Přehled základních vzorců

Zimní semestr

I. Teorie odhadu

1. Dvoustranný interval spolehlivosti pro průměr základního souboru

$$P(\bar{x}-\Delta<\mu<\bar{x}+\Delta)=1-\alpha$$

Výběr s vracením

a) známe-li rozptyl základního souboru $\Delta = u_\alpha \cdot \sqrt{\frac{\sigma^2}{n}}$ $\Delta' = u_\alpha \cdot \frac{V}{\sqrt{n}}$

b) neznáme-li rozptyl základního souboru $\Delta = t_{\alpha(n-1)} \cdot \sqrt{\frac{s^2}{n}}$ $\Delta' = t_{\alpha(n-1)} \cdot \frac{v'}{\sqrt{n}}$

Výběr bez vracení

a) známe-li σ^2 $\Delta = u_\alpha \cdot \sqrt{\frac{\sigma^2}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$ $\Delta' = u_\alpha \cdot \frac{V}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

b) neznáme-li σ^2 $\Delta = t_{\alpha(n-1)} \cdot \sqrt{\frac{s^2}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$ $\Delta' = t_{\alpha(n-1)} \cdot \frac{v'}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

Stanovení rozsahu výběru

Výběr s vracením $n = \frac{u_\alpha^2 \cdot \sigma^2}{\Delta^2}$ $n = \frac{t_\alpha^2 \cdot s^2}{\Delta^2}$

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

Výpočet spolehlivosti odhadu

Výběr s vracením $u_\alpha = \frac{\Delta\sqrt{n}}{\sqrt{\sigma^2}}$ $t_\alpha = \frac{\Delta\sqrt{n}}{\sqrt{s^2}}$

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

2. Dvoustranný interval spolehlivosti pro rozptyl

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

3. Dvoustranný interval spolehlivosti pro relativní četnost

$$P(f_i - \Delta < p < f_i + \Delta) = 1 - \alpha$$

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

Výběr bez vracení $\Delta = u_\alpha \cdot \sqrt{\frac{f_i(1-f_i)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

Stanovení rozsahu výběru

Výběr s vracením $n = \frac{u_\alpha^2 \cdot f_i(1-f_i)}{\Delta^2}$

Výběr bez vracení $n = \frac{u_\alpha^2 \cdot f_i(1-f_i) \cdot N}{\Delta^2 \cdot (N-1) + u_\alpha^2 \cdot f_i(1-f_i)}$

Výpočet spolehlivosti odhadu

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

Výběr bez vracení $u_\alpha = \sqrt{\frac{n \cdot (N-1) \cdot \Delta^2}{f_i(1-f_i) \cdot (N-n)}}$

II. Testování statistických hypotéz

Testy parametrické

1. Jednovýběrové testy

Test hypotézy o hodnotě rozptylu σ^2

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

Testové kritérium má χ^2-rozdělení pro počet stupňů volnosti f = n – 1.

Test hypotézy o hodnotě průměru μ

a) známe-li rozptyl základního souboru $u = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$

Testové kritérium se řídí normálním rozdělením.

b) neznáme-li rozptyl základního souboru $t = \dfrac{\overline{x} - \mu_0}{\sqrt{\dfrac{s^2}{n}}}$

Testové kritérium má Studentovo t-rozdělení pro f = n – 1.

Test hypotézy o hodnotě relativní četnosti p

$$u = \frac{\dfrac{m}{n} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}\ , \qquad \text{kde } f_i = \frac{m}{n}$$

Testové kritérium má normální rozdělení.

2. Dvouvýběrové testy

Test hypotézy o shodě dvou rozptylů (F-test)

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

Testové kritérium má F-rozdělení o f_1 = (m – 1) a f_2 = (n – 1) stupních volnosti.

Test hypotézy o shodě dvou průměrů

1) Test v případě dvou nezávislých výběrových souborů (t-test)

a) ***známe rozptyly základních souborů***

$$u = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}}$$

Testové kritérium se řídí normálním rozdělením.

b) ***neznáme rozptyly základních souborů, ale vycházíme z předpokladu, že rozptyly jsou shodné, tzn.*** $\sigma_1^2 = \sigma_2^2$

$$t = \frac{\overline{x} - \overline{y}}{s\sqrt{\dfrac{1}{m} + \dfrac{1}{n}}} \qquad s = \sqrt{\frac{1}{n+m-2} \cdot \left[(m-1)\cdot s_1^2 + (n-1)\cdot s_2^2\right]}$$

Testové kritérium má Studentovo t-rozdělení pro f = m + n – 2.

c) ***neznáme rozptyly základních souborů a předpokládáme, že rozptyly jsou různé*** $\sigma_1^2 \neq \sigma_2^2$

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

Kritická hodnota $t_\alpha^* = \dfrac{t_{\alpha(f_1)} \cdot \frac{s_1^2}{m} + t_{\alpha(f_2)} \cdot \frac{s_2^2}{n}}{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$, kde $t_{\alpha(f_1)}$ a $t_{\alpha(f_2)}$ jsou kritické hodnoty Studentova t-rozdělení pro $f_1 = (m - 1)$ a $f_2 = (n - 1)$ stupňů volnosti na zvolené hladině významnosti α.

