

Přehled vzorců

1. Základy počtu pravděpodobnosti

Pravděpodobnost klasická:

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

Pravděpodobnost statistická:

$$P(A) = \lim_{N \rightarrow \infty} \frac{M}{N}$$

Pravděpodobnost – vlastnosti: $0 \leq P(A) \leq 1$ $P(U) = 1$ $P(V) = 0$ $P(\bar{A}) = 1 - P(A)$

Věta o sčítání pravděpodobností:

Jevy slučitelné:

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

Jevy neslučitelné:

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

Podmínka nezávislosti jevu A na B:

$$P(A/B) = P(A)$$

Pravděpodobnost podmíněná:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad P(B/A) = \frac{P(A \cap B)}{P(A)}$$

Věta o násobení pravděpodobností:

Jevy závislé:

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

Jevy nezávislé:

$$P(A \cap B) = P(A) \times P(B)$$

Funkce distribuční – vlastnosti:

$$0 \leq F(X) \leq 1 \quad F(-\infty) = \lim_{x \rightarrow -\infty} F(X) = 0 \quad \text{a} \quad F(\infty) = \lim_{x \rightarrow \infty} F(X) = 1$$

$$F(x_1) \leq F(x_2) \quad P(x_1 \leq X < x_2) = F(x_2) - F(x_1)$$

Hustota pravděpodobnosti – vlastnosti: $f(x) \geq 0$ $\int_{-\infty}^{\infty} f(x) dx = 1$ $P(a < X < b) = F(b) - F(a)$

Funkce distribuční sdružená – vlastnosti:

$$0 \leq F(x, y) \leq 1 \quad F(x, -\infty) = F(-\infty, y) = F(-\infty, -\infty) = 0 \quad F(\infty, \infty) = 1$$

Střední hodnota náhodné veličiny:

Pro diskrétní náhodné veličiny:

$$E(X) = \sum_{i=1}^n x_i p_i$$

Pro spojitě náhodné veličiny:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

Rozptyl náhodné veličiny:

$$D(X) = E[X - E(X)]^2$$

Pro diskrétní náhodné veličiny:

$$D(X) = \sum_{i=1}^n [x_i - E(X)]^2 p_i$$

Pro spojitě náhodné veličiny:

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

Střední hodnota náhodné veličiny – vlastnosti:

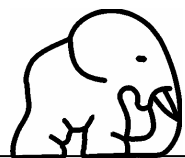
$$E(C) = C \quad E(aX + b) = aE(X) + b \quad E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

Rozptyl náhodné veličiny – vlastnosti:

$$D(C) = 0 \quad D(a + bX) = b^2 \cdot D(X) \quad D(X_1 \pm X_2 \pm \dots \pm X_n) = D(X_1) + D(X_2) + \dots + D(X_n)$$

Výpočetní tvar rozptylu: $D(X) = E(X^2) - [E(X)]^2$



Směrodatná odchylka: $\sigma(X) = \sqrt{D(X)}$

Rozdělení diskretních náhodných veličin:

Typ rozdělení	Parametry	E(X)	D(X)	Pravděpodobnostní funkce
Alternativní	p	p	$p(1-p)$	$P(X = k) = p^k (1-p)^{1-k}$ pro $k=0, 1$
Binomické	n, p	np	$np(1-p)$	$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ pro $k=0, 1, \dots, n$
Poissonovo	λ	$\lambda = np$	λ	$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$ pro $k=0, 1, \dots$
Hypergeometrické	M, N, n	np, kde $p = \frac{M}{N}$	$np(1-p) \frac{N-n}{N-1}$	$P(X = m) = \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}}$ pro $m=0, 1, \dots, \min n, M$

Rozdělení spojitých náhodných veličin:

Typ rozdělení	Parametry	E(X)	D(X)	Hustota pravděpodobnosti, distribuční funkce
Normální (obecné)	μ, σ^2	μ	σ^2	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ pro $x \in (-\infty, +\infty)$
Normované normální	0,1	0	1	$\varphi(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}}$ pro $u \in (-\infty, +\infty)$
Rovnoměrné	a, b	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$f(x) = \frac{1}{b-a}$ $F(x) = \frac{x-a}{b-a}$ pro $x \in \langle a, b \rangle$
Exponenciální	λ	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$f(x) = \lambda \cdot e^{-\lambda x}$ pro $x \geq 0$ $F(x) = 1 - e^{-\lambda x}$ pro $x \geq 0$ $f(x) = 0$ pro $x < 0$ $F(x) = 0$ pro $x < 0$

2. Náhodný výběr

Sturgesovo pravidlo:

Počet tříd:

$$k \approx 1 + 3,3 \log_{10} n$$

Nebo: $k = \sqrt{n}$

Délka intervalů:

$$h = \frac{R}{k} \text{ - na rozhraní intervalů pozorování do intervalu se sudým pořadovým číslem.}$$

- kde R je variační rozpětí.



Charakteristiky polohy:

Aritmetický průměr prostý:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Aritmetický průměr vážený:

$$\bar{x} = \frac{\sum_{i=1}^k x_i n_i}{\sum_{i=1}^k n_i} \quad \text{kde } n = \sum_{i=1}^k n_i$$

Medián:

Rozsah n je vyjádřen lichým číslem:

$$\tilde{x} = \frac{n+1}{2} - \text{bereme prostřední hodnotu.}$$

Rozsah n je vyjádřen sudým číslem:

Průměr z dvou prostředních hodnot.

Charakteristiky variability absolutní:

Výběrové variační rozpětí: $R = x_{\max} - x_{\min}$

Výběrový rozptyl prostý:

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

Výpočtové tvary výběrového rozptylu prostého:

$$s^2 = \frac{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i}{n-1} \quad s^2 = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n-1}$$
$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right]$$

Pracujeme s netříděným souborem

Výběrový rozptyl vážený:

$$s^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 n_i}{n-1} \quad s^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2}{n-1}$$

Pracujeme s utříděným souborem.

Výpočtový tvar výběrového rozptylu váženého:

$$s^2 = \frac{\sum_{i=1}^k x_i^2 n_i - \bar{x} \sum_{i=1}^k x_i n_i}{n-1}$$

Výběrová směrodatná odchylka: $s = \sqrt{s^2}$

Průměrná absolutní odchylka:

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

Průměrná relativní odchylka:

$$\bar{d}' = \frac{\bar{d}}{\bar{x}} \cdot 100[\%]$$

Charakteristiky variability relativní:

Výběrový variační koeficient: $V = \frac{s}{\bar{x}} \cdot 100[\%]$, kde s je směrodatná odchylka a \bar{x} aritmetický průměr.

Kvartilové rozpětí:

$$IQR = \tilde{x}_{0,75} - \tilde{x}_{0,25}$$

Kvartilová odchylka:

$$s_k = \frac{\tilde{x}_{0,75} + \tilde{x}_{0,25}}{2}$$

Rozdělení výběrových statistických charakteristik:

Střední hodnota náhodné veličiny \bar{x} :

$$E(\bar{x}) = E\left(\frac{1}{n} \cdot \sum_{i=1}^n x_i\right) = \frac{1}{n} \cdot \sum_{i=1}^n E(x_i) = \frac{1}{n} \cdot n \cdot \mu = \mu$$

Rozptyl výběrového aritmetického průměru \bar{x} :

$$D(\bar{x}) = D\left(\frac{1}{n} \cdot \sum_{i=1}^n x_i\right) = \frac{1}{n^2} \cdot \sum_{i=1}^n D(x_i) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$



Výběrová rozdělení:

χ^2 rozdělení (chí-kvadrát):

$$\chi^2 = U_1^2 + U_2^2 + \dots + U_n^2$$

Tabulky: hodnoty χ_α^2 definované

vztahem $P(\chi^2 > \chi_\alpha^2) = \alpha$.

Studentovo t-rozdělení:

$$t = \frac{U}{\sqrt{\frac{V}{n}}}$$

Tabulky: hodnoty $T_\alpha(n)$ vztahu

$$P(|t| > t_\alpha(n)) = \alpha$$

F-rozdělení:

$$F = \frac{V_1}{n_1} : \frac{V_2}{n_2}$$

Tabulky: hodnoty $F_\alpha(n_1, n_2)$ vztahu

$$P(F > F_\alpha(n_1, n_2)) = \alpha$$

3. Teorie odhadu

Označení charakteristik:

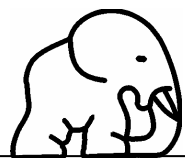
Charakteristika	Výběrový soubor	Základní soubor
Průměr	\bar{x}	μ
Rozptyl	s^2	σ^2
Směrodatná odchylka	s	σ
Medián	m/n	M
Relativní četnost	\bar{x}	p
Rozsah souboru	n	N

Interval spolehlivosti 100(1- α)% : $P(T_1 \leq 0 \leq T_2) = 1 - \alpha$

Intervalový odhad průměru základního souboru:

$P(\bar{x} - \Delta < \mu < \bar{x} + \Delta) = 1 - \alpha$, kde Δ je přípustná chyba odhadu.

