

Vysoká škola ekonomická
FAKULTA INFORMATIKY a STATISTIKY
Katedra statistiky a pravděpodobnosti

STATISTIKA

Vzorce pro 4ST201 a 4ST210
poznámky ke vzorcům:



STATISTICKY
NEKLASICKY

Verze: 4.5

Poslední změna: 17.2.2025

©KSTP, VŠE 2025

Popisná statistika

kvantily Excel: PERCENTIL.INC()

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

$$\tilde{x}_p, 0 < P < 1 \quad x_{(z_p)}, \quad nP < z_p < nP + 1,$$

$\tilde{x}$ - medián, Excel: MEDIAN()

označení kvantilů

$$\frac{x_{(z_p)} + x_{(z_p+1)}}{2}, nP = z_p$$

$\hat{x}$ - modus - nejčetnější hodnota, Excel: MODE.MULT()

průměry

Prostý arit. průměr

$$i = 1, 2, \dots, n \quad \text{prostý } \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Excel: PRŮMĚR()

vážený arit. průměr

$$\text{vážený } w_i \geq 0 \quad \bar{x}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

$x_j, j = 1, 2, \dots, k$, četnosti n_j

Absolutní četnost

relativní četnost

$$p_i = \frac{n_i}{n} \quad \sum_{i=1}^k n_i = n \quad \sum_{j=1}^k p_j = 1$$

celkový počet pozorování

Sočet všech relativních četností musí být 1 (100%)

vážený arit. průměr

$$\bar{x} = \frac{\sum_{j=1}^k n_j x_j}{n}$$

Vážený aritmetický průměr s relativními četnostmi

$$\bar{x} = \sum_{j=1}^k p_j x_j$$

Prostý harmonický průměr

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

Excel: HARMEAN()

Vážený harmonický průměr

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

Vážený harmonický průměr s relativními četnostmi

$$\bar{x}_H = \frac{1}{\sum_{i=1}^k \frac{p_i}{x_i}}$$

Geometrický průměr

$$\bar{x}_G = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}, x_i > 0, \quad \text{Excel: GEOMEAN()}$$

Mezikvartilové rozpětí

rozpětí $r = x_{(n)} - x_{(1)}, q = \tilde{x}_{0,75} - \tilde{x}_{0,25}$

variační rozpětí $\uparrow$ max $\uparrow$ min $\uparrow$ 3. kvartil $\uparrow$ 1. kvartil

Poznámka: Excel počítá kvantil jako (funkce PERCENTIL.INC())

$\lfloor h \rfloor$ je dolní celá část, $\lceil h \rceil$ je horní celá část

$$\tilde{x}_p, 0 < P < 1 \quad \tilde{x}_p = x_{(\lfloor h \rfloor)} + (h - \lfloor h \rfloor)(x_{(\lceil h \rceil)} - x_{(\lfloor h \rfloor)})$$



populační

rozptyl

Průměr

Prostý tvar

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

Výpočtový tvar

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

Excel: VAR.P()

Vážený rozptyl

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

Vážený průměr

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

vážený průměr

Vážený rozptyl s rel. četnostmi

$$s_x^2 = \sum_{i=1}^k p_i (x_i - \bar{x})^2$$

$$s_x^2 = \overline{x^2} - \bar{x}^2 = \sum_{i=1}^k p_i x_i^2 - \left[\sum_{i=1}^k p_i x_i \right]^2$$

Výpočtový tvar váženého rozptylu s relativními četnostmi

Směrodatná odchylka, variační koeficient, MAD

Směrodatná odchylka
Excel: SMODCH.P()

$$s_x = \sqrt{s_x^2}$$

$$v_x = \frac{s_x}{\bar{x}}, \bar{x} > 0$$

$$MAD = \text{median}(|x_i - \tilde{x}_{0.5}|)$$

Průměr ze skupinových průměrů, k skupin

relativní četnost

$$p_j = \frac{n_j}{n}, \sum_{j=1}^k n_j = n, \sum_{j=1}^k p_j = 1, j = 1, 2, \dots, k$$

$$\bar{x} = \frac{\sum_{j=1}^k n_j \bar{x}_j}{n}$$

$$\bar{x} = \sum_{j=1}^k p_j \bar{x}_j$$

vážený průměr

Vážený průměr s relativními četnostmi

Rozklad rozptylu

$$s_x^2 = \overline{s^2} + s_{\bar{x}}^2 = \frac{\sum_{j=1}^k n_j s_{xj}^2}{n} + \frac{\sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2}{n}$$

vnitroskupinový rozptyl, meziskupinový rozptyl, celkový průměr - s absolutními četnostmi

$$s_x^2 = \overline{s^2} + s_{\bar{x}}^2 = \sum_{j=1}^k p_j s_{xj}^2 + \sum_{j=1}^k p_j (\bar{x}_j - \bar{x})^2$$

