VYSOKÁ ŠKOLA EKONOMICKÁ V PRAZE
FAKULTA INFORMATIKY A STATISTIKY
Katedra statistiky a pravděpodobnosti

STATISTIKA

VZORCE PRO 4ST201

verze 2.1
poslední aktualizace: 22.9.2008

Dodatečná úprava
www.doucovanispetrem.cz
18.9.2021

Popisná statistika

Absolutní četnost ni
Relativní četnost pi

$$p_i = \frac{n_i}{n} \qquad \sum_{i=1}^{k} n_i = n \qquad \sum_{i=1}^{k} p_i = 1 \qquad i = 1, 2, \ldots, k$$

$\tilde{x}_p$ výpočet p% kvantilu

$$n \cdot \frac{P}{100} < z_P < n \cdot \frac{P}{100} + 1 \qquad n\,p < z_p < n\,p + 1$$

$$n\frac{P}{100} + 0,5 = z_p \qquad np + 0,5 = z_p$$

prosté tvary výpočtu průměrů

vážené tvary výpočtu průměrů (s absolutní či relativní četností)

aritmetický průměr

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

harmonický průměr

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

geometrický průměr

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

variační rozpětí

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

výpočty dle definice rozptylu

výpočty rychlejším odvozeným tvarem

průměr 2.mocnin (druhý obecný moment) $\overline{x^2}$

2.mocnina průměru $\bar{x}^2$

rozptyl pro základní soubor většinou značen σ^2

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

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

$$s_x^2 = \sum_{i=1}^{k} (x_i - \bar{x})^2 p_i \qquad s_x^2 = \overline{x^2} - \bar{x}^2 = \sum_{i=1}^{k} x_i^2\, p_i - \left(\sum_{i=1}^{k} x_i\, p_i\right)^2$$

výpočet celkového rozptylu pokud je celek rozdělen na skupiny, ve kterých známe četnost, průměr a rozptyl

$$s_x^2 = \overline{s^2} + s_{\bar{x}}^2 = \frac{\sum_{i=1}^{k} s_i^2\, n_i}{\sum_{i=1}^{k} n_i} + \frac{\sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2\, n_i}{\sum_{i=1}^{k} n_i} \qquad \bar{x} = \frac{\sum_{i=1}^{k} \bar{x}_i\, n_i}{\sum_{i=1}^{k} n_i}$$

výpočet celkového průměru, pokud je celek rozdělen na skupiny, ve kterých známe četnost a průměr

$$s_x^2 = \sum_{i=1}^{k} s_i^2 p_i + \sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2 p_i \qquad \bar{x} = \sum_{i=1}^{k} \bar{x}_i p_i$$

směrodatná odchylka

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

variační koeficient

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

Pravděpodobnost

Počet pravděpodobnosti

m - počet příznivých možností jevu A
n - počet všech možností

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

prst. sjednocení neslučitelných jevů

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

výpočet podmíněné prsti. jevu A za podmínky jevu B

prst. sjednocení slučitelných jevů

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

prst. průniku nezávislých jevů

$$P(A \cap B) = P(A)\, P(B) \qquad P(A \cap B) = P(A)\, P(B|A) = P(B)\, P(A|B)$$

prst. průniku závislých jevů

$$P(A) = \sum_{i=1}^{s} P(A \cap B_i) \qquad P(A) = \sum_{i=1}^{s} P(B_i)\, P(A|B_i)$$

věta o úplné pravděpodobnosti

Náhodné veličiny

pravděp. funkce

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

F(x) = distribuční funkce

$$P(x_1 < X \le x_2) = \sum_{x_1 < x \le x_2} P(x) = F(x_2) - F(x_1)$$

diskrétní veličiny

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt \qquad f(x) = F'(x) \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

spojité veličiny

$$P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x)dx = F(x_2) - F(x_1)$$

f(x)= hustota prsti

$$x_P \qquad F(x_P) = P \qquad x_P = F^{-1}(P)$$

Diskrétní / Spojité

Střední hodnota (průměr)

$$E(X) = \sum_x x\, P(x) \qquad E(X) = \int_{-\infty}^{\infty} x\, f(x)dx$$

Obecný rychlejší zápis výpočtu rozptylu

Rozptyl D(X)