Typ výběru	Interval spolehlivosti pro průměr	
známe σ^2 jedná se o výběr s opakováním	levostranný	$\bar{x} - u_{2\alpha} \frac{\sigma}{\sqrt{n}} < \mu$
	pravostranný	$\mu < \bar{x} + u_{2\alpha} \frac{\sigma}{\sqrt{n}}$
	oboustranný	$\bar{x} - u_\alpha \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + u_\alpha \frac{\sigma}{\sqrt{n}}$
neznáme σ^2 jedná se o výběr s opakováním	levostranný	$\bar{x} - t_{2\alpha} \frac{s}{\sqrt{n}} < \mu$
	pravostranný	$\mu < \bar{x} + t_{2\alpha} \frac{s}{\sqrt{n}}$
	oboustranný	$\bar{x} - t_\alpha \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_\alpha \frac{s}{\sqrt{n}}$
známe σ^2 jedná se o výběr bez opakování	levostranný	$\bar{x} - u_{2\alpha} \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} < \mu$
	pravostranný	$\mu < \bar{x} + u_{2\alpha} \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$
	oboustranný	$\bar{x} - u_\alpha \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} < \mu < \bar{x} + u_\alpha \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$
neznáme σ^2 jedná se o výběr bez opakování	levostranný	$\bar{x} - t_{2\alpha} \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} < \mu$
	pravostranný	$\mu < \bar{x} + t_{2\alpha} \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$
	oboustranný	$\bar{x} - t_\alpha \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} < \mu < \bar{x} + t_\alpha \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$



S vrácením:

$$\Delta = u_{\alpha} \sqrt{\frac{\sigma^2}{n}} \quad \Delta = t_{\alpha(n-1)} \sqrt{\frac{s^2}{n}}$$

Odhad průměru μ při známém průměru σ^2 :

Interval spolehlivosti pro průměr:

$$\mu - (\bar{x} - \Delta, \bar{x} + \Delta)$$

Bez vrácení:

$$\Delta = u_{\alpha} \sqrt{\frac{\sigma^2}{n}} \sqrt{\frac{N-n}{N-1}} \quad \Delta = t_{\alpha(n-1)} \sqrt{\frac{s^2}{n}} \sqrt{\frac{N-n}{N-1}}$$

Přípustná chyba:

$$\Delta = u_{\alpha} \frac{\sigma}{\sqrt{n}}, \text{ kde } u_{\alpha} \text{ je kritická hodnota}$$

=> u_{α} z tabulek str. 149 příloha 5.

Odhad průměru základního souboru μ při normálním rozptylu σ^2 :

Interval spolehlivosti pro průměr:

$$(\bar{x} - \Delta, \bar{x} + \Delta)$$

Přípustná chyba:

$$\Delta = u_{\alpha(n-1)} \frac{s}{\sqrt{n}}, \text{ kde } s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

=> $t_{\alpha(n-1)}$ z tabulek str. 153 příloha 7.

Intervaly:

Oboustranný

$$(\bar{x} - \Delta, \bar{x} + \Delta)$$

Jednostranný – pravostranný:

$$(-\infty, \bar{x} + \Delta)$$

Jednostranný – levostranný:

$$(\bar{x} - \Delta, +\infty)$$

Chyba u jednostranných intervalů:

Známe σ^2 :

$$\Delta = u_{2\alpha} \frac{\sigma}{\sqrt{n}}$$

Neznáme σ^2 :

$$\Delta = u_{2\alpha(n-1)} \frac{s}{\sqrt{n}}$$

Stanovení rozsahu souboru:

Rozsah výběrového souboru:

$$n = \frac{u_{\alpha}^2 \sigma^2}{\Delta^2} = \left(\frac{u_{\alpha} \sigma}{\Delta} \right)^2$$

Rozsah výběrového souboru (bez opakování):

$$n = \frac{u_{\alpha}^2 \sigma^2 \cdot N}{\Delta^2 (N-1) + u_{\alpha}^2 \sigma^2}$$

Náhodný výběr dvoufázový.

1. fáze – rozptyl:

$$s^2 = \frac{1}{m-1} \sum (x_i - \bar{x})^2$$

$$s = \sqrt{\frac{1}{m-1} \sum (x_i - \bar{x})^2}$$

1. fáze – rozsah výběrového souboru:

$$\text{S opakováním: } n = \frac{t_{\alpha}^2 (m-1) \cdot s^2}{\Delta^2} \quad n = \left(\frac{t_{\alpha} s}{\Delta} \right)^2$$

$$\text{Bez opakování: } n = \frac{t_{\alpha}^2 (m-1) \cdot s^2 \cdot N}{\Delta^2 (N-1) + t_{\alpha}^2 (m-1) \cdot s^2}$$

=> $t_{\alpha(m-1)}$ z tabulek str. 153 příloha 7.

2. fáze – závěr:

$m \geq n$ – není nutné provádět další šetření

$m < n$ – nutno doplnit předvýběr o $n-m$ jednotek na požadovaný rozsah n .

Intervalový odhad rozptylu σ^2 normálně rozděleného základního souboru:

100(1- α)% interval spolehlivosti pro rozptyl σ^2		
neznáme μ	levostranný	$\frac{(n-1) \cdot s^2}{\chi_{\alpha(n-1)}^2} < \sigma^2$
	pravostranný	$\sigma^2 < \frac{(n-1) \cdot s^2}{\chi_{1-\alpha(n-1)}^2}$
	oboustranný	$\frac{(n-1) \cdot s^2}{\chi_{\frac{\alpha}{2}(n-1)}^2} < \sigma^2 < \frac{(n-1) \cdot s^2}{\chi_{1-\frac{\alpha}{2}(n-1)}^2}$

Interval spolehlivosti: $\left(\frac{(n-1) \cdot s^2}{\chi_{\frac{\alpha}{2}(n-1)}^2}, \frac{(n-1) \cdot s^2}{\chi_{1-\frac{\alpha}{2}(n-1)}^2} \right)$



Vzhledem k hodnotě s^2 není interval spolehlivosti symetrický.

100(1-α)% interval spolehlivosti pro směrodatnou odchylku σ		
neznáme μ	levostranný	$s \sqrt{\frac{n-1}{\chi^2_{\alpha(n-1)}}} < \sigma$
	pravostranný	$\sigma < s \sqrt{\frac{n-1}{\chi^2_{1-\alpha(n-1)}}}$
	oboustranný	$s \sqrt{\frac{n-1}{\chi^2_{\frac{\alpha}{2}(n-1)}}} < \sigma < s \sqrt{\frac{n-1}{\chi^2_{1-\frac{\alpha}{2}(n-1)}}}$

Interval spolehlivosti: $P \left(\frac{s}{1 + \frac{u_\alpha}{\sqrt{2n}}} < \sigma < \frac{s}{1 - \frac{u_\alpha}{\sqrt{2n}}} \right) \approx 1 - \alpha$

Intervalový odhad rozptylu:

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

Intervalový odhad směrodatné odchylky:

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

Intervalový odhad parametru p alternativního rozdělení:

100(1-α)% interval spolehlivosti pro parametr p		
výběr s opakováním	levostranný	$\frac{m}{n} - u_{2\alpha} \cdot \sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n}} < p$
	pravostranný	$p < \frac{m}{n} + u_{2\alpha} \cdot \sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n}}$
	dvoustranný	$\frac{m}{n} - u_\alpha \cdot \sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n}} < p < \frac{m}{n} + u_\alpha \cdot \sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n}}$
výběr bez opakování	levostranný	$\frac{m}{n} - u_{2\alpha} \cdot \sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n} \cdot \frac{N-n}{N-1}} < p$
	pravostranný	$p < \frac{m}{n} + u_{2\alpha} \cdot \sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n} \cdot \frac{N-n}{N-1}}$
	dvoustranný	$\frac{m}{n} - u_\alpha \cdot \sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n} \cdot \frac{N-n}{N-1}} < p < \frac{m}{n} + u_\alpha \cdot \sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n} \cdot \frac{N-n}{N-1}}$
100(1-α)% interval spolehlivosti pro parametr p (s opravou pro spojitost)		
výběr s opakováním	levostranný	$\frac{m}{n} - \frac{1}{2n} - u_{2\alpha} \cdot \sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n}} < p$



	pravostranný	$p < \frac{m}{n} + \frac{1}{2n} + u_{2\alpha} \cdot \sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n}}$
	dvoustranný	$\frac{m}{n} - \frac{1}{2n} - u_{\alpha} \cdot \sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n}} < p < \frac{m}{n} + \frac{1}{2n} + u_{\alpha} \cdot \sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n}}$

Intervalový odhad relativní četnosti:

Malý soubor – $n < 100$:

$np(1-p) \geq 9$ – nutno ověřit podmínku $\Rightarrow N(\mu, \sigma^2)$ normálního rozdělení.

$np(1-p) < 9$ – tabulky str. 163 příloha 11.1 (až str. 170 příloha 11.8).

Velký soubor – $n \geq 100$: $P(f_i - \Delta < \pi < f_i + \Delta) = 1 - \alpha$

S vrácením:

$$\Delta = u_{\alpha} \sqrt{\frac{f_i(1-f_i)}{n}}$$

Bez vrácení:

$$\Delta = u_{\alpha} \sqrt{\frac{f_i(1-f_i)}{n}} \times \sqrt{\frac{N-n}{N-1}}$$

$$\sqrt{\frac{N-n}{N-1}} = \text{konečnostní násobitel}$$

$$n = \frac{u_{\alpha}^2 \times f_i(1-f_i)}{\Delta^2}$$

$$u_{\alpha} = \sqrt{\frac{\Delta^2 n}{f_i(1-f_i)}}$$

Interval pro neparametrický odhad mediánu základního souboru:

$P(x_k \leq M \leq x_{n-k+1}) \geq 1 - \alpha$, kde k z tabulek 148, příloha 3.