- s relativními četnostmi

Transformace

$x_1, x_2, \dots, x_n \rightarrow \bar{x}, s_x^2, v_x, c$ konstanta:

vlastnosti:

průměru rozptylu

$$y_i = c + x_i, i = 1, 2, \dots, n \rightarrow \bar{y} = c + \bar{x}, s_y^2 = s_x^2$$

$$z_i = c x_i, i = 1, 2, \dots, n \rightarrow \bar{z} = c \bar{x}, s_z^2 = c^2 s_x^2, v_z = v_x$$



Pravděpodobnost a náhodné veličiny

Teorie pravděpodobnosti

Výpočet pravděpodobnosti $P(A) = \frac{m}{n}$ *Doplňek pravděpodobnosti (Jev opačný)* $P(\bar{A}) = 1 - P(A)$ *m = příznivé jevy* *n = celkový počet jevů*

Slučitelné jevy $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ *Sjednocení* *Průnik*

Neslučitelné jevy $P(A \cup B) = P(A) + P(B)$ *Sjednocení* *Průnik*

Podmíněná pravděpodobnost $P(A|B) = \frac{P(A \cap B)}{P(B)}$ *Průnik* *Podmínka*

* $P(A \cap B) \leq \min(P(A), P(B))$ *Podmínka* *Průnik* *soačín* ** P(A) nezávisí na P(B) - podmínka B nemá vliv (protože jsou nezávislé)*

Nezávislé jevy $P(A \cap B) = P(A) \cdot P(B)$, $P(A|B) = P(A)$, $P(B|A) = P(B)$ ** Jen prokrožené A a B*

Závislé jevy $P(A \cap B) = P(A) P(B|A) = P(B) P(A|B)$ *Průnik - Vychází ze vzorce podmíněné pravděp.*

* *Průnik musí být vždy menší nebo roven minimu pravděpodobností*

Náhodná veličina (náhodná veličina X ; hodnota této náhodné veličiny x)

Diskrétní náhodné veličiny

Pravděpodobnostní funkce $P(x) = P(X = x)$ *Distribuční funkce* $F(x) = P(X \leq x) = \sum_{x_i \leq x} P(x_i)$

Pr. že X leží mezi x_1 a x_2 $P(x_1 < X \leq x_2) = \sum_{x_1 < x \leq x_2} P(x) = F(x_2) - F(x_1)$

Střední hodnota $E(X) = \sum_x x P(x)$ *Suma soačín x a P(x)*

Rozptyl $D(X) = \sum_x x^2 P(x) - \left[\sum_x x P(x) \right]^2$

Směrodatná odchylka $\sqrt{D(X)}$ *odmocnina rozptylu*

Spojité náhodné veličiny

Derivace dist. funkce $f(x) = F'(x)$ *Distribuční funkce* $F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$

hustota pravděp. $P(x_1 < X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x) dx$ *Výpočet přes dist. funkce* *Výpočet přes Integrál*

Pr. že X leží mezi x_1 a x_2 $x_p, 0 < P < 1$ $F(x_p) = P$ $x_p = F^{-1}(P)$ *Výpočet kvantila*

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

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

Poznámka: v anglické literatuře a také někdy v české se používá $Var(X)$ místo $D(X)$.