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

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

Směr.odchylka

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

$$X_i \;\; i = 1,..,n, \quad E(X_i) = \mu, \;\; D(X_i) = \sigma^2 \text{ nezávislé} \quad E(\sum X_i) = n\mu, \quad D(\sum X_i) = n\sigma^2$$

Limitní věty pro součet resp. pro průměr "n" veličin s danou střední hodnotou E(X) a rozptylem D(X)

$$E(\frac{1}{n}\sum X_i) = \mu, \quad D(\frac{1}{n}\sum X_i) = \frac{\sigma^2}{n}$$

Pravděpodobnostní rozdělení

Alternativní rozdělení $A[\pi]$

"pí" je prst. nastání sledovaného jevu v daném pokusu

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

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

Vybraná rozdělení pravděp. diskrétních veličin.
Vždy uvedena pravděpodobnostní funkce P(X), střední hodnota E(X) a rozptyl D(X)

Binomické rozdělení $Bi[n;\pi]$

$$P(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x} \qquad x = 0, 1, 2, ..., n, \quad n > 0, 0 < \pi < 1$$

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

Poissonovo rozdělení $Po[\lambda]$

$$P(x) = e^{-\lambda}\frac{\lambda^x}{x!} \qquad x = 0, 1, \ldots, \lambda > 0, \qquad E(X) = \lambda \qquad D(X) = \lambda$$

Hypergeometrické rozdělení $Hy[N;M;n]$

$$P(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \qquad x = max(0,\ M\text{-}N\text{+}n),\ ...,\ min(M,\ n),\ n > 0,\ N \ge n,\ M \le N$$

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

Od exponenciálního rozdělení se jedná o spojité veličiny. Opět E(X)=střední hodnota D(X)=rozptyl

Exponenciální rozdělení $E[A;\delta]$ $A \ge 0,\ \delta > 0$

$$F(x) = \begin{cases} 0 & x \le A \\ 1 - e^{-\frac{(x-A)}{\delta}} & x > A \end{cases} \qquad E(X) = A + \delta \qquad D(X) = \delta^2$$

Normální rozdělení $N[\mu;\sigma^2]$

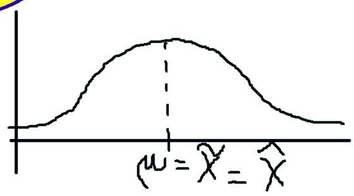
$$-\infty < x < \infty,\ -\infty < \mu < \infty,\ \sigma^2 > 0 \qquad E(X) = \mu \qquad D(X) = \sigma^2$$

Určení hodnoty z tabulky distrib. funkce normovaného norm. rozdělení

$$u = \frac{x-\mu}{\sigma} \qquad F(x) = \Phi(u) = \Phi(\frac{x-\mu}{\sigma}) \qquad x_P = \mu + \sigma\, u_P$$

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

Normální rozdělení - Gaussova křivka



Normované normální rozdělení $N[0;1]$

$$U = \frac{X-\mu}{\sigma} \qquad E(U) = 0 \qquad D(U) = 1$$

Vzorce pro určení hledané hodnoty z tabulek.

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

Logaritmicko-normální rozdělení $LN[\mu;\sigma^2]$

$$U = \frac{\ln X - \mu}{\sigma} \sim N[0;1] \qquad x > 0,\ -\infty < \mu < \infty,\ \sigma^2 > 0$$

$$F(x) = \Phi(\frac{\ln x - \mu}{\sigma}) \qquad x_P = \exp(\mu + \sigma\, u_P)$$

$$E(X) = e^{\mu+\sigma^2/2} \qquad D(X) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$$

$$\mu = E(\ln X) = \ln(E(X)) - \sigma^2/2 \qquad \sigma^2 = D(\ln X) = \ln(\frac{D(X)}{(E(X))^2} + 1)$$

Chí-kvadrát rozdělení $\chi^2[\nu]$ $x > 0$

Rozdělení odvozená od normálního rozdělení, využívaná v dalších částech statistiky.

