Pravděpodobnost

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

Věta o násobení pravděpodobností $P(A \cap B) = P(A) \,.\, P(B)$

Věta o sčítání pravděpodobností $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Podmíněná pravděpodobnost $P(A \cap B) = P(A) \,.\, P(B/A) = P(B) \cdot P(A/B)$

Úplná pravděpodobnost $P(A)=\sum P(B_i)\cdot P(A/B_i)$

Bayesův vzorec $P(B_i/A)=\frac{P(B_i)\cdot P(A/B_i)}{\sum P(B_i)\cdot P(A/B_i)}$

Distribuční funkce $P(X < x) = F(x)$

Normování náhodné veličiny $U=\frac{X-\mu}{\sigma}$

Doplnění k normálnímu rozdělení $F(-u) = 1 - F(u)$ $P(a < X < b) = F(b) - F(a)$

Binomické rozdělení $P(X=k)=\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}$

Poissonovo rozdělení $P(X=k)=\frac{e^{-\lambda}\cdot\lambda^k}{k!}$

Hypergeometrické rozdělení $P(X=k)=\frac{\binom{M}{k}\cdot\binom{N-M}{n-k}}{\binom{N}{n}}$

Rovnoměrné rozdělení $f(x)=\frac{1}{b-a}$ $F(x)=\frac{x-a}{b-a}$

Exponenciální rozdělení $f(x)=\lambda e^{-\lambda x}$ $F(x)=1-e^{-\lambda x}$

Základní statistické charakteristiky

Aritmetický průměr

Forma prostá $\bar{x}=\frac{\sum x_i}{n}$ Forma vážená $\bar{x}=\frac{\sum x_i n_i}{\sum n_i}$

Rozptyl

Forma prostá $s_0^2 = \frac{\sum (x_i - \bar{x})^2}{n}$ $s_0^2 = \frac{\sum x_i^2 - \bar{x}\sum x_i}{n}$ $s_0^2 = \frac{\sum x_i^2}{n} - (\bar{x})^2$

Forma vážená $s_0^2 = \frac{\sum (x_i - \bar{x})^2 \cdot n_i}{n}$ $s_0^2 = \frac{\sum x_i^2 n_i - \bar{x}\sum x_i n_i}{n}$ $s_0^2 = \frac{\sum x_i^2 n_i}{n} - \left(\frac{\sum x_i n_i}{n}\right)^2$

Směrodatná odchylka $s_0 = \sqrt{s_0^2}$

Variační koeficient $v = \frac{s_0}{\bar{x}} \cdot 100$ [%]

Teorie odhadu

Bodový odhad rozptylu

výběr s vracením $s^2 = s_0^2 \cdot \frac{n}{n-1}$

výběr bez vracení $s^2 = s_0^2 \cdot \frac{n}{n-1} \cdot \frac{N-1}{N}$

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

a) Známe-li rozptyl ZS:

u_α – kritická hodnota normálního rozdělení

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

výběr bez vracení $\Delta = u_\alpha \cdot \sqrt{\frac{\sigma^2}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

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

určení spolehlivosti odhadu $u_\alpha = \sqrt{\frac{n \cdot \Delta^2}{\sigma^2}}$ $u_\alpha = \sqrt{\frac{n \cdot (N-1) \cdot \Delta^2}{\sigma^2 \cdot (N-n)}}$

b) Neznáme-li rozptyl ZS:

t_α – kritická hodnota Studentova t-rozdělení pro (n – 1) stupeň volnosti

výběr s vracením $\Delta = t_{\alpha(n-1)} \cdot \sqrt{\frac{s^2}{n}}$

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

určení rozsahu výběru $n = \frac{t_\alpha^2 \cdot s^2}{\Delta^2}$ $n = \frac{t_\alpha^2 \cdot s^2 \cdot N}{\Delta^2 \cdot (N-1) + t_\alpha^2 \cdot s^2}$

určení spolehlivosti odhadu $t_\alpha = \sqrt{\frac{n \cdot \Delta^2}{s^2}}$ $t_\alpha = \sqrt{\frac{n \cdot (N-1) \cdot \Delta^2}{s^2 \cdot (N-n)}}$

Intervalový odhad rozptylu

χ^2_α – kritická hodnota χ^2 – rozdělení pro (n – 1) stupeň volnosti

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

Intervalový odhad relativní četnosti $P\left(f_i - \Delta < p < f_i + \Delta\right) = 1-\alpha$

u_α – kritická hodnota normálního rozdělení

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

určení rozsahu výběru $n = \frac{u^2_\alpha \cdot f_i(1-f_i)}{\Delta^2}$ $n = \frac{u^2_\alpha \cdot f_i(1-f_i)\cdot N}{\Delta^2\cdot(N-1) + u^2_\alpha\cdot f_i(1-f_i)}$

určení spolehlivosti odhadu $u_\alpha = \sqrt{\frac{n\cdot\Delta^2}{f_i(1-f_i)}}$ $u_\alpha = \sqrt{\frac{n\cdot(N-1)\cdot\Delta^2}{f_i(1-f_i)\cdot(N-n)}}$