Pravděpodobnostní rozdělení

Pravděpodobnostní rozdělení diskrétních náhodných veličin

Binomické rozdělení: $X \sim Bi(n, \pi)$, $n = 1, 2, \dots$

$P(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}$ $x = 0, 1, 2, \dots, n$, $0 < \pi < 1$, **Excel:** BINOM.DIST()

Pravděpodobnost (blue arrow from $\binom{n}{x}$ to $P(x)$)
velikost výběru (blue arrow from n to $\binom{n}{x}$)
konstantní pravděpodobnost (blue arrow from π to π^x)

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

Střední hodnota (blue arrow from $E(X)$ to $n\pi$)
Rozptyl (blue arrow from $D(X)$ to $n\pi(1-\pi)$)

Pravděpodobnostní rozdělení spojitých náhodných veličin

Normované normální rozdělení: $U \sim N(0, 1)$, $-\infty < u < \infty$

$E(U) = 0$ $D(U) = 1$

$E(x) = \text{střední hodnota (průměr)}$
 $D(x) = \text{rozptyl}$

$\Phi(u) = 1 - \Phi(-u)$, $u_p = -u_{1-p}$ **Excel:** NORM.S.INV()

Excel: NORM.S.DIST() (green arrow from $\Phi(u)$ to **Excel:** NORM.S.DIST())
Kvantil norm. normálního rozdělení (blue arrow from u_p to u_{1-p})

Normální rozdělení: $X \sim N(\mu, \sigma^2)$, $-\infty < x < \infty$ $-\infty < \mu < \infty$ $\sigma^2 > 0$

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

$U = \frac{X - \mu}{\sigma}$

$u = \frac{x - \mu}{\sigma}$

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

Distribuční funkce (blue arrow from $F(x)$ to $\Phi(u)$)
Excel: NORM.DIST() (green arrow from $F(x)$ to **Excel:** NORM.DIST())

Excel: NORM.INV()

$x_p = \mu + \sigma u_p$ $0 < P < 1$

Kvantil normálního rozdělení (blue arrow from x_p to u_p)

$P(x_1 < X \leq x_2) = P\left(\frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} \leq \frac{x_2 - \mu}{\sigma}\right) = P(u_1 < U \leq u_2) = \Phi(u_2) - \Phi(u_1)$

Chí-kvadrát rozdělení: $G \sim \chi^2(\nu)$, $\nu = 1, 2, \dots$, $g > 0$

Excel - Kvantil: CHISQ.INV()

t-rozdělení (Studentovo rozdělení): $T \sim t(\nu)$, $\nu = 1, 2, \dots$, $-\infty < t < \infty$ $t_p(\nu) = -t_{1-p}(\nu)$

Excel - Kvantil: T.INV()

F-rozdělení (Fisherovo – Snedecorovo rozdělení): $F \sim F(\nu_1, \nu_2)$, $\nu_1, \nu_2 = 1, 2, \dots$, $F > 0$

Excel - Kvantil: F.INV()

Poznámka: V anglické literatuře se používá Z místo U , rovněž kvantil se značí z_p místo u_p .



Statistická indukce: bodový a intervalový odhad parametrů

výběrový rozptyl

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

Průměr *výběv. směr. odchylka* *pop. směr. odchylka*
Počet pozorování *pop. rozptyl*

$\hat{S}^2$: Excel: VAR.S()
 S : Excel: SMODCH.VÝBĚR.S()

Bodové odhady parametrů

Střední hodnota (populační průměr) μ

$X \sim N(\mu, \sigma^2)$: $\hat{\mu} = \bar{x}$, směrodatná chyba $\frac{\hat{\sigma}}{\sqrt{n}} = \frac{S}{\sqrt{n}}$

normální rozdělení *odhad stř. hodnoty* *výběv. směr. odchylka* *počet pozorování*

Obecné rozdělení: $\widehat{E}(X) = \bar{x}$, směrodatná chyba $\frac{\hat{\sigma}}{\sqrt{n}} = \frac{S}{\sqrt{n}}$

očekávaná stř. hodnota *výběv. směr. odchylka* *počet pozorování*

Úhrn, N jednotek, $NE(X): \widehat{NE}(X) = N\bar{x}$, směrodatná chyba $N \frac{\hat{\sigma}}{\sqrt{n}} = N \frac{S}{\sqrt{n}}$

počet pozorování *počet pozorování ve VS* *počet pozorování v ZS*

(Populační) rozptyl σ^2

$X \sim N(\mu, \sigma^2)$: $\hat{\sigma}^2 = S^2$

normální rozdělení *odhad pop. rozptylu*

Pravděpodobnost (populační podíl) π

$\hat{\pi} = p = \frac{m}{n}$, směrodatná chyba $\sqrt{\frac{p(1-p)}{n}}$

počet výskytů jevu *počet pozorování ve VS* *výběv. podíl*

Úhrn (N jednotek) $N\pi$ - pop. podíl

počet pozorování v ZS

$\widehat{N\pi} = N\hat{\pi} = Np = \frac{Nm}{n}$, směrodatná chyba $N \cdot \sqrt{\frac{p(1-p)}{n}}$

počet výskytů jevu *výběv. podíl* *počet pozorování ve VS*

VS: výběrový soubor
 ZS: základní soubor



Statistická indukce: bodový a intervalový odhad parametrů

Intervalové odhady parametrů (spolehlivost intervalového odhadu $1 - \alpha$)