Rozdělení t (Studentovo) $t[\nu]$ $-\infty < x < \infty$ $t_P = -\,t_{1-P}$

F - rozdělení (Fisherovo – Snedecorovo) $F[\nu_1;\nu_2]$ $x > 0,$ $F_P(\nu_1,\nu_2) = \dfrac{1}{F_{1-P}(\nu_2,\nu_1)}$

Matematická statistika

výběrová směr.odchylka, většinou značena "s"

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

výpočtový tvar vzorce pro výběrový rozptyl

$$s^2 = \left(\overline{x^2} - \bar{x}^2\right)\cdot\frac{n}{n-1}$$

Odhady parametrů

střední hodnota est $\mu = \hat{\mu} = \bar{x}$ est $N\mu = N\bar{x}$

normální rozdělení

Spolehlivost odhadu = 1-alfa (např. 95%; 99%)

Pokud známe rozptyl základního souboru, používáme kvantily normálního rozdělení "u"

a) σ^2 známé

oboustranný odhad

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

levostranný odhad

pravostranný odhad

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

b) σ^2 neznámé

Pokud neznáme rozptyl základního souboru, používáme kvantily studentova rozdělení "t"

n-1 = stupně volnosti (řádky v tabulce studentova rozdělení)

$$P\left(\bar{x} - t_{1-\alpha/2}\frac{s'_x}{\sqrt{n}} < \mu < \bar{x} + t_{1-\alpha/2}\frac{s'_x}{\sqrt{n}}\right) = 1-\alpha \qquad t \sim t[n-1]$$

$$P\left(\bar{x} - t_{1-\alpha}\frac{s'_x}{\sqrt{n}} < \mu\right) = 1-\alpha \qquad P\left(\mu < \bar{x} + t_{1-\alpha}\frac{s'_x}{\sqrt{n}}\right) = 1-\alpha$$

obecné rozdělení, σ^2 neznámé, velký výběr ($n > 30$)

V případě velkého souboru lze nahradit studentovo rozdělení normálním rozdělením.

$$P\left(\bar{x} - u_{1-\alpha/2}\frac{s'_x}{\sqrt{n}} < E(X) < \bar{x} + u_{1-\alpha/2}\frac{s'_x}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(\bar{x} - u_{1-\alpha}\frac{s'_x}{\sqrt{n}} < E(X)\right) = 1-\alpha \quad P\left(E(X) < \bar{x} + u_{1-\alpha}\frac{s'_x}{\sqrt{n}}\right) = 1-\alpha$$

rozptyl σ^2 (normální rozdělení) est $\sigma^2 = \hat{\sigma}^2 = s'^2_x$

Bodový odhad rozptylu základ.souboru je výběrový rozptyl

Parametr π alternativního rozdělení (odhad relativní četnosti základního souboru)

est $\pi = \hat{\pi} = p$ est $N\pi = Np$

p = skutečná rel. četnost (podíl) sledované vlastnosti ve výběr.souboru

intervaly spolehlivosti pro odhad podílu opět oboustranný a jednostranné

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

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

"alfa" = hladina významnosti (5% , 1%..)

