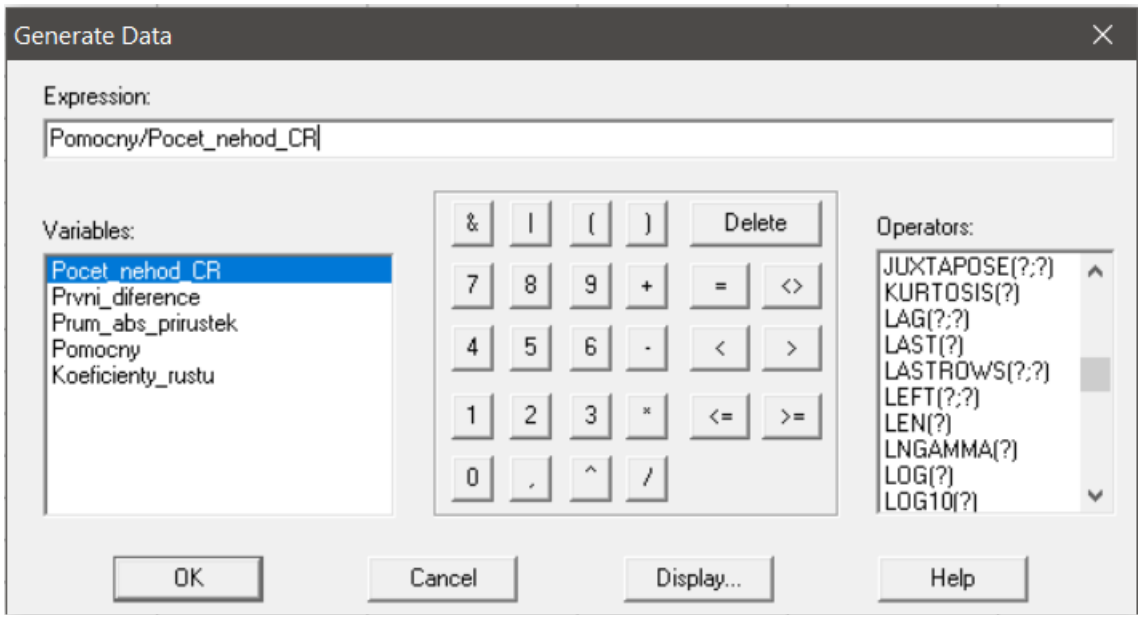


<untitled>

| | Pocet_nehod_CR | Prvni_diference | Prum_abs_prirustek | Pomocny |
|---|---|---|---|---|
| | | | | |
| | Numeric | Numeric | Numeric | Numeric |
| 1 | 182736 | | -7088,36363636 | 160376 |
| 2 | 160376 | -22360 | | 74815 |
| 3 | 74815 | -85561 | | 75522 |
| 4 | 75522 | 707 | | 75137 |
| 5 | 75137 | -385 | | 81404 |
| 6 | 81404 | 6267 | | 84398 |
| 7 | 84398 | 2994 | | 85859 |
| 8 | 85859 | 1461 | | 93067 |
| 9 | 93067 | 7208 | | 98864 |
| 10 | 98864 | 5797 | | 103821 |
| 11 | 103821 | 4957 | | 104764 |
| 12 | 104764 | 943 | | |

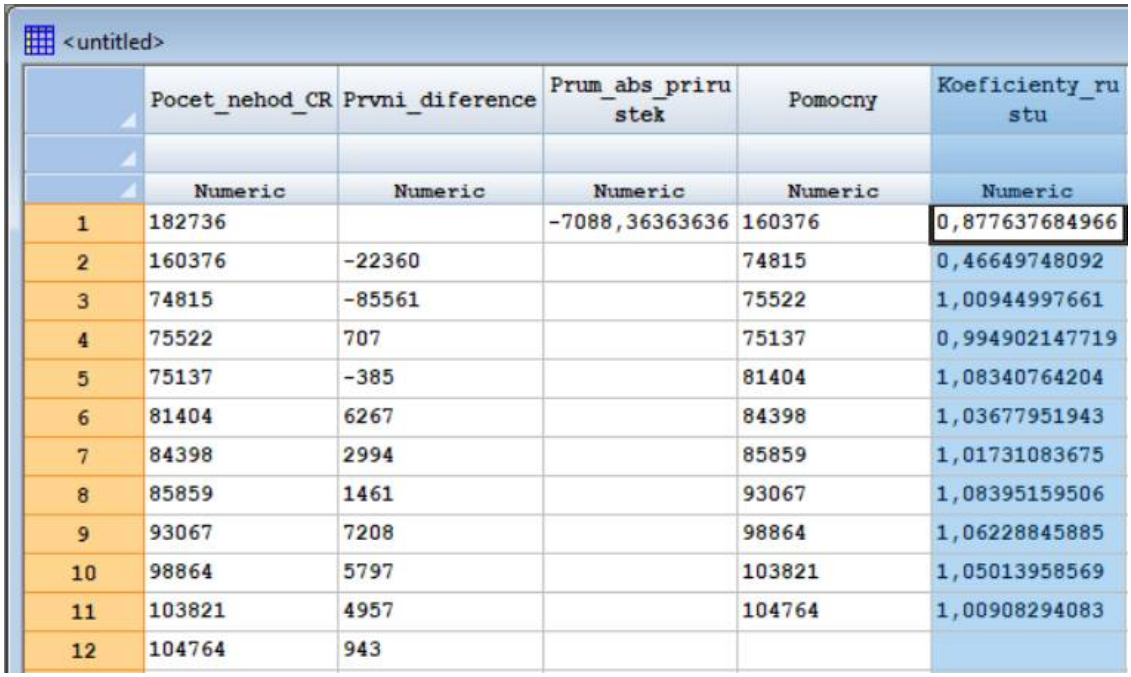
*Obrázek 6 – Vložení pomocné proměnné*

Kurzorem přeskočíme do dalšího volného sloupce, který podsvítíme, a v panelu *Generate Data* zadáme do řádku Expression vzorec „Pomocny/Počet_nehod_CR“, jak ukazuje Obr. 7.




*Obrázek 7 – Zadání vzorce pro výpočet koeficientů růstu v Generate Data*

Po stisknutí klávesy OK se hodnoty koeficientů růstu objeví v podsvíceném sloupci – viz Obr. 8 (už je zde přejmenován na Koeficienty_rustu).



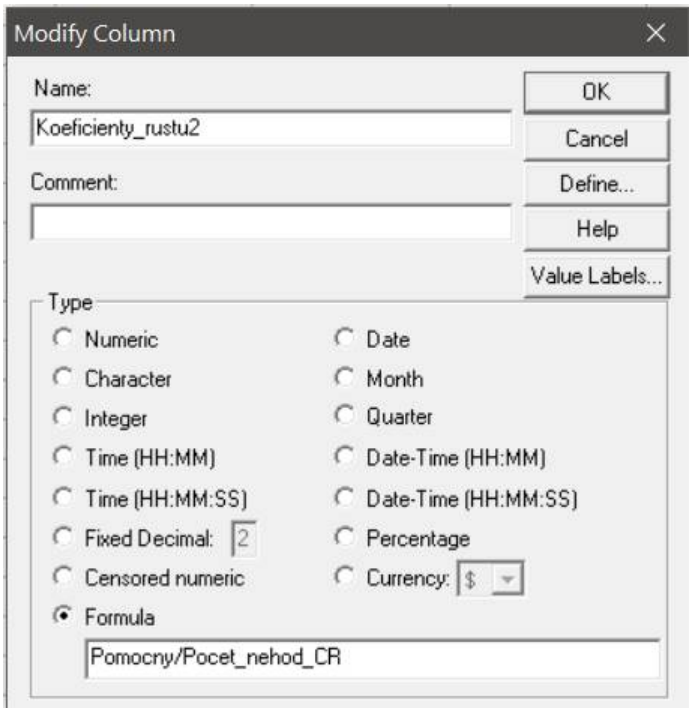
<untitled>

| | Pocet_nehod_CR | Prvni_diference | Prum_abs_prirustek | Pomocny | Koeficienty_rustu |
|---|---|---|---|---|---|
| | Numeric | Numeric | Numeric | Numeric | Numeric |
| 1 | 182736 | | -7088,36363636 | 160376 | 0,877637684966 |
| 2 | 160376 | -22360 | | 74815 | 0,46649748092 |
| 3 | 74815 | -85561 | | 75522 | 1,00944997661 |
| 4 | 75522 | 707 | | 75137 | 0,994902147719 |
| 5 | 75137 | -385 | | 81404 | 1,08340764204 |
| 6 | 81404 | 6267 | | 84398 | 1,03677951943 |
| 7 | 84398 | 2994 | | 85859 | 1,01731083675 |
| 8 | 85859 | 1461 | | 93067 | 1,08395159506 |
| 9 | 93067 | 7208 | | 98864 | 1,06228845885 |
| 10 | 98864 | 5797 | | 103821 | 1,05013958569 |
| 11 | 103821 | 4957 | | 104764 | 1,00908294083 |
| 12 | 104764 | 943 | | | |

*Obrázek 8 – Vypočítané hodnoty koeficientů růstu*

Interpretace vypočtených hodnot viz ruční řešení.

Druhý způsob výpočtu koeficientů růstu využívá nabídku *Modify Column*, kterou je možné vyvolat stejným způsobem jako *Generate Data*, tj. podsvítit další sloupec a pravým tlačítkem vyvolat příslušnou nabídku. V panelu *Modify Column* vyplníme nejprve název nové proměnné Koeficienty_rustu2 a pak zvolíme typ proměnné Formula. Do řádku Formula vepíšeme vzorec Pomocny/Počet_nehod_CR, jak je vidět na Obr. 9.



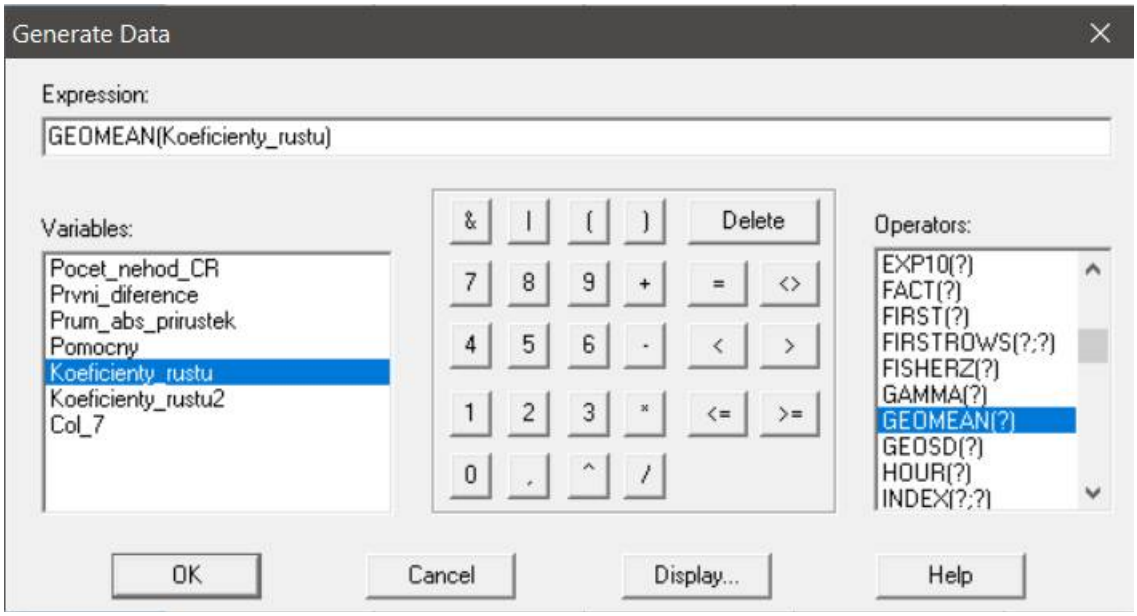

*Obrázek 9 – Zadání vzorce pro výpočet koeficientů růstu v Modify Column*

Po stisknutí klávesy OK se v podsvíceném sloupci objeví vypočítané hodnoty koeficientů růstu, jak ukazuje Obr. 10. Oproti předchozímu způsobu výpočtu jsou však šedé. To znamená, že pokud bychom např. udělali chybu ve vstupních datech, můžeme vstupní data opravit a daná oprava se promítne i do již vypočítaných hodnot.

| Koeficienty_rustu | Koeficienty_rustu2 |
|---|---|
| Numeric | Numeric |
| 0,877637684966 | 0,877637684966 |
| 0,46649748092 | 0,46649748092 |
| 1,00944997661 | 1,00944997661 |
| 0,994902147719 | 0,994902147719 |
| 1,08340764204 | 1,08340764204 |
| 1,03677951943 | 1,03677951943 |
| 1,01731083675 | 1,01731083675 |
| 1,08395159506 | 1,08395159506 |
| 1,06228845885 | 1,06228845885 |
| 1,05013958569 | 1,05013958569 |
| 1,00908294083 | 1,00908294083 |

*Obrázek 10 – Vypočítané hodnoty koeficientů růstu*

d) K výpočtu průměrného koeficientu růstu opět využijeme jeden z operátorů, protože víme, že je dán jako geometrický průměr koeficientů růstu. Podsvítíme nový sloupec a pravým tlačítkem myši opět zvolíme *Generate Data*, kde mezi operátory vybereme GEOMEAN (?). Vložíme tento operátor do řádku Expression a za otazník dosadíme proměnnou Koeficienty_rustu (nebo Koeficienty_rustu2) – viz Obr. 11. Potvrdíme OK.




*Obrázek 11 – Použití operátoru GEOMEAN v Generate Data*

V podsvíceném sloupci se objeví hodnota průměrného koeficientu růstu – viz Obr. 12.

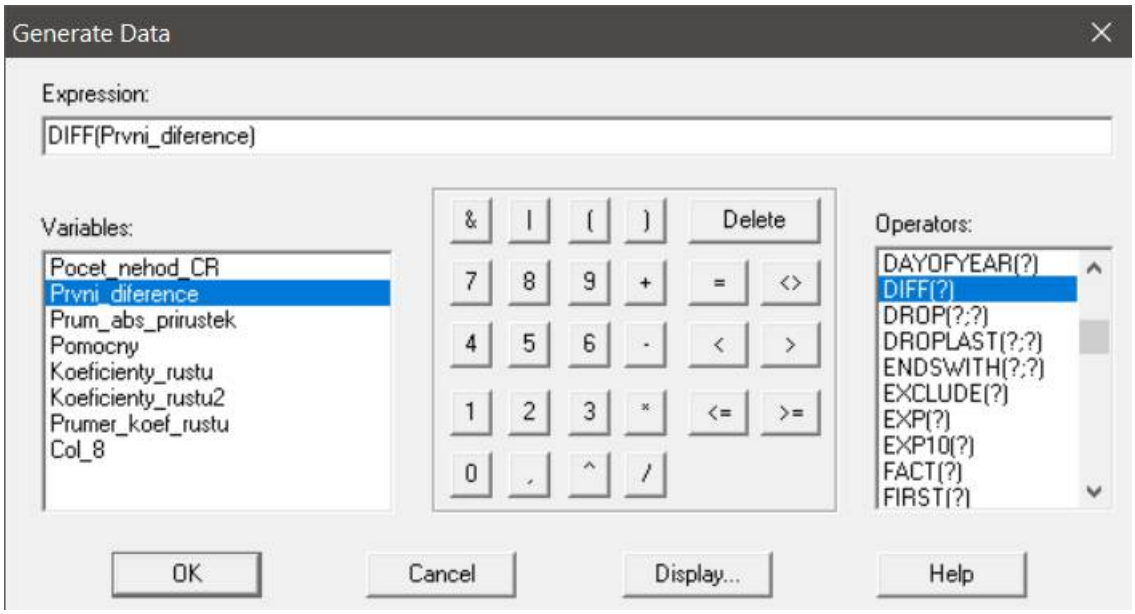
| Koeficienty_rustu | Koeficienty_rustu2 | Prumer_koef_rustu |
|---|---|---|
| Numeric | Numeric | Numeric |
| 0,877637684966 | 0,877637684966 | 0,950681995027 |
| 0,46649748092 | 0,46649748092 | |
| 1,00944997661 | 1,00944997661 | |
| 0,994902147719 | 0,994902147719 | |
| 1,08340764204 | 1,08340764204 | |
| 1,03677951943 | 1,03677951943 | |
| 1,01731083675 | 1,01731083675 | |
| 1,08395159506 | 1,08395159506 | |
| 1,06228845885 | 1,06228845885 | |
| 1,05013958569 | 1,05013958569 | |
| 1,00908294083 | 1,00908294083 | |

*Obrázek 12 – Vypočítaná hodnota průměrného koeficientu růstu*

Interpretace vypočtené hodnoty viz ruční výpočet.

## Druhé a třetí diference

Hodnoty druhých diferencí jsou definovány jako rozdíl sousedních prvních diferencí, proto k jejich výpočtu můžeme použít opět operátor DIFF. Podsvítíme tedy další sloupec, potvrdíme nabídku *Generate Data* a do řádku Expression zadáme operátor DIFF(?) a za otazník dosadíme proměnnou Prvni_diference – viz Obr. 13.




*Obrázek 13 – Použití operátoru DIFF v Generate Data*

Po potvrzení klávesou OK se v podsvíceném sloupci zobrazí vypočítané hodnoty druhých diferencí, jak je vidět na Obr. 14.

| Druhe_diference |
|---|
| Numeric |
| |
| |
| -63201 |
| 86268 |
| -1092 |
| 6652 |
| -3273 |
| -1533 |
| 5747 |
| -1411 |
| -840 |
| -4014 |

*Obrázek 14 – Vypočítané hodnoty druhých diferencí*

Hodnoty třetích diferencí získáme obdobným způsobem – jen za otazník k operátoru DIFF zadáme hodnoty druhých diferencí. Výsledné hodnoty můžeme vidět na Obr. 15.

| Treti_diference |
|---|
| Numeric |
| |
| |
| |
| 149469 |
| -87360 |
| 7744 |
| -9925 |
| 1740 |
| 7280 |
| -7158 |
| 571 |
| -3174 |

*Obrázek 15 – Vypočítané hodnoty třetích diferencí*

# 4 PRAVDĚPODOBNOST – 1. ČÁST

***Pravděpodobnost:***
- matematická disciplína, která se zabývá studiem zákonitostí, jimiž se řídí hromadné náhodné jevy
- vytváří pravděpodobnostní modely, pomocí nichž se snaží postihnout procesy, ovlivněné náhodou.

***Náhodné pokusy:*** procesy, jejichž výsledek nelze předem jednoznačně určit (je nejistý); závisí jednak na daných podmínkách, při kterých je prováděn, jednak na náhodě. Teorie pravděpodobnosti se zabývá pouze náhodnými pokusy, které jsou za stejných podmínek opakovatelné a u nichž je variabilita výsledků podstatná a vykazuje určitou zákonitost.

***Hromadné náhodné jevy:*** výsledky opakovatelných náhodných pokusů (symbolické značení – *A, B, C, …*).

***Pravděpodobnost náhodného jevu:*** pravděpodobnost náhodného jevu *A* je číslo *P(A)*, které lze interpretovat jako míru možnosti nastoupení náhodného jevu.

**Axiomatická teorie pravděpodobnosti:** pravděpodobnost je funkce, která každému náhodnému jevu přiřazuje reálné číslo, přičemž musí být splněny základní axiomy.

1. $P(A) \geq 0$
2. $P(A_1 \cup A_2 \cup ...) = P(A_1) + P(A_2) + ...$ (pro neslučitelné jevy)
3. $P(E) = 1$.

**Klasická definice pravděpodobnosti:** pravděpodobnost jevu *A* se rovná podílu případů příznivých nastoupení jevu *A* a počtu všech případů možných, jsou-li všechny stejně pravděpodobné.

$$P(A) = \frac{m}{n}$$

kde $m$ je počet případů příznivých
$n$ je počet případů možných.

**Statistická definice pravděpodobnosti:** pokud platí, že při rostoucím počtu opakování náhodného pokusu ($n$) kolísá relativní četnost $\frac{m}{n}$ ve stále užších mezích kolem určitého čísla, můžeme toto číslo považovat za pravděpodobnost jevu *A*.

relativní četnost jevu $A = \frac{m}{n}$

kde $m$ je počet nastoupení jevu *A*
$n$ je počet opakování pokusu.

- pravděpodobnosti náhodného jevu odhadujeme na základě výsledků, získaných při mnohonásobném opakování náhodného pokusu (aposteriorní charakter definice).

**Pravidla pro počítání s pravděpodobnostmi**

***Podmíněná pravděpodobnost:*** $P(A/B)$ je podmíněná pravděpodobnost jevu $A$ vzhledem k jevu $B$, tj. pravděpodobnost nastoupení jevu $A$ za předpokladu, že nastal jev $B$.

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \text{ pro } P(B) > 0$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)}, \text{ pro } P(A) > 0$$

***Pravidlo o násobení pravděpodobností:***
Pravděpodobnost současného nastoupení jevů $A$ a $B$ (tzn. jejich průniku) je rovna součinu nepodmíněné pravděpodobnosti jednoho jevu a podmíněné pravděpodobnosti druhého jevu vzhledem k prvnímu jevu.

$$P(A \cap B) = P(B) \cdot P(A/B)$$
$$P(A \cap B) = P(A) \cdot P(B/A)$$

*Zobecnění pravidla o násobení pravděpodobností pro dva a více jevů*:

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) \cdot \ldots \cdot P(A_n/A_1 \cap \ldots \cap A_{n-1}).$$

***Nezávislost jevů:***
Jestliže $P(A/B) = P(A)$, pak jev $A$ nezávisí na jevu $B$.
Jestliže $P(B/A) = P(B)$, pak jev $B$ nezávisí na jevu $A$.

*Nutná a postačující podmínka nezávislosti dvou jevů:* $P(A \cap B) = P(A) \cdot P(B)$.

*Zjednodušení pravidla o násobení pravděpodobností pro nezávislé jevy:*

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot \ldots \cdot P(A_n).$$

***Pravidlo pro sčítání pravděpodobností:***
Pravděpodobnost sjednocení jevů $A$ a $B$ je rovna součtu pravděpodobností těchto jevů, zmenšené o pravděpodobnost jejich průniku.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

***Disjunktní jevy:*** Jestliže $P(A \cap B) = 0$, pak jevy $A$ a $B$ jsou disjunktní.

*Zjednodušení pravidla pro sčítání pravděpodobností pro disjunktní jevy:*

$$P(A \cup B) = P(A) + P(B).$$

**Operace s náhodnými jevy**

- vztahy mezi náhodnými jevy graficky znázorňují tzv. ***Vennovy diagramy***.

**1.** $\boldsymbol{A \subset B}$
Jev *A* je částí jevu *B*; z jevu *A* plyne jev *B* (implikace); nastoupení jevu *A* má vždy za následek nastoupení jevu *B*.

**2.** $\boldsymbol{A = B}$
Jevy *A* a *B* jsou si rovny; $A \subset B$ a současně $B \subset A$.

**3.** $\boldsymbol{C = A \cap B}$
Jev *C* je průnik jevů *A* a *B* (logický součin); jev *C* nastane právě tehdy, nastane-li současně jev *A* i jev *B*.

**4.** $\boldsymbol{C = A \cup B}$
Jev *C* je sjednocení jevů *A* a *B* (logický součet); jev *C* nastane právě tehdy, nastane-li alespoň jeden z jevů *A* a *B*.

**5.** $\boldsymbol{C = A - B}$
Jev *C* je rozdíl jevů *A* a *B*; jev *C* nastane právě tehdy, když jev *A* nastane a současně jev *B* nenastane.

**6.** ***E*** je jev jistý, tj. jev, který musí nastat vždy.
∅ je jev nemožný, tj. jev, který nastat nemůže.

**Kombinatorika**

***Permutace***

$$P(n) = n!$$

***Variace bez opakování***

$$V_k(n) = \frac{n!}{(n-k)!}$$

***Variace s opakováním***

$$V_k'(n) = n^k$$

***Kombinace***

$$C_k(n) = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

# 5 PRAVDĚPODOBNOST – 2. ČÁST

***Náhodná veličina:***
- veličina, jejíž hodnota je jednoznačně určena výsledkem náhodného pokusu
- vlivem náhodných činitelů může nabýt různých hodnot, proto nelze její konkrétní hodnotu před provedením náhodného pokusu jednoznačně určit
- náhodnou veličinu pokládáme za danou, pokud známe všechny její možné hodnoty a pravděpodobnosti výskytu každé z nich
- symbolické značení – *X, Y, Z*, …
- příklady náhodných veličin: počet bodů, které padnou na hrací kostce, počet poruch určitého zařízení, doba čekání na obsluhu v určité prodejně, atd.

***Zákon rozdělení náhodné veličiny:*** pravidlo, které každé hodnotě nebo množině hodnot z každého intervalu přiřazuje pravděpodobnost, že náhodná veličina nabude této hodnoty nebo hodnoty z určitého intervalu. Je to pravděpodobnostní model empirické náhodné veličiny.

## 5.1 Popis rozdělení náhodné veličiny

### 5.1.1 Diskrétní náhodná veličina

***Distribuční funkce:*** pravděpodobnost, že NV nabude hodnoty menší nebo rovné *x*.

$$F(x) = P(X \le x) = \sum_{t \le x} P(t)$$

***Pravděpodobnostní funkce:*** pravděpodobnost, že NV nabude hodnoty rovné *x*.

$$P(x) = P(X = x)$$

$$\sum_{M} P(x) = 1,$$

kde $M$ je prostor hodnot NV X, tj. množina možných hodnot NV X.

### 5.1.2 Spojitá náhodná veličina

***Distribuční funkce***

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$$

***Hustota pravděpodobnosti***

$$f(x) = F'(x) = \frac{dF(x)}{dx}$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

## 5.2 Charakteristiky náhodných veličin

- číselné hodnoty, jejichž cílem je koncentrovat popis NV
- poskytují výstižný popis základních vlastností rozdělení NV.