Střední hodnota (populační průměr)

Normální rozdělení

dvoustranný interval spolehlivosti pro parametr μ :

$$\left(\underbrace{\bar{x} - t_{1-\alpha/2}(n-1)}_{\text{průměr}} \underbrace{\frac{S}{\sqrt{n}}}_{\text{státní volnosti}} \underbrace{\left[t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}} \right]}_{\text{směr odchylka vs}} \right) \rightarrow \text{Excel: CONFIDENCE.T()}$$

Kvantil Studentova rozdělení

jednostranné intervaly spolehlivosti:

$$\begin{array}{ll} \text{dolní mez} & \bar{x} - t_{1-\alpha}(n-1) \frac{S}{\sqrt{n}} \\ \text{horní mez} & \bar{x} + t_{1-\alpha}(n-1) \frac{S}{\sqrt{n}} \end{array}$$

levostranný interval pravostanný interval

Úhrn, N jednotek:

dvoustranný interval spolehlivosti

$$N \left(\bar{x} - t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}}, \bar{x} + t_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}} \right) = \left(N\bar{x} - Nt_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}}, N\bar{x} + Nt_{1-\alpha/2}(n-1) \frac{S}{\sqrt{n}} \right)$$

jednostranné intervaly spolehlivosti:

$$\begin{array}{ll} \text{dolní mez} & N \left(\bar{x} - t_{1-\alpha}(n-1) \frac{S}{\sqrt{n}} \right) \\ \text{horní mez} & N \left(\bar{x} + t_{1-\alpha}(n-1) \frac{S}{\sqrt{n}} \right) \end{array}$$

poče ugnásobíme finální intervaly hodnotou N (velikosti pop.)

Obecné rozdělení, velký náhodný výběr

$$\left(\bar{x} - u_{1-\alpha/2} \frac{S}{\sqrt{n}}, \bar{x} + u_{1-\alpha/2} \frac{S}{\sqrt{n}} \right) \rightarrow \text{Excel: CONFIDENCE.NORM()}$$

Kvantil norm. normálního rozdělení

jednostranné intervaly spolehlivosti:

$$\begin{array}{ll} \text{dolní mez} & \bar{x} - u_{1-\alpha} \frac{S}{\sqrt{n}} \\ \text{horní mez} & \bar{x} + u_{1-\alpha} \frac{S}{\sqrt{n}} \end{array}$$

levostranný interval pravostanný interval

Úhrn, N jednotek:

dvoustranný interval spolehlivosti

$$N \left(\bar{x} - u_{1-\alpha/2} \frac{S}{\sqrt{n}}, \bar{x} + u_{1-\alpha/2} \frac{S}{\sqrt{n}} \right) = \left(N\bar{x} - Nu_{1-\alpha/2} \frac{S}{\sqrt{n}}, N\bar{x} + Nu_{1-\alpha/2} \frac{S}{\sqrt{n}} \right) * \rightarrow \text{Výsledek stačí ugnásobit } N$$

jednostranné intervaly spolehlivosti:

$$\begin{array}{ll} \text{dolní mez} & N \left(\bar{x} - u_{1-\alpha} \frac{S}{\sqrt{n}} \right) \\ \text{horní mez} & N \left(\bar{x} + u_{1-\alpha} \frac{S}{\sqrt{n}} \right) \end{array}$$

levostranný interval pravostanný interval



Pravděpodobnost (populační podíl)

velký náhodný výběr

dvoustranný interval spolehlivosti:

$$* \left(\underset{\substack{\uparrow \\ \text{podíl vs}}}{p - u_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}}, p + \left[u_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right] \right) \rightarrow \text{Excel: CONFIDENCE.NORM()}$$

jednostranné intervaly spolehlivosti:

$$\begin{array}{ll} \text{dolní mez} & p - u_{1-\alpha} \sqrt{\frac{p(1-p)}{n}} \quad \text{horní mez} \quad p + u_{1-\alpha} \sqrt{\frac{p(1-p)}{n}} \\ \uparrow & \uparrow \\ \text{levostranný interval} & \text{pravostranný interval} \end{array}$$