jednovýběrové testy

Testování hypotéz

stejný způsob rozhodování jaký vybrat vzorec testového kritéria a k tomu kvantily rozdělení jako u odhadů střední hodnoty

kritická hodnota vždy z příslušné tabulky kvantilů

Střední hodnota normálního rozdělení

H_0	H_1	Testové kritérium	Kritický obor
$\mu = \mu_0$	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$ H1 postupně: pravostranný levostranný oboustranný tvar	σ^2 známé $U = \frac{\bar{x}-\mu_0}{\sigma}\sqrt{n}$ $\quad U \sim N[0;1]$	$W_\alpha=\{U \geq u_{1-\alpha}\}$ $W_\alpha=\{U \leq -u_{1-\alpha}\}$ $W_\alpha=\{\lvert U\rvert \geq u_{1-\alpha/2}\}$
		σ^2 neznámé $t = \frac{\bar{x}-\mu_0}{s'_x}\sqrt{n}$ $\quad t \sim t[n-1]$ opět v závorce stupně volnosti pro studentovo rozděl.	$W_\alpha=\{t \geq t_{1-\alpha}\}$ $W_\alpha=\{t \leq -t_{1-\alpha}\}$ $W_\alpha=\{\lvert t\rvert \geq t_{1-\alpha/2}\}$

Střední hodnota, obecné rozdělení, velký výběr

H_0	H_1	Testové kritérium	Kritický obor
$E(X)= \mu_0$	$E(X) > \mu_0$ $E(X) < \mu_0$ $E(X) \neq \mu_0$	σ^2 neznámé ($n > 30$) $U = \frac{\bar{x}-\mu_0}{s'_x}\sqrt{n}$ $\quad U \approx N[0;1]$	$W_\alpha=\{U \geq u_{1-\alpha}\}$ $W_\alpha=\{U \leq -u_{1-\alpha}\}$ $W_\alpha=\{\lvert U\rvert \geq u_{1-\alpha/2}\}$

Rozptyl v normálním rozdělení

stupně volnosti u chí kvadrát rozdělení

H_0	H_1	Testové kritérium	Kritický obor
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$ $\sigma^2 \neq \sigma_0^2$	$\chi^2 = \frac{(n-1)s_x'^2}{\sigma_0^2}$ $\quad \chi^2 \sim \chi^2[n-1]$	$W_\alpha=\{\chi^2 \geq \chi^2_{1-\alpha}\}$ $W_\alpha=\{\chi^2 \leq \chi^2_{\alpha}\}$ $W_\alpha=\{\chi^2 \leq \chi^2_{\alpha/2} \cup \chi^2 \geq \chi^2_{1-\alpha/2}\}$

Parametr π alternativního rozdělení (velké výběry)

test na relativní četnost (podíl) p

H_0	H_1	Testové kritérium	Kritický obor
$\pi = \pi_0$	$\pi > \pi_0$ $\pi < \pi_0$ $\pi \neq \pi_0$	$U = \frac{p-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$ $\quad U \sim N[0;1]$	$W_\alpha=\{U \geq u_{1-\alpha}\}$ $W_\alpha=\{U \leq -u_{1-\alpha}\}$ $W_\alpha=\{\lvert U\rvert \geq u_{1-\alpha/2}\}$

dvouvýběrové testy

Rovnost středních hodnot dvou rozdělení

normální rozdělení (nezávislé náhodné výběry z normálního rozdělení)

H_0	H_1	Testové kritérium	Kritický obor
$\mu_1 = \mu_2$ $\mu_1 - \mu_2 = 0$	$\mu_1 > \mu_2$ $\mu_1 < \mu_2$ $\mu_1 \neq \mu_2$	a) σ_1^2, σ_2^2 známé $U = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ $\quad U \sim N[0;1]$	$W_\alpha=\{U \geq u_{1-\alpha}\}$ $W_\alpha=\{U \leq -u_{1-\alpha}\}$ $W_\alpha=\{\lvert U\rvert \geq u_{1-\alpha/2}\}$
		σ_1^2, σ_2^2 neznámé, ale předpokládáme, že $\sigma_1^2 = \sigma_2^2$ $t = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\left(\frac{(n_1-1)s_1'^2+(n_2-1)s_2'^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ $t \sim t[n_1+n_2-2]$	$W_\alpha=\{t \geq t_{1-\alpha}\}$ $W_\alpha=\{t \leq -t_{1-\alpha}\}$ $W_\alpha=\{\lvert t\rvert \geq t_{1-\alpha/2}\}$ stupně volnosti