***Podle vlastnosti rozdělení, kterou popisují, rozeznáváme:***

- charakteristiky polohy
- charakteristiky variability
- charakteristiky šikmosti
- charakteristiky špičatosti.

### 5.2.1 Charakteristiky polohy

**Střední hodnota** $\boldsymbol{E(X)}$

- tzv. očekávaná hodnota (z latinského expectatis)

***a) Diskrétní NV***

$$E(X) = \sum_{M} x \cdot P(x)$$

***b) Spojitá NV***

$$E(X) = \int_{M} x \cdot f(x) dx$$

**Modus** $\boldsymbol{\hat{x}}$

***a) Diskrétní NV***

- $\hat{x}$ je hodnota NV, která má největší pravděpodobnost výskytu
  $\hat{x}$ … max $P(x)$.

***b) Spojitá NV***

- $\hat{x}$ je bod, v němž je hustota pravděpodobnosti maximální, tj. lokální maximum hustoty pravděpodobnosti *f(x)*
  $\hat{x}$ … $f'(x) = 0$.

**Kvantily**

- používají se především kvantily *spojité* náhodné veličiny
- $x_p$ je hodnota, kterou hodnoty NV nepřekročí s pravděpodobností 100p %.

*Hodnota $x_p$ je 100p% kvantilem NV X, jestliže pro ni platí:*

$$F(x_p) = P(X \le x_p) = p\,.$$

### 5.2.2 Charakteristiky variability

**Rozptyl** $\boldsymbol{D(X)}$

***a) Diskrétní NV***

$$D(X) = \sum_{M} [x - E(X)]^2 \cdot P(x) = \sum_{M} x^2 \cdot P(x) - \left[ \sum_{M} x \cdot P(x) \right]^2$$

***b) Spojitá NV***

$$D(X) = \int_{M} [x - E(X)]^2 f(x) dx = \int_{M} x^2 \cdot f(x) dx - \left[ \int_{M} x \cdot f(x) dx \right]^2$$

**Směrodatná odchylka** $\boldsymbol{\sigma(X)}$

$$\boldsymbol{\sigma(X) = \sqrt{D(X)}}$$

# 6 PRAVDĚPODOBNOST – 3. ČÁST

***Rozdělení náhodné veličiny:***
- pravděpodobnostní model chování náhodné veličiny
- existuje celá řada rozdělení pro diskrétní i spojité náhodné veličiny.

## 6.1 Některá rozdělení diskrétních náhodných veličin

### 6.1.1 Alternativní rozdělení $A(\pi)$

- NV $X$ je počet nastoupení jevu $A$ při realizaci náhodného pokusu
- rozdělení nula-jedničkové náhodné veličiny
- *rozdělení má jeden parametr:*
  $\pi$ … pravděpodobnost nastoupení sledovaného jevu při náhodném pokusu.

***Pravděpodobnostní funkce:***

$$P(x) = \pi^x (1-\pi)^{1-x},\ x = 0, 1,\ 0 < \pi < 1,$$
$$= 0 \qquad \text{jinak.}$$

$$E(X) = \pi$$
$$D(X) = \pi(1-\pi)$$

### 6.1.2 Binomické rozdělení $Bi(n;\ \pi)$

- NV $X$ je počet výskytů náhodného jevu $A$ v $n$ nezávislých náhodných pokusech, je-li pravděpodobnost nastoupení jevu $A$ ve všech pokusech stejná ($\pi$)
- *rozdělení má dva parametry:*
  $n$ … počet nezávislých pokusů
  $\pi$ … pravděpodobnost nastoupení sledovaného jevu v jednom pokusu.

***Pravděpodobnostní funkce:***

$$P(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x},\ x = 0, 1, \ldots, n,\ 0 < \pi < 1,$$
$$= 0 \qquad \text{jinak.}$$

$$E(X) = n\pi$$
$$D(X) = n\pi(1-\pi)$$

*Např.:* NV $X$ je počet „šestek“, které padnou při deseti hodech kostkou.

### 6.1.3 Poissonovo rozdělení $Po(\lambda)$

- NV X je počet výskytů náhodného jevu $A$ v určitém časovém intervalu délky $t$ (tzn. za jednotku času), v jednotce plochy nebo objemu (v prostorové jednotce)
- *rozdělení má jeden parametr:*
  $\lambda$ … střední hodnota rozdělení.

***Pravděpodobnostní funkce:***

$$P(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}, \; x = 0, 1, 2, \ldots, \lambda > 0,$$
$$= 0 \qquad \text{jinak.}$$

$$E(X) = \lambda$$
$$D(X) = \lambda$$

*Např.:* NV X je počet poruch stroje za směnu, počet telefonních hovorů za hodinu, počet vad na 1 $m^2$ koberce.

***Aproximace Binomického rozdělení rozdělením Poissonovým:***

- podmínky pro aproximaci: počet pokusů musí být dostatečně velký (alespoň $n > 30$) a pravděpodobnost $\pi$ velmi malá (alespoň $\pi \leq 0{,}1$)
- při aproximaci udává *P(x)* přibližnou pravděpodobnost, že ve velkém počtu *n* nezávislých náhodných pokusů se sledovaný jev *A* vyskytne *x*-krát, jestliže pravděpodobnost výskytu jevu v jednom pokusu je velni malá.

*Např.:* NV *X* je počet vadných výrobků ve velké sérii, je-li pravděpodobnost výroby zmetku velmi malá.

### 6.1.4 Hypergeometrické rozdělení $\boldsymbol{Hyp(N; M; n)}$

- používá se v případě závislých pokusů, tzn. při výběru bez vracení
- NV X je počet vybraných prvků se sledovanou vlastností při závislých pokusech
- *rozdělení má tři parametry:*
  *N* … rozsah souboru, z něhož vybíráme
  *M* … počet prvků v základním souboru, které mají sledovanou vlastnost
  *n* … počet závislých pokusů (rozsah výběru).

***Pravděpodobnostní funkce:***

$$P(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \; x = max[0,\; M - N + n], \ldots, min[M, n],$$
$$= 0 \qquad \text{jinak.}$$

$$E(X) = n \cdot \frac{M}{N}$$

$$D(X) = n \cdot \frac{M}{N}\left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1}$$

*Použití:* např. při kontrole jakosti u malého počtu výrobků nebo v případě, kdy kontrola má ráz destrukční zkoušky (výrobek je zničen).

# Příklad 1

## Zadání příkladu:

Na lístcích jsou napsána čísla od 1 do 10, přičemž každé číslo je napsáno na různém počtu lístků. Číslo 1 je na jednom lístku, číslo 2 na dvou lístcích, číslo 3 na třech lístcích atd. Číslo deset je tedy na deseti lístcích. Všechny lístky jsou vloženy do osudí a pečlivě promíchány. Stanovte pravděpodobnost, že:

a) Při třech náhodných po sobě jdoucích výběrech bez vracení bude první vytažené číslo 1, druhé vytažené číslo 2 a třetí vytažené číslo 3.
b) Při třech náhodných po sobě jdoucích výběrech s vracením bude první vytažené číslo 1, druhé vytažené číslo 2 a třetí vytažené číslo 3.

## Vypracování příkladu:

Celkový počet lístků v osudí = $1+2+3+\ldots+10 = 55$

a) Sledujeme celkem tři jevy, a to A, B a C. Vzhledem k tomu, že se jedná o výběr bez vracení, tedy o závislé pokusy, je třeba při výpočtu stanovit podmíněné pravděpodobnosti.

Jev A je vytažení lístku s číslem 1. Pravděpodobnost tohoto jevu není podmíněná. Lístek s číslem 1 je v osudí pouze jeden.

**1. krok:** $P(A) = \frac{1}{55} \doteq 0{,}018182$

Jev B je vytažení lístku s číslem 2. Stanovíme podmíněnou pravděpodobnost nastoupení jevu B za podmínky, že v prvním kroku nastal jev A, tedy byl vytažen lístek s číslem 1. Lístky s číslem 2 jsou v osudí celkem dva.

**2. krok:** $P(A/B) = \frac{2}{54} = 0{,}037037$

Jev C je vytažení lístku s číslem 3. Stanovíme podmíněnou pravděpodobnost nastoupení jevu C za podmínky, že v prvním kroku nastal jev A a v druhém kroku jev B. Lístky s číslem 3 jsou v osudí celkem tři.

**3. krok:** $P(C/A\cap B) = \frac{3}{53} \doteq 0{,}056604$

Pravděpodobnost, že při výběru bez vracení bude první vytažené číslo 1, druhé vytažené číslo 2 a třetí vytažené číslo 3 vypočteme jako součin pravděpodobností, stanovených pro jednotlivé kroky.

$$P(A\cap B\cap C) = P(A)\cdot P(B/A)\cdot P(C/A\cap B) = \frac{1}{55}\cdot\frac{2}{54}\cdot\frac{3}{53} \doteq 0{,}000038$$

## Interpretace:

Pravděpodobnost, že při třech náhodných po sobě jdoucích výběrech bez vracení bude první vytažené číslo 1, druhé vytažené číslo 2 a třetí vytažené číslo 3 je 0,0038 %. Tato pravděpodobnost je velmi malá.

b) Sledujeme celkem tři jevy, a to A, B a C. Vzhledem k tomu, že se jedná o výběr s vracením, tedy o nezávislé pokusy, je třeba stanovit nepodmíněné pravděpodobnosti jednotlivých jevů.

Jev A je vytažení lístku s číslem 1. Pravděpodobnost tohoto jevu není podmíněná. Lístek s číslem 1 je v osudí pouze jeden.

**1. krok:** $P(A) = \frac{1}{55} \doteq 0{,}018182$

Jev B je vytažení lístku s číslem 2. Pravděpodobnost tohoto jevu není podmíněná. Lístky s číslem 2 jsou v osudí celkem dva.

**2. krok:** $P(B) = \frac{2}{55} \doteq 0{,}036364$

Jev C je vytažení lístku s číslem 3. Pravděpodobnost tohoto jevu není podmíněná. Lístky s číslem 3 jsou v osudí celkem tři.

**3. krok:** $P(C) = \frac{3}{55} \doteq 0{,}054546$

Pravděpodobnost, že při výběru s vracením bude první vytažené číslo 1, druhé vytažené číslo 2 a třetí vytažené číslo 3 vypočteme jako součin pravděpodobností, stanovených pro jednotlivé kroky.

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C) = \frac{1}{55} \cdot \frac{2}{55} \cdot \frac{3}{55} \doteq 0{,}000036$$

## Interpretace:

Pravděpodobnost, že při třech náhodných po sobě jdoucích výběrech s vracením bude první vytažené číslo 1, druhé vytažené číslo 2 a třetí vytažené číslo 3 je 0,0036 %. Tato pravděpodobnost je velmi malá.

# Příklad 2

## Zadání příkladu:

Dodávku Kinder vajíček s překvapením obdržel supermarket v krabicích po 200 kusech. Jedná se o vajíčka s figurkami afrických zvířat, přičemž výrobce deklaruje, že 40 % vajíček obsahuje figurku zebry, 30 % figurku antilopy, 20 % figurku slona a 10 % figurku lva. Stanovte pravděpodobnost, že při náhodném výběru tří vajíček z plné krabice bude:

a) ve všech třech vajíčkách lev
b) ve všech třech vajíčkách zebra.

## Vypracování příkladu:

a) Sledujeme nastoupení jevu A, kterým je vybrání tří vajíček s překvapením s figurkou lva.

$$P(A) = \frac{m}{n}$$

***Případy příznivé nastoupení jevu A (m):***

V krabici je celkem 200 vajíček, z toho jich 20 obsahuje figurku lva. Příznivé případy jsou všechny trojice vajíček, které obsahují figurku lva.

$$C_3(20) = \frac{n!}{k!(n-k)!} = \frac{20!}{3!17!} = 1140$$

***Případy možné (n):***

V krabici je celkem 200 vajíček, možné případy jsou všechny trojice vajíček, které lze z 200 vajíček vytvořit.

$$C_3(200) = \frac{n!}{k!(n-k)!} = \frac{200!}{3!197!} = 1313400$$

$$P(A) = \frac{m}{n} = \frac{1140}{1313400} \doteq 0{,}000868$$

## Interpretace:

Pravděpodobnost, že při náhodném výběru tří vajíček z plné krabice bude ve všech vajíčkách figurka lva, činí 0,0868 %. Tato pravděpodobnost je velmi malá.

b) Sledujeme nastoupení jevu A, kterým je vybrání tří vajíček s překvapením s figurkou zebry.

$$P(A) = \frac{m}{n}$$

***Případy příznivé nastoupení jevu A (m):***

V krabici je celkem 200 vajíček, z toho jich 80 obsahuje figurku lva. Příznivé případy jsou všechny trojice vajíček, které obsahují figurku zebry.

$$C_3(80) = \frac{n!}{k!(n-k)!} = \frac{80!}{3!77!} = 82160$$

***Případy možné (n):***

V krabici je celkem 200 vajíček, možné případy jsou všechny trojice vajíček, které lze z 200 vajíček vytvořit.

$$C_3(200) = \frac{n!}{k!(n-k)!} = \frac{200!}{3!197!} = 1313400$$

$$P(A) = \frac{m}{n} = \frac{82160}{1313400} \doteq 0{,}062555$$

## Interpretace:

Pravděpodobnost, že při náhodném výběru tří vajíček z plné krabice bude ve všech vajíčkách figurka zebry, činí 6,2555 %. Tato pravděpodobnost je malá.

# Příklad 1

## Zadání příkladu:

Náhodná veličina $X$ je popsána následující pravděpodobnostní funkcí:

$$P(x) = 0{,}5^x \cdot 0{,}5 \qquad x = 0,1,2,3$$
$$= 0{,}5^4 \qquad x = 4$$
$$= 0 \qquad jinak$$

a) Vyjádřete tuto pravděpodobnostní funkci tabulkou.
b) Stanovte modus, střední hodnotu a rozptyl této náhodné veličiny.

## Vypracování příkladu:

a) Tabulka obsahuje všechny hodnoty, kterých může daná náhodná veličina nabýt, a jejich pravděpodobnosti. Jednotlivé pravděpodobnosti vypočteme dosazením příslušných hodnot náhodné veličiny do pravděpodobnostní funkce.

$P(0) = 0{,}5^0 \cdot 0{,}5 = 0{,}5$

$P(1) = 0{,}5^1 \cdot 0{,}5 = 0{,}25$

$P(2) = 0{,}5^2 \cdot 0{,}5 = 0{,}125$

$P(3) = 0{,}5^3 \cdot 0{,}5 = 0{,}0625$

$P(4) = 0{,}5^4 = 0{,}0625$

***Tabulka pravděpodobnostní funkce***

| **x** | 0 | 1 | 2 | 3 | 4 | Celkem |
|---|---|---|---|---|---|---|
| **P(x)** | 0,5 | 0,25 | 0,125 | 0,0625 | 0,0625 | 1,0000 |

## Interpretace:

Pravděpodobnost, že náhodná veličina nabude hodnoty 0, je 50 %.

Pravděpodobnost, že náhodná veličina nabude hodnoty 1, je 25 %.

Pravděpodobnost, že náhodná veličina nabude hodnoty 2, je 12,5 %.

Pravděpodobnost, že náhodná veličina nabude hodnoty 3, je 6,25 %.

Pravděpodobnost, že náhodná veličina nabude hodnoty 4, je 6,25 %.

b) ***Modus náhodné veličiny***

$\hat{x} = 0$

***Střední hodnota náhodné veličiny*** $E(X)$

$$E(X) = \sum_{M} x \cdot P(x) = 0{,}9375$$

***Rozptyl náhodné veličiny*** $D(X)$

$$D(X) = \sum_{M} [x - E(X)]^2 \cdot P(x) = \sum_{M} x^2 \cdot P(x) - \left[\sum_{M} x \cdot P(x)\right]^2 = 2{,}5125 - 0{,}9375^2 = 1{,}633594$$

Tabulka s výpočty

| ***x*** | *P(x)* | *xP(x)* | *$x^2$P(x)* |
|---|---|---|---|
| 0 | 0,5 | 0 | 0 |
| 1 | 0,25 | 0,25 | 0,25 |
| 2 | 0,125 | 0,25 | 1,00 |
| 3 | 0,0625 | 0,1875 | 0,5625 |
| 4 | 0,0625 | 0,25 | 1,000 |
| Celkem | 1,0000 | 0,9375 | 2,5125 |

## Interpretace:

Modus náhodné veličiny *X*, tedy hodnota s nejvyšší pravděpodobností nastoupení, je 0.

Střední hodnota náhodné veličiny *X* je 0,9375.

Rozptyl náhodné veličiny *X* je 1,633594.

# Příklad 2

## Zadání příkladu:

Spojitá náhodná veličina $X$ je popsána následující hustotou pravděpodobnosti:

$$f(x) = 3x^2 \qquad 0 < x < 1$$
$$= 0 \qquad \textit{jinak}$$

Vypočtěte kvartily této náhodné veličiny.

## Vypracování příkladu:

Nejdříve je třeba stanovit distribuční funkci náhodné veličiny $X$.

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} t^2 dt = x^3$$

$$F(x) = x^3 \qquad 0 < x < 1$$
$$= 0 \qquad \textit{jinak}$$

Kvartily stanovíme na základě následujícího vztahu:

$$F(x_p) = P(X \le x_p) = p$$

***První kvartil*** $(x_{0,25})$***:***

$$F(x_{0,25}) = P(X \le x_{0,25}) = 0,25$$

$$x_{0,25}^3 = 0,25$$

$$x_{0,25} = \sqrt[3]{0,25} \doteq 0,63$$

***Druhý kvartil*** $(x_{0,5})$***:***

$$F(x_{0,5}) = P(X \le x_{0,5}) = 0,5$$

$$x_{0,5}^3 = 0,5$$

$$x_{0,5} = \sqrt[3]{0,5} \doteq 0,794$$

***Třetí kvartil*** $(x_{0,5})$***:***

$$F(x_{0,75}) = P(X \le x_{0,75}) = 0,75$$

$$x_{0,75}^3 = 0,75$$

$$x_{0,75} = \sqrt[3]{0,75} \doteq 0,909$$

## Interpretace:

První (dolní) kvartil náhodné veličiny $X$ je 0,63, druhý (prostřední) kvartil, tedy medián, je 0,794, a třetí (horní) kvartil je 0,909.

# Příklad 1

## Zadání příkladu:

Pravděpodobnost, že spotřeba elektrické energie ve všední den sledovaného období přesáhne stanovenou normu je 0,3. Stanovte pravděpodobnost, že během pěti všedních dnů sledovaného období:

a) nebude norma překročena
b) bude norma překročena maximálně dvakrát
c) bude norma překročena nejméně čtyřikrát.

## Vypracování příkladu:

Jedná se o nezávislé náhodné pokusy, přičemž známe pravděpodobnost nastoupení sledovaného jevu (překročení normy) v jednom pokusu (spotřeba elektrické energie ve všední den sledovaného období).

Vhodným modelem je binomické rozdělení $Bi(n;\pi)$:

- $n$… počet nezávislých pokusů
- NV $X$… spotřeba elektrické energie
- jev $A$… překročení normy; $P(A)=0{,}3$

Parametry rozdělení:

$n = 5$

$\pi = 0{,}3$

Pravděpodobnostní funkce:

$$P(x)=\binom{n}{x}\pi^{x}(1-\pi)^{n-x}$$

a) $P(X=0)=P(0)=\binom{5}{0}0{,}3^{0}(1-0{,}3)^{5-0}\doteq 0{,}1681$

b) $P(X\le 2)=P(0)+P(1)+P(2)=\binom{5}{0}0{,}3^{0}(1-0{,}3)^{5-0}+\binom{5}{1}0{,}3^{1}(1-0{,}3)^{5-1}+$

$+\binom{5}{2}0{,}3^{2}(1-0{,}3)^{5-2}\doteq 0{,}8369$

c) $P(X\ge 4)=P(4)+P(5)=\binom{5}{4}0{,}3^{4}(1-0{,}3)^{5-4}+\binom{5}{5}0{,}3^{5}(1-0{,}3)^{5-5}\doteq 0{,}0308$

## Interpretace:

a) Pravděpodobnost, že během pěti všedních dnů sledovaného období nebude překročena norma, je 16,81 %. Tato pravděpodobnost je poměrně malá.

b) Pravděpodobnost, že během pěti všedních dnů sledovaného období bude norma překročena maximálně dvakrát, je 83,69 %. Tato pravděpodobnost je značně velká.

c) Pravděpodobnost, že během pěti všedních dnů sledovaného období bude norma překročena nejméně čtyřikrát, je 3,08 %. Tato pravděpodobnost je velmi malá.

## STATGRAPHICS:

Desribe – Distribution Fitting – Probability Distributions

V panelu Probability Distributions zaškrtneme Binomial.

V panelu Binomial Options zadáme oba parametry rozdělení.

V nabídce Tables and Graphs je třeba zaškrtnout položku Cumulative Distributions.

V proceduře Cumulative Distributions zvolíme nabídku Pane Options, ve které zadáme požadované hodnoty NV, jejichž pravděpodobnosti budeme počítat, tedy 0, 2 a 4.

**Cumulative Distribution**
Distribution: Binomial

Lower Tail Area (<)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0 | 0,0 | | | | |
| 2 | **0,52822** | | | | |
| 4 | 0,96922 | | | | |

Probability Mass (=)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0 | **0,16807** | | | | |
| 2 | **0,3087** | | | | |
| 4 | **0,02835** | | | | |

Upper Tail Area (>)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0 | 0,83193 | | | | |
| 2 | 0,16308 | | | | |
| 4 | **0,0024295** | | | | |

a) $P(X=0)=0{,}16807$

b) $P(X\le 2)=P(X<2)+P(X=2)=0{,}52822+0{,}3087=0{,}83692$

c) $P(X\ge 4)=P(X>4)+P(X=4)=0{,}0024295+0{,}02835=0{,}0307795$

## EXCEL:

Vzorce – Další funkce – Statistická

Zvolíme funkci BINOMDIST.

V panelu Argumenty funkce zadáme do jednotlivých řádků:

***Úspěch:*** hodnotu NV $X$

***Pokusy:*** počet nezávislých náhodných pokusů $n$

***Prst_úspěchu:*** parametr $\pi$

***Počet:*** NEPRAVDA pro pravděpodobnostní funkci

a) BINOMDIST(3; 5; 0,3; NEPRAVDA)

$P(X = 0) = 0{,}16807$

b) BINOMDIST(2; 5; 0,3; PRAVDA)

$P(X \leq 2) = 0{,}83692$

c) BINOMDIST(3; 5; 0,3; PRAVDA)

$P(X \geq 4) = 1 - P(X \leq 3) = 0{,}0307795$

# Příklad 2

## Zadání příkladu:

Bylo zjištěno, že na 1 000 m tkaniny je v průměru 5 kazů. Pro expedici jsou připraveny stometrové balíky tkaniny. Stanovte pravděpodobnost, že náhodně vybraný balík:

a) bude bez kazu
b) bude mít méně než tři kazy
c) bude mít nejméně dva kazy.

## Vypracování příkladu:

Náhodná veličina $X$ je počet kazů na 100 m tkaniny. Vhodným modelem pro tuto NV je Poissonovo rozdělení $Po(\lambda)$.

Stanovení parametru $\lambda$:

Na 1000 m tkaniny je v průměru 5 kazů. Na 100 m tkaniny je tedy v průměru 0,5 kazu.

$\lambda = 0{,}5$

Pravděpodobnostní funkce:

$$P(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$$

a) $\lambda = 0{,}5\,;\ x = 0$

$$P(X = 0) = P(0) = e^{-0{,}5} \cdot \frac{0{,}5^0}{0!} \doteq 0{,}6065$$

b) $\lambda = 0{,}5\,;\ x = 3$

$$P(X < 3) = P(0) + P(1) + P(2) = e^{-0{,}5} \cdot \frac{0{,}5^0}{0!} + e^{-0{,}5} \cdot \frac{0{,}5^1}{1!} +$$

$$e^{-0{,}5} \cdot \frac{0{,}5^2}{2!} == 0{,}6065307 + 0{,}3032653 + 0{,}0758163 \doteq 0{,}9856$$

c) $\lambda = 0{,}5\,;\ x = 2$

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - [P(0) + P(1)] = 1 - (0{,}6065307 + 0{,}3032653) \doteq 0{,}0902$$

## Interpretace:

a) Pravděpodobnost, že náhodně vybraný balík bude vzkazu, je 60,65 %. Tato pravděpodobnost je středně velká.

b) Pravděpodobnost, že náhodně vybraný balík bude mít méně než tři kazy, je 98,56 %. Tato pravděpodobnost je značně velká.

c) Pravděpodobnost, že náhodně vybraný balík bude mít nejméně dva kazy, je 9,02 %. Tato pravděpodobnost je malá.

## STATGRAPHICS:

Desribe – Distribution Fitting – Probability Distributions

V panelu Probability Distributions zaškrtneme Poisson.

V panelu Poisson Options zadáme parametr rozdělení.

V nabídce Tables and Graphs je třeba zaškrtnout položku Cumulative Distributions.

V proceduře Cumulative Distributions zvolíme nabídku Pane Options, ve které zadáme požadované hodnoty NV, jejichž pravděpodobnosti budeme počítat, tedy 0, 2, a 3.