Úhrn (N jednotek) velký náhodný výběr

dvoustranný interval spolehlivosti: $\rightarrow$ Výsledek * stačí vynásobit N

$$N \left(p - u_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}, p + u_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right) = \left(Np - Nu_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}, Np + Nu_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right)$$

jednostranné intervaly spolehlivosti:

$$\begin{array}{ll} \text{dolní mez} & N \left(p - u_{1-\alpha} \sqrt{\frac{p(1-p)}{n}} \right) \quad \text{horní mez} \quad N \left(p + u_{1-\alpha} \sqrt{\frac{p(1-p)}{n}} \right) \\ \uparrow & \uparrow \\ \text{levostranný interval} & \text{pravostranný interval} \end{array}$$



Statistická indukce: testování hypotéz

Testování statistických hypotéz o parametrech

Střední hodnota (populační průměr); normální rozdělení

H ₀	H ₁	Testová statistika	Kritický obor
$\mu = \mu_0$ <i>předpokládána střední hodnota</i>	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	$T = \frac{\bar{x} - \mu_0}{S / \sqrt{n}}$ <i>průměr</i> <i>směr. odchylka</i> <i>Studenťovo t-rozdělení</i> Pokud H ₀ platí $T \sim t(n-1)$	$W_\alpha = \{t; t \geq t_{1-\alpha}\}$ $W_\alpha = \{t; t \leq -t_{1-\alpha}\}$ $W_\alpha = \{t; t \geq t_{1-\alpha/2}\}$ <i>Kvantil standardov. t-rozdělení</i> <i>Excel: T.INV()</i>

Střední hodnota (populační průměr); obecné rozdělení náhodné veličiny, velký náhodný výběr

H ₀	H ₁	Testová statistika	Kritický obor
$E(X) = E(X)_0$ <i>předpokládána střed. hodnota</i>	$E(X) > E(X)_0$ $E(X) < E(X)_0$ $E(X) \neq E(X)_0$	$U = \frac{\bar{x} - E(X)_0}{S / \sqrt{n}}$ <i>průměr</i> <i>směr. odchylka</i> Pokud H ₀ platí $U \approx N(0,1)$	$W_\alpha = \{u; u \geq u_{1-\alpha}\}$ $W_\alpha = \{u; u \leq -u_{1-\alpha}\}$ $W_\alpha = \{u; u \geq u_{1-\alpha/2}\}$ <i>Kvantil norm. normálního rozdělení</i> <i>Excel: NORM.S.INV()</i>

Pravděpodobnost (populační podíl); velký náhodný výběr

H ₀	H ₁	Testová statistika	Kritický obor
$\pi = \pi_0$ <i>předpokládáný podíl</i>	$\pi > \pi_0$ $\pi < \pi_0$ $\pi \neq \pi_0$	$U = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$ <i>výb. podíl</i> Pokud H ₀ platí $U \approx N(0,1)$	$W_\alpha = \{u; u \geq u_{1-\alpha}\}$ $W_\alpha = \{u; u \leq -u_{1-\alpha}\}$ $W_\alpha = \{u; u \geq u_{1-\alpha/2}\}$ <i>Kvantil norm. normálního rozdělení</i> <i>Excel: NORM.S.INV()</i>

chyba ve vzorcích!
znaménko musí být $\geq$

Testování statistických hypotéz o pravděpodobnostním rozdělení v populaci

Chí-kvadrát (χ^2) test dobré shody (velký náhodný výběr, $n'_j = n\pi_{0,j}$, $n'_j \geq 5$)

H ₀	H ₁	Testová statistika	Kritický obor
$\pi_j = \pi_{0,j}$ $(j = 1, 2, \dots, k)$	non H ₀	$G = \sum_{j=1}^k \frac{(n_j - n'_j)^2}{n'_j}$ <i>na pozorované četnosti</i> <i>očekávané četnosti</i> <i>počet kategorií</i> Pokud H ₀ platí $G \approx \chi^2(k-1)$ <i>Stupně volnosti</i>	$W_\alpha = \{g; g \geq \chi^2_{1-\alpha}\}$ <i>Kvantil chí-kvadrát rozdělení</i> <i>Excel: CHISQ.INV()</i>