tzv. Welchův test stupně volnosti "v" pro studentovo rozdělení	σ_1^2, σ_2^2 neznámé, ale předpokládáme, že $\sigma_1^2 \neq \sigma_2^2$ $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1'^2}{n_1} + \dfrac{s_2'^2}{n_2}}} \qquad t \sim t[\nu]$ $\nu = \dfrac{\left(\dfrac{s_1'^2}{n_1} + \dfrac{s_2'^2}{n_2}\right)^2}{\dfrac{1}{n_1+1}\left(\dfrac{s_1'^2}{n_1}\right)^2 + \dfrac{1}{n_2+1}\left(\dfrac{s_2'^2}{n_2}\right)^2} - 2$	$W_\alpha = \{t \geq t_{1-\alpha}\}$ $W_\alpha = \{t \leq -t_{1-\alpha}\}$ $W_\alpha = \{\lvert t\rvert \geq t_{1-\alpha/2}\}$

velké nezávislé výběry

H_0	H_1	Testové kritérium	Kritický obor
$\mu_1 = \mu_2$ $\mu_1 - \mu_2 = 0$	$\mu_1 > \mu_2$ $\mu_1 < \mu_2$ $\mu_1 \neq \mu_2$	σ_1^2, σ_2^2 neznámé $U = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1'^2}{n_1} + \dfrac{s_2'^2}{n_2}}} \qquad U \approx N[0;1]$	$W_\alpha = \{U \geq u_{1-\alpha}\}$ $W_\alpha = \{U \leq -u_{1-\alpha}\}$ $W_\alpha = \{\lvert U\rvert \geq u_{1-\alpha/2}\}$

závislé výběry z normálního rozdělení (párový *t*-test)

H_0	H_1	Testové kritérium	Kritický obor
$\mu_1 = \mu_2$ $\mu_1 - \mu_2 = 0$	$\mu_1 > \mu_2$ $\mu_1 < \mu_2$ $\mu_1 \neq \mu_2$	$t = \dfrac{\bar{d}}{s_d'}\sqrt{n} \qquad t \sim t[n-1]$ $d_i = x_{1i} - x_{2i},\ i = 1,2,..,n$ di = diference (rozdíl) mezi dvojicí hodnot ve výběrech	$W_\alpha = \{t \geq t_{1-\alpha}\}$ $W_\alpha = \{t \leq -t_{1-\alpha}\}$ $W_\alpha = \{\lvert t\rvert \geq t_{1-\alpha/2}\}$

F test

Rovnost rozptylů dvou normálních rozdělení

H_0	H_1	Testové kritérium	Kritický obor
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$ $\sigma_1^2 < \sigma_2^2$ $\sigma_1^2 \neq \sigma_2^2$	$F = \dfrac{s_1'^2}{s_2'^2} \qquad F \sim F[n_1 - 1;\ n_2 - 1]$	$W_\alpha = \{F \geq F_{1-\alpha}\}$ $W_\alpha = \{F \leq F_\alpha\}$ $W_\alpha = \{F \leq F_{\alpha/2} \cup F \geq F_{1-\alpha/2}\}$

má dvoje stupně volnosti (první v tabulce nahoře, druhé na boku)