2) Test v případě dvou závislých souborů (párový t-test)

$$t = \frac{\bar{d}}{\sqrt{\frac{s_d^2}{n}}}$$

Testové kritérium má Studentovo t-rozdělení pro počet stupňů volnosti f = (n − 1).

$d_i = x_i - y_i$, $\quad \bar{d} = \dfrac{1}{n}\sum_{i=1}^{n} d_i = \bar{x} - \bar{y}$, $\quad s_d^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2$.

Test hypotézy o shodě dvou relativních četností

$$u = \frac{\frac{m_1}{n_1} - \frac{m_2}{n_2}}{\sqrt{\frac{\bar{p} \cdot \bar{q}}{n}}} \qquad \bar{p} = \frac{m_1 + m_2}{n_1 + n_2}, \qquad \bar{q} = 1 - \bar{p}, \qquad n = \frac{n_1 \cdot n_2}{n_1 + n_2}$$

Testové kritérium má normální rozdělení

3. Vícevýběrové testy

Testy o shodě více než dvou rozptylů

Cochranův test

$$G = \frac{s_{max}^2}{s_1^2 + s_2^2 + ... + s_k^2}$$

$G_{\alpha,k,f}$ – kritická hodnota pro Cochranův test (k – počet srovnávaných rozptylů, f = n – 1)

Bartlettův test

$$B = \frac{2{,}30259}{C}\left[(N-k)\log s^2 - \sum_{i=1}^{k}(n_i - 1)\log s_i^2\right],$$

kde $N = \sum_{i=1}^{m} n_i$, $\quad s^2 = \frac{1}{N-k}\sum_{i=1}^{m}(n_i - 1)\cdot s_i^2$, $\quad C = 1 + \frac{1}{3\cdot(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i - 1} - \frac{1}{N-k}\right)$.

Veličina B má přibližně χ^2-rozdělení o (k – 1) stupních volnosti, kde k je počet tříd.

Testy o shodě více než dvou průměrů (analýza rozptylu)

Analýza rozptylu jednoduchého třídění **– vyvážený model**

Variabilita	Součet čtverců	Stupně volnosti	Rozptyl	Testovací kritérium
Mezi třídami	$S_1 = \frac{1}{n}\sum_{i=1}^{m} x_{i\bullet}^2 - C$	m – 1	$s_1^2 = \frac{S_1}{m-1}$	$F = \frac{s_1^2}{s_r^2}$
Uvnitř tříd	$S_r = S - S_1$	m(n – 1)	$s_r^2 = \frac{S_r}{m(n-1)}$	
Celková	$S = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}^2 - C$	mn – 1		

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

Statistika F má Fisher-Snedecorovo F-rozdělení o f_1= (m – 1) a f_2 = m(n – 1) stupních volnosti.

Analýza rozptylu jednoduchého třídění **– nevyvážený model**

Variabilita	Součet čtverců	Stupně volnosti	Rozptyl	Testovací kritérium
Mezi třídami	$S_1 = \sum_{i=1}^{m}\frac{x_{i\bullet}^2}{n_i} - C$	m – 1	$s_1^2 = \frac{S_1}{m-1}$	$F = \frac{s_1^2}{s_r^2}$
Uvnitř tříd	$S_r = S - S_1$	$\sum_{i=1}^{m} n_i - m$	$s_r^2 = \frac{S_r}{\sum_{i=1}^{m} n_i - m}$	
Celková	$S = \sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij}^2 - C$	$\sum_{i=1}^{m} n_i - 1$		

kde $C = \frac{x_{\bullet\bullet}^2}{\sum n_i}$

Statistika F má Fisher-Snedecorovo F-rozdělení o f_1= (m – 1) a f_2 = $\left(\sum_{i=1}^{m} n_i - m\right)$ stupních volnosti.