**Cumulative Distribution**
Distribution: Poisson

Lower Tail Area (<)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0 | 0,0 | | | | |
| 3 | **0,985612** | | | | |
| 2 | 0,909796 | | | | |

Probability Mass (=)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0 | **0,606531** | | | | |
| 3 | 0,0126361 | | | | |
| 2 | **0,0758164** | | | | |

Upper Tail Area (>)

| *Variable* | *Dist. 1* | *Dist. 2* | *Dist. 3* | *Dist. 4* | *Dist. 5* |
|---|---|---|---|---|---|
| 0 | 0,393469 | | | | |
| 3 | 0,00175154 | | | | |
| 2 | **0,0143876** | | | | |

a) $P(X=0)=0{,}606531$

b) $P(X<3)=0{,}985612$

c) $P(X\geq 2)=P(X=2)+P(X>2)=0{,}0758164+0{,}0143876\doteq 0{,}0902$

## EXCEL:

Vzorce – Další funkce – Statistická

Zvolíme funkci POISSON.

V panelu Argumenty funkce zadáme do jednotlivých řádků:

***X:*** hodnotu NV $X$

***Střední:*** střední hodnota $\lambda$

***Součet:*** PRAVDA pro distribuční funkci

a) POISSON(0; 0,5; NEPRAVDA)
$P(X = 0) = 0{,}606531$

b) POISSON(2; 0,5; PRAVDA)
$P(X < 3) = P(X \leq 2) = 0{,}985612$

c) POISSON(1; 0,5; PRAVDA)
$P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0{,}909796 = 0{,}090204$

**Zadání příkladu:**

Velkoobchod dostává balíčky sušenek, o jejichž hmotnosti předpokládáme, že mám normální rozdělení s neznámou střední hodnotou $\mu$ a směrodatnou odchylkou $\sigma$ = 5g.

Z velké dodávky bylo náhodně vybráno 30 balíčků ke kontrole. Jejich průměrná hmotnost je 150.4 g.

Určete, v jakých mezích lze očekávat střední hmotnost balíčku se spolehlivostí 95 %?

Řešení:

$\sigma$ = 5  $\bar{x}$ = 150,4  1-α = 0,95  n = 30

$$P\left(\bar{x} - u_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + u_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(150{,}4 - u_{1-\frac{0{,}05}{2}} * \frac{5}{\sqrt{30}} < \mu < 150{,}4 + u_{1-\frac{0{,}05}{2}} * \frac{5}{\sqrt{30}}\right) = 1 - 0{,}05$$

Compute variable ➔ IDF.NORMAL(0.975,0,1)

$$P\left(150{,}4 - 1{,}96 * \frac{5}{\sqrt{30}} < \mu < 150{,}4 + 1{,}96 * \frac{5}{\sqrt{30}}\right) = 0{,}95 \text{ ➔ } P(148{,}61 < \mu < 152{,}19) = 0{,}95$$

Se spolehlivostí 95 % lze očekávat střední hodnotu balíčku v mezích 148,61 a 152,19 gramů.

**Zadání příkladu**

Z časového snímku určité montážní operace vyplývá, že bylo provedeno 12 pozorování a zjištěny tyto údaje:

47,9 38,4 45,2 46,1 43,5 39,3 49,7 48,3 41,9 43,8 37,2 46,7

Sestrojte 90% oboustranný interval spolehlivosti pro očekávánou délku montáže, předpokládáme normalitu rozdělení délky montáže

Poznámka: Očekáváná hodnota = střední hodnota!!!

Řešení:

Malý výběr s neznámým rozptylem $\sigma^2$

$$P\left(\bar{x} - t_{1-\frac{\alpha}{2}}(n-1) * \frac{s_x}{\sqrt{n}} < \mu < \bar{x} - t_{1-\frac{\alpha}{2}}(n-1) * \frac{s_x}{\sqrt{n}}\right) = 1 - \alpha$$

Směrodatná odchylka a průměr ➔ analyze ➔ descriptive ➔ frequencies ➔ std. Deviation + mean

$s_x$ = 4,08679 $\bar{x}$ = 44

kvantil studenta ➔ compute variable ➔ IDF.T(0.95,11) = 1,8

$$P\left(44 - 1{,}8 * \frac{4{,}08679}{\sqrt{12}} < \mu < 44 + 1{,}8 * \frac{4{,}08679}{\sqrt{12}}\right) = 0{,}9 \text{ ➔ } P(41{,}88 < \mu < 46{,}12) = 0{,}9$$

Dovětek: jaká by byla maximální očekáváná délka montáže se spolehlivostí 0,99

$$P\left(\mu < \bar{x} - t_{1-\alpha}(n-1) * \frac{s_x}{\sqrt{n}}\right) = 1 - \alpha \text{ ➔ } P\left(\mu < 44 - t_{1-0{,}01}(12-1) * \frac{4{,}08679}{\sqrt{12}}\right) = 1 - 0{,}01 \text{ ➔}$$

$$P\left(\mu < 44 - 2{,}72 * \frac{4{,}08679}{\sqrt{12}}\right) = 0{,}99 \text{ ➔ } P(\mu < 40{,}79) = 0{,}99$$

Se spolehlivostí 99 % bude maximální délka montáže rovna 40,79 vteřin

**Zadání příkladu:**

U 200 náhodně vybraných výrobků byla zjištěna průměrná spotřeba materiálu na 1 výrobek $\bar{x}$ = 150 a rozptyl $s'^2$ = 25. Víme, že spotřeba materiálu se řídí normálním rozdělením. Určete, v jakých mezích lze očekávat střední spotřebu materiálu na 1 výrobek se spolehlivostí 95 %.

Řešení:

n = 200 $\bar{x}$ = 150 $s'^2$ = 25 1-α = 0,95

$$P\left(\bar{x} - u_{1-\frac{\alpha}{2}} * \frac{s'}{\sqrt{n}} < \mu < \bar{x} + u_{1-\frac{\alpha}{2}} * \frac{s'}{\sqrt{n}}\right) = 1 - \alpha \rightarrow$$

$$P\left(150 - u_{1-\frac{0,05}{2}} * \frac{5}{\sqrt{200}} < \mu < 150 + u_{1-\frac{0,05}{2}} * \frac{5}{\sqrt{200}}\right) = 1 - 0,05 \rightarrow$$

$$P\left(150 - 1,96 * \frac{5}{\sqrt{200}} < \mu < 150 + 1,96 * \frac{5}{\sqrt{200}}\right) = 0,95 \rightarrow P(149,31 < \mu < 150,69) = 0,95$$

Se spolehlivostí 95 % bude střední spotřeba materiálu v mezích 149,31 a 150,69

Pokračování příkladu:

S 95 % spolehlivostí odhadněte, v jakých mezích se bude pohybovat **celková** spotřeba materiálu, jestliže se předpokládá, že se následujícím období se vyrobí 2000 ks tohoto výrobku

$$P(149,31 * \mathbf{2000} < \mu < 150,69 * \mathbf{2000}) = 0,95 \rightarrow P(298\,620 < \mu < 301\,380) = 0,95$$

Celková spotřeba materiálu se bude, se spolehlivostí 95 %, v mezích 298 620 a 301 380 jednotkami.

**Zadání příkladu:**

Určitá firma má v úmyslu zavést donáškovou službu. Náhodně vybrala 100 domácností a zjistila, že 30 domácností by o tuto službu mělo zájem. Sestrojte 95% interval spolehlivosti pro minimální podíl zájemců o tuto službu.

Řešení:

m = 30  n = 100  1-α = 0,95  N = 100  M = 30

minimální = levostranný

podmínka:

- $n^* \pi^*(1-\pi) > 9$ ➔ 100*0,3*(1-0,3) > 9 ➔ 21 > 9 ➔ **můžeme použít levostranný odhad**

$$P\left(p - u_{1-\alpha} * \sqrt{\frac{p*(1-p)}{n}} < \pi\right) = 1 - \alpha \rightarrow \quad p = \frac{m}{n} = 0{,}3 \rightarrow P\left(0{,}3 - u_{0{,}95} * \sqrt{\frac{0{,}3*(1-0{,}3)}{100}} < \pi\right) = 0{,}95 \rightarrow$$

$$P\left(0{,}3 - 1{,}64 * \sqrt{\frac{0{,}3*(1-0{,}3)}{100}} < \pi\right) = 0{,}95 \rightarrow P(0{,}224846 < \pi) = 0{,}95$$

Se spolehlivostí 95 % je minimální podíl zájemců o tuto službu 22,4846 %

Pokračování:

Jaký by byl se spolehlivostí 95 % minimální domácností, které mají zájem o donáškovou službu, jestliže předpokládáme dosah na 5000 domácností?

5000 * 0,224846 = 1124,23 ➔ 1125 domácností

**Zadání příkladu:**

Při průjezdu přes most bylo kontrolována rychlost jízdy 100 vozidel. Bylo zjištěno, že 24 vozidel překročilo nejvyšší povolenou rychlost. Dále bylo z výsledků vypočtena průměrná rychlost jízdy 65 km/h a výběrová směrodatná odchylka 7 km/h. Předpokládáme že rychlost jízdy má normální rozdělení.

Sestrojte 95% interval spolehlivosti:

a) Pro průměrnou rychlost vozidel přejíždějících most.
b) Pro podíl vozidel, překračujících na mostě nejvyšší povolenou rychlost.

N = 100 M = 24 n = 100 m = 24 $\bar{x}$ = 65 s‘ = 7 α = 0,05

Řešení pro a):

$$P\left(\bar{x} - u_{1-\frac{\alpha}{2}} * \frac{s'}{\sqrt{n}} < \mu < \bar{x} + u_{1-\frac{\alpha}{2}} * \frac{s'}{\sqrt{n}}\right) = 1 - \alpha \rightarrow$$

$$P\left(65 - u_{0,975} * \frac{7}{\sqrt{100}} < \mu < 65 + u_{0,975} * \frac{7}{\sqrt{100}}\right) = 0,95 \rightarrow u_{0,975} = 1,96$$

$P(63,628 < \mu < 66,372) = 0,95$ ➔ se spolehlivostí 95 % bude průměrná rychlost vozidel mezi….

Řešení pro b):

π = p = 24/100 = 0,24

*n* π*(1 – π) > 9* ➔ 18,24 > 9 ➔ podmínka splněna

$$P\left(p - u_{1-\frac{\alpha}{2}} * \sqrt{\frac{p*(1-p)}{n}} < \pi < p + u_{1-\frac{\alpha}{2}} * \sqrt{\frac{p*(1-p)}{n}}\right) = 1 - \alpha \rightarrow$$

$$P\left(0,24 - 1,96 * \sqrt{\frac{0,24*(1-0,24)}{100}} < \pi < 0,24 + 1,96 * \sqrt{\frac{0,24*(1-0,24)}{100}}\right) = 0,95 \rightarrow$$

$P(0,156291 < \pi < 0,323708) = 0,95$ ➔ se spolehlivostí 95 % můžeme konstatovat že podíl vozidel, které na mostě překročí maximální povolenou rychlost se bude pohybovat v intervalu 15,63 % a 32,37 %

**Zadání příkladu:**

Při sociologickém průzkumu je třeba odhadnou průměrnou výši úspor dvojic uzavírajících manželství. Požadujeme spolehlivost 95 %, maximální přípustná chyba odhadu je 200 Kč. Máme informaci o variabilitě úspor, která říká, že směrodatná odchylka úspor činí 2500 kč.

Stanovte potřebný rozsah výběru pro provedení takového průzkumu.

α = 0,05 Δ = 200 $\sigma$ = 2500

řešení:

$n \geq \frac{u^2_{1-\frac{\alpha}{2}} * \sigma^2}{\Delta^2}$ ➔ $n \geq \frac{1{,}96^2 * 2500^2}{200^2}$ ➔ $n \geq$ 600,25 ➔ $n \geq 601$

K zajištění požadované přesnosti a spolehlivosti je potřebný rozsah výběru roven 601 dvojicím uzavírající manželství

**Zadání příkladu**

Určete rozsah výběru osob, kterého by bylo zapotřebí ke zjištění – se spolehlivostí 0,95 a chybou 0,02 – podílu kuřáků, předpokládá-li se, že tento podíl bude okolo 30 %

Řešení:

α = 0,05 Δ = 0,02 p = 0,3

$n \geq \frac{u^2_{1-\frac{\alpha}{2}} * p * (1-p)}{\Delta^2}$ ➔ $n \geq \frac{1{,}96^2 * 0{,}3 * (1-0{,}3)}{0{,}2^2}$ ➔ $n \geq 2016{,}84$ ➔ $n \geq 2017$

Od minimálně 2017 respondentů musíme získat odpovědi pokud požadujeme zmíněnou přesnost a spolehlivost

**zadání příkladu**

sestrojte oboustranný 95% interval spolehlivosti pro rozptyl normálně rozdělené náhodné veličiny na základě následujících údajů – jsou zadány tyto hodnoty:

42; 48; 60; 43; 36; 50; 52; 38; 56; 45

Řešení:

$\alpha = 0{,}05 \quad n = 10 \quad s_x^2 = 59{,}11$

$$P\left(\frac{(n-1)s_x^2}{\chi^2_{1-\frac{a}{2}}(n-1)} < \sigma^2 < \frac{(n-1)s_x^2}{\chi^2_{\frac{a}{2}}(n-1)}\right) = 0{,}95 \rightarrow$$

$$P\left(\frac{(10-1)*59{,}11}{\chi^2_{0{,}975}(10-1)} < \sigma^2 < \frac{(10-1)*59{,}11}{\chi^2_{0{,}025}(10-1)}\right) = 0{,}95 \rightarrow$$

$$P\left(\frac{(10-1)*59{,}11}{19{,}02} < \sigma^2 < \frac{(10-1)*59{,}11}{2{,}7}\right) = 0{,}95 \rightarrow \quad P(27{,}97 < \sigma^2 < 197{,}03) = 0{,}95$$

Se spolehlivostí 95 % bude rozptyl normálně rozdělené náhodné veličiny podle zmíněných údajů v intervalu mezi 27,97 a 197,03

**Zadání příkladu**

Pevnost proužků textilie je náhodná veličina s normálním rozdělením. Tato pevnost byla změřena u 25 náhodně vybraných vzorků. Z tohoto výběru byly vypočteny tyto výběrové charakteristiky:

$\bar{x}$ = 85,1  s‘ = 3,25

Určete se spolehlivostí 99 %, v jakých mezích se bude pohybovat rozptyl a směrodatná odchylka pevnosti.

Řešení:

n = 25  $\bar{x}$ = 85,1  s‘ = 3,25  α = 0,01

$$P\left(\frac{(n-1)s_x^2}{\chi^2_{1-\frac{a}{2}}(n-1)} < \sigma^2 < \frac{(n-1)s_x^2}{\chi^2_{\frac{a}{2}}(n-1)}\right) = 0{,}99 \rightarrow$$

$$P\left(\frac{(25-1)*3{,}25^2}{\chi^2_{0{,}995}(25-1)} < \sigma^2 < \frac{(25-1)*3{,}25^2}{\chi^2_{0{,}005}(25-1)}\right) = 0{,}99 \rightarrow$$

$$P\left(\frac{(25-1)*3{,}25^2}{45{,}56} < \sigma^2 < \frac{(25-1)*3{,}25^2}{9{,}89}\right) = 0{,}99 \rightarrow \; P(5{,}5641 < \sigma^2 < 25{,}6320) = 0{,}99$$

$P(2{,}3588 < \sigma < 5{,}0628) = 0{,}99$

**Zadání příkladu**

Pevnost proužků textilie je náhodná veličina s normálním rozdělením. Tato pevnost byla změřena u 20 náhodně vybraných vzorků. Z tohoto výběru byly vypočteny tyto výběrové charakteristiky:

$\bar{x}$ = 85,1 s‘ = 3,25

Testujte na hladině významnosti 5 %, zda je průměrná pevnost proužků textilie 88 jednotek.

Řešení:

n = 20 $\bar{x}$ = 85,1 s‘ = 3,25 α = 0,05

1) hypotézy

$H_0$: $\mu = 88$ $H_1$ : $\mu \neq 88$

2) testové kritérium

$t = \frac{\bar{x}-\mu_0}{\frac{s'}{\sqrt{n}}}$ ➔ $t = \frac{85,1-88}{\frac{3,25}{\sqrt{20}}}$ ➔ t = -3,99

3) kritický obor

$W \equiv \left\{t; t \leq t_{\frac{\alpha}{2}}(n-1) \cup t \geq t_{1-\frac{\alpha}{2}}(n-1)\right\}$ ➔ $W \equiv \left\{t; t \leq t_{0,025}(20-1) \cup t \geq t_{0,975}(20-1)\right\}$ ➔

$W \equiv \{t; t \leq -2,09 \cup t \geq 2,09\}$ ➔ $t \in W$ ➔ zamítáme $H_0$ a přijímáme $H_1$

Na hladině významnosti 5 % jsme prokázali, že průměrná pevnosti proužků textilie není 88 jednotek

**zadání příkladu**

máme výběrový soubor o rozsahu 20 jednotek, jehož prvky jsou plochy o rozměru 1 $m^2$, na nichž se zjišťuje hmotnost nesklizeného zrní v gramech. Hmotnost nesklizeného zrní je náhodná veličina, která má normální rozdělení. Posuďte na hladině významnosti 5 %, zda je střední hmotnost nesklizeného zrní nižší než 7 g na $m^2$.

data v gramech:

3,2 8,7 6,5 5,6 6,1 9,5 5,5 6,3 8,0 6,2 5,5 6,2 3,2 8,5 5,6 6,8 6,2 6,6 7,8 6,0

Řešení příkladu:

n = 20 α = 0,05 $s'^2$ = 2,516 $\mu_0$ = 7 $\bar{x}$ = 6,4 malý rozsah, neznáme rozptyl

1) hypotézy

$H_0$: $\mu = \mu_0$ $H_1$: $\mu < \mu_0$

2) testové kritérium

$t = \frac{\bar{x} - \mu_0}{\frac{s'}{\sqrt{n}}}$ ➔ $t = \frac{6{,}4 - 7}{\frac{1{,}5862}{\sqrt{20}}}$ ➔ t = -1,6916

3) kritický obor

$W \equiv \{t; t \leq t_\alpha(n-1)\}$ ➔ $W \equiv \{t; t \leq t_{0{,}05}(20-1)\}$ ➔ $W \equiv \{t; t \leq -1{,}73\}$ ➔

$t \notin W$ ➔ nezamítáme $H_0$ a nepřijímáme $H_1$

Na hladině významnosti 5 % jsme neprokázali, že by byla střední hmotnost nesklizeného zrní nižší než 7 g.

**Zadání příkladu**

Výrobce stanovil záruční dobu výrobku na dva roky, přičemž předpokládal, že během záruční doby bude nutno opravovat 15 % výrobků. Na hladině významnosti 1 % ověřte, zda výrobce vycházel ze správného předpokladu, jestliže z 900 prodaných výrobků bylo reklamováno 150.

Řešení příkladu:

$\pi = 0{,}15 \quad \alpha = 0{,}01 \quad n = 900 \quad m = 150 \quad p = \frac{1}{6}$

$n^* \pi^*(1 - \pi) > 9$ ➔ 114,75 > 9

1) hypotézy

$H_0$: $\pi = 0{,}15$ $H_1$: non $H_0$

2) testové kritérium

$$U = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 * (1 - \pi_0)}{n}}} \rightarrow U = \frac{\frac{1}{6} - 0{,}15}{\sqrt{\frac{0{,}15 * (1 - 0{,}15)}{900}}} \rightarrow U = 1{,}4$$

3) kritický obor

$$W \equiv \left\{u; u \leq u_{\frac{\alpha}{2}} \cup u \geq u_{1-\frac{\alpha}{2}}\right\} \rightarrow W \equiv \left\{u; u \leq u_{0{,}005} \cup u \geq u_{0{,}995}\right\} \rightarrow$$

$W \equiv \{u; u \leq -2{,}58 \cup u \geq 2{,}58\}$ ➔ $U \notin W$ ➔ nezamítáme $H_0$ a nepřijímáme $H_1$

Na hladině významnosti 1 % nezamítáme předpoklad o tom, že podíl opravovaných výrobků bude 15 %.

**Zadání příkladu**

Zácvik laboranta na optickém přístroji považujeme za ukončený, jestliže při opakovaném měření určitého objektu dosahuje směrodatné odchylky menší než 0,14 mm. Posuďte na hladině významnosti 5 %, zda můžeme zácvik laboranta považovat za ukončený, jestliže výsledky jeho měření jsou následující:

6,42 6,44 6,38 6,21 6,38 6,6 6,51

Předpokládáme normalitu rozdělení měření

Řešení:

n = 7 $\sigma = 0{,}14$ $\sigma^2 = 0{,}0196$ $\alpha = 0{,}05$ $s_x^2 = 0{,}015$

1) hypotézy

$H_0$: $\sigma^2 = 0{,}14^2$ $H_1$: $\sigma^2 < 0{,}14^2$

2) testové kritérium

$$\chi^2 = \frac{(n-1)s_x^2}{\sigma_0^2} \rightarrow \chi^2 = \frac{(7-1)*0{,}015}{0{,}0196} \rightarrow \chi^2 = 4{,}5918$$

3) kritický obor

$W \equiv \{\chi^2;\ \chi^2 \leq\ \chi_\alpha^2(n-1)\} \rightarrow W \equiv \{\chi^2;\ \chi^2 \leq\ \chi_{0{,}05}^2(7-1)\} \rightarrow W \equiv \{\chi^2;\ \chi^2 \leq\ 1{,}64\} \rightarrow$

$\chi^2 \notin W$ ➔ nezamítáme $H_0$ a nepřijímáme $H_1$

Na hladině významnosti 5 % jsme neprokázali, že by byl zácvik laboranta ukončený.

**Zadání příkladu:**

Pro porovnání rozptylů stipendií mezi studenty 2. a 4. ročníku VŠ byly provedeny náhodné výběry ve 2. ročníku 10 studentů a ve 4. ročníku 12 studentů. Posuďte na hladině významnosti 5 %, zda jsou rozptyly jejich stipendií v obou ročnících stejné. Předpokládáme normalitu rozdělení stipendií.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Stipendium 2. ročník | 2 600 | 0 | 0 | 3 200 | 0 | 1 260 | 4 400 | 4 400 | 0 | 1 520 | | |
| Stipendium 4. ročník | 0 | 3 600 | 2 400 | 9 800 | 0 | 2 400 | 0 | 500 | 0 | 500 | 2 400 | 500 |

Řešení:

$n_1 = 12$ $n_2 = 12$ $\alpha = 0{,}05$ $s_1'^2 = 3267951{,}111$ $s_2'^2 = 7848106{,}060606$

1) hypotézy

$H_0$: $\sigma_1^2 = \sigma_2^2$ $H_1$ : non $H_0$

2) testové kritérium

$$F = \frac{s_1'^2}{s_2'^2} \rightarrow F = \frac{3267951{,}111}{7848106{,}060606} \rightarrow F = 0{,}4164$$

3) kritický obor

$$W \equiv \left\{F;\ F \leq F_{\frac{\alpha}{2}}(n_1 - 1;\ n_2 - 1) \cup F;\ F \geq F_{1-\frac{\alpha}{2}}(n_1 - 1;\ n_2 - 1)\right\} \rightarrow$$

$$W \equiv \left\{F;\ F \leq F_{\frac{0{,}05}{2}}(10 - 1;\ 12 - 1) \cup F;\ F \geq F_{1-\frac{0{,}05}{2}}(10 - 1;\ 12 - 1)\right\} \rightarrow$$

$W \equiv \{F;\ F \leq 0{,}26 \cup F;\ F \geq 3{,}59\}$ ➔ $F \notin W$ ➔ nezamítáme $H_0$ a nepřijímáme $H_1$

Na hladině významnosti 5 % nezamítáme předpoklad o tom, že by rozptyly stipendií v obou třídách byly stejné

**Zadání příkladu**

Pro určení rozdílů v průměrné spotřebě benzínu u 2 typů automobilů bylo náhodně vybráno 10 automobilů od každého typu a při rychlosti 90 km/h byly naměřeny tyto hodnoty spotřeby (l/100 km)

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Typ 1 | 6,2 | 7,3 | 6,3 | 5,5 | 6,8 | 6,5 | 6,3 | 6,6 | 7,1 | 6,5 |
| Typ 2 | 5,7 | 5,0 | 5,3 | 5,6 | 6,1 | 5,3 | 5,8 | 5,7 | 5,4 | 5,5 |

Na hladině významnosti 5 % rozhodněte, zda je rozdíl v průměrné spotřebě obou typů automobilů statisticky významný. Předpokládáme normalitu rozdělení spotřeby benzínu.

Řešení příkladu:

$n_1 = 10$ $n_2 = 10$ $\alpha = 0,05$ testuje se shoda dvou průměrů, neznáme rozptyly, musíme udělat test shody rozptylů abychom si vybrali, která metoda z testová shody dvou průměrů je vhodná.