Podmínka na oček. četnosti



Analýza závislosti

Analýza rozptylu

rozklad čtverců: hodnoty faktoru (k možností),

počet pozorování n_j , $j = 1, 2, \dots, k$; y_{ji} , $i = 1, 2, \dots, n_j$ pozorování Y

celkový součet čtverců $S_y = \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ji} - \bar{y})^2 = S_{y.m} + S_{y.v}$ vnitroskupinový součet čtverců $S_{y.v} = \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_j)^2$

$S_{y.m} = \sum_{j=1}^k n_j (\bar{y}_j - \bar{y})^2$ mezikupinový součet čtverců

Poměr determinace $p^2 = \frac{S_{y.m}}{S_y}$

H ₀	H ₁	Testové kritérium	Kritický obor
$\mu_1 = \mu_2 = \dots = \mu_k$ rovnost str. hodnot	non H ₀ alespoň 1 str. hodnota je odlišná	$F = \frac{\frac{S_{y.m}}{k-1}}{\frac{S_{y.v}}{n-k}}$ Pokud H ₀ platí $F \sim F(k-1, n-k)$	$W_\alpha = \{F; F \geq F_{1-\alpha}\}$ Kvantil F-rozdělení Excel: F.INV()

Kontingenční tabulka ($r \times s$)

$n_{i.} = \sum_{j=1}^s n_{ij}$ $n_{.j} = \sum_{i=1}^r n_{ij}$ $n'_{ij} = \frac{n_{i.} n_{.j}}{n}$ $n'_{ij} \geq 5$

očekávané četnosti Podmínka na oček. četnosti

Test nezávislosti v kontingenční tabulce

H ₀	H ₁	Testové kritérium	Kritický obor
znaky jsou nezávislé	non H ₀ závislost	$G = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}}$ Pokud H ₀ platí $G \approx \chi^2((r-1)(s-1))$	$W_\alpha = \{g; g \geq \chi^2_{1-\alpha}\}$ Kvantil chí-kvadrát rozdělení Excel: CHISQ.INV()

$C = \sqrt{\frac{G}{n+G}} \in \left(0, \sqrt{\frac{m-1}{m}}\right)$ $V = \sqrt{\frac{G}{n(m-1)}}$, $m = \min(r, s)$

Pearsonův koef. kontingence Cramerův koef. kontingence $< 0,1$

$r = s = 2$ $G = n \frac{(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}}$ ← Zjednodušený vzorec pro tabulky 2x2



Regrese a korelace

regresní přímka $y = \beta_0 + \beta_1 x + \varepsilon$
konstanta β_0 *směrnice* β_1 *reziduum* ε

$$\hat{\beta}_0 = b_0, \hat{\beta}_1 = b_1 : \min_{b_0, b_1} \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 \quad Y = b_0 + b_1 x, \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$e_i = y_i - b_0 - b_1 x_i = y_i - Y_i = y_i - \hat{y}_i$$

Kovariance $s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n} = \overline{xy} - \bar{x} \cdot \bar{y}$ *napozorovaná hodnota* x_i *významná hodnota* y_i **Excel:** COVARIANCE.P()

$$\hat{\beta}_1 = b_1 = b_{yx} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{s_{xy}}{s_x^2} \quad \text{Excel: SLOPE()}$$

$$\hat{\beta}_0 = b_0 = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \bar{y} - b_{yx} \bar{x} \quad \text{Excel: INTERCEPT()}$$

Jiné regresní funkce

Regresní parabola

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \rightarrow Y = b_0 + b_1 x + b_2 x^2, \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2$$

Vícenásobná regrese

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \rightarrow Y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k, \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k$$

Rozklad čtverců

Teoretický součet čtverců $S_T = \sum_{i=1}^n (Y_i - \bar{y})^2$ *Reziduální součet čtverců* $S_R = \sum_{i=1}^n (y_i - Y_i)^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2$

celkový součet čtverců $S_y = \sum_{i=1}^n (y_i - \bar{y})^2$ $S_y = S_R + S_T$ $s_R^2 = \frac{S_R}{n-p}$ *Reziduální rozptyl* $I^2_{ADJ} = R^2_{ADJ} = 1 - (1 - I^2) \frac{n-1}{n-p}$ *počet pozorování* $n-1$ *počet parametrů* p