4. Testování statistických hypotéz

Testy parametrické jednovýběrové:

Test hypotézy o rozptylu normálního rozdělení:

Testuje se nulová hypotéza:

$$H_0: \sigma^2 = \sigma_0^2$$

Testové kritérium:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

Nulová hypotéza	Alternativa	Kritický obor
$H_0: \sigma^2 = \sigma_0^2$	$H_1: \sigma^2 \neq \sigma_0^2$	$K = \{\chi^2 > \chi_{\frac{\alpha}{2}(n-1)}^2 \text{ nebo } \chi^2 < \chi_{1-\frac{\alpha}{2}(n-1)}^2\}$
	$H_1: \sigma^2 > \sigma_0^2$	$K = \{\chi^2 > \chi_{\alpha(n-1)}^2\}$
	$H_1: \sigma^2 < \sigma_0^2$	$K = \{\chi^2 < \chi_{\alpha(n-1)}^2\}$

$\chi_{\alpha(n-1)}^2$ dle tabulky str. 150 příloha 6.1 (až str. 153 příloha 6.3).

Test hypotézy o průměru normálního rozdělení = jednovýběrový t-test:

✓ Neznáme rozptyl základního souboru:

Testuje se nulová hypotéza:

$$H_0: \mu = \mu_0$$

Testové kritérium:

$$t = \frac{\bar{x} - \mu_0}{s} \sqrt{n}, \text{ kde } s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$$

Nulová hypotéza	Alternativa	Kritický obor
$H_0: \mu = \mu_0$	$H_1: \mu \neq \mu_0$	$K = \{ t > t_{\alpha(n-1)}\}$
	$H_1: \mu > \mu_0$	$K = \{t > t_{2\alpha(n-1)}\}$
	$H_1: \mu < \mu_0$	$K = \{t < -t_{2\alpha(n-1)}\}$

$t_{\alpha(n-1)}$ dle tabulky str. 153 příloha 7.



- ✓ Známe rozptyl základního souboru:

Testuje se nulová hypotéza:

$$H_0: \mu = \mu_0$$

Testové kritérium:

$$U = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n}$$

Nulová hypotéza	Alternativa	Kritický obor
$H_0: \mu = \mu_0$	$H_1: \mu \neq \mu_0$	$K = \{ U > \mu_\alpha\}$
	$H_1: \mu > \mu_0$	$K = \{U > \mu_{2\alpha}\}$
	$H_1: \mu < \mu_0$	$K = \{U < -\mu_{2\alpha}\}$

Test hypotézy o parametru p alternativního rozdělení:

(Test hypotézy o hodnotě relativní četnosti \Rightarrow varianta s π je ze cvičení)

Testuje se nulová hypotéza:

Testové kritérium:

$$H_0: p = p_0$$

$$H_0: \pi = \pi_0$$

$$u = \frac{\frac{m}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad u = \frac{f_i - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

$$\text{V případě absolutních četností } u = \frac{m - np_0}{\sqrt{np_0(1-p_0)}}$$

Nulová hypotéza	Alternativa	Kritický obor
$H_0: p = p_0$	$H_1: p \neq p_0$	$K = \{ u > u_\alpha\}$
$H_0: \pi = \pi_0$	$H_1: p > p_0$ $H_0: \pi > \pi_0$	$K = \{u > u_{2\alpha}\}$
	$H_1: p < p_0$ $H_0: \pi < \pi_0$	$K = \{u < -u_{2\alpha}\}$

u_α dle tabulky str. 149 příloha 5.

Testy parametrické dvouvýběrové:

Srovnání rozptylů dvou normálních rozdělení = F-test: (neznámé rozptyly základního souboru σ_1^2 a σ_2^2)

Testuje se nulová hypotéza:

$$H_0: \sigma^2 = \sigma_2^2$$

Testové kritérium:

$$F = \frac{s_1^2}{s_2^2}, \quad s_1^2 \geq s_2^2$$

Nulová hypotéza	Alternativa	Kritický obor
$H_0: \sigma^2 = \sigma_2^2$	$H_1: \sigma^2 \neq \sigma_2^2$	$K = \{F > F_{\frac{\alpha}{2}(m-1, n-1)}\}$
	$H_1: \sigma^2 > \sigma_2^2$	$K = \{F > F_{\alpha(m-1, n-1)}\}$

F_α dle tabulky str. 154 příloha 8.1 (až str. 159 příloha 8.6).

Porovnání průměrů dvou normálních rozdělení:

- ✓ **Dvouvýběrový t-test – test hypotézy při stejných rozptylech:** (známé rozptyly zákl. souboru σ_1^2 a σ_2^2)

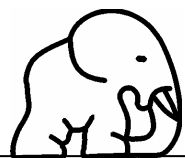
Testuje se nulová hypotéza:

Testové kritérium:

$$H_0: \mu_1 = \mu_2$$

$$t = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{m} + \frac{1}{n}}}, \quad \text{kde } s^2 = \frac{1}{m+n-2} [(m-1)s_1^2 + (n-1)s_2^2]$$

Nulová hypotéza	Alternativa	Kritický obor
$H_0: \mu_1 = \mu_2$	$H_1: \mu_1 \neq \mu_2$	$K = \{ t > t_{\alpha(m-n+2)}\}$
	$H_1: \mu_1 > \mu_2$	$K = \{t > t_{2\alpha(m-n+2)}\}$
	$H_1: \mu_1 < \mu_2$	$K = \{t < -t_{2\alpha(m-n+2)}\}$



✓ **Welchův test – hypotézy při nesterjých rozptylech:** (známé rozptyly zákl. souboru σ_1^2 a σ_2^2)

Testuje se nulová hypotéza:

$$H_0 : \mu_1 = \mu_2$$

Testové kritérium:

$$t = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

Počet stupňů volnosti:

$$f = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n} \right)}{\frac{\left(\frac{s_1^2}{m} \right)^2}{m-1} + \frac{\left(\frac{s_2^2}{n} \right)^2}{n-1}}$$

Zaokrouhlit na celé číslo.

Párový t-test

Diference

$$d_i = x_i - y_i$$

Aritmetický průměr difference:

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i = \bar{x} - \bar{y} = \frac{\sum d_i}{n} = \frac{\sum (x_i - y_i)}{n}$$

Rozptyl:

$$s_d^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2 = \frac{\sum (d_i - \bar{d})^2}{n-1}$$

Testuje se nulová hypotéza:

$$H_0 : \mu_d = 0$$

Testové kritérium:

$$t = \frac{\bar{d}}{s_d} \sqrt{n}$$

Nulová hypotéza	Alternativa	Kritický obor
$H_0 : \mu = \mu_0$	$H_1 : \mu \neq \mu_0$	$K = \{ t > t_{\alpha(n-1)} \}$
	$H_1 : \mu > \mu_0$	$K = \{ t > t_{2\alpha(n-1)} \}$
	$H_1 : \mu < \mu_0$	$K = \{ t < -t_{2\alpha(n-1)} \}$

Test hypotézy o parametrech p_1 a p_2 dvou alternativních rozdělení: (o shodě dvou relativních četností)

Testuje se nulová hypotéza:

$$H_0 : p_1 = p_2$$

Testové kritérium:

$$u = \frac{\frac{m_1}{n_1} - \frac{m_2}{n_2}}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \text{ kde } \bar{p} = \frac{m_1 + m_2}{n_1 + n_2}, q = 1 - \bar{p} \text{ a } n = \frac{n_1 n_2}{n_1 + n_2}$$

Nulová hypotéza	Alternativa	Kritický obor
$H_0 : p_1 = p_2$	$H_1 : p_1 \neq p_2$	$K = \{ u > u_{\alpha} \}$
	$H_1 : p_1 > p_2$	$K = \{ u > u_{2\alpha} \}$
	$H_1 : p_1 < p_2$	$K = \{ u < -u_{2\alpha} \}$

Testy parametrické vícevýběrové:

Porovnání průměrů více než dvou normálních rozdělení = analýza rozptylu

Součet hodnot:

$$x_{i\bullet} = \sum_{j=1}^{n_i} x_{ij}$$

Průměr hodnot:

$$\bar{x}_{i\bullet} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} = \frac{x_{i\bullet}}{n_i}$$

Celkový součet:

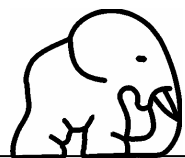
$$x_{\bullet\bullet} = \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij}$$

Celkový průměr:

$$\bar{x} = \frac{x_{\bullet\bullet}}{n}$$

Tečka • nahrazuje index, přes který sčítáme. Celkový součet se značí $x_{\bullet\bullet}$.

Testuje se nulová hypotéza: $H_0 : \mu_1 = \mu_2 = \dots = \mu_n$



Variabilita	Součet čtverců	Stupeň volnosti	Rozptyl	Testové kritérium
Mezi třídami	$S_1 = n \sum_{i=1}^m (\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet})^2 = \sum_{i=1}^m \frac{x_{i\bullet}^2}{n_i} - C = \frac{1}{n} \sum_{i=1}^m x_{i\bullet}^2 - C$	m-1	$S_1^2 = \frac{S_1}{m-1}$	$F = \frac{s_1^2}{s_r^2}$
Uvnitř tříd (reziduální)	$S_r = \sum_{i=1}^m \sum_{j=1}^{n_i} (\bar{x}_{ij} - \bar{x}_{i\bullet})^2 = S - S_1$	m(n-1) n-m	$S_r^2 = \frac{S_r}{n-m}$	
Celková	$S = \sum_{i=1}^m \sum_{j=1}^{n_i} (\bar{x}_{ij} - \bar{x}_{\bullet\bullet})^2 = \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij}^2 - C = \sum_{i,j} x_{ij}^2 - C$	mn-1		

$$C = \frac{x_{\bullet\bullet}^2}{n}$$

Výsledkem je statistika F – podíl rozptylu mezi třídami s_1^2 a rozptylu reziduálního s_r^2 .