1a) hypotéza

$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$

1b) testové kritérium

$F = \frac{s_1'^2}{s_2'^2} \rightarrow F = \frac{0,252111}{0,096000} \rightarrow F = 2,6262$

1c) kritický obor

$W \equiv \left\{F;\ F \leq F_{\frac{\alpha}{2}}(n_1 - 1;\ n_2 - 1) \cup F;\ F \geq F_{1-\frac{\alpha}{2}}(n_1 - 1;\ n_2 - 1)\right\} \rightarrow$

$W \equiv \left\{F;\ F \leq F_{0,025}(10 - 1;\ 10 - 1) \cup F;\ F \geq F_{0,975}(10 - 1;\ 10 - 1)\right\} \rightarrow$

$W \equiv \{F;\ F \leq 0,25 \cup F;\ F \geq 4,03\} \rightarrow$

$F \notin W$ ➔ nezamítáme $H_0$ a nepřijímáme $H_1$ ➔ považujeme rozptyly za stejné

2a) hypotézy

$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$

2b) testové kritérium

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)s_1'^2 + (n_2-1)s_2'^2}{n_1+n_2-2}} * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightarrow$$

$s_1'^2 = 0,252111$ $s_2'^2 = 0,096$

$\bar{x}_1 = 6,51$ $\bar{x}_2 = 5,54$

$n_1 = 10$ $n_2 = 10$

$$t = \frac{6{,}51 - 5{,}54}{\sqrt{\frac{(10-1)0{,}252111 + (10-1)0{,}096}{10+10-2}} * \sqrt{\frac{1}{10} + \frac{1}{10}}}$$ ➔ t = 5,1989

2c) kritický obor

$W \equiv \left\{t;\ t \leq\ t_{\frac{\alpha}{2}}(n_1 +\ n_2 - 2) \cup t;\ t \geq\ t_{1-\frac{\alpha}{2}}(n_1 +\ n_2 - 2)\right\}$ ➔

$W \equiv \left\{t;\ t \leq\ t_{\frac{0{,}05}{2}}(18) \cup t;\ t \geq\ t_{1-\frac{0{,}05}{2}}(18)\right\}$ ➔ $W \equiv \left\{t;\ t \leq\ t_{\frac{0{,}05}{2}}(18) \cup t;\ t \geq\ t_{1-\frac{0{,}05}{2}}(18)\right\}$ ➔

$W \equiv \{t;\ t \leq\ -2{,}1 \cup t;\ t \geq\ 2{,}1\}$ ➔

$t \in W$ ➔ zamítáme $H_0$ a přijímáme $H_1$

Na hladině významnosti 5 % jsme prokázali, že existuje statisticky významný rozdíl v průměrné spotřebě dvou automobilů

**<u>Zadání stejné příkladu ale máme zadané testové charakteristiky</u>**

Z dat byly vypočteny výběrové průměry $\bar{x}_1$ = 6,51 a $\bar{x}_2$ = 5,54 a výběrové rozptyly $s'^2_1 = 0{,}2521$ a $s'^2_2$ = 0,096. Na hladině významnosti 5 % rozhodněte, zda je průměrná spotřeba druhého typu automobilu významně nižší než u prvního typu. Předpokládáme normalitu rozdělení spotřeby benzínu.

$n_1$ = 10 $n_2$ = 10 $\bar{x}_1$ = 6,51 $\bar{x}_2$ = 5,54 $s'^2_1 = 0{,}2521$ $s'^2_2$ = 0,096 α = 0,05

1a) hypotéza

$H_0$: $\sigma_1^2 = \sigma_2^2$ $H_1$: $\sigma_1^2 \neq \sigma_2^2$

1b) výpočet shodnosti rozptylů v SPSS

analyze ➔ compare means ➔ summary Independent-Samples T-Test ➔ vyplnit tabulku, podle dat, které máme zadané (standard deviation = směrodatná odchylka ➔ vypočítám ze zadaného výběrového rozptylu $s'^2_i$ = odmocníme oba $s'^2_i$ ) ➔

➔ vyleze nám tabulka a pod ní Hartleyho test ➔ rozhodují jestli jsou rozptyly shodné nebo ne ➔ F = 2,622 a Sig. = 0,0721 (Sig. = P-value) ➔ P-V > α ( 0,0721 > 0,05) ➔

➔ Sig. (P-value) > α ➔ platí $H_0$ ➔ nezamítáme $H_0$ a nepřijímáme $H_1$

Na hladině významnosti 5 % jsme prokázali, že rozptyly jsou shodné

2a) hypotéza o tom jestli je průměrná spotřeba druhého automobilu významně nižší než u druhého

$H_0$: $\mu_1 = \mu_2$ $H_1$: $\mu_1 < \mu_2$

2b) testové kritérium

Podle předešlého výpočtu v SPSS se $t$ = 5,199

2c) kritický obor – dle $H_1$

$W \equiv \{t;\ t \geq\ t_{1-\alpha}(n_1 +\ n_2 - 2)\}$ ➔ $\equiv \{t;\ t \geq\ 1{,}73\}$ ➔ $t \in W$ ➔ $H_0$ zamítáme a $H_1$ přijímáme

Na hladině významnosti 5 % jsme prokázali, že průměrná spotřeba druhého automobilu je významně nižší než průměrná spotřeba prvního automobilu.

**Zadání příkladu**

výrobek je dodáván ve třech provedeních – A, B, C. výrobce očekává zájem zákazníků ve struktuře:

- A = 46 %
- B = 32 %
- C = 22 %

Ověřte na hladině významnosti 5 %, zda je předpoklad výrobce správný a to na základě údajů o prodejích za první měsíc, kdy bylo prodání 600 ks výrobků ve struktuře, kterou uvádí následující tabulka:

| výrobek | Struktura prodeje |
|---|---|
| A | 280 |
| B | 190 |
| C | 130 |
| celkem | 600 |

Řešení:

1) hypotézy

$H_0$: $\pi_1 = 0{,}46$ $\pi_2 = 0{,}32$ $\pi_3 = 0{,}22$

$H_1$ : non $H_0$

2) testové kritérium

$$G = \sum_{i=1}^{k} \frac{\left(n_i - n\pi_{0\,;\,i}\right)^2}{n\pi_{0\,;\,i}} \rightarrow$$

| Výrobek | Struktura prodeje – pozorovaná (empirická) **($n_i$)** | Teoretické četnosti **($\pi_{0,i}$)** | Teoretická struktura prodeje **($n*\pi_{0,i}$)** | $\frac{\left(n_i - n\pi_{0\,;\,i}\right)^2}{n\pi_{0\,;\,i}}$ |
|---|---|---|---|---|
| A | 280 | 0,46 | 276 | 0,06 |
| B | 190 | 0,32 | 192 | 0,02 |
| C | 130 | 0,22 | 132 | 0,03 |
| celkem | 600 | 1,00 | 600 | **0,11** |

Podmínka: **$n*\pi_{0,i} \geq 5$ ➔ splněno**

G = 0,11

3) kritický obor (k = počet skupin)

$$W \equiv \{G; G \geq \chi^2_{1-\alpha}(k-1)\} \rightarrow W \equiv \{G; G \geq \chi^2_{1-0{,}05}(3-1)\} \rightarrow W \equiv \{G; G \geq 5{,}99\} \rightarrow$$

$G \notin W$ ➔ nezamítáme $H_0$ a nepřijímáme $H_1$

Na hladině významnosti 5 %, nezamítáme předpoklad o tom, že výrobce to odhadl správně

Řešení v SPSS:

1) Najebat tam zadaný sloupce

| VAR00001 | VAR00002 |
|---|---|
| 1 | 280,00 |
| 2 | 190,00 |
| 3 | 130,00 |

2) zvážení

Data ➔ weight cases ➔ weight cases by (cetnosti = druhý sloupec)

3a)

Analyze ➔ non parametric test ➔legacy dialogs ➔ chi-square ➔ test variable list (typ = první sloupec) + values (276 + 192 + 132) + exact (confidence level 95)

3b) vyhodnocení

Chi-Square = G = 0,109

Sig. (P-value) = 0,947 > α = 0,05 ➔ nezamítáme $H_0$ a $H_1$ nepřijímáme

**Zadání příkladu:**

Pracovníci na poště zpracovávají přehledy o zkontrolovaných poukázkách. Každý list přehledu nese záznam o 200 poukázkách. Výkon pracovníků se poměřuje počtem závad na zpracovaných listech. Pro normování práce je nutné získat představu o rozdělení listů podle počtu závad. Máme k dispozici údaje o 500 listech (viz tabulka).

Rozhodněte na hladině významnosti 1 %, zda by v této situaci vyhovovalo Poissonovo rozdělení se střední hodnotou $\lambda = 4$

| Počet závad ($x_i$) | Počet listů ($n_i$) |
|---|---|
| 0 | 11 |
| 1 | 37 |
| 2 | 65 |
| 3 | 100 |
| 4 | 103 |
| 5 | 80 |
| 6 | 49 |
| 7 | 33 |
| 8 | 10 |
| 9 | 8 |
| 10 | 3 |
| 11 a více | 1 |
| celkem | 500 |

Řešení:

1) hypotézy

$H_0$ : počet závad se řídí Po (4)

$H_1$ : non $H_0$

2a) testové kritérium

$$G = \sum_{i=1}^{k} \frac{\left(n_i - n\pi_{0;i}\right)^2}{n\pi_{0;i}} \rightarrow$$

| **Počet závad ($x_i$)** | **Počet listů ($n_i$)** | **$\pi_{0;i}$ = Po (4)** | $n * \pi_{0;i}$ | $\frac{\left(n_i - n\pi_{0;i}\right)^2}{n\pi_{0;i}}$ |
|---|---|---|---|---|
| 0 | 11 | 0,0183 | 9,15 | 0,37 |
| 1 | 37 | 0,0733 | 36,65 | 0,00 |
| 2 | 65 | 0,1465 | 73,25 | 0,93 |
| 3 | 100 | 0,1954 | 97,7 | 0,05 |
| 4 | 103 | 0,1954 | 97,7 | 0,29 |
| 5 | 80 | 0,1563 | 78,15 | 0,04 |
| 6 | 49 | 0,1042 | 52,1 | 0,18 |
| 7 | 33 | 0,0595 | 29,75 | 0,36 |
| 8 | 10 | 0,0298 | 14,9 | 1,61 |
| 9 | 8 | 0,0132 | 6,6 | 0,30 |
| 10 | 3 | 0,0053 | 2,65 | 0,05 |
| 11 a více | 1 | 0,0028 (součet P(11) + P(12) + P(13)) | 1,4 | 0,11 |
| **celkem** | **500** | **1,00** | **500,00** | **4,29** |

2b) Třetí sloupec:

- Buďto podle vzorce: $P(x) = e^{-\lambda} * \frac{\lambda^x}{x!}$
- Nebo: Transform ➔ compute variable ➔ PDF & noncentral PDF + Pdf. Poisson ➔ PDF. POISSON (první sloupec, 4)
- **Poslední řádek = 1 – suma předchozích pravděpodobností!!!!**

Podmínka: **$n*\pi_{0,i} \geq 5$ ➔ nesplněno**

- Druhá podmínka: ***$n\pi_{0,i}$* > 1 a alespoň 80 % tříd musí platit *$n\pi_{0,i}$* ≥ 5 ➔ splněno**

G = 4,29

3) kritický obor

$W \equiv \{G; G \geq \chi^2_{1-\alpha}(k-p-1)\}$ ➔ $W \equiv \{G; G \geq \chi^2_{0,99}(12-0-1)\}$ ➔

$W \equiv \{G; G \geq 24{,}725\}$ ➔ $G \notin W$ ➔ nezamítáme $H_0$ a nepřijímáme $H_1$

Na hladině významnosti 1 % nezamítáme předpoklad o tom, že se počet závad řídí Poissonovým rozdělením se střední hodnotou $\lambda = 4$.

Řešení v SPSS:

1) zadání tabulky + vážení

| VAR00001 | VAR00007 |
|---|---|
| ,00 | 11,00 |
| 1,00 | 37,00 |
| 2,00 | 65,00 |
| 3,00 | 100,00 |
| 4,00 | 103,00 |
| 5,00 | 80,00 |
| 6,00 | 49,00 |
| 7,00 | 33,00 |
| 8,00 | 10,00 |
| 9,00 | 8,00 |
| 10,00 | 3,00 |
| 11,00 | 1,00 |

Data ➔ weight cases by (druhý sloupec)

2) výpočet dalších dvou sloupců

| **$\pi_{0;i}$ = Po (4)** | $n * \pi_{0;i}$ |
|---|---|
| 0,0183 | 9,15 |
| 0,0733 | 36,65 |
| 0,1465 | 73,25 |
| 0,1954 | 97,7 |
| 0,1954 | 97,7 |
| 0,1563 | 78,15 |
| 0,1042 | 52,1 |
| 0,0595 | 29,75 |
| 0,0298 | 14,9 |
| 0,0132 | 6,6 |
| 0,0053 | 2,65 |
| 0,0028 (součet P(11) + P(12) + P(13)) | 1,4 |
| **1,00** | **500,00** |

3)

Analyze ➔ non parametric test ➔legacy dialogs ➔ chi-square ➔ test variable list (typ = první sloupec) + values (9,15 + 36,65 + 73,25 + …) + exact (confidence level 99)

- Chi-square = G = 4,3
- Sig. = 0,96 ➔ 0,96 > α (0,01) ➔ $H_0$ nezamítáme a $H_1$ nepřijímáme

Na hladině významnosti 1 % nezamítáme předpoklad o tom, že se počet závad na zpracovaných listech řídí Poissonovým rozdělením se střední hodnotou $\lambda = 4$

**Zadání příkladu**

Skupina hráčů, hrající pravidelně hru v kostky, pojala podezření, že jedna z hracích kostek není souměrná. Byla proto podrobena zkoušce 60 hodům s těmito výsledky:

| **Čísla** | $\mathbf{n_i}$ |
|---|---|
| 1 | 13 |
| 2 | 10 |
| 3 | 9 |
| 4 | 11 |
| 5 | 11 |
| 6 | 6 |
| **celkem** | **60** |

Ověřte na hladině významnosti 1 % hypotézu o souměrnosti kostky

Řešení:

1) hypotézy

$H_0$: P(x) = 1/6  $H_1$: non $H_0$

2) testové kritérium

$$G = \sum_{i=1}^{k} \frac{\left(n_i - n\pi_{0\,;\,i}\right)^2}{n\pi_{0\,;\,i}} \rightarrow$$

| **Čísla** | $\mathbf{n_i}$ | $\boldsymbol{\pi_{0,i}}$ | $\mathbf{n*\pi_{0,i}}$ | $\frac{\left(n_i - n\pi_{0\,;\,i}\right)^2}{n\pi_{0\,;\,i}}$ |
|---|---|---|---|---|
| 1 | 13 | 1/6 | 10 | 0,9 |
| 2 | 10 | 1/6 | 10 | 0 |
| 3 | 9 | 1/6 | 10 | 0,1 |
| 4 | 11 | 1/6 | 10 | 0,1 |
| 5 | 11 | 1/6 | 10 | 0,1 |
| 6 | 6 | 1/6 | 10 | 1,6 |
| **celkem** | **60** | **1,00** | **60** | **2,8** |

G = 2,8

3) kritický obor

$$W \equiv \{G; G \geq \chi^2_{1-\alpha}(k-1)\} \rightarrow W \equiv \{G; G \geq \chi^2_{1-0,01}(6-1)\} \rightarrow W \equiv \{G; G \geq 15{,}09\} \rightarrow$$

$G \notin W$ ➔ nezamítáme $H_0$ a nepřijímáme $H_1$

Na hladině významnosti 1 % nezamítáme předpoklad o tom, že kostka je souměrná

Řešení v SPSS:

1)

Zadám data z tabulky (první dva sloupce) ➔ data weight cases ➔ weight cases by (četnosti = druhý sloupec)

2)

Analyze ➔ non parametric test ➔ legacy ➔ chi-square ➔ test variable list (čísla 1 2 3 …) ➔ all categories equal ➔

G = 2,8

Sig. (P-value) = 0,731 ➔ 0,731 > $\alpha$ = 0,01 ➔ nezamítáme $H_0$ a nepřijímáme $H_1$

### *Statistický znak*

- označení (odraz) určité vlastnosti, kterou má v té či oné míře každá jednotka daného statisti
- u souboru osob např. věk, váha, výška, atd.
- ekvivalentní označení: statistická proměnná.

### *Hodnota statistického znaku ( = pozorování)*

- míra dané vlastnosti (statistického znaku) u každé jednotky statistického souboru.

! *Počet hodnot (pozorování) = rozsah souboru.*

### *Obměna ( = varianta) statistického znaku*

- hodnota ve smyslu vyjádření různého stupně dané vlastnosti.

! *Počet variant ≤ rozsah souboru.*

**Statistický znak shodný:** v daném statistickém souboru nabývá pouze jedné varianty.
**Statistický znak proměnný:** v daném statistickém souboru nabývá více než jedné varianty.
Ekvivalentní označení = **statistická proměnná.**

### *Tabulka rozdělení četností*

- Je to tabulka prostého rozdělení četností; uspořádání. vzestupně
- Takováto tabulka je výsledkem zpracování diskrétní proměnné s několika málo obměnami
- V případě zpracování diskrétní proměnné s mnoha obměnami nebo spojité proměnné není použitelná, pak je třeba sestavit tabulku intervalového rozdělení četností.

### *Statistické charakteristiky (míry)*

- shrnují informaci, obsaženou v datech (vyjadřují ji v koncentrované formě); charakterizují základní rysy zkoumaného souboru dat;
- umožňují porovnávání více souborů

1. charakteristiky polohy (úrovně),
2. charakteristiky variability,
3. charakteristiky šikmosti,
4. charakteristiky špičatosti.

$$\bar{x}_H \le \bar{x}_G \le \bar{x} \le \bar{x}_K$$

**1)** a) *charakteristiky, které jsou funkcí všech hodnot* – průměry (u všech prostý i vážený)

- **Aritm. Průměr** –použití: tam, kde má smysl součet hodnot proměnné
- **Harmonic. Průměr** –použití: tam, kde má smysl součet převrácených hodnot – zaměstnanci plní úkoly současně
- **Geometrický průměr** - použití: … součin hodnot proměnné – průměrný koef. růstu v časových řadách
- **Kvadratický průměr** - … součin čtverců hodnot proměnné – hodnoty jsou již odchylky od pův. hodnot

b) modus, kvantily (median, tercil, kvantil, kvintily, sextily, septily, oktávily, notily, decily, percentily)

**2)** a) *míry absolutní variabl.*- var. , kvantilové rozpětí; kvant. Odchylka, průměrná absolutní odchylka, rozptyl, směrodatná odchylka

b) *míry relativní variability* – variační koeficient

…3), 4)

### *Náhodná veličina*

= veličina, jejíž hodnota je jednoznačně určena výsledkem náhodného pokusu;

- vlivem náhodných činitelů může nabýt různých hodnot, proto nelze její konkrétní hodnotu před provedením náhodného pokusu jednoznačně určit př: počet bodů, které padnou na hrací kostce, počet poruch určitého zařízení, …

*Zákon rozdělení náhodné veličiny:* pravidlo, které každé hodnotě nebo množině hodnot z každého intervalu přiřazuje pravděpodobnost, že náhodná veličina nabude této hodnoty nebo hodnoty z určitého intervalu. Je to pravděpodobnostní model empirické náhodné veličiny.

Popis náhodné veličiny
1) Diskrétní náhodná veličina
2) Spojitá náhodná veličina

**2) Typ vztahů mezi obměnami a hodnotami proměnné**

- ***nominální (jmenné, názvové):*** slovní proměnné, jejichž obměny nelze hierarchicky uspořádat, tzn. že nelze jednoznačně stanovit, která je nižší, a která vyšší. O jejích obměnách lze pouze konstatovat, zda jdou stejné nebo různé. Např.: pohlaví, jméno, rodinný stav, atd.

- ***ordinální (pořadové):*** slovní i číselné proměnné, jejichž obměny lze jednoznačně seřadit od nejnižší k nejvyšší nebo obráceně. Např.: nejvyšší dokončené vzdělání, hodnosti v armádě, pořadí v soutěži, atd.

- ***metrické (měřitelné):*** vždy číselné, jsou udány v určitých měrných jednotkách – vyjadřují tedy velikost měřených vlastností. Nabývají jak kladných, tak nekladných hodnot. Lze změřit, o kolik je jedna obměna větší (resp. menší) než druhá. Obměny lze porovnávat rozdílem, někdy také podílem (ne vždy – pokud jsou některé obměny záporné či nulové, není to možné). Např.: teplota vzduchu, zisk podniku, atd.

***kardinální (stěžejní):*** ty metrické proměnné, které nabývají pouze kladných hodnot, jejich obměny lze porovnávat jak rozdílem, tak podílem. Je tedy možno změřit, o kolik měrných jednotek je jedna obměna větší (resp. menší) než druhá, a také kolikrát je jedna obměna větší (resp. menší) než druhá. Např.: věk, váha, výška, atd.

**3) Počet variant, kterých proměnné nabývají**

- ***alternativní:*** nabývají pouze dvou obměn;
- ***množné:*** nabývají více než dvou obměn.

.

**4) Počet hodnot, kterých proměnné nabývají**

- ***diskrétní (nespojité):*** nabývají spočetně mnoha hodnot z konečného či nekonečného intervalu. Např.: počet dětí v rodině, atd.

- ***spojité (kontinuální):*** nabývají všech hodnot z konečného či nekonečného intervalu. Např.: spotřeba elektrické energie, příjem, atd.

### *Operace s náhodnými jevy*

- vztahy mezi náhodnými jevy graficky znázorňují tzv. Vennovy diagramy;

| | | |
|---|---|---|
| 1. | $A \subset B$ | Jev A je částí jevu B; z jevu A plyne jev B (implikace); nastoupení jevu A má vždy za následek nastoupení jevu B. |
| 2. | $A = B$ | Jevy A a B jsou si rovny; $A \subset B$ a současně $B \subset A$. |
| 3. | $C = A \cap B$ | Jev C je průnik jevů A a B (logický součin); jev C nastane právě tehdy, nastane-li současně jev A i jev B. |
| 4. | $C = A \cup B$ | Jev C je sjednocení jevů A a B (logický součet); jev C nastane právě tehdy, nastane-li alespoň jeden z jevů A a B. |
| 5. | $C = A - B$ | Jev C je rozdíl jevů A a B; jev C nastane právě tehdy, když jev A a současně jev B nenastane. |
| 6. | E je jev jistý<br>Ø je jev nemožný | Jev, který musí nastat vždy.<br>Jev, který nastat nemůže. |

### *Kombinatorika*

- permutace, variace bez opakování, variace s opakováním, kombinace

### *Pravděpodobnost*

- matematická disciplína, která se zabývá studiem zákonitostí, jimiž se řídí hromadné náhodné jevy; vytváří pravděpodobnostní modely, pomocí nichž se snaží postihnout procesy, ovlivněné náhodou.

*Náhodné pokusy*: procesy, jejichž výsledek nelze předem jednoznačně určit (je nejistý); závisí jednak na daných podmínkách, při kterých je prováděn, jednak na náhodě

$$P(A) = \frac{m}{n} \quad ,$$

*Stat. definice:* Jestliže při rostoucím počtu opakování náhodného pokusu (n) relativní četnost $\frac{m}{n}$ kolísá ve stále užších mezích kolem určitého čísla, můžeme toto číslo považovat za pravděpodobnost jevu A.

Podmíněná pravděpodobnost

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad , \quad \text{pro } P(B) \neq 0 \qquad P(B|A) = \frac{P(A \cap B)}{P(A)} \quad , \quad \text{pro } P(A) \neq 0$$

Pravidlo o násobení pravděpodobností

1. jevy A a B jsou na sebe závislé

$$P(A \cap B) = P(B) \cdot P(A/B)$$
$$P(A \cap B) = P(A) \cdot P(B/A).$$

2. jevy A a B jsou nezávislé

$$P(A \cap B) = P(A) \cdot P(B)$$

3.

Pravidlo pro sčítání pravděpodobností

1. jevy A a B jsou slučitelné $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$P(A \cup B) = P(A) + P(B).$$

2. jevy A a B jsou neslučitelné

***Charakteristiky náhodných veličin***

- číselné hodnoty, jejichž cílem je koncentrovat (zestručnit) popis NV; výstižný popis základních vlastností rozdělení NV.

1. Charakteristiky polohy

*Střední hodnota E(X )* = očekávaná hodnota (z lat. expectatis).

*Modus x^*

a) Diskrétní NV = hodnota NV, která má největší pravděpodobnost výskytu (nejpravděpodobnější hodnota). x^.....max P(x)

b) Spojitá NV = bod, v němž je hustota pravděpodobnosti maximální, tj. lokální maximum hustoty pravděpodobnosti f(x). x^..... f ¢(x) = 0

*Kvantily* = používají se především kvantily SPOJITÉ náhodné veličiny.

2. Charakteristiky variability

*Rozptyl* D(x)

*Směrodatná odchylka* $\sigma$ (X )

***Matematiká statistika***

- Na základě výběrových dat usuzujeme na obecnější skutečnosti, týkající se základního souboru (dále ZS), tj.provádíme zevšeobecňující (induktivní) úsudek.

**Teorie odhadu**

1. Bodový odhad

- nahrazení neznámé hodnoty parametru ZS hodnotou vhodné výběrové charakteristiky, která bude sloužit jako dobrá náhrada neznámého parametru
- posuzujeme dle vlastností: 1. nevychýlenost (nestrannost, nezkreslenost),
  2. konzistence,
  3. vydatnost,
  4. výběrová charakteristika má být postačující.

2. Interval.odhad = odhad neznámého parametru ZS pomocí intervalu, který s pravděpodobností P = 1 – $\alpha$ bude obsahovat skutečnou hodnotu odhadované charakteristiky ZS $\theta$

spolehlivost odhadu: pravděpodobnost $0 < \alpha < 1$.

riziko odhadu = udává kolika případech ze 100 (v jakém % případů) nebude IS pokrývat odhadovaný parametr $\theta$

intervaly konstruovány jako: oboustranné, jednostranné (pravo-, levo-)

**Odhad parametru μ (relativní četnosti) alternativního rozdělení**

Je třeba mít k dispozici výběr dostatečně velkého rozsahu; to je zajištěno splněním podmínky: $n\,\pi\,(1-\pi) > 9$

1. Bodový odhad – Bodovým odhadem relativní četnosti $\pi$ je výběrová relativní četnost (výběrový podíl) p

2. Interval.odhad = oboustranný, pravostranný, levostranný

**Odhad parametru $\sigma^2$ (rozptylu) normálního rozdělení**

Je třeba mít k dispozici výběr dostatečně velkého rozsahu; to je zajištěno splněním podmínky: $n\,\pi\,(1-\pi) > 9$

1. Bodový odhad – Bodovým odhadem rozptylu $\sigma^2$ je výběrový rozptyl $s'^2_x$

2. Interval.odhad – Při konstrukci IS pro parametr $\sigma^2$ rozlišujeme 2 případy - buď známe parametr μ nebo ho neznáme. V praxi je častější případ, kdy parametr μ neznáme – oboustranný, pravostranný, levostranný

***Testování statistických hypotéz***

- Testování hypotéz je postup, sloužící k ověření předpokladů o ZS (hypotéz) na základě výběrových dat (tj. hodnot z výběrového souboru).
- Hypotéza = určitý předpoklad o základním souboru; tvrzení o neznámém parametru ZS (parametrické testy) nebo o různých vlastnostech ZS (neparametrické testy).
- Testování umožňuje rozhodnout, zda určitou hypotézu zamítneme (event. přijmeme), a to s malým, předem zvoleným rizikem.