Analýza rozptylu dvojného třídění s jedním pozorováním v každé podtřídě

Zdroje variability	Součet čtverců odchylek hodnot x_{ij}	Stupně volnosti f	Rozptyl	Testovací kritérium
Mezi řádky	$S_1 = \frac{1}{n}\sum_{i=1}^{m} x_{i\bullet}^2 - C$	m – 1	$s_1^2 = \frac{S_1}{m-1}$	$F_1 = \frac{s_1^2}{s_r^2}$
Mezi sloupci	$S_2 = \frac{1}{m}\sum_{j=1}^{n} x_{\bullet j}^2 - C$	n – 1	$s_2^2 = \frac{S_2}{n-1}$	$F_2 = \frac{s_2^2}{s_r^2}$
Reziduální	$S_r = S - S_1 - S_2$	(m – 1)(n – 1)	$s_r^2 = \frac{S_r}{(m-1)(n-1)}$	
Celková	$S = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}^2 - C$	mn – 1		

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

Metody mnohonásobného srovnávání

Duncanova metoda uspořádání průměrů

Kritická hodnota diferencí $s_{\bar{x}} \cdot R_{p(f)\alpha}$

$$s_{\bar{x}} = \sqrt{\frac{s_r^2}{n}}$$

$R_{p,f,\alpha}$ – tabulková hodnota pro Duncanův test, kde:

p – počet srovnávaných průměrů

f – počet stupňů volnosti reziduálního rozptylu

α – hladina významnosti

Kramerova metoda $\left|\bar{x}_i - \bar{x}_j\right|$ .... $\sqrt{\frac{1}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \cdot s_r \cdot R_{p(f)\alpha}$

Scheffého S-metoda $\left|\bar{x}_i - \bar{x}_j\right|$ .... $\sqrt{\left(\frac{1}{n_i} + \frac{1}{n_j}\right)(m-1) \cdot s_r^2 \cdot F_{\alpha(f_1, f_r)}}$

Počet stupňů volnosti F-rozdělení je $f_1 = (m - 1)$, $f_r = m(n - 1) = \left(\sum_{i=1}^{m} n_i - m\right)$.

Tuckeyova T-metoda $\left|\bar{x}_i - \bar{x}_j\right|$ .... $q_{\alpha, f_r, m} \cdot \sqrt{\frac{s_r^2}{n}}$.

Veličina $q_{\alpha, f_r, m}$ je kritická hodnota studentizovaného rozpětí pro hladinu významnosti α, stupně volnosti reziduálního rozptylu $f_r = [m(n - 1)]$ a m počet srovnávaných průměrů.

Testy neparametrické

1. Testy dobré shody

χ^2 – test dobré shody

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

kde: n_j ….. empirické (skutečné) četnosti v j-té třídě

np_j … teoretické četnosti v j-té třídě (j = 1, 2, ….., k)

k ….. počet tříd (skupin)

Testovací kritérium má za předpokladu, že provádíme dostatečně velký výběr, přibližně χ^2-rozdělení pro f = (k – c – 1) stupně volnosti (k je počet intervalů a c je počet parametrů, které nejsou hypotézou H_0 specifikovány).

Kolmogorov - Smirnovův test

$$D = \frac{1}{n} \max \left| N_j - H_j \right|$$

kde: N_j ………… kumulativní empirické četnosti

H_j ………… kumulativní teoretické četnosti

n …………. rozsah sledovaného souboru

Tabulka kritických hodnot D_α je sestavena pouze pro $n \leq 40$. Pro výběry větších rozsahů se kritické hodnoty určí podle vztahů:

$$D_{0,05} = \frac{1{,}36}{\sqrt{n}} \qquad D_{0,01} = \frac{1{,}63}{\sqrt{n}}$$

2. Klasické neparametrické testy (pořadové)

Wilcoxonův-Whiteův test

$T = \min(T_x, T_y)$

$T_{\alpha(m,n)}$ – kritická hodnota pro m a n – rozsahy souborů a α – hladinu významnosti

Znaménkový test

$Z = \min(Z^+, Z^-)$

$Z_{\alpha(n)}$ – kritická hodnota, kde n je počet nenulových diferencí

Wilcoxonův test

$W = \min(W^-, W^+)$

$W_{\alpha(n)}$ – kritická hodnota pro α a n, kde n je počet nenulových diferencí

Kruskal - Wallisův test

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

Testové kritérium se řídí χ^2-rozdělením o (k − 1) stupních volnosti (kde k je počet úrovní třídícího znaku).

3. Ostatní neparametrické testy

Dixonův test

$$Q_1 = \frac{x_2 - x_1}{x_n - x_1} \qquad Q_n = \frac{x_n - x_{n-1}}{x_n - x_1}$$

$Q_{\alpha(n)}$ – tabulková hodnota pro Dixonův test, hladinu významnosti α a n (kde n je rozsah souboru)

III. Analýza závislosti kvantitativních znaků

Jednoduchá korelace a regrese

funkce lineární (přímka)

$$y' = a + b \cdot x_i$$

Soustava normálních rovnic

$$na_{yx} + b_{yx} \cdot \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$a_{yx} \sum_{i=1}^{n} x_i + b_{yx} \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