Kritický obor na hladině významnosti α učen nerovností $F > F_{\alpha(m-1, m(n-1))}$.

Bodové odhady parametrů: $\mu = \hat{\mu} = \bar{x}_{\bullet\bullet}$, $\hat{\mu}_i = \bar{x}_{i\bullet}$, $\hat{\sigma}_i^2 = s_i^2$, $\hat{\sigma}^2 = s_r^2$.

Oboustranný 100(1- α)% interval spolehlivosti pro μ_i : $\bar{x}_i \pm t_{\alpha(n-m)} \frac{s_r}{\sqrt{n_i}}$

Mnohonásobné porovnávání = podrobnější hodnocení výsledků analýzy rozptylu

✓ Scheffého metoda = S-metoda:

Zamítá se hypotéza $\mu_i = \mu_j$, jestliže $|\bar{x}_{i\bullet} - \bar{x}_{j\bullet}| > \left[\left(\frac{1}{n_i} + \frac{1}{n_j} \right) (m-1) s_r^2 F_{\alpha(m-1, n-m)} \right]^{\frac{1}{2}}$.

Je-li $\bar{x}_{i\bullet} - \bar{x}_{j\bullet}$ větší než pravá strana nerovnosti, μ_i je významně větší než μ_j .

Je-li $\bar{x}_{i\bullet} - \bar{x}_{j\bullet}$ menší než záporně vzatá pravá strana nerovnosti, μ_i je významně menší než μ_j .

✓ Tukeyova metoda = T-metoda:

Vyžaduje, aby pokusný plán byl vyvážený: $|\bar{x}_{i\bullet} - \bar{x}_{j\bullet}| > T s_r$

Porovnání rozptylů více než dvou normálních rozdělení:

✓ Bartlettův test:

Testuje se nulová hypotéza:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2$$

Testové kritérium:

$$B = \frac{1}{C} \left[(n-m) \ln s^2 - \sum_{i=1}^m (n_i - 1) \ln s_i^2 \right], \text{ kde:}$$

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} x_{ij}^2 - n_i \bar{x}_{i\bullet}^2$$

$$s^2 = \frac{1}{n-m} \sum_{i=1}^m (n_i - 1) s_i^2$$

$$C = 1 + \frac{1}{3(m-1)} \left(\sum_{i=1}^m \frac{1}{n_i - 1} - \frac{1}{n-m} \right)$$

Jestliže vypočtená hodnota B překročí hodnotu $\chi_{\alpha(m-1)}^2$, zamítneme nulovou hypotézu na hladině významnosti α .

✓ Hartleyův test:

Testuje se nulová hypotéza:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2$$

Testové kritérium:

$$F_{\max} = \frac{s_{\max}^2}{s_{\min}^2}$$

Kritický obor: $K = \{ F_{\max} > F_{\max; \alpha(k, f)} \}$

Tabulky str. 184 příloha 22.1 a 22.2.

**Cochranův test:**

Testové kritérium:

$$C = \frac{s_{\max}^2}{s_A^2 + s_B^2 + s_C^2}$$

Tabulky str. 177 příloha 14.1 a 14.2.

Kritický obor:

$$C > C_{(m, m-1)} \Rightarrow \text{zamítáme } H_0.$$

Testy dobré shody: **χ^2 -test shody:**Distribuční funkce předpokládaného rozdělení $F_0(x)$: $p_j = P(x_j \leq X < x_{j+1}) = F_0(x_{j+1}) - F_0(x_j)$.Posouzení shody empirického a teoretického rozdělení: $\chi^2 = \sum_{j=1}^k \frac{(n_j - np_j)^2}{np_j}$ - za platnosti H_0 má χ^2 o $f=k-c-1$

stupních volnosti.

Kritický obor: $K = \{\chi^2 > \chi_{\alpha(k-c-1)}^2\}$ **Davidův test normality:**

$$\text{Testové kritérium: } T = \frac{R}{s} = \frac{x_{\max} - x_{\min}}{s} \quad T_d \leq T \leq T_h \Rightarrow H_0$$

Tabulky str. 183 příloha 21.

Testy neparametrické:**Dvouvýběrový Wilcoxonův test:**

Součty:

$$T_x = R_{x_1} + \dots + R_{x_m}$$

$$T_y = R_{y_1} + \dots + R_{y_n}$$

Součty pořadových čísel

Pomocné veličiny:

$$U_x = mn + \frac{m(m+1)}{2} - T_x$$

$$U_y = mn + \frac{n(n+1)}{2} - T_y$$

Testové kritérium:

$$U_0 = \frac{U_x - \frac{1}{2}mn}{\sqrt{\frac{mn}{12}(m+n+1)}} \quad U = \min(U_x, U_y)$$

Tabulky str. 181 příloha 20.1 a str. 182 příloha 20.2.

Wilcoxonův test:Diference: $d_i = x_i - y_i$ Přiřazení pořadových čísel: $R(|d_i|)$ Součty: $W = \min(W^+, W^-) \leq W_\alpha$

Větší rozsah (není v tabulkách):

$$U_w = \frac{W^+ - \frac{1}{4}n(n+1)}{\sqrt{\frac{1}{24}n(n+1)(2n+1)}}$$

Tabulky str. 180 příloha 18.

Kruskal-Wallisův test:

Testové kritérium:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^m \frac{T_i^2}{n_i} - 3(n+1)$$

Kritický obor:

$$H > \chi_{\alpha(k-1)}^2 \Rightarrow \text{zamítáme } H_0.$$

Neparametrické metody mnohonásobného porovnávání – Neményiho metoda:

Doplnění Kruskal-Wallisova testu o Neményiho metodu mnohonásobného pozorování.

$$|T_i - T_j| > \frac{u_{2\alpha}}{m(m-1)} \sqrt{\frac{n(n-1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

Tabulka str. 185 příloha 23.

**Test náhodnosti:**

Číslo x_i ($2 \leq i \leq n-1$) nazveme bodem zvratu v posloupnosti různých čísel, jestliže platí buď $x_{i-1} < x_i > x_{i+1}$ nebo $x_{i-1} > x_i < x_{i+1}$.

Testové kritérium:
$$U = \frac{Z - \frac{2n-4}{3}}{\sqrt{\frac{16n-29}{90}}}$$

Z – celkový počet zvratů v dané posloupnosti.

Dixonův test – test extrémních odchylek:

Testové kritérium:

Prověřujeme minimální hodnotu:

$$Q_1 = \frac{x_{(2)} - x_{(1)}}{x_{(n)} - x_{(1)}}$$

Prověřujeme maximální hodnotu:

$$Q_2 = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}}$$

Kritický obor:

$$Q_1 > Q_{2\alpha} \Rightarrow \text{zamítáme } H_0.$$

$$Q_n > Q_{n\alpha} \Rightarrow \text{zamítáme } H_0.$$

$Q_{1\alpha} = Q_{n\alpha}$ - tabulky str. 180 příloha 19.

Wilcoxon-Whiteův test:

$$T = \min(T_x, T_y)$$

$$T_y = \frac{N(N+1)}{2} - T_x$$

$$T > T_{\alpha(m,n)} \Rightarrow \text{přijímáme } H_0.$$

Tabulky str. 178 příloha 16.1 a str. 179 příloha 16.2

5. Korelační a regresní analýza – statistická analýza vztahů mezi veličinami**Jednoduchá lineární regrese:****Regresní funkce:**

$y_i = f(x_i) + e_i$, kde $f(x)$ je regresní funkce, e_i náhodná (reziduální) odchylka (náhodná chyba).

Regresní funkce lineární:

$$y_i = \alpha + \beta x_i + e_i$$

Regresní přímka (teoretická):

$$f(x) = \alpha + \beta x$$

Rozptýlenost bodů kolem přímky:
$$s_{y \cdot x}^2 = \frac{\sum e_i^2}{n-2} = \frac{\sum (y_i - \hat{y}_i)^2}{n-2}$$

Metoda nejmenších čtverců – bodové odhady a , b parametrů α , β regresní přímky:

$$S = \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \text{ – hledáme minimum, tedy } \frac{\partial S}{\partial \alpha} = 0 \text{ a } \frac{\partial S}{\partial \beta} = 0.$$

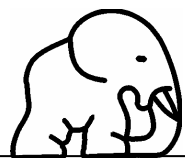
$$s_r^2 = \sum e_i^2 = \sum (y_i - a - bx_i)^2 \Rightarrow \text{optimální hodnoty } a = \bar{y} - b\bar{x} \text{ a } b = r \frac{s_y}{s_x},$$

kde r je korelace, s_x a s_y směrodatné odchylky

$$f(a, b) = \sum_{i=1}^n [y_i - (a + bx_i)]^2 = \min$$

$$\frac{\partial f(a, b)}{\partial a} = \sum_{i=1}^n 2(y_i - a - bx_i)(0 - 1 - 0) = -2 \sum_{i=1}^n (y_i - a - bx_i) \quad -2 \sum_{i=1}^n (y_i - a - bx_i) = 0$$

$$\frac{\partial f(a, b)}{\partial b} = \sum_{i=1}^n 2(y_i - a - bx_i)(0 - 0 - x_i) = -2 \sum_{i=1}^n (y_i - a - bx_i)x_i \quad -2 \sum_{i=1}^n (y_i - a - bx_i)x_i = 0$$



Soustava normálních rovnic:

$$na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$

$$a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i$$

Řešení soustavy normálních rovnic:

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}$$

$$a = \bar{y} - b\bar{x}$$

Rezidua:

$d_i = y_i - y'_i$
 y_i – zjištěné hodnoty Y
 y'_i – vypočtené hodnoty

Reziduální součet čtverců:

$$S_r = \sum_{i=1}^n d_i^2$$

Reziduální rozptyl:

$$s_r^2 = \frac{S_r}{n-2} = \frac{\sum_{i=1}^n (y_i - a - bx_i)^2}{n-2}$$

Oboustranná závislost – závislost Y na X:

$$y' = a_{yx} + b_{yx}x$$

Oboustranná závislost – závislost X na Y:

$$x' = a_{xy} + b_{xy}y$$

Vztah mezi směnicemi regresních přímek b_{yx} a b_{xy} : $r = \sqrt{b_{yx} * b_{xy}}$

Pás spolehlivosti: $y'_i \pm t_{1-\frac{\alpha}{2}} \times s_{y.x}$, kde $t_{1-\frac{\alpha}{2}}$ jsou 100(1- α /2)% kvantily Studentova t-rozdělení s (n-2) stupni

volnosti a $s_{y.x}$ je směrodatná chyba $s_{y.x} = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^n (y_i - y'_i)^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^n y_i^2 - \sum_{i=1}^n y_i y'_i}{n-2}}$, přičemž

$$\sum_{i=1}^n y_i y'_i = \sum_{i=1}^n y_i (a_{yx} + b_{yx} x_i) = a_{yx} \sum_{i=1}^n y_i + b_{yx} \sum_{i=1}^n x_i y_i$$

Testy hypotéz o parametrech lineární regrese a intervalový odhad:

Test hypotézy o hodnotě regresního koeficientu:

Testuje se nulová hypotéza:

$$H_0 : \beta = \beta_0$$

Testové kritérium:

$$t = \frac{b - \beta}{s_b}, \text{ kde } s_b = \sqrt{\frac{s_r^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Nulová hypotéza	Alternativa	Kritický obor
$H_0 : \beta = \beta_0$	$H_1 : \beta \neq \beta_0$	$K = \{ t > t_{\alpha(n-2)} \}$
$(H_0 : \beta_0 = 0)$	$H_1 : \beta > \beta_0$	$K = \{ t > t_{2\alpha(n-2)} \}$
	$H_1 : \beta < \beta_0$	$K = \{ t < -t_{2\alpha(n-2)} \}$

Pokud speciálně $\beta_0 = 0$:

$$H_0 : \beta_0 = 0$$

Testové kritérium:

$$t = \frac{b}{s_b} = \frac{b}{s_r} \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Intervalový odhad regresního koeficientu:

Postup jako u intervalového odhadu.

Oboustranný 100(1- α)% interval spolehlivosti pro regresní koeficient β : $(b - t_{\alpha(n-2)} s_b, b + t_{\alpha(n-2)} s_b)$



Konfidenční pás pro regresní přímku:

Bodový odhad:

$$a + bx_k \text{ pro } \alpha + \beta x_k$$

Intervalový odhad:

$$\left(a + bx_k - t_{\alpha(n-2)} s_y, a + bx_k + t_{\alpha(n-2)} s_y \right), \text{ kde } s_y = s_r \sqrt{\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Test rovnoběžnosti dvou regresních přímek:

Testuje se nulová hypotéza:

$$H_0 : \beta_1 = \beta_2 \text{ proti alternativě } H_0 : \beta_1 \neq \beta_2$$

Testové kritérium:

$$t = \frac{(b_1 - b_2) \sqrt{n_1 + n_2 - 4}}{\sqrt{[(n_1 - 2)s_{1r}^2 + (n_2 - 2)s_{2r}^2] \left[\frac{1}{\sum x_{1i}^2 - n_1 \bar{x}_1^2} + \frac{1}{\sum x_{2i}^2 - n_2 \bar{x}_2^2} \right]}}$$

Kritický obor: $K = \{ |t| > t_{\alpha(n_1+n_2-4)} \}$

Test významnosti regresního koeficientu:

Testuje se nulová hypotéza:

$$H_0 : \beta_{yx} = 0$$

$|t| > t_{\alpha(n-2)} \Rightarrow H_0 \text{ se zamítá}$

Testové kritérium:

$$t = \frac{b_{yx}}{s_{b_{yx}}}, \text{ kde } s_{b_{yx}} = \frac{s_y}{s_x} * \sqrt{\frac{1-r^2}{n-2}}$$

Intervalový odhad regresního koeficientu:

$$P(b_{yx} - t_{\alpha(n-2)} \times s_{b_{yx}} \leq \beta_{yx} \leq b_{yx} + t_{\alpha(n-2)} \times s_{b_{yx}}) = 1 - \alpha$$

Test významnosti regresní přímky:

Variabilita	Součet čtverců	Stupně volnosti	Rozptyl	Testovací kritérium
Regrese	$S_1 = \sum_{i=1}^n (y'_i - \bar{y})^2$	p - 1	$s_1^2 = \frac{S_1}{p-1}$	$F = \frac{s_1^2}{s_r^2}$
Kolem regrese	$S_r = \sum_{i=1}^n (y_i - y'_i)^2$	n - p	$s_r^2 = \frac{S_r}{n-p}$	
Celková	$S = \sum_{i=1}^n (y_i - \bar{y})^2$			

$F > F_{\alpha[(p-1); (n-p)]} \Rightarrow \text{zamítáme } H_0.$

Intervalový odhad regresní přímky:

$$P\left(y'_i - u_{1-\frac{\alpha}{2}} \times s_{\bar{y}} \sqrt{1 + \frac{(x_i - \bar{x})^2}{s_x^2}} \leq y'_j \leq y'_i + u_{1-\frac{\alpha}{2}} \times s_{\bar{y}} \sqrt{1 + \frac{(x_i - \bar{x})^2}{s_x^2}} \right) = 1 - \alpha, \text{ kde } s_x^2 \text{ je rozptyl proměnné } X,$$

$$s_y \text{ směrodatná odchylka proměnné } Y, \quad s_{\bar{y}} = \frac{s_y}{\sqrt{n}}, \quad y'_{j(H,D)} = y'_i \pm t_{1-\frac{\alpha}{2}} \times s_{\bar{y}} \sqrt{1 + \frac{(x_i - \bar{x})^2}{s_x^2}}.$$

Jednorovnicové regresní modely:

- ✓ Model zcela lineární – $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$
- ✓ Model racionální celistvé a lomené funkce:
 - Model regresní paraboly s-tého stupně – $Y = \beta_0 + \beta_1 X^1 + \beta_2 X^2 + \dots + \beta_s X^s + \varepsilon$
 - Model regresní hyperboly s-tého stupně – $Y = \beta_0 + \beta_1 X^{-1} + \beta_2 X^{-2} + \dots + \beta_s X^{-s} + \varepsilon$
- ✓ Model lineární v parametrech – $Y = \beta_0 + \beta_1 f_1 + \dots + \beta_r f_r + \varepsilon$

**✓ Modely převoditelné transformací na lineární model:**

- Model regresní obecně součinnový – $Y = \varepsilon\eta$
- Model regresní lineární exponenciální – $\eta = \beta_0\beta_1X$ nebo $\eta = \exp(\beta_0 + \beta_1X)$
- Model kvadratické exponenciály – $\eta = \exp(\beta_0 + \beta_1X + \beta_2X^2 + \varepsilon)$
- Model regresní obecný lineárně-exponenciální s k vysvětlujícími proměnnými – $\exp(\beta_0 + \beta_1X + \dots + \beta_kX_k + \varepsilon)$.

Nelineární regrese – metoda linearizující transformace:

Základní tvar:

$$f(x) = \alpha + \beta g(x)$$

Aditivní typ funkcí:

Kvadratická – parabola 2. stupně:

$$y'_i = a + bx_i + cx_i^2$$

Kubická – parabola 3. stupně:

$$y'_i = a + bx_i + cx_i^2 + dx_i^3$$

Lineární lomená – hyperbola 1. stupně:

$$y'_i = a + \frac{b}{x_i} \quad f(x) = \alpha + \frac{\beta}{x}$$

Kvadratická lomená – hyperbola 2. stupně:

$$y'_i = a + \frac{b}{x_i} + \frac{c}{x_i^2}$$

Iracionální:

$$y'_i = a + bx_i + \sqrt{cx_i}$$

Logaritmická:

$$y'_i = a + b \log x_i \quad f(x) = \alpha + \beta \ln x$$

Multiplikativní typ funkcí:

Exponenciální:

$$y'_i = a \cdot b^{x_i} \quad f(x) = \alpha \beta^x$$

Mocninná:

$$y'_i = a \cdot x_i^b \quad f(x) = \alpha x^\beta$$

Charakteristiky korelace u nelineární regrese:

- ✓ Rozptyl empirických (skutečně zjištěných) hodnot y: $s_y^2 = \frac{1}{n} \sum (y_i - \bar{y})^2$.
- ✓ Rozptyl vyrovnaných hodnot (teoretický): $s_{y'}^2 = \frac{1}{n} \sum (y'_i - \bar{y})^2$.
- ✓ Rozptyl skutečně zjištěných hodnot kolem regresní čáry = rozptyl empirických hodnot od hodnot vyrovnaných (reziduální): $s_{(y-y')}^2 = \frac{1}{n} \sum (y_i - y'_i - \overline{y - y'})^2 = \frac{1}{n} \sum (y_i - y'_i)^2$.

$$s_y^2 = s_{y'}^2 + s_{(y-y')}^2$$

Podíl složek na empirickém rozptylu:

- ✓ Teoretický rozptyl $s_{y'}^2 = 0$, takže $s_y^2 = s_{(y-y')}^2$.
- ✓ Reziduální rozptyl $s_{(y-y')}^2 = 0$, takže $s_y^2 = s_{y'}^2$.
- ✓ Teoretický rozptyl $s_{y'}^2 \neq 0$ a $s_{(y-y')}^2 \neq 0$, takže $s_y^2 = s_{y'}^2 + s_{(y-y')}^2$.