Chí-kvadrát test dobré shody

H_0 a H_1	Testové kritérium	Kritický obor
H_0: $\pi_i = \pi_{0,i} \quad i = 1, .., k$ H_1: non H_0	$G = \displaystyle\sum_{i=1}^{k} \frac{(n_i - n\pi_{0,i})^2}{n\pi_{0,i}} \qquad \chi^2 \approx \chi^2[k-1]$	$W_\alpha = \{\chi^2 \geq \chi^2_{1-\alpha}\}$ $n\pi_{0,i} \geq 5$

chová se vždy jako pravostranná hypotéza

test shody rozdělení kateg. veličiny s předpokládaným (%) modelem

test. kritérium často značeno χ^2

ni = skutečná četnost kategorie
n = celkový počet hodnot
$\pi_{0,i}$ = očekávaná pravděp. kategorie

k = počet kategorií. Pokud je třeba odhadovat počet parametrů "p" testovaného rozdělení, pak počet stupňů volnosti bude k-1-p

pravidlo pro velikost očekáv. četností, jinak slučování kategorií

Analýza závislostí

Test chí kvadrát nezávislosti pro 2 kateg.veličiny

Kontingenční tabulka (*r* x *s*)

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

ni. = součet i-tého řádku
n.j = součet j-tého sloupce
n = celkový počet hodnot

H0: nezávislost veličin
H1: závislost veličin

H_0	H_1	Testové kritérium		Kritický obor
$\pi_{ij} = \pi_{i.}\, \pi_{.j}$, $1 \leq i \leq r$, $1 \leq j \leq s$	non H_0	$G = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}}$	$G \approx \chi^2[(r-1)(s-1)]$	$W_\alpha = \{G \geq \chi^2_{1-\alpha}\}$

stupně volnosti, kde
r= počet řádků
s= počet sloupců kont.tab.

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

Pearsonův koef. kontingence

Cramerův koef. kontingence

Tabulka 2 x 2
$$G = n\frac{(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}}$$

Rychlejší výpočet test. kritéria u asociační tabulky 2*2

ANOVA

Analýza rozptylu

skupinový průměr

celkový průměr

$$S_y = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = S_{y.m} + S_{y.v} \qquad S_{y.m} = \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2 n_i \qquad S_{y.v} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

Sy = celkový součet čtverců
Sy.m = meziskupinový součet čtverců
Sy.v = vnitroskupinový součet čtverců

determinační poměr

$$P^2 = \frac{S_{y,m}}{S_y} \qquad P = \sqrt{P^2}$$

k = počet skupin
n= počet pozorování celkem

H_0	H_1	Testové kritérium		Kritický obor
$\mu_1 = \mu_2 = .. = \mu_k$	non H_0	$F = \dfrac{\frac{S_{y.m}}{k-1}}{\frac{S_{y.v}}{n-k}}$	$F \sim F[k-1; n-k]$	$W_\alpha = \{F \geq F_{1-\alpha}\}$

Regrese a korelace

regresní přímka $y = \beta_0 + \beta_1 x + \varepsilon$, $Y = b_0 + b_1 x$ $\quad$ $\text{minimum}_{b_0, b_1} \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$

princip metody nejmenších čtverců

kovariance
$$s_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n} = \overline{xy} - \bar{x}.\bar{y}$$

regresní koeficient
$$b_1 = b_{yx} = \frac{n\sum y_i x_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} = \frac{\overline{xy} - \bar{x}.\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{s_{xy}}{s_x^2}$$

regresní konstanta
$$b_0 = \frac{\sum y_i \sum x_i^2 - \sum y_i x_i \sum x_i}{n\sum x_i^2 - (\sum x_i)^2} = \bar{y} - b_{yx}\bar{x}$$

kvadratická funkce (parabola)

vícenásobná lineární regrese

Jiné regresní funkce $Y = b_0 + b_1 x + b_2 x^2$ $\quad Y = b_0 + b_1 x_1 + b_2 x_2 + .. + b_k x_k$

celkový součet čtverců

$$S_y = \sum_{i=1}^{n} (y_i - \overline{y})^2$$

teoretický součet čtverců

$$S_T = \sum_{i=1}^{n} (Y_i - \overline{y})^2$$

$$s_y^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^2 = \frac{S_y}{n} \qquad s_Y^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \overline{y})^2 = \frac{S_T}{n}$$