Výpočtové vzorce

$b_{yx} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$	$b_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum y_i^2 - \left(\sum y_i\right)^2}$
$b_{yx} = \frac{\operatorname{cov} xy}{s_x^2}$	$b_{xy} = \frac{\operatorname{cov} xy}{s_y^2}$
$b_{yx} = r \cdot \frac{s_y}{s_x}$	$b_{xy} = r \cdot \frac{s_x}{s_y}$
$a_{yx} = \bar{y} - b_{yx} \cdot \bar{x}$	$a_{xy} = \bar{x} - b_{xy} \cdot \bar{y}$
$\operatorname{cov} xy = \overline{xy} - \bar{x} \cdot \bar{y}$	

Síla závislosti – korelační koeficient

$$r_{xy} = r_{yx} = \frac{\operatorname{cov} xy}{s_x \cdot s_y} \qquad r_{yx} = r_{xy} = \pm\sqrt{b_{yx} \cdot b_{xy}}$$

$$r_{yx} = r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{\left[n\sum x_i^2 - \left(\sum x_i\right)^2\right] \cdot \left[n\sum y_i^2 - \left(\sum y_i\right)^2\right]}}$$

Síla závislosti – Spearmanův koeficient pořadové korelace

$$r_S = 1 - \frac{6 \cdot \sum d_i^2}{n \cdot (n^2 - 1)},$$ kde d_i = rozdíly pořadí R_x a R_y

Maticový počet

Regresní funkce $\vec{y} = X\vec{b} + \vec{\varepsilon}$

Soustavu normálních rovnic lze zapsat: $X^T \cdot \vec{y} = X^T \cdot X \cdot \vec{b}$

Index korelace

$$I_{yx} = \sqrt{\frac{\vec{b}^T \cdot X^T \cdot \vec{y} - \frac{1}{n}(\Sigma y_i)^2}{\vec{y}^T \cdot \vec{y} - \frac{1}{n}(\Sigma y_i)^2}}$$

Testování výběrových charakteristik korelace a regrese

Test korelačního koeficientu

$$t_r = \frac{|r_{yx}|}{\sqrt{\frac{1 - r_{yx}^2}{n - 2}}}$$

t_r má Studentovo t-rozdělení pro (n – 2) stupně volnosti

Test o regresním koeficientu

$$t_b = \frac{|b_{yx}|}{s_{b_{yx}}} \qquad s_{b_{yx}} = \frac{s_y}{s_x} \cdot \sqrt{\frac{1 - r_{yx}^2}{n - 2}}$$

$$t_b = \frac{|b_{xy}|}{s_{b_{xy}}} \qquad s_{b_{xy}} = \frac{s_x}{s_y} \cdot \sqrt{\frac{1 - r_{yx}^2}{n - 2}}$$

t_b má Studentovo t-rozdělení pro (n – 2) stupně volnosti

Testování regresní funkce pomocí analýzy rozptylu

Variabilita	Součet čtverců	Stupně volnosti	Rozptyl	Testovací statistika
regrese	$S_1 = \sum_{i=1}^{n}(y_i' - \bar{y})^2$	p – 1	$s_1^2 = \dfrac{S_1}{p-1}$	$F = \dfrac{s_1^2}{s_r^2}$
kolem regrese	$S_r = \sum_{i=1}^{n}(y_i - y_i')^2$	n – p	$s_r^2 = \dfrac{S_r}{n-p}$	
celková	$S = \sum_{i=1}^{n}(y_i - \bar{y})^2$	n – 1		

p = počet parametrů regresní funkce

Statistika F má Snedecorovo F-rozdělení o $[(p-1),(n-p)]$ stupních volnosti.

Bodový odhad korelačního koeficientu $\hat{\rho}_{yx}$

$$\hat{\rho}_{yx} = \sqrt{1-(1-r)^2 \cdot \frac{n-1}{n-2}}$$

Intervalový odhad korelačního koeficientu ρ_{yx}

je-li n < 100, potom Fisherova z-transformace

$$z \in (z_r \pm u_\alpha \cdot s_z)$$

$$s_z = \frac{1}{\sqrt{n-3}}$$

je-li n > 100, potom $r_{yx} \pm u_\alpha \cdot s_r$

$$s_r = \frac{1-r_{yx}^2}{\sqrt{n-2}}$$

Intervalový odhad regresního koeficientu β_{yx}:

$$b_{yx} \pm t_{\alpha(n-2)} \cdot s_{b_{yx}}, \qquad \text{kde } s_{byx} = \frac{s_y}{s_x} \cdot \sqrt{\frac{1-r_{yx}^2}{n-2}}$$

$$b_{xy} \pm t_{\alpha(n-2)} \cdot s_{b_{xy}}, \qquad \text{kde } s_{bxy} = \frac{s_x}{s_y} \cdot \sqrt{\frac{1-r_{yx}^2}{n-2}}$$