Měření těsnosti závislosti – korelace:**Výběrový korelační koeficient:**

$$r = \frac{s_{xy}}{s_x s_y}$$

Výběrová kovariance:

$$s_{xy} = \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y})$$

Vztah mezi korelačním koeficientem a regresními koeficienty:

$$r = b_{ys} \frac{s_x}{s_y} \quad r = b_{xy} \frac{s_y}{s_x}$$

Regresní koeficienty:

$$b_{yx} = r \frac{s_y}{s_x} \quad b_{xy} = r \frac{s_x}{s_y}$$

Absolutní hodnota korelačního koeficientu:

$$|r| = \sqrt{b_{yx} b_{xy}}$$

Koeficient determinance:

$$r^2 = \frac{\sum (y_i' - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

**Pearsonův korelační koeficient:**

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\text{cov}(x, y)}{s_x s_y}, \text{ kde } s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Výpočetní tvar korelačního koeficientu:

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] \left[n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right]}}$$

Korelační koeficient – vlastnosti:

$$-1 \leq r \leq 1$$

$|r|=1$ pokud existuje lineární funkční závislost, všechny body leží na přímce.

$r=0$ pokud jsou veličiny lineárně nezávislé, X a Y nazýváme nekorelované proměnné

$r < 0$ – Y se v průměru zmenšuje (platí i obráceně)

Těsnost lineární závislosti:

$0 < |r| \leq 0,3$ slabá závislost

$0,3 < |r| \leq 0,8$ mírná (střední) závislost

$0,8 < |r| \leq 1$ silná závislost

$r < 0,3$ těsnost získá

$0,3 \leq r < 0,5$ těsnost mírná

$0,5 \leq r < 0,7$ těsnost význačná

$0,7 \leq r < 0,9$ těsnost velká

$0,9 \leq r < 1,0$ těsnost velmi vysoká

Rozptyl vyrovnávaných hodnot (teoretický) – vysvětlená variabilita

$$s_{y'}^2 = \frac{1}{n} \sum (y_i' - \bar{y})^2$$

Rozptyl empirických hodnot (zjištěných) – celková variabilita

$$s_y^2 = \frac{1}{n} \sum (y_i - \bar{y})^2$$

Spearmanův korelační koeficient pořadí – koeficient korelace pořadových čísel r_s ($-1 \leq r_s \leq 1$):

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

Index korelace:

$$I_{yx} = \frac{\sqrt{\sum_{i=1}^n (y_i' - \bar{y})^2}}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \sqrt{1 - \frac{\sum_{i=1}^n (y_i - y_i')^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} = \sqrt{\frac{s_{y'}^2}{s_y^2}}$$

Není symetrickou mírou těsnosti závislosti.

Síla korelační závislosti:

$I_{yx} < 0,3$ slabá závislost s malou výstižností

$I_{yx} > 0,8$ velmi silná závislost s velkou výstižností

Index determinance

$$I_{yx}^2 = \frac{s_{y'}^2}{s_y^2} = 1 - \frac{s_{(y-y')}^2}{s_y^2}$$

Poměr determinance:

$$\frac{s_{\bar{y}}^2}{s_y^2} = \frac{s_y^2 - s^2}{s_y^2} = 1 - \frac{s^2}{s_y^2}$$

Korelační poměr:

$$\eta_{yx} = \sqrt{\frac{s_{\bar{y}}^2}{s_y^2}} \quad \eta_{yx} = \sqrt{\frac{s_y^2 - s^2}{s_y^2}} = \sqrt{1 - \frac{s^2}{s_y^2}}$$

Testy hypotéz o korelačním koeficientu a jeho intervalový odhad:**Test významnosti korelačního koeficientu:**

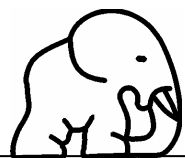
Testuje se nulová hypotéza:

$H_0: \rho = 0$ – zamítnutí znamená, že výběrový koeficient r se statisticky významně liší od nuly \Rightarrow mezi veličinami existuje lineární závislost.

Kritický obor $K = \{|t| > t_{\alpha(n-2)}\}$ pro $H_0: \rho \neq 0$

Testové kritérium:

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2}$$

**Intervalový odhad korelačního koeficientu:** ($n < 100$)

$$P(r - u_{\alpha} * s_r \leq \rho \leq r + u_{\alpha} * s_r) = 1 - \alpha, \text{ kde } s_r = \frac{1 - r^2}{\sqrt{n}}$$

Fischerova transformace: ($n > 100$)

$$r \rightarrow z_r = \arctan h(r) = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) \Rightarrow \text{interval se rozšíří na } -\infty \leq z_{rv} \leq +\infty$$

$$\text{Průměr: } \mu_{z_r} = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) \quad \text{Směrodatná odchylka: } s_{z_r} = \frac{1}{\sqrt{n-3}}$$

$$\text{Dvoustranný interval spolehlivosti: } P(z_r - t_{\alpha(n-2)} * s_{z_r} \leq \mu_{z_r} \leq z_r + t_{\alpha(n-2)} * s_{z_r}) = 1 - \alpha$$

$$\text{Inverzní transformace: } z_r^{-1} : r = \frac{e^{2z} - 1}{e^{2z} + 1}$$

Test hypotézy $H_0: \rho = \rho_0$:

Testuje se nulová hypotéza:

$$H_0 : \rho = \rho_0$$

Testové kritérium:

$$u = (Z - Z_0) \sqrt{n-3}, \text{ kde } Z = \frac{1}{2} \ln \frac{1+r}{1-r} \text{ a } Z_0 = \frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0}$$

Nulová hypotéza	Alternativa	Kritický obor
$H_0 : \rho = \rho_0$	$H_0 : \rho \neq \rho_0$	$K = \{ u > u_{\alpha}\}$
	$H_1 : \rho > \rho_0$	$K = \{u > u_{2\alpha}\}$
	$H_1 : \rho < \rho_0$	$K = \{u < -u_{2\alpha}\}$

Interval spolehlivosti pro korelační koeficient ρ :

$$Z_1 = Z - u_{\alpha} \frac{1}{\sqrt{n-3}}$$

$$Z_2 = Z + u_{\alpha} \frac{1}{\sqrt{n-3}}$$

Test shody dvou koeficientů ρ_1 a ρ_2 :

Testuje se nulová hypotéza:

$$H_0 : \rho_1 = \rho_2$$

Testové kritérium:

$$u = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n+3}}}, \text{ kde } Z_i = \frac{1}{2} \ln \frac{1+r_i}{1-r_i}$$

Nulová hypotéza	Alternativa	Kritický obor
$H_0 : \rho_1 = \rho_2$	$H_0 : \rho_1 \neq \rho_2$	$K = \{ u > u_{\alpha}\}$
	$H_1 : \rho_1 > \rho_2$	$K = \{u > u_{2\alpha}\}$
	$H_1 : \rho_1 < \rho_2$	$K = \{u < -u_{2\alpha}\}$

Pořadová korelace:

$$\text{Spearmanův koeficient pořadové korelace: } r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2-1)} = 1 - A_R \sum_{i=1}^n d_i^2, \text{ kde } d_i = p_i - q_i \text{ a } A_R = \frac{6}{n(n^2-1)}$$

Pokud $|r_s|$ překročí kritickou hodnotu R_{α} , zamítáme hypotézu o nezávislosti pozorovaných náhodných veličin na hladině významnosti α .

Mnohonásobná regrese a korelace:**Model závislosti Y na $X_1 \dots X_k$:**

$y_i = f(x_{i1}, \dots, x_{ik}) + e_i$, $f(\dots)$ je regresní funkce a e_i náhodná chyba i-tého pozorování veličiny Y.

**Regresní funkce:**

$$f(x_{i1}, \dots, x_{ik}) = \eta_1 = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

B – parciální (dílčí) regresní koeficienty

Model mnohonásobné lineární regrese:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + e_i$$

Metoda nejmenších čtverců:

$$S = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_k x_{ik})^2$$

$$y' = b_0 + b_1 x_1 - \dots - \beta_k x_{ik}$$

Maticový zápis:

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_k x_{1k} + e_1$$

$$y = X\beta + e$$

$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_k x_{2k} + e_2$$

...