Symbolika: $H_0$, $H_1$

***Testové kriterium (t):*** - vhodná charakteristika, která má při platnosti $H_0$ známé pravděpodobnostní rozdělení;
- prostor hodnot testového kriteria rozdělíme na dva disjunktní obory (W a V).

***Kritický obor (W):*** je tvořen hodnotami TK, které jsou při platnosti $H_0$ tak extrémní, že pravděpodobnost jejich výskytu je velmi malá. Jde tedy o oblast přijetí $H_1$.

***Obor přijetí (V):*** je tvořen těmi hodnotami TK, které svědčí ve prospěch $H_0$.

***Hladina významnosti (α)= pravděpodobnost chyby 1. druhu:*** pravděpodobnost že zamítneme $H_0$, ačkoli platí.

***Pravděpodobnost chyby 2. druhu (β):*** pravděpodobnost, že nezamítneme $H_0$, ačkoli neplatí.

***Síla testu (1- β):*** pravděpodobnost správného zamítnutí $H_0$ (schopnost testu zamítnout neplatnou $H_0$)

***Některá rozdělení náhodných veličin*** = pravděp. model chování náhodné veličiny

1. Diskrétní náh. veličiny

**Binomické rozdělení Bi (n; $\pi$)**

- NV X je počet výskytů náhodného jevu A v n nezávislých náhodných pokusech, je-li pravděpodobnost nastoupení jevu A ve všech pokusech stejná ($\pi$);
- rozdělení má 2 parametry :
  n ... počet nezávislých náhodných pokusů;
  π ... pravděpodobnost nastoupení sledovaného jevu v 1 pokusu.

př: počet „šestek“, které padnou při deseti hodech kostkou.

**Poissonovo rozdělení Po ($\lambda$)**

- NV X je počet výskytů náhodného jevu A v určitém časovém intervalu délky t (tzn. za jednotku času),
v jednotce plochy nebo objemu (v prostorové jednotce);
- Rozdělení má 1 parametr : $\lambda$ ... střední hodnota rozdělení.:

př: počet poruch stroje za směnu, počet telefonních hovorů za hodinu, počet vad na $m^2$ koberce.

Aproximace Binomického rozdělení rozdělením Poissonovým

Podmínky: Počet pokusů n musí být dostatečně velký (alespoň n > 30) a pravděpodobnost p velmi malá (alespoň $\pi \leq 0{,}1$).

Při aproximaci udává P(x) přibližnou pravděpodobnost, že ve velkém počtu n nezávislých náhodných pokusů se sledovaný jev A vyskytne x-krát, je-li pravděpodobnost výskytu jevu v jednom pokusu velmi malá.

př: počet vadných výrobků ve velké sérii, je-li pravděpodobnost výroby zmetku velmi malá

**Hypergeometrické rozdělení H (N; M; n)**

Používá se v případě závislých pokusů, tzn. při výběru bez vracení.

- NV X je počet vybraných prvků se sledovanou vlastností při závislých pokusech.
- má 3 parametry : N … rozsah souboru, z něhož vybíráme;
  M ... počet prvků v základním souboru, které mají sledovanou vl.
  n ... rozsah výběru ( = počet závislých pokusů).

př: při kontrole jakosti u malého počtu výrobků nebo v případě, kdy kontrola má ráz destrukční zkoušky (výrobek je zničen).

2. Rozdělení spojitých náhodných veličin

**Exponenciální rozdělení E (A; $\delta$)**

- NV X je doba čekání do nastoupení sledovaného jevu, může-li tento jev nastat v kterémkoli okamžiku.
- Parametr A = počáteční doba, během které sledovaný jev nastat nemůže.

př: doba čekání zákazníka na obsluhu v prodejně, doba realizace dvou po sobě jdoucích telefonních hovorů, doba životnosti zařízení, u nichž dochází k poruše z náhodných příčin (ne v důsledku opotřebení).

použití: V teorii spolehlivosti a životnosti, v teorii hromadné obsluhy (tzv. teorii front), v teorii obnovy.

**Normální rozdělení N ($\mu$; $\sigma^2$)**

- Je vhodné tam, kde kolísání NV je způsobeno velkým počtem nepatrných a vzájemně nezávislých vlivů.
- Klasickým typem veličin, které se řídí tímto rozdělením, jsou náhodné chyby.
- Pomocí N ($\mu$; $\sigma^2$) lze za jistých podmínek aproximovat řadu jiných rozdělení, a to i nespojitých.

**Normované normální rozdělení N (0;1)**

- Původní NV X, která má N (m; s2) normujeme, tzn. transformujeme na NV U, která má N (0; 1).
- Je tak zavedena normovaná veličina U, která má nulovou střední hodnotu a jednotkový rozptyl.
- Hodnoty distribuční funkce a kvantilůN (0; 1) je možno tabelovat.

***Rozdělení některých funkcí náhodných veličin***

**Rozdělení $\chi 2$ $\chi 2$ (ν)**

- NV $\chi 2$ je součtem ν nezávislých NV s normovaným normálním rozdělením.
- Rozdělení má 1parametr : ν ... počet stupňů volnosti.
- Kvantily jsou tabelovány pro ν = 1, 2, ... , 30 a pro vybrané pravděpodobnosti P.

**Studentovo rozdělení (t) t(ν)**

- NV t je podílem dvou nezávislých NV: NV U s rozdělením N (0; 1) a NV $\chi^2$ s rozdělením $\chi^2(\nu)$.
- Rozdělení má 1 parametr : ν ... počet stupňů volnosti.
- Kvantily jsou tabelovány pro ν = 1, 2, ... , 30 a pro vybrané pravděpodobnosti P.
- Používá se především pro výběry malého rozsahu (n < 30).
- Rozdělení je symetrické podle bodu t = 0, pro kvantily proto platí vztah $t_p = -t_{1-p}$.

***Některé neparametrické shody***

**χ2- test dobré shody**

= Slouží k ověření shody mezi teoretickým a empirickým rozdělením.
- Je použitelný jen v případě velkých výběrů.
- Předpokladem testu je možnost roztřídit výsledky náhodného výběru do určitého počtu (k) disjunktních tříd podle nějakého znaku.
- Nemáme-li k dispozici dostatečně velký výběr, použijeme místo tohoto testu Kolmogorovův-Smirnovův test

***Analýza rozptylu (jednofakorová. a více faktorová)***

**Jednofaktorová**

= Zkoumá, zda lze změny hodnot měřitelné proměnné y vysvětlovat faktorem x, který může být slovní i číselná proměnná.
použití: ověření významnosti rozdílu mezi výběrovými průměry většího počtu náhodných výběrů.
předpoklady: náh. výběry jsou nezávisléd; každý z výběrů má norm. rozdělení s neznám. střední hod. $\mu$ a neznámým rozptylem $\sigma^2$; stejné ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \ldots$); n > k; y je číselné, x může být čísůs. i slovní

***χ2 – Test nezávislosti v kontingenční tabulce***

**Kontingenční tabulka** = dvourozměrná tabulka, do které jsou uspořádány údaje o dvou slovních proměnných (příp. jedné proměnné slovní a druhé číselné)..
- sdružené (simultánní) četnosti – uvnitř tabulky
- okrajové (marginální) četnosti - na okrajích tabulky
předpoklad: všechna políčka jsou obsazena, $n'_{ij} \geq 5$

**Asociační tabulka** – čtyřpolní dvourozměrná tabulka, která obsahuje 2 slovní proměnné, které nabývají pouze 2 hodnot
používáme koeficient asociace $r_{AB}$ $\quad r_{AB} \in \langle -1,1 \rangle$

***Analýza závislostí (regresní a korelační analýza)***

=zkoumání závislosti dvou event. více proměnných, měření síly této závislosti atd.; cílem je hlubší vniknutí do podstaty sledovaných jevů a procesů, přiblížení k tzv. příčinným souvislostem
Korelační tabulka = dvourozměrná tabulka, ve které jsou uspořádány numerické proměnné

**Regresní analýza** – zkoumání jednostranné závislosti proměnné y (závislá, vysvětlovaná) na proměnné x (nezávislá, vysvětlující)
-nezávislá proměnná = příčina, závislá proměnná = důsledek;

***Postup regresní analýzy:***

1. Volba typu regresní funkce.
2. Odhad parametrů zvolené regresní funkce.
3. Testování hypotéz o parametrech regresní funkce.
4. Ověření vhodnosti zvoleného regresního modelu

(1) Regresní funkce = různé typy, nejčastěji přímka, každá má určitý počet parametrů (2) – neznámé konstanty, značíme ($\beta_0$, $\beta_1$,…), hodnoty odhadujeme z výběrových dat
funkce z lineár. hlediska p. = přímka, parabola, hyperbola, rovina
funkce z nelineár. hlediska p. = exponen., mocninná, Tőrnquistova
I. Jednoduchá lineární regrese – reg. funkce lineární, 1 proměnná x
Teoretická – nutno provést odhad neznámých parametrů,
Empirická – odhad parametrů se provádí metodou nejmenších čtverců = princip: parametry odhadujeme tak, aby pro ně byl minimální součet čtverců reziduí
1. Stanovíme parciální derivace a položíme je rovny 0.
2. Vznikne soustava dvou rovnic (tzv. normální rovnice).
3. Vyřešíme je a získáme vzorce pro výpočet 0 1 b a b .

II. Nelineární regrese - kde nejde odhadovat, využití různých metod častá je metoda linearizující transformace
(3) t- testy = dílčí testy o nulových hodnotách jednotlivých regresních parametrů.
(4) F-test = testujeme vhodnost modelu jako celku; analýza rozptylu.
kritéria pro posouzení:
**Index determinace** - ↑ hodnota $I^2$, udává, jaký podíl variability proměnné y lze vysvětlit zvolenou regresní funkcí (lze udávat i v %). $\quad I^2 \in \langle 0;1 \rangle$
- je zároveň mírou těsnosti závislosti proměnné y na proměnné x.

***Index korelace:*** $\quad I = \pm\sqrt{I^2}\,; \quad I \in \langle -1;1 \rangle$

**Test. krit. F** - ↑ hodnota F
**Reziduální součet čtv.** - ↓ hodnota $S_R$
**Reziduální rozptyl** - ↓ hodnota $s_R^2$

***Korelační analýza***

- zabývá se především intenzitou vzájemného vztahu proměnných;
- je základní metodou měření síly lineární závislosti číselných proměnných;

Sdružené regresní přímky
Míry těsnosti lineární závislosti:
Koeficient determinace: $r_{xy}{}^2 \in \langle 0;1 \rangle$
Koeficient korelace: $r_{xy}{}^2 \in \langle -1;1 \rangle$ ;
- měří sílu lineární závislosti, nikoli závislosti obecně (lineární závislost = korelovanost).

***Interpretace:***

1. znaménko +/– udává směr závislosti:
   $r_{xy} > 0 \Rightarrow$ přímá závislost
   $r_{xy} < 0 \Rightarrow$ nepřímá závislost
2. $|r_{xy}|$ udává sílu závislosti:
   $r_{xy} = 0 \Rightarrow$ lineární nezávislost
   $|r_{xy}| = 1 \Rightarrow$ funkční (úplná) závislost
   $|r_{xy}| \to 0 \Rightarrow$ slabá lineární závislost
   $|r_{xy}| \to 1 \Rightarrow$ silná lineární závislost

Pořadová korelace
- využití: o síle závislosti mezi 2 kvantitativními znaky nebo určit závislost mezi pořadími znaků
Spearmanův koeficient pořadové korelace:
- varianta korelačního koeficientu;
- měří sílu lineární závislosti dvou pořadí.

***ČASOVÉ ŘADY***

- posloupnost chronologicky uspořádaných pozorování; rozdělení:
a) dle časového hlediska zjišt. údajů:
intervalové – (př- počet rozvodů za rok) intervalový ukazatel = ukazatel, jehož velikost závisí na délce intervalu, za který je sledován. lze vytvořit součty (resp. průměry), je nutná stejná délka intervalů
okamžikové – (př. počet obyvatel k 31.12) okamžikový ukazatel = ukazatel, jehož hodnoty se vztahují ke konkrétnímu časovému okamžiku.
- součet nedává reálný smysl, průměr nelze stanovit běžným způsobem
- pro charakterizaci používáme průměr chronologický:
okamžiky stejně dlouhé – prostý chronolog. průměr
okamžiky různě dlouhé – vážený chronolog. průměr
b) dle periodicity – roční/dlouhodobé (rok a více), krátkodobé (čtvrtletní, měsíční, …)
c) dle způsobu vyjádření ukazatelů – naturálních a peněžních

Základní charakteristiky časových řad
*Absolutní diference* (přírůstky) – charakterizují absolutní změnu (přírůstek nebo úbytek) hodnoty ukazatele v časovém okamžiku t oproti období předcházejícímu (t-1).
*Průměrný absolutní přírůstek* – aritmetický průměr jednotlivých prvních diferencí
*Koeficient růstu* – udává, kolikrát vzrostla hodnota ČR v časovém okamžiku t proti období předcházejícímu.
*Průměrný koeficient růstu* – geometrický průměr jednotlivých koeficientů růstu

Základní koncepce modelování časových řad
cíl: analýza a prognóza vývoje, snaha pochopit na základě abstrakce mechanismus chování ČŘ
1) Jednorozměrný model –
**Klasický model** - zkoumá vliv časového faktoru na analyzovaný ukazatel.
- nezkoumá věcné příčiny dynamiky ČŘ.; - jeden z možných klasických přístupů je dekompozice ČŘ, tj. rozklad ČŘ na čtyři složky časového pohybu; popis jednotlivých forem pohybu

$T_t$ … ***trendová složka (trend)*** = dlouhodobá tendence ve vývoji hodnot analyzovaného ukazatele v čase.
$S_t$ … ***sezónní složka*** = pravidelně se opakující odchylka od trendu; periodicita maximálně 1 rok.
$C_t$ … ***cyklická složka*** = kolísání okolo trendu v důsledku dlouhodobého vývoje; periodicita delší než 1 rok.
$\varepsilon_t$ … ***náhodná složka*** = náhodné výkyvy, které nemají systematický charakter.

aditivní a multiplikativní
2) Vícerozměrný model – je založen na předpokladu, že vývoj analyzovaného ukazatele není ovlivněn pouze časovým faktorem, ale rovněž skupinou jiných, souvisejících ukazatelů.

Trendová analýza - - popis trendu (vyrovnání, vyhlazení) pomocí matematické (analytické, „hladké“) funkce = trendová funkce.
- tzv. modely s konstantními parametry.
- informace o charakteru hlavní tendence ve vývoji ukazatele
typy: Lineární – nejvíce používaný, Parabolický, Exponenciální, Modifikov. expon., Logistický

***Klouzavé průměry***
- adaptivní přístup k modelování trendu ČŘ., tzv. modely s proměnlivými parametry
- posloupnost empirických pozorování nahradíme řadou průměrů z těchto pozorování vypočtených
- při postupném výpočtu průměrů postupujeme (kloužeme) vždy o jedno pozorování kupředu, přičemž nejstarší (tj. první) pozorování z dané skupiny vypouštíme.

Klouzavá část období interpolace (m): časový interval délky m = 2 p +1, který se posunuje po časové ose vždy o jednotku.

**Prosté klouzavé průměry** – na částech o rozsahu m je definován lineární trend.
**Vážené klouzavé průměry** –.na částech o rozsahu m je definován parabolický trend
**Centrované klouzavé průměry** - v případě, že rozsah klouzavé části je číslo *sudé.*, třední body klouzavých částí již nejsou celá čísla, proto nelze přímo přiřadit hodnoty klouzavých průměrů k empirickým pozorováním ČŘ.

Popis sezónní složky:
= periodicky se opakující obousměrné odchylky hodnot ČŘ od trendu.
- délka periody je maximálně jeden rok.
- nejprve je třeba zjistit, zda ČŘ reálně vykazuje sezónní výkyvy
1. Model konstantní sezónnosti (aditivní model)
2. Model proporcionální sezónnosti (multiplikativní model)

Sezónní očišťování:
1. vyrovnání ČŘ klouzavými průměry vhodného typu (možno i jiným způsobem, např. trendovou funkcí).
2. výpočet sezónních faktorů (rozdílových nebo indexních).
3. očištění údajů původní ČŘ:
   a) adit.: od hodnot ČŘ odečteme příslušný rozdílový sezónní faktor.
   b) mult.: hodnoty ČŘ vydělíme příslušným indexním sezónním faktorem.

***Srovnávání hodnot statistických ukazatelů*** (zabývá se hospodářská statistika)
- cílem je nalézt způsoby měření ekonomické skutečnosti (ve formě ukazatelů) a jejího vyhodnocení (např. měření inflace, dynamiky produkce, vývoje kurzů akcií, atd.);
- ukazatele jsou veličiny, s nimiž se denně setkáváme (tisk, TV, rozhlas,…) – např. HDP, průměrná mzda,dovoz, vývoz, produktivita práce, atd.;

*Statistický ukazatel* = veličina, která kvantitativně popisuje určitou sociálně-ekonomickou hromadnou skutečnost; proměnná veličina

*Údaj* – konkrétní hodnota ukazatele; vzniká konkrétním vymezením času a prostoru

**Základní typy ukazatelů**

- členění ukazatelů lze provádět z mnoha různých hledisek, která se vzájemně mohou prolínat.

**1. Ukazatele primární:** - jsou přímo zjišťované, neodvozené.
- např. stav zásob, počet pracovníků k 31. 12., ….

**2. Ukazatele sekundární**: - jsou odvozené, jde o funkci ukazatelů primárních.
- např. časové průměry, produktivita práce na pracovníka, zisk, ….

NEBO

**1. Ukazatele absolutní:** vyjadřují velikost jevu bez vztahu k jinému jevu.

**2. Ukazatele relativní:** vyjadřují velikost jednoho jevu na měrnou jednotku jiného jevu.

NEBO

**1. Ukazatele okamžikové**

**2. Ukazatele intervalové**

NEBO

**1. Ukazatele extenzitní:** - měří extenzitu (množství, objem, rozsah) sledovaného jevu
- vždy absolutní čísla (získáme spočtením, změřením, zvážením)
- standardní symbolické značení **q a Q**.

**2. Ukazatele intenzitní:** - měří intenzitu (úroveň) sledovaných jevů
- lze je vyjádřit jako poměr dvou extenzitních ukazatelů, je to poměrné číslo
- standardní symbolické značení **p**.

Vlastnosti ukazatelů
- stejnorodost – relativní, závisí na způs. vymezení souboru jednotek pro daný účel zkoumání
- srovnatelnost – srovnatelné jsou ukazatele, jejichž srovnáním získáme smysluplnou veličinu (relativní ukazatel, resp. index).
- shrnovatelnost – vyjadřuje schopnost ukazatele určit jeho celkovou hodnotu na základě hodnot dílčích;

**Způsoby srovnávání hodnota ukazatelů**
1) **Absolutní rozdíl (**diference, přírůstek) – rozměrové číslo, - udává, o kolik měrných jednotek se hodnoty vzájemně liší.
2) **Index** - bezrozměrné číslo; udává, kolikrát je jedna hodnota větší (menší) než druhá

Individuální a jednoduché indexy
=slouží k bezprostřednímu srovnávání dvou hodnot téhož ukazatele, který není složen z dílčích částí; prostý podíl (rozdíl) hodnot ukazatele

**Index množství** - harakterizuje změnu hodnoty sledovaného extenzitního ukazatele (q resp. Q) v běžném období proti období základnímu.
**Index hodnoty**
**Index úrovně** - charakterizuje změnu hodnoty sledovaného intenzitního ukazatele (p) v běžném období proti období základnímu
**Časové indexy a rozdíly** – časové indexy individuální jednoduché bývají často sdružené do delších časových řad.

*Řetězové in. a rozdíly* - charakterizují změny hodnot vzhledem k předcházejícímu období; - indexy (rozdíly) s měnícím se základem
*Bazické indexy a rozdíly* - charakterizují změny hodnot vzhledem k určitému, pevně stanovenému období; - indexy (rozdíly) se stálým základem; - důležitá je volba základního období, zvolit nějakou „normální" hodnotu (ne extrémní, atypickou).

vztahy baz. a řetěz. index. - - umožňují přepočet jedněch na druhé; - používáme je v případě, že nemáme k dispozici jednotlivé údaje, ale pouze řadu indexů.

**Index úrovně (tj. index proměnlivého složení)**
- je konstruován jako podíl dvou průměrů; udává změnu průměrné hodnoty intenzitního ukazatele způsobenou daným činitelem za předpokladu konstantní hodnoty druhého činitele.

$$I_{\bar{p}} = \frac{\bar{p}_1}{\bar{p}_0} = \frac{\frac{\sum Q_1}{\sum q_1}}{\frac{\sum Q_0}{\sum q_0}} = \frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_0 q_0}{\sum q_0}} = \frac{\frac{\sum Q_1}{\sum \frac{Q_1}{p_1}}}{\frac{\sum Q_0}{\sum \frac{Q_0}{p_0}}}$$

ovlivňují: změna dílčích hodnot intenzitního ukazatele, tj. hodnot v dílčích částech celku a změna složení (struktury) hodnot extenzitního ukazatele, tj. změna vah.

***Rozklad $I_{\bar{p}}$ metodou postupných změn***

A. $I_{\bar{p}} = I_{SS}(q_0) \cdot I_{STR}(p_1)$

nebo

B. $I_{\bar{p}} = I_{SS}(q_1) \cdot I_{STR}(p_0)$

**Index stálého složení $I_{SS}$**
=- charakterizuje vliv změny intenzitního ukazatele při stálém složení (v běžném či základním období) na změnu průměrné hodnoty intenzitního ukazatele; - slouží ke zjištění vlivu samotných změn dílčích hodnot intenzitního ukazatele na změnu vyjádřenou p I ;
**Index struktury $I_{STR}$**
=- charakterizuje vliv změny struktury při stálé hodnotě intenzitního ukazatele (v běžném či základním období) na změnu průměrné hodnoty intenzitního ukazatele; - slouží ke zjištění vlivu změn ve struktuře extenzitního ukazatele

*Souhrnné indexy* - slouží ke srovnávání hodnot nestejnorodých extenzitních a intenzitních ukazatelů;

**Indexy úrovně (cenové)** - Loweův, Laspeyresův, Paascheho, Fishereův srovnávání hodnot nestejnorodých intenzitních ukazatelů (např. změny cen různých druhů výrobků);
**Indexy množství (objemu)** - Loweův, Laspeyresův, …
slouží ke srovnávání hodnot nestejnorodých extenzitních ukazatelů (např. výroby, prodeje, spotřeby, nákupu apod. různých druhů výrobků).
**Souhrnný index hodnoty** - vyjadřuje změnu hodnoty produkce, tj. jak změnu objemu, tak změnu cen;

1) **Definujte pojem statistika.**
   - věda o sběru dat a zpracování hromadných údajů, zabývá se jevy, které mají hromadný charakter
   - hromadnost studována na statistických souborech
2) **Co je to popisná statistika?**
   - elementární metody sběru a zpracování informací
   - jednotkou je statistický soubor (osob, podniků, institucí, zvířat, zemí, atd.).
   - statistické soubory jsou tvořeny statistickými jednotkami, mají vlastnosti jednoznačně vymezeny.
3) **Co je matematická statistika a jak se dělí.**
   - moderní – zabývá se složitějšími metodami sběru a zpracování hromadných údajů; vytváří zvláštní druh matematických modelů – tzv. pravděpodobnostní modely – teorie pravděpodobnosti
4) **Typy statistických ukazatelů**
   - okamžikové, intervalové, primární, sekundární, extenzivní, intenzivní, stejnorodé, nestejnorodé
5) **Druhy statistických vlastností (2)**

6) **Statistické jednotky**
   - elementární jednotky stat. pozorování, jsou nositeli znaků
7) **Statistické znaky**
   - vlastnosti jednotek, která je předmětem zkoumání
   - kvalitativní – slovně vyjádřené – alternativní (2 obměny znaku); množné (více než 2 obměny)
   - kvantitativní – číselně vyjádřené, diskrétní (celočíselné), spojité (desetinné číslo, logaritmy)
8) **Statistický soubor**
   - množina jedinců, na kterých je prováděno statistické šetření
   - základní soubor – všechny jednotky s danou vlastností
   - výběrový soubor – vybrán ze základního, podmnožina je menší
9) **Rozdíl mezi ZS a VS**
   - VS je vlastně část ZS
   - VS je menší než ZS (úplné zjišťování, tvořen všemi jednotkami), VS(neúplné zjišťování)
10) **Základní etapy statistických prací**
    - statistické šetření (zjišťování) - získávání neznámých informací o znacích statistických jednotek , výsledkem statistického zjišťování jsou neuspořádané údaje
    - statistické zpracování
    - statistická analýza
11) **Co je statistické zjišťování?**
    - získávání neznámých informací o znacích statistických jednotek
    - výsledkem statistického zjišťování jsou neuspořádané údaje
    - ankety, dotazníky, experiment, výsledek vědeckého experimentu
    - pro zpřehlednění se data třídí
12) **Základní míry polohy rozdělení a k čemu slouží**
    - průměry: aritmetický; vážený ar. prům.; harmonický; geometrický; celkový ar. prům.; chronologický
    - ostatní střední hodnoty: medián, modus
    - měly by jedním číslem popsat střední úroveň hodnoty statistického znaku a umožnit jeho hlubší analýzy
    - reprezentují vhodnou střední hodnotu daného souboru kolem níž se soustřeďují hodnoty tohoto souboru
13) **Prosté x vážené charakteristiky polohy – rozdíl (3)**
    - prosté – u nesetříděných dat, máme-li relativní četnosti
    - vážené – u setříděných dat (tabulka rozdělení četností
14) **Průměr aritmetický**
    - součet všech hodnot znaků dělený počtem znaků
15) **Průměr geometrický**
    - n-tá odmocnina ze součinu znaků
16) **Jaké znáte míry založené na geometrickém průměru?**
    - všude tam, kde má smysl násobit hodnoty, např. průměrný koeficient růstu nebo Fisherův index
17) **V jakém oboru statistiky se můžeme setkat s geometrickým průměrem**
    - používá se u časových řad – průměrné tempo růstu (koeficient růstu)
18) **Průměr harmonický**
    - podíl počtu pozorování a sumy převrácených hodnot znaků
19) **Kdy a k čemu používáme harmonický průměr (3)**
    -v indexní analýze; průměr převrácených hodnot
20) **Průměr chronologický**
    - použití v okamžikové časové řadě
    - prostá forma – tam, kde délka mezi rozhodnými obdobími je stejná)
    - vážená forma – kde vahami jsou počty dní v měsíci,…

