

# Vzorečník

## Základní vzorce

varianční rozpětí:  $X_{max} - X_{min}$       průměr:  $m = \frac{X_1 + X_2 + \dots + X_n}{n}$   
výběrový rozptyl:  $s^2 = \frac{\sum(X_i - m)^2}{n-1}$       z-skor:  $z_i = \frac{X_i - m}{s}$   
mezikvartilové rozpětí  $IQR = Q_3 - Q_1$       percentil:  $X_{percentil} = \frac{c+0,5f}{N}$   
T-skor:  $T = 50 + 10.z$

## Pravděpodobnost

pravděpodobnost jevu A:  $P(A) = \frac{n}{m}$   
permutace n prvků:  $n!$       kombinace r prvků z n-prvkové množiny:  $\frac{n!}{r!(n-r)!}$   
 $P(A \cap B) = P(A) \cdot P(B)$        $P(A \cup B) = P(A) + P(B) - P(A \cap B)$   
 $P(A|B) = \frac{P(A \cap B)}{P(B)}$       šance:  $O(A) = \frac{P(A)}{P(\neg A)}$   
Bayesův teorém:  $P(A|B) = \frac{P(A) \cdot P(B|A)}{P(A) \cdot P(B|A) + P(A') \cdot P(B|A')}$

## Korelace

kovariance:  $c_{xy} = \frac{\sum x_i y_i}{n-1} = \frac{(X_1 - m_x) \cdot (Y_1 - m_y) + \dots + (X_i - m_x) \cdot (Y_i - m_y)}{n-1}$   
Pearsonův korelační koeficient:  $r_{xy} = \frac{c_{xy}}{s_x \cdot s_y} = \frac{\sum z_x z_y}{n-1}$   
Spermanův korelační koeficient:  $r_s = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)}$   
Kendallov  $\tau$ :  $\tau = \frac{K - D}{\frac{n \cdot (n-1)}{2}}$   
parciální korelace:  $r_{AB.C} = \frac{r_{AB} - r_{AC} r_{BC}}{\sqrt{1 - r_{AC}^2} \sqrt{1 - r_{BC}^2}}$   
semiparciální korelace:  $r_{A(B.C)} = \frac{r_{AB} - r_{AC} r_{BC}}{\sqrt{1 - r_{BC}^2}}$   
vnitřní konzistence:  $r_{tt} = \frac{k}{k-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2}\right)$

## Lineární regrese

$Y' = a + bX$        $b = r_{xy} \cdot \left(\frac{s_y}{s_x}\right)$        $a = m_y - b \cdot m_x$   
 $s_{reg}^2 = \frac{\sum(m_y - Y')^2}{n-1}$        $s_{res}^2 = \frac{\sum(Y - Y')^2}{n-1} = s_y^2(1 - r^2)$        $s_y^2 = s_{reg}^2 + s_{res}^2$   
koeficient determinace:  $R^2 = \frac{s_{reg}^2}{s_y^2}$

## Testování hypotéz

### Intervaly spolehlivosti

výběrová chyba průměru:  $\sigma_m = \frac{\sigma}{\sqrt{n}}$        $s_m = \frac{s}{\sqrt{n}}$   
interval spolehlivosti pro průměr:  $m - z \cdot \sigma_m$  [nebo]  $t \cdot s_m < \mu < m + z \cdot \sigma_m$  [nebo]  $t \cdot s_m$   
výběrová chyba korelace:  $s_0 = \frac{1}{\sqrt{n-3}}$

interval spolehlivosti pro korelaci:  $fischer_z - z_{krit} \cdot s_0 < \rho < fischer_z + z_{krit} \cdot s_0$

z-skor pro testování signifikance korelace:  $z_r = \frac{fischer_z}{s_0}$

## T-testy

	t-test pro jeden výběr		t-test pro dva nezávislé výběry	t-test pro dva závislé výběry
$s_d$	$\sigma_m = \frac{\sigma}{\sqrt{n}}$	$s_m = \frac{s}{\sqrt{n}}$	$s_d = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$	$\sqrt{\frac{1}{n} \cdot (s_1^2 + s_2^2 - 2s_1s_2r)}$
df/v	-	$df = n - 1$	$df = n_1 + n_2 - 2$	$df = n - 1$
t	$z = \frac{m-\mu}{\sigma_m}$	$t = \frac{m-\mu}{s_m}$	$t = \frac{m_1-m_2}{s_d} = \frac{d}{s_d}$	$t = \frac{m_1-m_2}{s_d} = \frac{d}{s_d}$
funkce	NORMDIST	TDIST	TDIST(t,v,2)	TDIST(t,v,2)
ES			Cohenovo d = $\frac{d}{s_{pooled}}$	Cohenovo d = $\frac{d}{s_{pooled}}$

t-test pro dvě nezávislé skupiny:  $s_{pooled} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$ ;  $s_d = s_{pooled} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

t-test pro dvě závislé skupiny:  $s_{pooled} = \sqrt{s_1^2 + s_2^2 - 2s_1s_2r} = \sqrt{\frac{\sum(d_i-d)^2}{n-1}}$ ;  $s_d = \frac{s_{pooled}}{\sqrt{n}}$

síla testu nulové hypotézy o průměrech:  $(1 - \beta) = Fn.(-1, 96 + \frac{|\delta| \cdot \sqrt{N}}{\sigma})$

(Fn je distribuční funkce normálního rozložení;  $\delta$  je:  $\mu - \mu_0$  v situaci 1 výběru,  $\mu_d$  u párového testu,  $\mu_1 - \mu_2$  pro 2 nezávislé výběry)

pravděpodobnost výskytu alespoň 1 chyby I. typu u k nezávislých srovnání:  $p = 1 - (1 - \alpha)^k$

Cohenovo d':  $d' = \frac{m_1-m_2}{s_{control}}$

Cohenovo d a r:  $r = \sqrt{\frac{d^2}{d^2+4}}$   $d = \frac{2r}{\sqrt{1-r^2}}$

Interval spolehlivosti pro Cohenovo d:  $d \pm t \cdot s_d$

## Chi-kvadrát

$\chi^2$  test dobré shody:  $\chi^2 = \sum \frac{(n_i - np_i)^2}{np_i}$ ;  $df = k - 1$

(n - sledovaná četnost; np - očekávaná četnost)

směrodatná chyba  $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$

95% interval spolehlivosti:  $p \pm 2\sigma_p$

$\chi^2$  test homogenity:  $m_{sr} = \frac{n_s n_r}{n}$ ;  $\chi^2 = \sum \sum \frac{(n_{sr} - m_{sr})^2}{m_{sr}}$ ;  $df = (r - 1) \cdot (s - 1)$

( $m_{sr}$  - očekávaná četnost v konkrétní bunce tabulky;  $n_{s/r}$  - zjištěná četnost v konkrétním sloupci/řádku;

$n_{sr}$  - zjištěná četnost v konkrétní bunce tabulky)

Standardizovaná rezidua:  $R = \frac{n_{rs} - m_{rs}}{\sqrt{m_{rs}}}$