reziduální součet čtverců

$$S_R = \sum_{i=1}^{n} (y_i - Y_i)^2 = \sum_{i=1}^{n} e_i^2 \qquad s_{(y-Y)}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - Y_i)^2 = \frac{S_R}{n} \qquad s_R^2 = \frac{S_R}{n-p}$$

reziduální rozptyl

$$S_y = S_R + S_T \qquad s_y^2 = s_Y^2 + s_{(y-Y)}^2$$

Koeficient resp. Index determinace

Upravená hodnota (adjustovaná) indexu determinace

reziduální směr. odchylka

$$s_R = \sqrt{\frac{S_R}{n-p}} = \sqrt{s_R^2} \qquad I_{yx}^2 = R^2 = \frac{S_T}{S_y} \qquad I_{yx} = \sqrt{I_{yx}^2} \qquad I_{ADJ}^2 = R_{ADJ}^2 = 1 - (1 - I_{yx}^2)\frac{n-1}{n-p}$$

Index korelace

Test hypotézy o regresních parametrech

testy významnosti regresních parametrů, resp. celého regresního modelu

H_0	H_1	Testové kritérium		Kritický obor
$\beta_i = 0$	$\beta_i \neq 0$	$t = \dfrac{b_i}{s(b_i)}$	$t \sim t[n-p]$	$W_\alpha = \{\lvert t \rvert \geq t_{1-\alpha/2}\}$

H0: parametr. není stat. významný
H1: je významný

n = počet pozorování
p = počet parametrů regresní funkce

Test o modelu $\quad p = k + 1$

H_0	H_1	Testové kritérium		Kritický obor
$\beta_0 = c$, $\beta_1 = 0$, …, $\beta_k = 0$	non H_0	$F = \dfrac{\frac{S_T}{p-1}}{\frac{S_R}{n-p}}$	$F \sim F[p-1; n-p]$	$W_\alpha = \{F \geq F_{1-\alpha}\}$

H0: model není stat. významný (všechny koef. = 0)
H1: model je stat.významný

stupně volnosti Fisherova rozdělení podobně jako u ANOVY

Pearsonův

korelační koeficient

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

Test významnosti korelačního koeficientu

H_0	H_1	Testové kritérium		Kritický obor
$\rho_{yx} = 0$	$\rho_{yx} \neq 0$	$t = \dfrac{r_{yx}\sqrt{n-2}}{\sqrt{1 - r_{yx}^2}}$	$t \sim t[n-2]$	$W_\alpha = \{\lvert t \rvert \geq t_{1-\alpha/2}\}$

H0: korel. koef. není stat. významný
H1: je významný

chronologický průměr u okamžikových ČŘ prostý a vážený tvar

Časové řady

aritmetický průměr u intervalových ČŘ

$$\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n} \qquad \bar{y} = \frac{\frac{1}{2}y_1 + \sum_{t=2}^{n-1} y_t + \frac{1}{2}y_n}{n-1} \qquad \bar{y} = \frac{\frac{y_1+y_2}{2}d_1 + \frac{y_2+y_3}{2}d_2 + \ldots + \frac{y_{n-1}+y_n}{2}d_{n-1}}{d_1 + d_2 + \ldots + d_{n-1}}$$

absol. přírůstek (1.diference)

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

průměrný absol. přírůstek

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

koef. růstu (řetězový index)

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

průměrný koef. růstu

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

míry dynamiky vývoje ukazatele (+ např. relativní přírůstek)

Klouzavé průměry

lichá délka

$$m = 2p + 1 \qquad \bar{y}_t = \frac{\sum_{i=-p}^{p} y_{t+i}}{m} = \frac{y_{t-p} + \ldots + y_{t-1} + y_t + y_{t+1} + \ldots + y_{t+p}}{m}$$

sudá délka (centrované kl.pr.)