**21) Tempo a průměrný koeficient tempa růstu**
- počítání geometrického průměru

**22) Medián**
- x s vlnovkou - prostřední hodnota znaku v souboru uspořádaná podle velikosti
- lichý počet hodnot v souboru - střední hodnota
- sudý počet hodnot - průměr střední hodnoty

**23) Modus**
- hodnota, která se nejčastěji vyskytuje, hodnota znaku s největší četností

**24) Jak vypočítáte modus a medián spojité náhodné veličiny, znáte-li její distribuční funkci?**

**25) Uveďte situaci, kdy může medián popsat polohu statistického souboru lépe než průměr.**
- medián může popsat polohu statistického souboru lépe, pokud je nějaká hodnota hodně vychýlená, tzn., že se hodně liší od ostatních
- pak je průměr zkreslený a medián je lepší měrou polohy statistického souboru. př.: 4, 5, 5, 5, 5, 7, 48

**26) Pro která pravděpodobnostní rozdělení je jejich střední hodnota rovna mediánu a zároveň modu? Vysvětlete a uveďte příklady.**
Pro symetrická (Normální, studentovo)

**27) K čemu se používají podmíněné průměry**
– je to nejjednodušší způsob určení regresní závislosti (přímka podmíněných průměrů)- nelze však na jejich základě provádět odhady

**28) Co se stane s průměrem, rozptylem, směrodatnou odchylkou, mediánem a rozpětím statistického souboru, jestliže každá hodnota statistického souboru se:**
**a) zvětší dvakrát** - průměr a medián se zdvojnásobí, rozptyl se zvýší čtyřikrát; směrodatná odchylka a rozpětí statistického souboru se zvýší dvakrát
**b) zvětší o čtyři** – průměr a medián se zvětší o čtyři; rozptyl se nezmění; směrodatná odchylka a rozpětí statistického souboru se nezmění

**29) Rozdělení četnosti a co je intervalové rozdělení četnosti**
- rozdělení četností - u nespojitých znaků původně neuspořádané údaje roztřídit do rozdělení četností

**30) Jak se stanovuje interval relativní četnosti ZS u malých VS**
- výběr relativní četnosti se řídí binomickým rozdělením v případě výběru bez vracení se řídí hyperbolickým rozdělením
- výpočet vede ke složitým variacím, proto máme sestaveny tabulky a přímo odečítáme meze intervalu z tabulek

**31) Definice pojmu kumulativní četnost**
- absolutní a relativní
- vznikají postupným načítáním

**32) Druhy grafů**
- spojnicové, sloupcové(polygon, histogram), bodové, výsečové, speciální (kvartogram)

**33) Histogram**
- sloupcový graf - u intervalového rozdělení četností

**34) Které charakteristiky statistického souboru můžete přibližně zjistit z histogramu četnosti, aniž byste prováděli výpočet?**
- počet intervalů a jejich šířku, absolutní četnost intervalu a pokud jsou intervaly stejně dlouhé i modus

**35) Jaký graf používáme u jednorozměrných četností.**
- sloupcový

**36) Základní míry variability**
- absolutní: rozptyl, směrodatná odchylka, variační rozpětí, prům. odchylka
- relativní: variační koeficient, relativní průměrná odchylka

**37) Rozptyl**
- aritmetický průměr čtverců individuálních odchylek jednotlivých hodnot znaku od aritmetických průměrů
- nedostatek – jednotky jsou druhou mocninou původních jednotek

**38) Směrodatná odchylka v souboru výběrových průměrů**
- měří abs. Variabilitu
- je uvedena ve stejných měrných jednotkách jako zkoumaný stat. znak; s=odm.s na2
- prostá: S0=odm. z((sum(xi-x)na 2)/n)
- vážená: S0=odm. z((sum(xi-x)na 2*ni)/(sum.ni))
- informuje o proměnlivosti jednotlivých hodnot znaku kolem výběr. aritm. průměru

**39) Variační rozpětí**
- jednoduchá míra adaptability
- pouze odchylky mezi sebou - orientační

**40) Relativní ukazatele variability.**
- variační koeficient, relativní průměrná odchylka

**41) K čemu slouží variační koeficient? Jaká je jeho přednost?**
- variační koeficient je zákl. mírou relativní variability

- může se použít i tehdy pokud se znaky liší svou úrovní, což je výhoda
- počítá se jako podíl směrodatné odchylky a průměru

**42) Jak se změní variační koeficient, přičteme-li ke všem hodnotám souboru stejnou konstantu?**

Směrodatná odchylka v čitateli zůstane stejná a průměr ve jmenovateli se zvětší tuto konstantu => variační

**43) Lze vždy vypočítat variační koeficient souboru dat? Názor zdůvodněte.**

Ne. Variační koeficient se počítá jako podíl směr. odchylky a průměru => je-li např. průměr nulový, Variační koeficient vypočítat nelze.

**44) Kvantil**

- je hodnota, která rozděluje soubor hodnot na dvě části

**45) Kvartil**

- dělí soubor po 25%

**46) Rozdíl mezi charakteristikami šikmosti a špičatosti (3)**

- charakteristika šikmosti (nesouměrnosti)– ukazuje, jak soubor vypadá, stupeň koncentrace malých a velkých hodnot v souboru
- charakteristika špičatosti - ukazuje, jak jsou hodnoty nahloučeny kolem průměru

**47) Význam výběrového šetření v praxi (3)**

- pořizujeme výběrový soubor, aby nám poskytl informace o celém souboru
- hlavním nedostatkem je, že jsou zatíženy výběrovou chybou

**48) Výhody úplného zjišťování oproti neúplnému výběrovému zjišťování**

- úplné – při práci se základním souborem, nákladné, zdlouhavé, občas nemožné
- neúplné – při práci s výběrovým souborem, výběrový soubor musí být dobrým reprezentantem

**49) Vysvětlete pojmy oblastní a vícestupňový náhodný výběr**

- vícestupňový - výběr provádíme na více stupních (města – školy – fakulty – ročníky – studenti)
- oblastní - dvoustupňový výběr; v 1. stupni vybíráme oblast a ve 2. stupni vybíráme z oblasti jednotku

**50) Kvótní výběr - v čem spočívá**

- typ mechanického výběru při náhodném výběru

**51) Jaké znáte techniky pořízení náhodného výběru?**

- losování – opora výběru – výběr zastoupíme lístky
- tabulky náhodných čísel – generátor náhodných čísel
- mechanický výběr – systematické, každá n-tá jednotka v náhodně uspořádané posloupnosti speciální výběr

**52) Existuje rozdíl mezi stanovením intervalu u vracení a bez vracení?**

- s vracením – jednotku po výběru vracíme zpět
- bez vracení – rozsah ZS se zmenšuje, pravděpodobnost vybrání se zvětšuje
- u velkých souborů zbytečné zbytečné pracovat s vracením

**53) Jaký test k ověřování náhodnosti výběrového souboru? (3)**

**54) Metoda základního masivu**

- kdy se soubor skládá z několika velkých a mnoha malých jednotek
- zjišťování provádíme na velkých jednotkách

**55) Záměrný výběr**

- značná míra subjektivity toho, kdo vybírá
- vybere ty, o kterých si myslí, že dobře zastoupí soubor, ty blízké průměru, nelze vyvodit chyba

**56) Dělení (druhy) náhodného výběru**

- s vracením, bez vracení
- s nestejnou pravděpodobností vybrání
- prostý – se stejnými pravděpodobnostmi

**57) Náhodný jev**

- jev, který může nastat nebo nenastane v závislosti na náhodě a je výsledkem náhodného pokusu (charakterizuje výsledek náhodného pokusu kvalitativně)

**58) Náhodný pokus**

- realizace podmínek a vlivů, z nichž některé jsou známé a jiné náhodné

**59) Jev jistý, náhodný, nemožný**

- jev jistý - takový, který vždy nastane při každém provedení náhodného pokusu
- jev náhodný - jevy, které v závislosti na náhodě mohou, ale nemusí při uskutečňování daného komplexu podmínek nastat
- jev nemožný - náhodný jev, který nenastane při žádném provedení náhodného pokusu

**60) Klasické a statistické definice pravděpodobnosti**

- klasická - může li určitý pokus vykázat konečný počet n různých výsledků, které jsou stejně možné a jestliže m těchto výsledků má za následek nastoupení jevu A, kdežto zbylých n-m vylučuje: potom P(A)=m/n
- statistická - spojena s pojmem relativní četnosti; s rostoucím počtem pokusů se relativní četnost stabilizuje a přibližuje se k určitému konstrukčnímu číslu. P(A)= lim při n ku nekonečnu * M/N

**61) Matematická charakteristika pravděpodobnosti**

**62) Rozdíl mezi náhodnou veličinou a náhodným jevem (2)**

- náhodný jev – takový jev, který v závislosti na náhodě může, ale nemusí při uskutečňování daného komplexu podmínek nastat; charakterizuje výsledek náhodného pokusu kvalitativně (slovně)
- náhodná veličina – libovolná kvantitativní charakteristika náhodného pokusu; proměnná, která nabývá konkrétních hodnot, či hodnot z různých intervalů v závislosti na náhodě

**63) Zákon rozdělení náhodné veličiny**

- pravidlo, které každé hodnotě, nebo množině hodnot z každého intervalu přiřazuje pravděpodobnost, že náhodná veličina nebude této hodnoty, nebo hodnoty z tohoto intervalu
- tento zákon může být vyjádřen různou formou:

jako řada rozdělení pravděpodobností (grafem je polygon, diskrétní veličiny)
distribuční fce (univerzální zákon rozdělení, diskrétní i náhodné veličiny)
hustota pravděpodobnosti (spojité náhodné veličiny)

**64) Druhy rozdělení náhodných veličin**

- spojité (normální, exponenciální, chí-kvadratické,Studentovo t-rozdělení, F-rozdělení, rovnoměrné rozdělení)
- nespojité - diskrétní (Alternativní, Binomické, Poissonovo, Hypergeometrické, Geometrické)

**65) Charakterizujte normální rozdělení**

- náhodná veličina se řídí normálním rozdělením, je-li její střední hodnota μ a rozptyl ∂2
- grafem hustoty pravděpodobnosti je Gaussova křivka
- speciálním případem normálního rozdělení je normované normální rozdělení

**66) Binomické rozdělení je (možnosti) (2)**

- nejdůležitější typ rozdělení diskrétní náhodné veličiny
- rozdělením náhodné veličiny, která představuje počet výskytů jevu A při n nezávislých pokusech, přičemž pravděpodobnost výskytu jevu A je v každém pokusu konstantní

**67) Jaký je vztah binomického a Poissonova rozdělení?**

- má-li náhodná veličina X binomické rozdělení takové, že počet pokusů n je dostatečně veliké (nad 30), pravděpodobnost výskytu sledovaného jevu v jednom pokuse pod 0,1 a n konečné číslo, je možno toto rozdělení aproximovat Poissonovým rozdělením

**68)** Pravidlo tří sigma

- i když náhodná veličina X, která má normální rozdělení, může nabývat hodnot z intervalu od (- ∞,∞), je téměř nemožné, aby se pozorované hodnoty této veličiny odchylovaly od střední hodnoty o více než 3 sigma

**69) Co vyjadřuje zákon velkých čísel?**

- se zvyšováním počtu náhodných pokusů dochází k přibližování se empirické charakteristiky popisující výsledky těchto pokusů k charakteristice teoretické

**70) Co vyjadřuje centrální limitní věta?**

- vyjadřuje konvergenci pravděpodobnostních rozdělení k normálnímu rozdělení při dostatečně velkém rozsahu souboru.

**71) Co je normovaná náhodná veličina, jaké má charakteristiky a jaký má význam?**

- má normální rozdělení se střední hodnotou 0 a rozptylem 1
- význam je ve výpočtu distribuční funkce, která se z normálního rozdělení počítá obtížně

**72) V čem spočívá z pohledu teorie pravděpodobnosti průnik jevů A,B a sjednocení jevů A,B**

- průnik jevu A,B - spočívá v současné realizaci jak jevu A, tak jevu B
- sjednocení jevů A,B - spočívá v nastoupení alespoň jednoho z jevů a nebo B

**73) Je možné, aby existovaly 2 náhodné jevy, že pravděpodobnost jejich průniku je větší než pravděpodobnost jejich sjednocení?**

- Ne. Protože pravd. průniku může být max. rovná pravděp. sjednocení, když množiny splývají nebo je menší nebo množiny nemají průnik.

**74) Při výpočtu pravděpodobnosti projiti třemi zkouškami, z nichž každá má svou pravděpodobnost úspěchu používáme:**

a) sčítání
b) násobení
c) rozdíl

**75) Podmínky pro sčítání pravděpodobností a vzorec**

- jsou-li jevy A a B slučitelné, potom pravděpodobnosti jejich sjednocení se rovná součtu pravděpodobností jednotlivých jevů zmenšenému o pravděpodobnost jejich průniků
- v případě neslučitelných jevů je průnik těchto jevů jev nemožný

**76) Co jsou kvalitativní znaky, jak se dělí, příklady**

- kvalitativní znaky jsou znaky slovní, získané z anket, dotazníků
- dělíme na znaky alternativní -

**77) Jak ověříte nezávislost dvou kvalitativních znaků?**

Kontingenční tabulkou

**78) Jakým způsobem můžete zjistit, jestli existuje závislost mezi dvěma kvalitativními znaky a jakým v případě kvantitativních znaků?**

- závislost mezi 2 kvalitativními znaky ověřujeme kontingenční tabulkou a testem chi-kvadrát.
- závislost mezi 2 kvantitativními znaky měříme klasicky pomocí regresní a korelační analýzy, celkový F-test, Test t pro

jednotlivé parametry a koeficienty
- u 1 kvalitativního a 1 kvantitativního znaku se používá jednofaktorová analýza rozptylu

**79) Koeficient asociace slouží k:**
- vyjádření těsnosti 2 alternativních znaků

**80) Jak určíme nejvhodnější typ funkce při měření závislosti dvou kvantitativních znaků**
- zkušenost, logika, emp. metoda-korelační pole
- zkoušet = počítat - zpětně vybrat ten s nejvyšším korelačním charakterem

**81) Jaké jsou hlavní úlohy při měření závislosti 2 kvantitativních znaků**
- vystihnout průběh závislosti závisle proměnné na nezávisle proměnné, tak abychom mohli provádět odhady závisle proměnné na základě daných hodnot nezávisle proměnné
- změřit sílu závislosti, abychom mohli posoudit její sílu, intenzitu a abychom mohli zároveň posoudit přesnost odhadů z 1 bodu
- 1. úkol - regrese, 2. úkol – korelace

**82) Teoretický soubor výběrových průměrů**
- ze základního souboru vybereme všechny teoreticky možné VS, těch je nekonečně mnoho; v každém výběrovém souboru si vypočítáme výběrový průměr, všechny tyto průměry nám vytvoří teoretický soubor výběrových průměrů

**83) Statistická indukce**
- nejprve pořídíme výběrový soubor, na základě VS si spočítáme výběrové charakteristiky, na základě výběrových charakteristik odhadujeme charakteristiky ZS

**84) Jaké známe odhady (3)**
- bodový – jedno konkrétní číslo, které vybereme z VS, aby nám nahradilo ZS
- intervalový – stanovení intervalu, ve kterém ta neznámá charakteristika bude ležet, a určitou pravděpodobností

**85) Co je bodový odhad?**
- bodový – jedno konkrétní číslo, které vybereme z VS, aby nám nahradilo ZS

**86) Jaké znáte vlastnosti bodových odhadů a co vyjadřují, jaké jsou na ně kladeny požadavky?**
- nezkreslenost, nestrannost odhadu (střední hodnota výběrové statistiky = odhadované charakteristice)
- konzistence (odhad se s rostoucím rozsahem výběru blíží odhadované charakteristice základního souboru)
- vydatnost (co nejmenší rozptyl)
- postačujícnost (mimo ní neexistuje žádná jiná statistika poskytující další doplňující informace o odhadované charakteristice základního souboru)

**87) Proč musí být bodový odhad v základním souboru vydatný**
- můžeme použít více charakteristik odhadu, za nejvydatnější je ta, která má nejmenší rozptyl

**88) Jak spolu souvisí přesnost odhadu a spolehlivost odhadu?**
- čím širší interval spolehlivosti, tím ale menší přesnost
- spolehlivost = pravděpodobnost, se kterou bude odhadovaná charakteristika ležet v tom vymezeném intervalu; = maximální chyba, které se při odhadu s danou spolehlivostí můžeme dopustit

**89) Co znamená, že odhad je vychýlený? Znáte některé vychýlené odhady?**
- E(g) - G je tzv. zkreslení neboli vychýlení
- takový odhad vede k systematickému nadhodnocování či podhodnocování odhadované charakteristiky ZS

**90) Přípustná chyba u intervalového odhadu a k čemu jí používáme**
- chyba, které se při odhadu můžeme dopustit, aby hodnota padla do intervalu
- přesnost intervalového odhadu je charakterizována přípustnou chybou odhadu delta, která představuje polovinu délky intervalu spolehlivosti

**91) Přesnost odhadu: (2)**
- pravděpodobnost, s jakou se charakteristika nachází v intervalu
- vyneseme kritickou hodnotu příslušného rozdělení
- max chyba, které se při odhadu s danou spolehlivostí dopustíme vyjádřena hodnotou směrodatné odchylky souboru výběrových prům.
- ani jedna správně

**92) Jaké znáte metody pro získání odhadů parametrů regresních funkcí lineárních v parametrech? Napište princip metod.**
- požadavek kompenzace kladné a záporné odchylky empirických hodnot od hodnot vyrovnaných a metoda nejmenších čtverců (aby součet čtverců popsaných odchylek byl minimální)

**93) Jaké znáte metody pro získání odhadů parametrů regresních funkcí nelineárních v parametrech?**
- pro funkce nelineární v parametrech používáme linearizující transformaci
- pak použijeme metodu nejmenších čtverců, parciální derivace, dále dostaneme soustavu normálních rovnic a nakonec pomoci Cramerova pravidla (determinanty) vyjádříme $b_0$, $b_1$, ...

**94) U kterých z uvedených regresních funkcí lze k odhadu parametrů použít metodu nejmenších čtverců: přímka, parabola, exponenciála? Názor vysvětlete.**
- u přímky a paraboly, protože jsou lineární v parametrech na rozdíl od exponenciály, která není - tam je nutná lineární transformace.

**95) Pojmy: (3)**
a) alternativní hypotéza – popírá platnost nulové hypotézy

b) testovací kritérium – míra nesouhlasu výsledků pokusu s testovanou hypotézou (odpovídají-li data nulové hypotéze- testovací kritérium je rovno nule; čím více se výběrové hodnoty blíží k alternativní hypotéze, tím roste i testovací kritérium)

c) hladina významnosti – pravděpodobnost chyby 1. druhu; udává výši rizika, s jakým se $H_0$ zamítá, i když platí

**96) Alternativní hypotéza**

- popírá platnost nulové hypotézy
- přijímáme ji jestliže jsme nulovou hypotézu zamítli jako nesprávnou

**97) Vysvětlete pojem hladina významnosti a kritický obor.**

- hladina významnosti - pevná, předem zvolená pravděpodobnost chyby I. Druhu
- kritický obor - podprostor výběrového prostoru obsahující hodnoty svědčící ve prospěch alternativní hypotézy

**98) Vysvětlete pojem "síla testu" a "hladina významnosti".**

- síla testu - s jakou pravděpodobností se nedopustíme chyby II. druhu
- hladina významnosti - pevná, předem zvolená pravděpodobnost chyby I. druhu

**99) Co udává hladina významnosti alfa**

- udává výši rizika s jakým H0 zamítá, i když platí pravděpodobnost chyby 1. druhu; alfa=P(T je elem. K\H0)

**100) Testovací kriterium je**

- veličina vypočtená z výběrových hodnot, řídí se rozdělením(norm. studentovým), porovnáváme s krit. hodnotou

**101) Interval spolehlivosti a k čemu slouží**

- **i**nterval, ve kterém leží neznámé charakteristiky s určitou předem známou pravděpodobností. Prostřednictvím intervalu spol. posuzujeme přísnost odhadu. s rostoucí šířkou intervalu spol. klesá přesnost odhadu.

**102) Čím můžeme ovlivnit velikost intervalu spolehlivosti?**

- zvolenou hladinou spolehlivosti. Čím širší interval spolehlivosti, tím ale menší přesnost.

**103) Vysvětlete pojmy chyba I. druhu, chyba II. druhu a síla testu.**

a) Zamítnutí nulové hypotézy, ačkoliv ve skutečnosti platí

b) přijmutí nulové hypotézy, ačkoliv ve skutečnosti platí hypotéza alternativní

c) s jakou pravděpodobností se nedopustíme chyby II. druhu

**104) Jakým způsobem můžeme snížit pravděpodobnost chyby I. druhu a jak pravděpodobnost chyby II. druhu (při dané chybě I. druhu)?**

- chybu I. druhu snížíme tak, že místo 95% hladiny významnosti volíme 99%
- chybu II. druhu snížíme tím, že zvýšíme rozsah souboru

**105) Rozdíl mezi dvoustranným a jednostranným intervalem**

- dvoustranný - charakteristiky ZS jsou omezeny zdola i shora
- jednostranný - charakteristiky ZS jsou omezeny shora nebo zdola

**106) Obecný postup statistického testování**

- vybereme vhodný test (parametrický a neparametrický)
- formulace nulové a alternativní hypotézy
- volba hladiny významnosti
- volba testovacího kriteria
- určení kritického oboru
- výpočet hodnoty testovacího kriteria z výběrových hodnot
- rozhodnutí: jestliže vyp. hodnota t.k. padne do kritického oboru H0 se zamítá jinak se H0 nezamítá

**107) Rozdíl mezi parametrickými a neparametrickými testy**

- Parametrické - testy sloužící k ověřování hypotéz, které se týkají hodnot parametrů rozdělení - spolehlivé, vyžadují znalost ZS a hodnotu parametrů
- Neparametrické - místo původních hodnot pracujeme s pořadovými čísly - jednoduché, menší síla testu, nevyžadují znalost ZS

**108) Jaký závěr lze učinit, vede-li celkový F-test o regresní funkci k zamítnutí testované hypotézy (nejdříve uveďte, jak jsou při tomto testu formulovány testovaná a alternativní hypotéza).**

- testovaná = neexistence vlivu faktoru (o nezávislosti znaku Y na zkoumaném faktoru); Alternativní = existence
- závěr = podobnost skupinových průměrů na celý soubor

**109) Lze zvětšením rozsahu výběru ovlivnit šíři intervalu spolehlivosti odhadu střední hodnoty? Pokud ano, vysvětlete jak.**

- pokud máme větší rozsah souboru, vede to k užšímu intervalu spolehlivosti. Vzorek je více interpretativní.