$R^2 = I^2 = \frac{S_T}{S_y}$ *Koeficient/Index determinace <0,1>* $s_R = \sqrt{\frac{S_R}{n-p}} = \sqrt{s_R^2}$ *Reziduální směrná odchylka*

Upravený koeficient determinace



Test hypotézy o regresních parametrech

H ₀	H ₁	Testové kritérium	Kritický obor
$\beta_j = 0$	$\beta_j \neq 0$	$T = \frac{b_j}{s(b_j)}$ $\leftarrow$ regresní koef. $\leftarrow$ standardní chyba reg. koef. Pokud H₀ platí $T \sim t(n-p)$	$W_\alpha = \{t; t \geq t_{1-\alpha/2}\}$ $\leftarrow$ kvantil standardního t-rozdělení Excel: T.INV()

$\uparrow$ stupně volnosti $\leftarrow$ počet parametrů



Test o modelu ($p = k + 1$)

počet parametrů
počet koef. + (1) konstanta

H ₀	H ₁	Testové kritérium	Kritický obor
$\beta_0 = c$ $\beta_1 = 0$ $\dots$ $\beta_k = 0$	non H ₀	$F = \frac{\frac{S_T}{p-1}}{\frac{S_R}{n-p}}$ <i>počet parametrů</i> Pokud H ₀ platí $F \sim F(p-1, n-p)$	$W_\alpha = \{F; F \geq F_{1-\alpha}\}$ <i>Kvantil F-rozdělení</i> Excel: F.INV()

Korelační koeficienty

Pearsonův korelační koeficient

$$r_{yx} = r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2}} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}} = \frac{s_{xy}}{s_x s_y}$$

Kovariance
Korelační koeficient <-1,1>
Excel: CORREL()
směr. odchylky

Spearmanův pořadový korelační koeficient

$$r_s = 1 - \frac{6 \sum_{i=1}^n (\text{pořadí } x_i - \text{pořadí } y_i)^2}{n(n^2 - 1)}$$

počet pozorování
případně pomocí Pearsonova korelačního koeficientu
<-1,1>
 proměnných pořadí x_i a pořadí y_i

Test o nulovosti korelačního koeficientu

H ₀	H ₁	Testové kritérium	Kritický obor
$\rho_{yx} = 0$ <i>nevýznamný kor. koef.</i>	$\rho_{yx} \neq 0$ <i>významný kor. koef.</i>	$T = \frac{r_{yx} \sqrt{n-2}}{\sqrt{1-r_{yx}^2}}$ <i>statistická volnost</i> Pokud H ₀ platí $T \sim t(n-2)$	$W_\alpha = \{t; t \geq t_{1-\alpha/2}\}$ <i>Kvantil standardního t-rozdělení</i> Excel: T.INV()



Analýza časových řad

Průměrné hodnoty časové řady období $t = 1, 2, \dots, n$

aritmetický průměr

$$\bar{y} = \frac{\sum_{t=1}^n y_t}{n}$$

$$\bar{y} = \frac{\frac{1}{2} y_1 + \sum_{t=2}^{n-1} y_t + \frac{1}{2} y_n}{n-1}$$

chronologický průměr

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} d_1 + \frac{y_2 + y_3}{2} d_2 + \dots + \frac{y_{n-1} + y_n}{2} d_{n-1}}{d_1 + d_2 + \dots + d_{n-1}}$$

chronologický vážený průměr

$$\Delta_t = y_t - y_{t-1}$$

absolutní přírůstek

$$\bar{\Delta} = \frac{1}{n-1} \sum_{t=2}^n \Delta_t = \frac{y_n - y_1}{n-1}$$

průměrný abs. přírůstek

Řetězové a bazické indexy

$$k_t = I_{t/t-1} = \frac{y_t}{y_{t-1}}$$

Koeficient růsta

$$\bar{k} = \sqrt[n-1]{k_2 k_3 \dots k_n} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

průměrný koef. růsta

$$I_{t/1} = \frac{y_t}{y_1} = I_{2/1} I_{3/2} \dots I_{t/t-1}$$

Bazický index

$$I_{t/t-1} = \frac{y_t}{y_{t-1}} = \frac{I_{t/1}}{I_{(t-1)/1}}$$

Řetězový index

Klouzavé průměry

$$m = 2p + 1$$

prosté

$$\bar{y}_t = \frac{\sum_{i=-p}^p y_{t+i}}{m} = \frac{y_{t-p} + \dots + y_{t-1} + y_t + y_{t+1} + \dots + y_{t+p}}{m}$$

$$m = 2p$$

centrované

$$\bar{y}_t = \frac{1}{2m} (y_{t-p} + 2y_{t-p+1} + \dots + 2y_{t-1} + 2y_t + 2y_{t+1} + \dots + 2y_{t+p-1} + y_{t+p})$$