$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_k x_{nk} + e_n$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{bmatrix} \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}$$

$$S = (y - X\beta)'(y - X\beta) = e'e$$

$$X'Xb = X'y$$

$$b = (X'X)^{-1}(X'y)$$

Reziduální rozptyl:

$$s_r^2 = \frac{\sum_{i=1}^n (y_i - y'_i)^2}{n - k - 1} = \frac{y'y - b'X'y}{n - k - 1}$$

Koeficient mnohonásobné determinance

$$R^2_{y \cdot x_1 \dots x_k} = \frac{\sum_{i=1}^n (y'_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{- odmocnina je koeficient mnohonásobné korelace}$$

Test významnosti:

Testuje se nulová hypotéza:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

Testové kritérium:

$$F = \frac{R^2}{1 - R^2} \times \frac{n - k - 1}{k}$$

Kritický obor:

$$K = \{F > F_{\alpha(k; n-k-1)}\}$$

Koeficient parciální korelace:

$$r_{xy \cdot z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{yz}^2)}}$$

- písmena před tečkou označují veličiny, jejichž závislost sledujeme
- písmena za tečkou označují veličiny, jejichž vliv byl vyloučen

Testování průkaznosti koeficientu vícenásobné korelace:

Testuje se nulová hypotéza:

$$H_0: \rho_{y, x_1 x_2 \dots x_k} = 0$$

Testové kritérium:

$$F = \frac{r_{y \cdot x_1 x_2 \dots x_k}^2 (n - k - 1)}{(1 - r_{y \cdot x_1 x_2 \dots x_k}^2) k}$$

Testování průkaznosti dílčí korelace:

Testuje se nulová hypotéza:

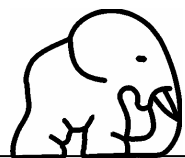
$$H_0: \rho_{y x_r \cdot x_1 x_2 \dots x_{r-1} x_{r+1} \dots x_k} = 0$$

Testové kritérium:

$$t_{(n-k-1)} = \left| r_{y x_r \cdot x_1 x_2 \dots x_{r-1} x_{r+1} \dots x_k} \right| \cdot \sqrt{\frac{n - k - 1}{1 - r_{y x_r \cdot x_1 x_2 \dots x_{r-1} x_{r+1} \dots x_k}^2}}$$

Testování průkaznosti vícenásobné regresní funkce:

$$F_{(k; n-k-1)} = \frac{s_1^2}{s_r^2}$$



Zdroj variability	Součet čtverců	Stupně volnosti	Rozptyl
Regrese	$S_1 = S_c - S_r$	k	$s_1^2 = \frac{S_1}{k}$
Reziduum	$S_r = \sum_{i=1}^n (y_i - y'_i)^2$	n - k - 1	$s_r^2 = \frac{S_r}{n-k-1}$
Celkem	$S_c = \sum_{i=1}^n (y_i - \bar{y})^2$	n - 1	

Vícenásobná exponenciální funkce:

$$y'_i = b_0 \prod_{r=1}^k b_r^{x_{ri}} \quad y'_i = b_0 \times b_1^{x_{1i}} \times b_2^{x_{2i}} \times \dots \times b_k^{x_{ki}} \quad y'_i = \log b_0 + x_{1i} \log b_1 + x_{2i} \log b_2 + \dots + x_{ki} \log b_k$$

Soustava normálních rovnic:

Aditivní tvary:

$$\sum_{i=1}^n (y_i - y'_i)^2 = \min$$

Multiplikativní tvary:

$$\sum_{i=1}^n (\log y_i - \log y'_i)^2 = \min$$

Kritérium M:

$$M = \frac{\frac{F}{\sum_{j=1}^k t_j^2} - 1}{\frac{F}{\sum_{j=1}^k t_j^2} + 1}$$

Farrarův-Glauberův test:

$$B = - \left[(n-1) - \frac{1}{6}(2k+5) \right] \times \ln |R|, \text{ kde}$$

$|R|$ - determinant korelační matice

6. Analýza kategoriálních dat

Porovnání relativní četnosti s teoretickou hodnotou:

Testuje se nulová hypotéza:

$$H_0: p = p_0$$

Testové kritérium:

$$u = \frac{\frac{m}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Nulová hypotéza	Alternativa	Kritický obor
$H_0: p = p_0$	$H_1: p \neq p_0$	$K = \{ u > u_{\alpha} \}$
	$H_1: p > p_0$	$K = \{ u > u_{2\alpha} \}$
	$H_1: p < p_0$	$K = \{ u < -u_{2\alpha} \}$

Intervalový odhad relativní četnosti:

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



Porovnání dvou relativních četností:

Testuje se nulová hypotéza:

$$H_0: p_1 = p_2$$

$|u| > u_\alpha \Rightarrow H_0$ zamítáme

$H_0: (p_1 - p_2) = \Delta$, příp. $= 0$

$H_1: (p_1 - p_2) \neq \Delta$, příp. $\neq 0$

$$u = \frac{(p_1 - p_2) - \Delta}{s_{(p_1 - p_2)}}, \text{ kde:}$$

Testové kritérium:

$$u = \frac{\frac{m_1}{n_1} + \frac{m_2}{n_2}}{\sqrt{\bar{p} \cdot (1 - \bar{p}) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Podmínka	Směrodatná odchylka	Hypotéza
$\Delta \neq 0$	$s_{(p_1 - p_2)} = \sqrt{\frac{f_1(1 - f_1)}{n_1} + \frac{f_2(1 - f_2)}{n_2}}$	$ u > u_\alpha \Rightarrow H_0$ zamítáme
$\Delta = 0$	$s_{(p_1 - p_2)} = \sqrt{pq \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}, \text{ kde}$	$p = \frac{m_1 + m_2}{n_1 + n_2}$

Dvoustranný interval spolehlivosti: $(p_1 - p_2) \in (f_1 - f_2) \pm u_\alpha \cdot s_{(p_1 - p_2)}$

Arcussinová transformace: $\varphi(p) = \arcsin \sqrt{p}$

$$\text{Hypotéza o rovnosti pravděpodobností: } z = \frac{\varphi(p_1) - \varphi(p_2)}{28,648 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

χ^2 – test dobré shody:

Testuje se nulová hypotéza:

$$H_0: F(x) = F_0(x)$$

Testové kritérium:

$$\chi^2 = \sum_{j=1}^k \frac{(n_j - np_j)^2}{np_j}$$

Pearsonův koeficient kontingence:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

Normalizovaný koeficient kontingence:

$$C_n = \frac{C}{C_{\max}}, \text{ kde } C_{\max} = \sqrt{\frac{r-1}{r}}$$

Síla závislosti:

Cramerův koeficient = Cramerovo V:

$$V = \sqrt{\frac{\chi^2}{n(h-1)}}$$

Čuprovův koeficient kontingence:

$$K = \sqrt{\frac{\chi^2}{n\sqrt{(r-1)(s-1)}}}$$

Asociační přímká:

Rovnice asociační přímky: $\frac{(a+c)}{n} = A_{BA} + B_{BA} \cdot \frac{(a+b)}{n}$, kde A_{BA} je absolutní člen, B_{BA} je regresní koeficient.

A je nezávisle proměnná, B je závisle proměnná:

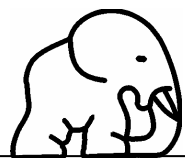
$$B_{BA} = \frac{n \cdot a - [(a+b) \cdot (a+c)]}{(a+b) \cdot (c+d)}$$

$$A_{BA} = \frac{(a+c)}{n} - B_{BA} \cdot \frac{(a+b)}{n}$$

A je závisle proměnná, B je nezávisle proměnná:

$$B_{AB} = \frac{n \cdot a - [(a+b) \cdot (a+c)]}{(a+c) \cdot (b+d)}$$

$$A_{AB} = \frac{(a+b)}{n} - B_{AB} \cdot \frac{(a+c)}{n}$$

**Koeficient asociace (V, rab):**

$$V = \frac{n \cdot a - [(a+b)(a+c)]}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} = \frac{a \cdot d - b \cdot c}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \quad V = \sqrt{B_{AB} \cdot B_{BA}}$$

Yuleův koeficient asociace:

$$Q = \frac{a \cdot d - b \cdot c}{a \cdot d + b \cdot c} \quad Q \in < -1; 1 >$$

Koeficient koligace Y:

$$Y = \frac{1 - \sqrt{\frac{b \cdot c}{a \cdot d}}}{1 + \sqrt{\frac{b \cdot c}{a \cdot d}}} \quad Y \in < -1; 1 >$$

Analýza párových dichotomických proměnných – McNemarův test:

Testuje se nulová hypotéza:

$$H_0: p_{\cdot 1} = p_1$$

Testové kritérium:

$$\chi^2 = \frac{(b - c)^2}{b + c}$$

Cochranův test:

$$Q = \frac{t(t-1) \sum_i T_i^2 - (t-1)N^2}{tN - \sum_j B_j^2}$$

Bowkerův test:

$$\chi^2 = \frac{\sum (n_{ij} - n_{ji})^2}{(n_{ij} + n_{ji})}$$

7. Analýza časových řad**Časová řada intervalová – průměr:**

- ✓ Prostý aritmetický průměr jednotlivých hodnot – všechny intervaly jsou stejně dlouhé.
- ✓ Vážený aritmetický průměr – intervaly mají různou délku.