**110) "Jestliže hypotézu □ = □0 zamítneme při jednostranném testu na určité hladině významnosti, pak jí zcela určitě zamítneme i při oboustranné alternativě při stejné hladině významnosti." Souhlasíte? Vysvětlete proč.**

- neplatí vždy <= dvoustranný test má při stejné hladině významnosti poloviční alfu a hodnota se do něj může, ale nemusí vejít

**111) Postup při použití testu minimální průkazové definice**

- u vyváženého modelu: dělíme diference mezi průměry ale neseřazujeme je podle velikosti; spočítáme hodnotu Dmin=g=odmocnina z (sr na 2/n); porovnáme diference s touto hodnotou; jestli je dif. větší než Dmin tak je rozdíl mezi průměrem statisticky významný

**112) Při testování spotřeby automobilu určité značky používáme:**

a) F-test
b) t-test
c) Bartlettův test

**113) Při testování více než dvou rozptylů se nepoužívá**

a) Hartleyův test
b) Bartlettův test
c) F-test

**114) Bartelův test**

- test rozptylu více souborů, H0 rozptyly se rovnají
- testuje, zda můžeme udělat analýzu rozptylů - podmínka - shodné rozptyly z normálního rozdělení
- pokud nesplňuje: a) transformace hodnot; b) doměření hodnot; vyloučení ext. hodnot

**115) Cochranův test**

- stejné rozsahy n

**116) Hartleyův test**

- řídí se f rozdělením

**117) Ferrarův – Glauberův test, k čemu se používá**

**118) Rozdílové testy**

**119) Pořadové testy**

**120) Rozptyl a jak jej definujeme**

- aritmetický průměr čtverců individuálních odchylek jednotlivých hodnot znaku x od aritmetických průměrů
- charakteristika variability
- informuje o proměnlivosti jednotlivých hodnot znaku kolem výběr. ar. prům.
- výběrový rozptyl s na 2 souboru: pozorované x1, x2....xn

**121) Jak provádíme rozklad celkového rozptylu? (4)**

- rozklad celkového rozptylu všech hodnot na dílčí rozptyly (vyjadřují podíl jednotlivých faktorů na celkovém rozptylu) a na rozptyl reziduální (zbytkový, vyjadřuje podíl faktorů působících náhodně )

**122) Kde prakticky využíváme rozklad rozptylu?**

- při počítání vnitro a meziskup. variability
- u regres. a korel. analýzy ke konstrukci indexů

**123) K čemu slouží analýza rozptylu a jaké jsou předpoklady jejího použití**

- zobecnění t-testů na více než 2 VS rozklad výběr. rozptylu na něk. částí, které jsou příslušné jednotlivým uvažovaným zdrojům variability, zkoumáme li vliv jednoho či více faktorů ne výsledky kvantitavního znaku
- analýza rozptylu je založena na předpokladu, že ze zjištěných hodnot lze provést odhad rozptylů - 2 způsoby: uvnitř tříd, mezi třídami
- typické použití v zemědělství - plemenictví

**124) Analýza rozptylu je (možnosti) (2)**

- je obecný postup, který patří do oblasti statistických metod – metodologický nástroj
- technika pro zakládání a vyhodnocování experimentu
- je více výběrový test, který testuje rozdíl mezi průměry
- zkoumá vliv jednoho i více faktorů na výsledný znak kvantitativní, každý faktor je sledován na několika úrovních (třídách), každá úroveň představuje 1 výběrový soubor; faktorem může být znak kvantitativní i kvalitativní
- pokud je úroveň každého faktoru pevně fixována – model analýzy rozptylu s pevnými efekty
- pokud jsou úrovně faktorů náhodně vybrány – model s náhodnými efekty

**125) Jaké míry jsou založeny na rozkladu rozptylu?**

- korelační koeficienty

**126) Jakou metodu byste použili k posouzení, jestli je prodej určitého zboží v průběhu dne rovnoměrně rozložen?**

Možná rozptyl - nebo řetězový index

***127)* Kde prakticky využíváme rozklad rozptylu?**

- při počítání vnitro- a meziskup. variability a u regres. a korel. analýzy ke konstrukci indexů.

**128) Metody podrobnějšího hodnocení výsledků analýzy rozptylu**

- metody mnohonásobných porovnání, s-metoda (Scheffeho metoda), t-metoda (metoda minimální důkazové diference), Cramerova metoda, Duncanova metoda

**129) Duncanova metoda**

**130) Postup při použití Duncanova testu**

- hodnotám přiřadíme pořadové číslo a spočítáme diferenci
- uspořádáme do tabulky, určíme krit. hodnotu diference a porovnáme diference s kritickými, jeli dif. větší než krit. pak je rozdíl statisticky významný

**131) Kramerova metoda a k čemu slouží**

- pomocí ní se provádí podrobnější hodnocení výsledku v analýze rozptylu u neváženého modelu
- rozdíl mezi průměry je statisticky významný, jestliže jejich diference je větší než výraz na pravé straně nerovnosti

**132) Scheffeho metoda**

- rozdíl mezi průměry je stat. významný, jestliže jejich diference je větší než vypočítaná hodnota

**133) Tukeyův test**

**134) Rozdíl mezi vyváženým a nevyváženým modelem analýzy rozptylu**

- vyvážený - stejný počet opakování v každém řádku
- nevyvážený - různý počet opakování v ...

**135) Rozdělení F při analýze rozptylu**

- F,alfa (m-1; n-m)
- rozdělení F pro f1=m-1 a f2=n-m stupňů volnosti

**136) K čemu se používá F-test**

- test o rozdílu 2 výběrových rozptylů

**137) Jaké testy předcházejí analýze rozptylu**

a) Bartletův
b) Cochranův
c) Hartleyův)

**138) Kdy používáme analýzu rozptylu + podmínky použitelnosti**

(1) používá se, když zkoumáme vliv jednoho či více faktorů na výsledný znak kvantitativní
(2) pomínky použitelnosti - normalita rozdělení ZS z kt. jsou pořízeny VS; statistická nezávislost náhodných chyb; existence shodných rozptylů ZS ze kt. jsou pořízeny VS

**139) Co je interakce u AR?**

**140) Jak se provádí a proč podrobnější vyhodnocení AR?**

**141) Modely AR dvojného třídění**

- zobecnění t-testů na více než 2 VS rozklad výběr. rozptylu na něk. částí, které jsou příslušné jednotlivým uvažovaným zdrojům variability, zkoumáme li vliv jednoho či více faktorů ne výsledky kvantitavního znaku
- analýza rozptylu je založena na předpokladu, že ze zjištěných hodnot lze provést odhad rozptylů - 2 způsoby: uvnitř tříd, mezi třídami
- typické použití v zemědělství - plemenictví

**142) Vypište neparametrické testy**

- testy dobré shody – $\chi 2$ test, Kalgor- Smirnovův test
- vlastní neparametrické testy – Wilcox- Whiteův test (parametrický t-test), znaménkový test (parametrický pár. t-test), Wilcoxonův test (parametrický t-test), Kruskal-Wallisův test (parametrická analýza rozptylu), Dixonův test, Friedmanův test, test iterací

**143) Výhody a nevýhody neparametrických testů (3)**

- v praxi se často setkáváme s výběry poměrně malých rozsahů, které pocházejí z výrazně nenormálních souborů, nebo ze souborů, o jejichž rozdělení nic nevíme
- nepředpokládáme specifikované rozdělení základního souboru, z něhož se získává náhodný výběr
- výhoda – nezávislost na tvaru rozdělení studovaných náhodných veličin, použitelnost pro studium jak znaků kvantitativních, tak znaků kvalitativních (obecnější použití); výpočetní jednoduchost
- nevýhody – menší síla (schopnost odhalit nesprávnost testované hypotézy)

**144) Které testy používáme jako neparametrickou obdobu párového t-testu a jak je provádíme**

- znaménkový: spočítáme diference mezi páry hodnot; počet kladných hodnot z+; počet záporných hodnot z-; menší z oboru je testovací kritérium z: z je větší než z,alfa (sečtou se počty ve skupinách a min. výsledná hodnota je testové kriterium, které se porovná s tabulkovou hodnotou)
- wilcoxonův: spočítáme diference; k nenulovým diferencím přiřadíme vzestupně pořadová čísla; pořadová čísla roztřídíme zpětně do dvou skupin podle znamének diferencí; v obou skupinách pořadová čísla sečteme a dostaneme W+ a W-; Wmin(W+,W-) menší z obou čísel je test. krit.; W je menší než W,alfa - H0 se zamítá;

**145) Který test používáme jako neparametrickou obdobu jednoduché analýzy rozptylu a jak se tento test provádí**

- Kruskal-Wallisův test: hodnoty sloučíme do jednoho výběru; seřadíme od největší do nejmenší; očíslujeme vzestupně; sečteme pořadová čísla pro každý řádek zvlášť; výpočet test. kritérium, porovnáme s kritickou hodnotou

**146) Postup při Wilcoxon-Whiteova testu**

- hodnoty sloučíme do jednoho výběru
- seřadíme od nejmenší do největší
- očíslujeme hodnoty vzestupně
- zpětně roztřídíme do VS
- sečteme pořadová čísla; T=min(Tx;Ty); T je menší než T,alfa - H0 se zamítá
- u souboru n větší než 20 normální aproximace odhadnem

**147) Test o shodě průměru základního souboru (2)**

- je určen k testování rozdílů mezi průměry u nezávislých výběrů
- používá se v případě párových výběrů
- používá se k testování rozdílu mezi výběrovým prům. a předpokládanou hodnotou prům. základního souboru
- určuje, zda prům. zákl. souboru je stat. významný
- ani jedna správně

**148) K čemu je Friedmanův test? (4)**

- máme více než 2 závislé soubory, tzn. k >2

**149) Uveďte aspoň 3 testy, u nichž použijete jako kritické hodnoty kvantity rozdělení chí-kvadrát. Uveďte, čeho se týkají a co říkají.**

- chi-kvadrát dobré shody
- test o parametru delta exponenciálního rozdělení
- test hypotézy o rozptylu v základním souboru

**150) Jaké znáte testy shody? K čemu slouží?**

- Chi-kvadrát test dobré shody, shoda 2 průměrů, rozptylů
- porovnávají dva různé soubory

**151) Uveďte alespoň dva případy, kdy použijete při testování hypotézy testované kriterium, které má rozdělení chí-kvadrát.**

- chi-kvadrát dobré shody, test o parametru delta expon. rozdělení a test hypotézy o rozptylu v zákl. souboru.

**152) Kdy používáme Fisherův test**

- když je n menší než 20 nebo n je prvkem (20,40) a jednina rel. čet. 15

**153) V kterých tabulkách se provádí $\chi 2$ test?**

**154) Jaké jsou základní předpoklady Kolmogorov-Smirnovova testu**

**155) Kdy používáme chí-kvadrát test?**

**156) Jak dělíme metody vícerozměrné stat. analýzy z hlediska klasifikace jednotek**

- shluková analýza
- diskriminační analýza
- faktorová analýza...

**157) Shluková analýza**

- vícerozměrná statistika
- souhrnný název pro řadu výpočetních postupů, jejichž cílem je rozklad daného souboru na několik relativně homogenních množin (shluků) a to tak, aby jednotky uvnitř jednotlivých shluků si byly co nejvíce podobné a jednotky patřící do jiných shluků co nejvíce nepodobné
- míra vzdálenosti - je jejím určením
- vytváří celky metody - nejbližšího souseda, nejvzdálenějšího souseda, průměrná vzdálenost - centrální, mediánová, warolová(graf)

**158) Diskriminační analýza**

- řeší problematiku vícerozměrné klasifikace
- předmětem je nalezení statisticky nejvhodnějšího způsobu rozlišení mezi 2 či více soubory statistických jednotek
- klasická úloha diskriminační analýzy spočívá v tom, že jsou předem známy 2 či více skupin jednotek a o každé víme do které skupiny patří
- na základě naměřených údajů se pro každou skupinu vypočítá diskriminační fce sloužící k dodatečnému zařazování nových jednotek
- do určité skupiny je zařazena ta jednotka, pro níž je pravděp. příslušnost této jednotky ke skupině největší
- slouží k dodatečnému zařazení znaků do předem stanovených skupin

**159) Faktorová analýza**

- umožňuje objasnit strukturu pozorovaných závislostí, redukuje počet výchozích proměnných pomocí hypotetických faktorů při minimální ztrátě informací a odhaduje skryté vztahy mezi proměnnými
- použití při zpracování dotazníků

**160) Kovarianční analýza**

- umožňuje rozčlenění variability proměnné y na části, které jsou přiřaditelné jednak kvalitativním a jednak kvantitativním vlivům

**161) Analýza hlavních komponent**

- podstatou je transformace souboru napozorovaných proměnných do nových proměnných, tzv. hlavních komponent, které jsou vzájemně nezávislé a jsou seřazeny dle velikosti svého příspěvku ke vysvětlení celkového rozptylu napozorovaných proměnných
- tato metoda citlivá na delta jednotek větší, než hodnoty normují
- použití: odhalení skrytých vztahů mezi proměnnými
- založena na bezezbytkovém vysvětlení celkového rozptylu proměnných

**162) Kanonická korelační analýza**

- vychází z logického členění proměnných do 2 skupin (výchozí + cílové veličiny) - každá skupina je charakterizována 1 soubornou veličinou = kanonickou proměnnou = lineární kombinace původních proměnných dané skupiny

**163) Jak se provádí odhad pro celou regresní přímku?**

**164) Rozdíl mezi jednoduchým a vícenásobným modelem regrese a korelace**

- počet nezávisle proměnných
- v modelu jednoduché stat. závislosti je předpokládáno, že změny závisle proměnných jsou vyvolány změnami jediné nezávislé proměnné, ostatní vlivy jsou považovány za náhodné
- teorie vícenásobné regrese a korelace je zobecněním teorie jednoduché regrese a korelace

**165) Korelace x regrese (2)**

- korelace –těsnost, síla, míra závislosti mezi kvantitativními statistickými znaky
- regrese – popis průběhu závislosti mezi kvantitativními znaky pomocí regresního modelu ( tímto modelem je regresní funkce

**166) Základní podmínka metody nejmenších čtverců**

- $\sum d_i^2=0$ - vede k jednomu závěru
- $\sum d_i=min$ - můžeme spočítat parametry

**167) Metoda nejmenších čtverců a k čemu se používá**

- matematická metoda s jejíž pomocí můžeme vypočítat parametry regresních fcí(s výjimkou fcí, které nemůžeme převést na aditivní tvar)

**168) Pás spolehlivosti regresní přímky – vysvětlit (3)**

**169) Jaká je regresní funkce nelineární v parametrech, jak parametry určujeme (3)**

**170) Co se dá zjistit z regresní přímky?**

**171) Co vyjadřuje regresní index a koeficient?**

**172) Test regresního koeficientu.**

**173) Podle jakých kritérií se vybírá nejvhodnější funkce pro vícenásobnou nelineární regresi?**

**174) Postup při metodě stupňovité regrese (4)**

**175) Napište rovnici regresní funkce a korelační koeficient, když máš zadáno x s pruhem (celkový průměr), y s pruhem (celkový průměr), s x, s y a s xy**

**176) Lze testovat významnost regresní funkce a určit její interval spolehlivosti? Pokud ano, napište stručný postup.**

**177) Určení závislosti u nelineární regrese**

**178) Vícenásobná regresní funkce a jednoduchá regresní funkce – rozdíl**

**179) Regrese na hlavních komponentách**

- vícenásobný regresní model předpokládá nezávislost nezávislé proměnné
- tento předpoklad bývá v praxi často nahrazen a proto se doporučuje do vícenásobní regrese nahradit nezávislé proměnou hlavními komponentami

**180) Regresní koeficient u lineární regrese**

- byx - vyjadřuje o kolik se změní závisle proměnná y, jestliže se nezávisle proměnná x změní o jednotku
- použití k odhadům změny

**181) Aditivita vícenásobného regresního modelu**

- funkci pro vícenásobnou regresi získáme jako součet jednoduchých regres. fcí

**182) Jakým způsobem určujeme parametry regresních funkcí? (3)**

- metoda nejmenších čtverců – obecná soustava normálních rovnic

**183) Beta koeficienty a k čemu se používají**

- normované přepočtení regresní koeficienty
- k určení podílu jedn. nezávisle proměnné u regres. odhadu závisle proměnné.(vícenásobná lineární regrese)

**184) Jak postupujeme při určení nejvhodnějšího typu regresní funkce**

- na základě zkušenosti, logické posouzení, empirické posouzení (korel. pole), výpočet mnoha funkcí-výběr nejlepší

**185) Jak provádíme výběr nejvhodnější nelineární regresní funkce? (3)**

- podle toho, která nelineární regresní funkce má nejvyšší hodnotu indexu korelace

**186) Co jsou to sdružené regresní přímky**

- hodnoty korelačních koef. převádí v hodnoty z a vždy se řídí alespoň přibližně normálním rozdělením

**187) Typy nelineárních regresních fcí**

- polynom, parabola 2. st., exponenciála, mocninná funkce, hyperbola, růstová funkce

**188) Interakce ve vícenásobné regresní analýze (2)**

- podíl nezávisle proměnných na regresním odhadu závisle proměnné
- posuzuje vhodnost regresní funkce
- představuje vzájemné působení proměnných (nejde to slovo přečíst)
- vyjmenuj vztahy mezi koeficienty
- ani jedna možnost

**189) Korelační pole**

- empirická metoda pro nalezení vhodné regresní funkce, vyjádření párové závislosti
- vyjadřuje vzdálenost mezi teoretickou a skutečnou hodnotou u vybrané funkce

**190) Korelační tabulka**

- kombinační tabulka, která vyjadřuje průběh závislosti 2 proměnných (kvalitativní znaky)
- na úhlopříčce je pevná závislost
- čím jsou více soustředěné kolem úhlopříčky, tím je silnější závislost a naopak
- těsnost závislostí v ní měříme korelačním poměrem

**191) Korelační tabulka**

- kombinační tabulka. která vyjadřuje průběh závislosti 2 proměnných(kvalit znaky)
- těsnost závislostí v ní měříme korelačním poměrem

**192) K čemu slouží Spearmanův koeficient korelace pořadových čísel**

- charakterizuje těsnost jednoduché závislosti kvantitativních znaků
- nabývá hodnot 0-1 původní hodnoty nahradíme pořadovými čísly je to u korel.
- jedná se o neparametrickou charakteristiku
- měří těsnost jakékoli statistické závislosti, která je monotónní
- poskytuje rychlou a dostatečně přesnou informaci o těsnosti sledované závislosti

**193) Totální koeficient korelace a determinace**

- totální koeficient korelace - vyjadřuje těsnost závislosti Y na všech nezávislých prom. X
- totální koeficient determinace - druhá mocnina korelace; udává z kolika % je nez. proměnná Y ovlivněna všemi uvažovanými nezávisle proměnnými X

**194) Koeficient vícenásobné totální korelace**

-měří těsnost závislosti závisle proměnných na všech uvažovaných nezávisle proměnných

**195) Koeficient parciální korelace**

-měří těsnost závisle proměnných na jedné z uvažovaných nezávisle proměnných

**196) Korelační index je (možnosti) (2)**

- měříme jím těsnost závislosti u nelineárních funkcí
- je v intervalu od 0 do 1, čím více se blíží 1 tím je závislost silnější, čím blíže 0 tím je závislost slabší

**197) Jak ověřujeme statistickou významnost korelačního indexu a korelačního koeficientu? (3)**

- korelační koeficient – podle t-Studentova rozdělení, tabulky kritických hodnot korelačního koeficientu , neprovádíme test, pouze porovnáváme hodnotu korel. koeficientu s tabulkovou , je-li vypočtená > tabulková, můžeme rovnou korel. koeficient považovat za statisticky významný
- korelační index - netestujeme

**198) Co měří dílčí korelační koeficient a co vícenásobný korelační koeficient?**

- dílčí korelační koeficient vyjadřuje těsnost lineárního vztahu mezi dvěma proměnnými při vyloučení vlivu jedné nebo více dalších proměnných
- vícenásobný koeficient korelace charakterizuje těsnost závislosti jedné proměnné na lineární kombinaci jiných proměnných

**199) Koeficient determinace - $r^2$; vyjádření v %**

- udává z kolika % je závisle proměnná ovlivněna uvažovanou nezáv. Proměnnou (u lineární regrese a korelace)

**200) Totální koeficient determinace**

- vyjadřujeme v % a udává nám z kolika % je závislá proměnná ovlivněna uvažovanými nezávisle proměnnými

**201) K čemu lze prakticky využít determinační index? O čem na jeho základě můžeme rozhodnout?**

- index determinace vynásobený 100 udává relativně v procentech tu část rozptylu závisle proměnné z, kterou se podařilo vysvětlit použitou regresní funkcí
- čím bližší je jedné, tím je daná závislost silnější.

**202) Může být vícenásobný korelační koeficient menší než některý z jednoduchých korelačních koeficientů?**

- nemůže <= vícenásobný je složen z jednoduchých

**203) Jakou metodu byste použili k posouzení, jestli je prodej určitého zboží v průběhu dne rovnoměrně rozložen?**

- rozptyl nebo řetězový index

**204) Posuďte, která varianta vztahů mezi koeficienty je možná:**

**a) byx=-5 , bxy=-0,2**

**b) ryx=-0,72 , bxy=-2,1**

**c) byx=0,5 , bxy=2,1**

**d) byx=-0,1 , bxy=0,7**

Možná je varianta A - Nesmí mít opačná znaménka a jejich součin musí být menší nebo roven jedné.

**205) Je možné, aby se korelační koeficient rovnal korelačnímu poměru? Kdy?**

- ano. Když je daná závislost naprosto shodná se svou regresní přímkou.

**206) Na jakém principu je založena konstrukce měr těsnosti závislosti?**

- na principu rozkladu rozptylů

**207) U kterých z uvedených regresních funkcí lze k odhadu parametrů použít metodu nejmenších čtverců: přímka, parabola, exponenciála ? Názor vysvětlete.**

- u přímky a paraboly, protože jsou lineární v parametrech na rozdíl od exponenciály, která není - tam je nutná lineární transformace.

**208) Jak určíme ve vícenásobné lineární regresní funkci podíl jednotlivých nezávisle proměnných v regresním odhadu závisle proměnné**

– **β koeficienty regresní fce** (dílčí regresní koeficienty na které jsou přepočítány nezávislé odchylky)- o kolik směrodatných odchylek se změní závisle proměnná jestliže se nezávisle proměnná změní o 1 odchylku

**209) Co je multikolinearita a jak ji zjistíš? (5)**

- je to závislost mezi vysvětlujícími proměnnými
- informace o ní čerpáme z matice korelačních koeficientů a jejího determinantu
- jestliže jsou párově nekorelované, multikolinearita neexistuje a D=1, pokud D=0, potom mluvíme o úplné multikolinearita
- většinou je škodlivá od hodnoty 0,75
- jestliže multikolinearita existuje, tak je některá proměnná x v modelu zbytečná

**210) Co je zdánlivá korelace?**

- zdánlivá korelace spočívá v tom, že je někdy možné pozorovat silnou závislost mezi proměnnými i v případě, kdy mezi proměnnými ve skutečnosti závislost buď skoro, nebo vůbec neexistuje.

**211) Doprovodná regrese**

- u analýzy určujeme, zda kromě regrese na kvantitativním znaku existuje i regrese na kvalitativním znaku - f-test

**212) Čím měříme těsnost závislosti**

- u lineárních závislostí - koef korelace, determinace
- u nelineárních závislostí - index korelace, determinace

**213) Kanonická korelační analýza**

- vycházejí z logického členění proměnných do dvou skupin
- každá skupina proměnných je charakterizována 1 souhrnnou veličinou, tzv. kanonickou proměnnou = lineární kombinace pův. prom. dané skup.

**214) Jak zjistíme zda je rozdíl korelačních koeficientů statisticky významný**

- tabulky - pomocí testu o korelačním koeficientu
- tr=ryx/(odm(1-r na 2xy)/(n-2))
- tr - má studentovo rozdělení pro n-2 stupňů volnosti
- t větší než t alfa - statisticky významný

**215) z-transformace, kdy ji používáme, její význam**

- hodnoty r1 převádíme na z1 a r2 na z2 u testu o rozdílu 2 korelačních koeficientů
- metoda pro intervalový odhad korelačního koeficientu, jestliže n<100
- při intervalovém odhadu korelačního koef. v případě že n je menší než 100 není možná normální aproximace

**216) Fischerova Z-transformace**

- speciální postup, kdy hodnoty korelačního koeficientu převádíme(pomocí tabulky) na hodnotu Z(r1=Z1, r2=Z2)
- testovací kriterium - u=(Z1/Z2)/(odm(1/(n1-3))+(1/(n2-3)))
- u je větší než u alfa ... H0 se zamítá

**217) Koeficient korelace – co vyjadřuje, jaké známe (4)**

**218) Jak určíme parametry mocninné funkce**

**219) Kdy se používá korelační poměr, jakých nabývá hodnot**

**220) Rozdíl mezi dílčím a totálním korelačním koeficientem**

**221) Rozdíl mezi volnou a pevnou závislostí**

**222) Index korelace – co to je, co vyjadřuje, jak je vymezen, kdy se používá**

**223) Jakých hodnot nabývá korel. index a jak ho interpretujeme; co vyjadřuje index determinace**

- korelační index nabývá hodnot<0;1>, čím více se blíží 1 tím je závislost silnější, čím blíže 0 tím je závislost slabší
- index determinace - I na 2xy - vyjádřen v % udává z kolika % je závisle proměnná ovlivněna uvažovanou nezávisle proměnnou

**224) Uveďte vztah mezi korelačním koeficientem a regresními koeficienty – ryx = √byx.bxy**

**225) Čím měříme těsnost závislosti**

– korelačním koeficientem (lineární regrese), korelačním indexem (nelineární regrese)

**226) Jak se vyjadřuje doprovodná regrese u kovariance?**

**227) Je možné, aby se korelační koeficient rovnal korelačnímu poměru? Kdy?**

Ano. Když je daná závislost naprosto shodná se svou regresní přímkou.

**228) Množné znaky - kontingenční tabulka**

- alespoň 1 znak většinou více než 2, měříme pouze test závislosti; Pearsenův koeficient kontingence vyzívá chí-test

**229) K čemu slouží normovaný koeficient kontingence**

- kontingenční tabulky mají různý počet řádků a sloupců - používáme ho pro porovnávání

**230) Koeficient asociace slouží k:**

- vyjádření těsnosti 2 alternativních znaků

**231) Čím určujeme těsnost závislosti v asociační tabulce**

- koeficient asociace <-1;1>, obdoba koeficientu korelace

**232) Asociační tabulka**

- vyjadřuje vztah 2 alternativních znaků
- asociační přímka, těsnost, koeficient asociace (Yuelův koeficient asociace) v tabulce prvky a a b b
- používají se zde upravené chí-testy

**233) Jaké funkce popisují průběh závislosti v asociační tabulce**

- lineární fce

**234) Jak měříme těsnost závislosti v kontingenční tabulce**

- Pearssonův koef kontingence - C (0,1)
- čupronův koef kont. - K (0,1)
- normovaný koef kontingence - Cn (0,1)
- vyjádřit průběh závislosti v kontingenční tabulce neumíme

**235) Jak ověřujeme závislost znaků v asociační tabulce**

- chí-kvadrát test, fischerův test

**236) Jaké funkce lze použít k popisu průběhu závislosti v asociační tabulce**

- pouze lineární asoc. přímku

**237) Rozdíl mezi korelační a kontingenční tabulkou?**

**238) Kontingenční tabulka, nakreslit obecné schéma**

**239) Fischer-pravděpodobnostní test tabulek**

**240) K čemu slouží normovaný koeficient kontingence**

- kontingenční tabulky mají různý počet řádků a sloupců - používáme ho pro porovnávání
- aby byl dostačující - postačující odhad

**241) Jaké testy se používají v asociační tabulce**

- upravený chý-kvadrát test; fisherův test
- jakými způsoby lze vyjádřit trend časové řady
- graficky, mechanicky (klouzavé prům.), analyticky(trend. fce.)