Dekompozice časové řady

$$y_t = T_t + S_t + C_t + \varepsilon_t$$

Additivní

$$y_t = T_t S_t C_t \varepsilon_t$$

Multiplikativní

$$T_t = \beta_0 + \beta_1 t$$

$$\hat{T}_t = b_0 + b_1 t$$

přímka

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2$$

$$\hat{T}_t = b_0 + b_1 t + b_2 t^2$$

parabola

$$e_t = y_t - \hat{y}_t$$

reziduum

$$MSE = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n} = \frac{\sum_{t=1}^n e_t^2}{n}$$

průměrná čtvercová chyba

$$D = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

← Durbin-Watsonův koeficient $< 0,47$

Regresní metoda s umělými proměnnými (lineární trend, sezónnost délky 4)

$$y_t = T_t + S_t + \varepsilon_t = \beta_0 + \beta_1 t + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \varepsilon_t$$

nutné přepočítat

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\alpha}_1 x_{1t} + \hat{\alpha}_2 x_{2t} + \hat{\alpha}_3 x_{3t} = b_0 + b_1 t + a_1 x_{1t} + a_2 x_{2t} + a_3 x_{3t}$$

$$\bar{a} = \frac{a_1 + a_2 + a_3}{4}$$

$$S_{i+4j} = a_i - \bar{a}, i = 1, 2, 3$$

i-tá sezónní složka

$$S_{4+4j} = -\bar{a}$$

čtvrtá sez. složka

$$\hat{T}_t = (b_0 + \bar{a}) + b_1 t$$

30



Indexní analýza

$$Q = pq \quad IQ = Ip \cdot Iq$$

Tržba cena množství index tržeb

$$Ip = \frac{p_1}{p_0} \quad \Delta p = p_1 - p_0 \quad Iq = \frac{q_1}{q_0} \quad \Delta q = q_1 - q_0 \quad IQ = \frac{Q_1}{Q_0} \quad \Delta Q = Q_1 - Q_0$$

cenný index množství index index tržeb

Složené individuální indexy

množství: $I(\Sigma q) = \frac{\sum q_1}{\sum q_0} = \frac{\sum Iq \cdot q_0}{\sum q_0} = \frac{\sum q_1}{\sum \frac{q_1}{Iq}}$

$$\Delta(\Sigma q) = \sum q_1 - \sum q_0$$

tržeb/hodnoty: $I(\Sigma Q) = \frac{\sum Q_1}{\sum Q_0} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum IQ \cdot Q_0}{\sum Q_0} = \frac{\sum Q_1}{\sum \frac{Q_1}{IQ}}$

$$\Delta(\Sigma Q) = \sum Q_1 - \sum Q_0$$

cenný: $I\bar{p} = \frac{\bar{p}_1}{\bar{p}_0} = \frac{\frac{\sum Q_1}{\sum q_1}}{\frac{\sum Q_0}{\sum q_0}} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum \frac{Q_1}{p_1}}{\sum \frac{Q_0}{p_0}}$

$$\Delta\bar{p} = \bar{p}_1 - \bar{p}_0 = \frac{\sum p_1 q_1}{\sum q_1} - \frac{\sum p_0 q_0}{\sum q_0}$$

Souhrnné indexy

$$Ip^{(L)} = \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{\sum Ip \cdot p_0 q_0}{\sum p_0 q_0} = \frac{\sum Ip \cdot Q_0}{\sum Q_0}$$

$$Ip^{(P)} = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{Ip}} = \frac{\sum Q_1}{\sum \frac{Q_1}{Ip}}$$

$$Iq^{(L)} = \frac{\sum p_0 q_1}{\sum p_0 q_0} = \frac{\sum Iq \cdot p_0 q_0}{\sum p_0 q_0} = \frac{\sum Iq \cdot Q_0}{\sum Q_0}$$

$$Iq^{(P)} = \frac{\sum p_1 q_1}{\sum p_1 q_0} = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{Iq}} = \frac{\sum Q_1}{\sum \frac{Q_1}{Iq}}$$