Testování statistických hypotéz

Jednovýběrové testy

Test hypotézy o hodnotě průměru

známe-li rozptyl ZS: $u = \frac{\bar{x}-\mu_0}{\sqrt{\frac{\sigma^2}{n}}}$ neznáme-li rozptyl ZS: $t = \frac{\bar{x}-\mu_0}{\sqrt{\frac{s^2}{n}}}$

kritická hodnota: u – normální rozdělení, t – Studentovo t-rozdělení pro (n – 1) stupeň volnosti

Test hypotézy o hodnotě rozptylu

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

kritická hodnota: χ^2-rozdělení o (n – 1) stupni volnosti

Test hypotézy o hodnotě relativní četnosti

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

kritická hodnota: u – normální rozdělení

Dvouvýběrové testy

Test hypotézy o shodě dvou rozptylů

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

kritická hodnota: F – rozdělení pro (m – 1) a (n – 1) stupně volnosti

Test hypotézy o shodě dvou průměrů

a) dvouvýběrový t – test při shodných rozptylech

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

kritická hodnota: Studentovo t-rozdělení pro (m + n – 2) stupně volnosti

b) t – test při nestejných rozptylech (Welchův test)

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

kritická hodnota: Studentovo t-rozdělení pro f stupňů volnosti, kde

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

c) t – test při nestejných rozptylech (Behrens – Fisherův test)

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

kritická hodnota: přepočítaná hodnota Studentova t-rozdělení

$$t_\alpha^* = \frac{t_{\alpha(m-1)} \cdot \frac{s_1^2}{m} + t_{\alpha(n-1)} \cdot \frac{s_2^2}{n}}{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

d) t – test pro závislé výběry

$$t = \frac{\bar{d}}{\sqrt{\frac{s_d^2}{n}}} \qquad \bar{d} = \frac{\sum d_i}{n} = \frac{\sum (x_i - y_i)}{n} \qquad s_d^2 = \frac{1}{n-1}\sum \left(d_i - \bar{d}\right)^2$$

kritická hodnota: Studentovo t-rozdělení pro (n – 1) stupeň volnosti

<u>Test hypotézy o shodě dvou relativních četností</u>

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

kritická hodnota: u – normální rozdělení

<u>Regresní a korelační analýza</u>

Soustava normálních rovnic – přímka

$$na + b\sum x_i = \sum y_i$$
$$a\sum x_i + b\sum x_i^2 = \sum x_i y_i$$

Výpočtové vzorce

$$b_{yx} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \qquad a_{yx} = \bar{y} - b_{yx}\cdot\bar{x}$$

$$b_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum y_i^2 - \left(\sum y_i\right)^2} \qquad a_{xy} = \bar{x} - b_{xy}\cdot\bar{y}$$

Korelační koeficient

$$r = \frac{\mathrm{cov}(xy)}{s_x \cdot s_y} \qquad \mathrm{cov}(xy) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} \qquad r = \pm\sqrt{b_{yx}\cdot b_{xy}}$$

$$r = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{\left[n\sum x_i^2 - \left(\sum x_i\right)^2\right]\cdot\left[n\sum y_i^2 - \left(\sum y_i\right)^2\right]}} \qquad r = b_{yx}\cdot\frac{s_x}{s_y} = b_{xy}\cdot\frac{s_y}{s_x}$$

Spearmanův koeficient

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

Testy a odhady v regresní a korelační analýze

a) Test korelačního koeficientu

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

kritická hodnota: Studentovo t-rozdělení pro (n – 2) stupně volnosti

b) Test regresního koeficientu

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

kritická hodnota: Studentovo t-rozdělení pro (n – 2) stupně volnosti

c) Intervalový odhad regresního koeficientu

$$P\left(b_{yx} - t_{\alpha(n-2)} \cdot s_{b_{yx}} \le \beta_{yx} \le b_{yx} + t_{\alpha(n-2)} \cdot s_{b_{yx}}\right) = 1-\alpha$$

d) Bodový odhad korelačního koeficientu

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

e) Intervalový odhad korelačního koeficientu

n > 100: $P\left(r - u_\alpha \cdot s_r \le \rho \le r + u_\alpha \cdot s_r\right) = 1-\alpha$ $\qquad s_r = \frac{1-r^2}{\sqrt{n}}$

n < 100: $r \rightarrow z_r = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$

$$P\left(z_r - u_\alpha \cdot s_{z_r} \le Z \le z_r + u_\alpha \cdot s_{z_r}\right) = 1-\alpha \qquad s_{z_r} = \frac{1}{\sqrt{n-3}}$$

$$z_r \rightarrow r = \frac{e^{2Z}-1}{e^{2Z}+1}$$