Časová řada okamžiková – průměr:

$$\text{Prostý: } \bar{y} = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + \dots + \frac{y_{n-1} + y_n}{2}}{n-1} = \frac{\frac{1}{2}y_1 + y_2 + y_3 + \dots + y_{n-1} + \frac{1}{2}y_n}{n-1}$$

$$\text{Vážený: } \bar{y} = \frac{\frac{y_1 + y_2}{2}(t_2 - t_1) + \frac{y_2 + y_3}{2}(t_3 - t_2) + \dots + \frac{y_{n-1} + y_n}{2}(t_n - t_{n-1})}{t_n - t_1}$$

Elementární charakteristiky časových řad:

$$\text{První absolutní difference: } \Delta_t^1 = y_t - y_{t-1} \quad t = 2, 3, \dots, n$$

$$\text{Druhé absolutní difference: } \Delta_t^2 = \Delta_t^1 - \Delta_{t-1}^1 = (y_{t+2} - y_{t+1}) - (y_{t+1} - y_t) \quad t = 3, 4, \dots, n$$

$$\text{Tempo růstu (řetězové indexy): } k_t = \frac{y_t}{y_{t-1}} \quad t = 2, 3, \dots, n$$

$$\text{Průměrný index růstu: } \bar{k} = \sqrt[n-1]{k_1 \cdot k_2 \cdot \dots \cdot k_{n-1}} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

$$\text{Tempo přírůstku: } r_t = \frac{\Delta_t^1}{y_{t-1}} = \frac{y_t - y_{t-1}}{y_{t-1}}$$

$$\text{Koeficienty zrychlení: } z_t = \frac{\Delta_t^2}{\Delta_{t-1}^1}$$

$$\text{Bazické indexy: } BI = \frac{y_t}{y_0}$$

**Analytické vyrovnávání pomocí trendových funkcí:**

Lineární:	$T_t = a + b \cdot t$	Mocninná:	$T_t = a \cdot t^b$
Kvadratická:	$T_t = a + b \cdot t + c \cdot t^2$	Odmocninná:	$T_t = a + b \cdot \sqrt{t}$
Logaritmická:	$T_t = a + b \cdot \log t$	Kombinovaná:	$T_t = a + b \cdot t + c \cdot \sqrt{t}$
Exponenciální:	$T_t = a \cdot b^t$	Logistická:	$T_t = \frac{k}{1 + e^{a+bt}}$

Trend lineární:

Trendová přímka:

$$y'_t = a + bt$$

Parametry trendu:

$$a = \frac{\sum y_t}{n} - b \cdot \frac{\sum t}{n} = \bar{y} - b \cdot \bar{t} \quad b = \frac{n \sum t \cdot y_t - \sum t \sum y_t}{n \sum t^2 - (\sum t)^2}$$

Trend parabolický:

Trendová přímka:

$$y'_t = a + bt + ct^2$$

Parametry trendu:

$$\begin{aligned} \sum y_t &= na + b \sum t + c \sum t^2 & a &= \frac{\sum y_t \sum t'^4 - \sum t'^2 \sum y_t \cdot t'^2}{n \sum t'^4 - (\sum t'^2)^2} \\ \sum t \cdot y_t &= a \sum t + b \sum t^2 + c \sum t^3 & b &= \frac{\sum y_t \cdot t'}{\sum t'^2} \\ \sum t^2 \cdot y_t &= a \sum t^2 + b \sum t^3 + c \sum t^4 & c &= \frac{n \sum y_t \cdot t'^2 - \sum t'^2 \sum y_t}{n \sum t'^4 - (\sum t'^2)^2} \end{aligned}$$

Trend exponenciální:

Trendová přímka:

$$y'_t = a \cdot b^t$$

Parametry trendu:

$$\log y'_t = \log a + t \cdot \log b$$

$$\sum \log y_i = n \cdot \log a + \log b \sum t$$

$$\sum t \cdot \log y_i = \log a \sum t + \log b \sum t^2$$

$$\log a = \frac{\sum \log y_i}{n}$$

$$\log b = \frac{\sum t' \cdot \log y_i}{\sum t'^2}$$

Trend exponenciální modifikovaný – trendová přímka: $T_t = k + ab_t$

$$\text{Trend logistický: } T_t = \frac{k}{1 + a \cdot e^{-bt}} \quad T_t = \frac{k}{1 + e^{a+bt}} \quad T_t = \frac{k}{1 + e^{-bt}} \quad T_t = \frac{k}{1 + a \cdot b^t}$$

Gompertzova křivka: $T_t = k \cdot a^{b^t}$ **Posuzování kvality modelu:**

Směrodatná odchylka, index determinance, index korelace...

Střední chyba odhadu ME:

$$ME = \frac{\sum (y_t - y'_t)}{n}$$

Střední čtvercová chyba MSE:

$$MSE = \frac{\sum_{t=1}^n (y_t - y'_t)^2}{n} = \sum \frac{e_t^2}{n-k}$$

RMSE:

$$RMSE = \sqrt{MSE}$$

Střední absolutní chyba MAE:

$$MAE = \frac{1}{n} \sum |y_t - y'_t|$$



Střední procentuální chyba MPE:

$$MPE = \frac{100}{n} \sum \left(\frac{y_t - y'_t}{y_t} \right)$$

Střední absolutní procentuální chyba MAPE: $MAPE = \frac{100}{n} \sum \left| \frac{y_t - y'_t}{y_t} \right|$

$$H = \frac{\sqrt{MSE}}{\bar{Y}}$$

Statistická verifikace trendového modelu:

$T_t = a + bt \Rightarrow$ odhad trendu $\mu_t = \alpha + \beta t$

Testuje se nulová hypotéza:

$H_0: \alpha = 0$

Testové kritérium:

$$t = \frac{a}{s_a}, \text{ kde } s_a = s \sqrt{\frac{\sum_{t=1}^n t^2}{n \sum_{t=1}^n t^2 - (n \cdot \bar{t})^2}}$$

$H_0: \beta = 0$

$$t = \frac{b}{s_b}, \text{ kde } s_b = \frac{s}{\sqrt{\sum_{t=1}^n t^2 - n \cdot (\bar{t})^2}}$$

$H_0: \beta = 0: |t| > t_{\alpha}$ pro $(n-2)$ stupně volnosti $\Rightarrow H_0$ se zamítá ve prospěch alternativní hypotézy $H_0: \beta \neq 0$

Test hypotézy o existenci sezónnosti:

Testuje se nulová hypotéza: Testové kritérium:

$H_0: \beta_j = 0$

$$F = \frac{m \sum_{j=1}^r (\bar{y}_{\cdot j} - \bar{y})^2}{(r-1) \cdot \sigma^2}, \text{ kde } \sigma^2 = \frac{\sum_{i=1}^m \sum_{j=1}^r (y_{ij} - \bar{y})^2 - r \sum_{i=1}^m (\bar{y}_{i \cdot} - \bar{y})^2 - m \sum_{j=1}^r (\bar{y}_{\cdot j} - \bar{y})^2}{(r-1)(m-1)}$$

Chyby předpovědi:

$$\Delta = \Delta_m - \Delta_p$$

Chyba ex ante – modelová chyba předpovědi:

$$\Delta_m = P_{t+i} + Y_{t+i}$$

Chyba ex post - chyba vlastního prognostického modelu:

$$\Delta_p = Y_{t+i} - y_{t+i}$$

Posouzení kvality vyrovňávání:

Reziduální odchylka:

$$s_{y'_t} = \sqrt{\frac{\sum_{t=1}^n (y_t - y'_t)^2}{n - p - 1}}$$

Variační koeficient:

$$v_{y'_t} = \frac{s_{y'_t}}{\bar{y}} \cdot 100$$

Předpověď:

Intervalová:

$$P(y'_{t+i} - t_{\alpha} \cdot s_{y'_{t+i}} \leq y_{t+i} \leq y'_{t+i} + t_{\alpha} \cdot s_{y'_{t+i}}) = 1 - \alpha$$

Bodová:

$$s_{y'_{t+i}} = s_y \cdot \sqrt{(1 - I^2) \cdot \frac{n(n^2 - 1) + 12i^2}{(n^2 - 1) \cdot (n - 2)}}$$

Absolutní chyba předpovědi:

Rozdíl mezi předpovídanou (P_{t+i}) a skutečnou hodnotou (y_{t+i}).

$\Delta_{t+i} < 0$ podceňující předpověď

$\Delta_{t+i} > 0$ nadceňující předpověď

$\Delta_{t+i} = 0$ bezchybná předpověď



Průměrná chyba předpovědi:

$$\bar{D} = \frac{1}{m} \sum_{t=1}^m (P_{t+i} - y_{t+i}), \text{ kde}$$

$$\bar{P} = \frac{1}{m} \sum_{t=1}^m P_{t+i}, \quad \bar{y} = \frac{1}{m} \sum_{t=1}^m y_{t+i}$$

Čtvercová chyba předpovědi:

$$S_{\Delta}^2 = \frac{1}{m} \sum_{t=1}^m (P_{t+i} - y_{t+i})^2$$

Průměrná hodnota chyby předpovědi:

$$S_{\Delta} = \sqrt{S_{\Delta}^2} = \sqrt{\frac{1}{m} \sum_{t=1}^m (P_{t+i} - y_{t+i})^2}$$

Relativní chyba předpovědi:

$$\delta_{t+i} = \frac{\Delta_{t+i}}{y_{t+i}} = \frac{P_{t+i} - y_{t+i}}{y_{t+i}}$$

Theilův koeficient nesouladu:

$$T^2 = \frac{\sum_{t=1}^m (P_{t+i} - y_{t+i})^2}{\sum_{t=1}^m y_{t+i}^2}$$

Relativní chyba předpovědi v procentech:

$$T = \sqrt{T^2} = \sqrt{\frac{\sum_{t=1}^m (P_{t+i} - y_{t+i})^2}{\sum_{t=1}^m y_{t+i}^2}}$$

Durbin-Watsonův test autokorelace:

$$d = \frac{(e_2 - e_1)^2 + \dots + (e_{n-1} - e_{n-2})^2 + (e_2 - e_{n-1})^2}{e_1^2 + e_2^2 + \dots + e_n^2} = \frac{\sum_{t=1}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

$4 - d_L < DW < 4$	H_0 se zamítá – autokorelace
$4 - d_U < DW < 4 - d_L$	Neumíme rozhodnout, je třeba zvýšit n
$2 < DW < 4 - d_U$	Přijímá se H_0 – autokorelace není
$d_U < DW < 2$	Přijímá se H_0 – autokorelace není
$d_L < DW < d_U$	Neumíme rozhodnout, je třeba zvýšit n
$0 < DW < d_L$	H_0 se zamítá – autokorelace