**242) Charakteristické rysy časových řad**

- uspořádaná řada údajů, které odlišují v čase
- rysy – trend (vývojové tendence v čr-vzestupný, sestupný, stacionární), kolísání (odchylky od rovnoměrného vývoje-periodické(cyklické, sezónní), náhodné(okamžikové, intervalové))

**243) Elementární charakteristiky čas. Řad**

- slouží k rychlé informaci o charakteru a chování ukazatele v čas. řadě

**244) Harmonická analýza, podstata, k čemu se používá a kde**

- soubor postupů používaných při analýze vlnitých pohybů u přírodních jevů

- v některých ČŘ se přesně pravidelně odchylují skutečné hodnoty od celkové tendence
- graf připomíná spojité periodické funkce goniometrického typu

**245) Co je autokorelace? Kde se s ní setkáváme a jak ji měříme?**

- je korelace mezi sousedními odchylkami od trendu
- ověřujeme ji pomocí Durbin- Watsonova testu
- závislost mezi 2 po sobě jdoucími členy v čř
- u takové čř nelze provést měření těsnosti závislosti
- zjišťuje se pomocí koeficient autokorelace (nízké hodnoty - neexistuje autokorelace)

**246) Jak se měří průměr v okamžikové a intervalové ČŘ?**

**247)** Analýza časových řad

- trend x periodická složka x náhodná složka

**248) Kdy používáme klouzavé průměry**

- při mechynickém popisu trendu - očišťuje čř od náhodného a period. kolísání

**249) K čemu používáme klouzavé průměry?**

- k vyrovnávání ČŘ
- nevyrovnáme ji celou najednou, ale po částech

**250) Co si představujete pod pojmem "dekompozice časové řady"?**

- rozklad čas. řady na 4 složky: trend, sezónní kolísání, cyklické kolísání a náhodná složka
- u aditivního modelu to sčítáme, u multiplikativního násobíme

**251) Jaké jsou předpoklady o náhodné složce časové řady? Proč ověřujeme jejich splnění a co to znamená, když se přesvědčíme, že nejsou splněny?**

- kompenzace v rámci časové řady, homoskedasticita (jejich variabilita se v čase nemění), jsou vzájemně nekorelované
- nesplnění = nepodařilo se v časové řadě eliminovat systematickou složku beze zbytku

**252) Co je extrapolace časové řady a jaký má význam?**

- dělání prognóz do budoucnosti
- nejčastější způsob použití extrapolačních kritérií je založen na simulaci spočívá v tom, že z analýzy řady oddělíme určitou část pozorování a na vhodný T funkce usuzujeme podle toho, jak dobře extrapoluje tato pozorování.

**253) Pomocí jakých měr posoudíte přesnost extrapolace?**

- Index korelace a determinace, u lin. záv. pomocí korel. koeficientu. Nebo pokud nemůžu model popsat žádnou závislosí tak poměrem determinace.

**254) Jak vyberete vhodnou metodu pro extrapolaci časové řady?**

- nakreslím graf a posoudím průběh, potom analyzuji růstové charakteristiky a diference

**255) Z údajů časových řad dvou ukazatelů y a x byl vypočten korelační koeficient ryx = 0,8. Jaké závěry z této hodnoty můžeme učinit?**

- mezi y a x existuje středně silná lineární závislost

**256) Z údajů časových řad ukazatele y a ukazatele x byl vypočten korelační koeficient ryx = 0,6. O čem vypovídá?**

Středně silná závislost

**257) Časová řada čtvrtletních údajů je rostoucí a sezónní indexy rostou úměrně trendu. Jak byste posoudili sezónnost takové časové řady?**

- jedná se o proporcionální sezónnost

**258) Co chápeme pod pojmem paralelismus časových řad a co způsobuje?**

Souběh časových řad: Systematické složky časových řad, zejm. celkový vývoj tendence, mohou mít velmi podobný průběh v čase, což může vést k pozorování silné závislosti mezi proměnnými, jež ve skutečnosti vůbec závislé nejsou. Proto je třeba zjistit, zda existuje nějaký vztah mezi náhodnými složkami, a pak teprve pak usuzovat na souv, mezi ČŘ.

**259) Popište základní princip exponenciálního vyrovnávání časové řady.**

Pozorování blízká současnosti jsou pro odhad parametrů důležitější, než pozorování vzdálenější, starší, a měla by jim proto být přisuzována větší váha.

**260) Jaké znáte metody pro získání odhadů parametrů trendových funkcí nelineárních v parametrech?**

- linearizovat model vhodnou transformací, pak požadavek kompenzace kladné a záporné odchylky empirických hodnot od hodnot vyrovnaných a metoda nejmenších čtverců (aby součet čtverců popsaných odchylek byl minimální)

**261) Náhodné kolísání časové řady, jak ho měříme**

- výsledek náhodných jevů, které nelze předurčit
- pomocí abs. prům. odch. a rel. prům. odch.

**262) Jak se určí průměr časové řady okamžikové a intervalové**

- okamžiková: y s pruhem=((y1/2+y2+...+yn-1*yn/2)/(n-1))
- intervalová - počítá se jako prostý ar. průměr

**263) Jakými zp. lze vyjádřit trend časové řady**

- graficky, mechanicky(klouz. prům.), analyticky(trend. fce.)

**264) Jak počítáme v čas. řadě koef. růstu a prům. koef. růstu**

- koef. růstu - Ki=(yi/(yi-1))*100
- prům. koef. růstu - k=ntá odm. z (k1*k2*...*kn)

**265) Sezónní index je**

- poměr skutečných a teoretických hodnot

**266) Extrapolace**

- jedná se o statistické prognózování kdy pomocí trendové fce a sezónních indexů je možno odhadnout budoucí vývoj daného ukazatele

**267) Interpolace**

- přibližné určení chybějící hodnoty sledováného ukazatele v ČR za předpokladu, že známe sousední hodnoty
- provádí se 2 způsoby: prostřednictvím 2 sousedních údajů; prostřednictvím všech hodnot v ČŘ

**268) Popište základní princip exponenciálního vyrovnávání časové řady.**

- Pozorování blízká současnosti jsou pro odhad parametrů důležitější, než pozorování vzdálenější, starší, a měla by jim proto být přisuzována větší váha.

**269) Co jsou odvozené časové řady a aspoň dva typy**

(z druhotných ukazatelů, např. HDP/na obyvatele)

**270) Čím vyjadřujeme vliv sezónního kolísání**

**271) Co je stacionární časová řada**

**272) Odvozené časové řady**

**273) Postup při korelaci časové řady.**

**274) Co jsou bazické a co řetězové indexy a jaký je mezi nimi vztah?**

- jsou to indexy, které se týkají časové posloupnosti
- bazický index má základ v zákl. období, řetězový vždy v předchozím období
- bazické se dají konstruovat pomocí násobení řetězových a řetězové zase dělením bazických

**275) Jaký index (indexy) použijete k posouzení změny cen vyráběné produkce u podniku, který vyrábí 7 různých výrobků?**

- Souhrnný index - různé druhy výrobků nemůžeme sčítat - Laspeyesův, Paascheho, Fischerův...

**276) Která znáte rozdělení spojitých náhodných veličin? Uveďte oblasti jejich použití.**

Normální, Logaritmicko-normální, Exponenciální, Chí-kvadrát, Studentovo, Rozdělení F

**277) Jaká je hlavní myšlenka rozkladu indexů metodou postupných změn?**

Index hodnot analyzovaného ukazatele rozkládáme na dílčí analytické indexy, které se skládají součinem.

**278) Může nastat situace, že index ceny hovězího masa stoupne, i když ve všech prodejnách cena poklesla? Zdůvodněte svůj názor.**

Z ekonomického hlediska ano. Asice v případě, že by indexy cen ostatního zboží a služeb stouply více.

**279) Jak budete postupovat, chcete-li zjistit, jak se na změně průměrné dovozní ceny určitého výrobku podílely změny cen z jednotlivých dovozních zemí a jak změna struktury dovozu podle dovozních zemí?**

Indexem proměnlivého složení

**280) Charakterizujte situaci, kdy použijete složený individuální cenový index a kdy souhrnný cenový index.**

Složený index: Chceme-li zjistit, jak se na změně průměrné ceny určitého výrobku podílely změny jednotlivých cen, a jak změna struktury Souhrnný index: nemůžeme-li sčítat různé druhy výrobků - Laspeyesův, Paascheho, Fischerův...

**281) Podle jakých hledisek dělíme indexy**

- věcné, časové, prostorové

**282) Sezónní index**

- podíl pův. hodnot a hodnot očištěných od sez. vlivů

**283) Souhrnné indexy úrovně - (2)**

- Indexy stejnorodý intenzity ukazatele; nestejnorodý extenzity ukazatele; srovnávající nest. prod.; nest. intenzity ukazatele; ani jedna

**284) Výsledný ukazatel intenzity je stejnorodý pokud:**

- **o**ba intenzivní ukazatele jsou stejnorodé; oba extenzivní ukazatele jsou stejnorodé; oba nestejnorodé; jeden extenziv. uk. je stejnorodý, druhý nestejnorodý

**285) Stejnorodé a nestejnorodé ukazatele**

- stejnorodé - lze li srovnat ext. ukaz. součtem v přir. měrných jedn.
- nestejnorodé - nelze je mezi sebou spočítat

**286) Indexy**

- podíl dvou hodnot téhož ukazatele; individuální a souhrnný; věcný, prostorový, časový
- individuální - indexy stejnorodých ukazatelů
- souhrnné - indexy nestejnorodých ukazatelů
- obě tyto skupiny se dělí na indexy množství a indexy úrovně

**287) Typy stat. ukazatelů**

- okamžikové, intervalové, primární, sekundární, extenzivní, intenzivní, stejnorodé, nestejnorodé

**288) Rozklad indexu hodnoty.**

**289) Rozdíl mezi souhrnným indexem množství a indexem úrovně.**

**290) Fischerův index – k čemu se používá.**

**291) Montgomeryho index – k čemu (3)**

**292) Stejnorodé a nestejnorodé ukazatele**

- stejnorodé - lze li srovnat ext. ukaz. součtem v přir. měrných jedn.
- nestejnorodé - nelze je mezi sebou spočítat

**293) K čemu Guttmanův koeficient prediktability (3)**

**294) Co je to vydatnost?**

**295) K čemu využíváme Theilův koeficient?**

**296) Co vyjadřuje Bayesův vzorec?**

- v případě, že jsou známy nejen nepodmíněné pravděpodobnosti P(Bi), ale i podmíněné pravděp. P(A/Bi) a je známo, že výsledkem je nastoupení jevu A, lze P(Bi) spočítat podle vzorce: P(Bi) = P(Bi)*P(A/Bi) / sum P(Bi)*P(A/Bi)

**297) Zemědělský závod**

- souhrn pozemků větších než 0,1 ha na kterých se hospodaří
- souhrn pozemků menší než 0,1 ha když:

**298) Jak odhadujeme v zemědělství ha výnosy**

- subjektivní odhad - celé území se rozdělí na obvody a v každém se odhaduje prům. výnos
- objektvní metoda - výběrové šetření, přímé měření; sledují se ha výnosy - klouz. prům.

**299) Kultura**

- pozemek s trvale určeným způsobem využívání půd(orná půda, vinice), plodina - brambory rané...

**300) Rozdíl mezi sklizňovou plochou a sklizenou plochou.**

**301) Statistická síť pro zemědělství.**

**302) Co je šetření Agrocenzus a kdy se toto šetření naposled uskutečnilo**

**303) Systém rodinných účtů**

**304) Vyjmenovat sčítací testy**

**305) Agrocenz**

- vychází z prahových hodnot; výběrová šetření; vysoké náklady

**306) Jaká měření provádí statistika v zemědělské rostlinné výrobě**

## Určení kvantilů (známky z testu)

- Describe -> numeric data -> one-variable analysis -> tab: col_1 ▶ data -> ok -> tab ✔ percentiles -> ok -> pk (pravý klik) -> pane option -> tab: percentiles: vl. Hodnoty -> ok
- *Describe – Distribution fitting – Probability distributions* – **specifikace rozdělení** – *Inverse CDF*

## Rozdělení četností -> rozepsat tab. Četností

- Přepsat tab. Do stat. -> ozn. 1. sloupec -> pk -> generate data -> tab: operators: REP[?,?] -> REP[col_1,col_2] -> dále práce jen s 1. sloupcem!

## Míra mutability (barva očí/auta…)

- Nepoužívá se statgraphic -> dosazení do vzorečku

ZNÁME ROZSAH SOUBORU

míra mutability

$$M=\frac{n^2-\sum n_i^2}{n\cdot(n-1)}\in\langle 0,1\rangle$$

NEZNÁME ROZSAH SOUBORU (REL. ČETNOSTI)

nominální variance

$$NOMVAR=1-\sum p_i^2\in\langle 0,1)$$

## Pravděpodobnost (tahání míčků)

a) URČENÍ PRAVDĚPODOBNOSTI S VRACENÍM – BINOMICKÉ ROZDĚLENÍ

- Describe -> distribution fitting -> prob. dist. -> tab: ✔ binominal
  - event prob. – pravděpodobnost nastání jevu A (např. pravděpodobnost, že padne šestka na kostce)
  - trials (π) – počet pokusů
- -> tab: cumulative dist. -> ok -> pane option -> tab: hodnoty výpočtu
- *Právě 2 => x = 2*
- *Nejméně 2 => x ≥ 2*
- *Více než 3 max 5 => 3 < x ≤ 5*

binomické

$$P(x)=\binom{n}{x}\cdot p^x\cdot q^{n-x}$$

$$E(X)=n\cdot p$$

$$D(X)=n\cdot p\cdot q$$

b) URČENÍ PRAVDĚPODOBNOSTI BEZ VRACENÍ – HYPERGEOMETRICKÉ ROZDĚLENÍ

- Describe -> distribution fitting -> prob. dist. -> tab: ✔ hypergeometric -> ok
  - event. Prob. – pravděpodobnost nastání jevu A (např. výběr černého míčku z mixu bílých a černých)
  - trials – počet pokusů
  - pop. size – velikost souboru, ze kterého to počítáme

hypergeometrické

$$P(x)=\frac{\binom{M}{x}\cdot\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$E(X)=n\cdot\frac{M}{N}$$

$$D(X)=n\cdot\frac{M}{N}\cdot\left(1-\frac{M}{N}\right)\cdot\frac{N-n}{N-1}$$

## Pravděpodobnost (návštěva obchodu)

- pravděpodobnost frekvence (jak moc za jakou dobu) - počet poruch stroje za jednu směnu, počet telefonních hovorů na ústředně za jednu hodinu, počet vad na 1m$^2$ koberce
- Describe -> distribution fitting -> probability distribution -> tab: ✔ pisson -> ok ✔ -> tab: mean -> ok -> tab: cumulative dist. -> ok -> počet zákazníků
  - mean (λ lambda) – střední hodnota (počet zákazníků za určitou dobu)

poissonovo

$$P(x)=\frac{\lambda^x}{x!}\cdot e^{-\lambda}$$

$$E(X)=\lambda$$

$$D(X)=\lambda$$

Během 3min nepříjde žádný zákazník
- Jev A... příchod zákazníka
- NV X... počet příchodů zákazníků do prodejny během 3min
- 60min.....20 zákazníků
  3min.....1 zákazník
- lambda = 1

varační koeficient

$$V_x=\frac{s_x}{\bar{x}}$$

- směrodatná odchylka se podílí na průměru x %, jedná se spíše o niž/střed/vyš variabilitu

směrodatná odchylka

$$s_x=\sqrt{s_x^2}=\sqrt{\frac{\sum(x_i-\bar{x})^2}{n}}$$

- Standard deviation (statgraphic) vynásobit $\sqrt{\frac{n-1}{n}}$

## Pravděpodobnost se směrodat. Odch. (váha kapra / plat) – NORMÁLNÍ ROZDĚLENÍ

- Describe -> dist. Fitting -> prob. dist. -> tab:✔ normal -> ok -> tab: mean std. Dev.
  - Mean (μ - ný)– střední hodnota
  - Std. Dev. (σ) – směrodatná odchylka
- -> ok -> pane option -> tab: sledované hodnoty
- *P(x ≥ 2,5) => x > 2,5*
- *P( 2,1 < x < 2,4) => x < 2,4 – x < 2,1*

normální

$$f(x)=\frac{1}{\sigma\cdot\sqrt{2\cdot\pi}}\cdot e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}}$$

$$F(x)=\frac{1}{\sigma\cdot\sqrt{2\cdot\pi}}\cdot\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}}$$

$$E(X)=\mu$$

$$D(x)=\sigma^2$$

## Pravděpodobnost s čas. Horizontem (prac. Úraz/ obsluha zákazníka ) - EXPONENCIÁLNÍ

- např. doba pro obsloužení překročí 5min
- Describe -> dist. Fitting -> prob. dist. -> tab:✔ exponential -> ok -> tab: mean
  - mean – střední hodnota (např. doba na obsloužení zákazníka = 3min)
- -> ok -> tab:✔cumulative dist. -> ok -> pk -> pane option -> hledané obd.
  - P(X>5)= Doba překročí 5min -> pane options: 5

exponenciální

$$f(x)=\frac{1}{\delta}\cdot e^{-\frac{(x-A)}{\delta}}$$

x>A jinak f(x)=0

$$F(x)=1-e^{-\frac{(x-A)}{\delta}}$$

$$E(X)=A+\delta$$

$$D(X)=\delta^2$$

## Rozsah výběru (spolehlivost rozsahu)

- Describe -> distribution fitting -> prob. dist. -> tab:✔ normal -> ok -> tab: std. Dev. Mean
  - std. Dev. – směrodatná odchylka
  - mean – střední hodnota
- -> ok -> tab:✔ inverse CDF

odhad minimálního rozsahu výběru

podle průměru

$$n\geq\frac{u^2_{1-\frac{\alpha}{2}}\cdot\sigma^2}{\Delta^2}$$

podle relativní četnosti

$$n\geq\frac{u^2_{1-\frac{\alpha}{2}}\cdot p(1-p)}{\Delta^2}$$

## Ověření hypotézy (kvalifikace zaměstnanců / rychlost výroby)

- ověření změny – jestli bude v dalším období vyšší % změna
- Describe -> numeric data -> hypotesis test -> tab:✔ binomial proportion -> ok -> tab:
- Ok -> tab: *less than (ověřuju, že změna je menší)*
  *Grather than (ověřuju, že změna je větší)*
  *Not equal (jenom porovnání bez rozdílu lepší/horší)*
  *Alpha: hladina významnosti (%)*
- Když: *P-value > alpha – $H_0$ nezamítám a $H_1$ nepotvrzuji -> neprokáže se*
  *P-value < alpha – $H_0$ zamítáme a $H_1$ přijímáme*

**Null hypotesis:** původní hypotéza
**Sample proportion:** vypočítat ručně m/n
**Sample size:** rozsah souboru

hypotéza o průměru
výběr velký H0: μ = μ0

$$U=\frac{\bar{x}-\mu_0}{\sigma}\cdot\sqrt{n}$$

normované normální
W{u, u1-α/2 ≤ u ≤ uα/2}

hypotéza o průměru
výběr malý

$$t=\frac{\bar{x}-\mu_0}{\acute{s}_x}\cdot\sqrt{n}$$

t s v=n-1 stupni volnosti

hypotéza o rel. četnosti
výběr n>9/π(1-π) H0: π = π0

$$U=\frac{p-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

rozdělení normované normální

hypotéza o rozptylu
H0: σ= σ0

$$\chi^2=\frac{(n-1)\acute{s}^2_x}{\sigma^2_0}$$

χ2 s v=n-1 stupni volnosti

## Potvrzení předpokladu (domníváme se že… , ověřte… )

- Describe -> num. Data -> hypotesis test -> tab:✔ normal mean -> ok
- Tab: *not equal*

Ověřte na hladině významnosti 5% předpoklad, že vozidla vykonají denně v průměru 6 jízd denně. (průměr = 5 jízd, celkem vozidel = 100

**Null hypotesis:** doměnka
**Sample mean:** dřívější prům.
**Sample size:** dřívější rozsah souboru
**Sample sigma:** směrodatná odchylka

*Alpha: míra významnosti*
*Use Z-test: (jen u x > 30)*

## Kontingenční tabulky (porovnávání zaměstnání a ovl. Vlivů)

- např. zjistit, jestli závisí výpověd na dosaženém vzdělání
- Describe -> Categorical data -> Contingency tables
  - Přeházet sloupce
- Výsledek: tests of independence podle P-value
- Výsledek: summary statistics

Contingency coeff.
Cramer's V
} **+** = přímo závisí X **-** = **NE**přímo závisí

G=statistic(test of independence)
asociační tabulka
χ2[1]

$$G=n\cdot\frac{(n_{11}n_{22}-n_{12}n_{21})^2}{n_{.1}n_{.2}n_{1.}n_{2.}}$$

Pearson's R
koeficient asociace

$$r_{AB}=\frac{(n_{11}n_{22}-n_{12}n_{21})}{\sqrt{n_{.1}n_{.2}n_{1.}n_{2.}}}$$

G=statistic(test of independence)
χ2-test nezávislosti

$$G=\sum\sum\frac{(n_{ij}-\acute{n}_{ij})^2}{\acute{n}_{ij}}\qquad \acute{n}_{ij}=\frac{n_{i.}\cdot n_{.j}}{n}$$

χ2 s (r-1)·(s-1) st. v W{G, G ≥ χ21-α(r-1)·(s-1)}

Crammer's V
cramerův koeficient

$$C_C=\sqrt{\frac{G}{n\cdot h}}$$

Contingency coeff
paersonův

$$C_P=\sqrt{\frac{G}{G+n}}$$

**Analýza rozptylu (tabulka bez buněk s číslama v řádku – podobná kontingenční)**

- řádky do sloupců
- označení tabulky
  - Compate – Multiple Samples – Multiple Sample Comparison –OK
    - Multiple Data Columns – OK
    - přeházet sloupce – OK
    - ANOVA Table, Variance check
- Bartletův test
  - $H_0$: $\sigma_1^2 + \sigma_2^2 + \sigma_3^2$ (sigma $1^2$ + …)
  - $H_1$: non $H_0$
  - P-Value (variance check) ≲ α (hladina významnosti)
    - potvrdit a zamítnout hypotézy
  - poměr determinace = $p^2 = \frac{s_{ym}}{s_{yv}} = \frac{between\ groups}{total\ (Corr.)}$ ] obojí ze „Sum of squares“ v „ANOVA table“
    - vyjádření v %

| analýza rozptylu | testové kritérium |
|---|---|
| rozdělení F (k-1) (n-k) | $F = \frac{\frac{S_{ym}}{k-1}}{\frac{S_{yv}}{n-k}}$ |

F = „F-ratio“ v ANOVA table

**TEORIE**

- **parametrické testy** – hypotézy se týkají neznámého parametru základního souboru a při testování vycházíme ze známého pravděpodobnostního rozdělení náhodné veličiny v ZS (např. hod kostkou)
- **neparametrické testy** – hypotézy se týkají různých vlastností základního souboru a test nevyžaduje znalost konkrétního rozdělení v ZS (např. klasifikace známek)
- **koeficienty kontingence** – vyjadřují míru závislosti a její přímost (přímá, nepřímá)
  - Crammerův $C_{Cr} = \sqrt{\frac{G}{n \times h}}$ $C_{Cr} \in \langle 0; 1 \rangle$
  - Pearsonův $C_P = \sqrt{\frac{G}{G+n}}$ $C_P \in \langle 0; 1 \rangle$
- **směrnice regresní přímky** – udává průměrnou změnu proměnné „y“ odpovídající zvýšení proměnné „x“ o jednotku (např. Pokud se „x“ zvýší o jednotku, tak se proměnná „y“ v průměru zvýší o 0,95)
- **absolutní člen regresní přímky** je ve stejných jednotkách, jako závislá proměnná „y“ (domníváme se, čistě teoreticky to nemusí být správně)
- **poměr determinace** – měření intenzity závislosti proměnné „y“ na faktoru „x“
  - $P^2 = \frac{S_{ym}}{S_y}$ $P^2 \in \langle 0; 1 \rangle$
- **index determinace** – udává, jaký podíl variability proměnné „y“ lze vysvětlit zvolenou regresní funkcí – udává se v %
  - $I^2 = \frac{S_T}{S_y}$ $I^2 \in \langle 0; 1 \rangle$
- **spojitá NV** – veličina, jejíž hodnota je určena výsledkem náhodného pokusu a má nekonečně mnoho hodnot (tj. v desetinných číslech)
- **diskrétní NV** – veličina, jejíž hodnota je určena výsledkem náhodného pokusu a hodnoty tvoří čísla z množiny celých čísel
- **Bartlettův test** – slouží k ověření shody rozptylů a setkáme se s ním u analýzy rozptylu
- **kdy je vhodné použít Binomické rozdělení** – při zkoumání pravděpodobnosti náhodného jevu A, kdy je tato pravděpodobnost konstantní (např. výběr s vracením)
- **kdy je vhodné použít Hypergeometrické rozdělení** – při zkoumání pravděpodobnosti s výběrem bez vracení
- **graf intervalového rozdělení četnosti** – histogram
- **graf diskrétního rozdělení četnosti** – polygon četnosti
- **bezrozměrné charakteristiky variability** – variační koeficient, giniho koeficient
- **nominální proměnná** – nelze charakterizovat pomocí průměru
- **kardinální proměnná** – lze charakterizovat pomocí průměru