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

Dekompozice časové řady

aditivní rozklad

multiplikativní rozklad

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

lineární trendová funkce

kvadratická trendová funkce

$T_t = \beta_0 + \beta_1 t \qquad \hat{T}_t = b_0 + b_1 t \qquad T_t = \beta_0 + \beta_1 t + \beta_2 t^2 \qquad \hat{T}_t = b_0 + b_1 t + b_2 t^2$

$T_t = \beta_0 t^{\beta_1} \qquad \ln T_t = \ln \beta_0 + \beta_1 \ln t \qquad \ln(\hat{T}_t) = \ln b_0 + b_1 \ln t$

exponenciální trendová funkce s linearizující transformací

$$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - T_t)^2$$

průměrná čtvercová chyba

Analýza sezónní složky

1. Metoda empirických indexů (délka sezónnosti r)

$\bar{y}_t$ klouzavé průměry délky r (například $r = 4$) sezónní index$_t = \frac{y_t}{\bar{y}_t}$

$$\text{průměrný sezónní index}_i = \frac{\sum_{t\ z\ i-\text{té sezóny}} \text{sezonní index}_t}{\text{počet hodnot z } i-\text{té sezóny}} \qquad i = 1, 2, ., r \text{ (např. } r = 4)$$

$$\text{standardizovaný sezónní index} = \frac{r}{\sum_{j=1}^{r} \text{průměrný sezonní index}_j} \cdot \text{průměrný sezonní index}_i$$

aby se suma sez. indexů rovnala r

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

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

$$\bar{a} = \frac{a_1 + a_2 + a_3}{4} \qquad S_{i+4j} = a_i - \bar{a} \quad i=1,2,3 \qquad S_{4+4j} = -\bar{a} \qquad \hat{T}_t = (b_0 + \bar{a}) + b_1 t$$

průměrná sezónní hodnota

i-tá sezónní složka

4. sezónní složka

odhad trendové hodnoty v období "t"

Indexní analýza

bazický index

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

řetězový index

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

Všude kromě indexu také vzorec pro rozdíl (delta)

Q = tržby
p = cena
q= množství

$$Q = pq$$

individuální jednoduchý cenový index

$$Ip = \frac{p_1}{p_0} \qquad \Delta p = p_1 - p_0$$

individuální jednoduchý množstevní index

$$Iq = \frac{q_1}{q_0} \qquad \Delta q = q_1 - q_0$$

individuální jednoduchý index tržeb

$$IQ = \frac{Q_1}{Q_0} \qquad \Delta Q = Q_1 - Q_0$$

individuální složený index množství

$$I(\Sigma q) = \frac{\sum q_1}{\sum q_0} = \frac{\sum Iq.q_0}{\sum q_0} = \frac{\sum q_1}{\sum \frac{q_1}{Iq}} \qquad \Delta(\Sigma q) = \sum q_1 - \sum q_0$$

individuální složený index tržeb

$$I(\Sigma Q) = \frac{\sum Q_1}{\sum Q_0} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum IQ.Q_0}{\sum Q_0} = \frac{\sum Q_1}{\sum \frac{Q_1}{IQ}} \qquad \Delta(\Sigma Q) = \sum Q_1 - \sum Q_0$$

individuální složený index ceny

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

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

souhrnný Laspeyresův cenový index

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

souhrnný Paascheho cenový index

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

Fisherův cenový index

$$Ip^{(F)} = \sqrt{Ip^{(L)}\ .\ Ip^{(P)}}$$

souhrnný Laspeyresův množstevní index

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

souhrnný Paascheho množstevní index

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

Fisherův množstevní index

$$Iq^{(F)} = \sqrt{Ip^{(L)}\ \ Iq^{(P)}}$$

souhrnný index tržeb

$$I(\Sigma Q) = \frac{\sum Q_1}{\sum Q_0} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum IQ.Q_0}{\sum Q_0} = \frac{\sum Q_1}{\sum \frac{Q_1}{IQ}} \qquad \Delta(\Sigma Q) = \sum p_1 q_1 - \sum p_0 q_0$$