Test hypotézy o střední hodnotě normálního rozdělení.

Věta. Nechť základní soubor X má normální rozdělení se střední hodnotou μ_0, nechť $\alpha \in (0, 1)$ a nechť $(X_1, \dots, X_n)$ je náhodný výběr z X. Nechť

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Pak

$$\frac{\overline{X} - \mu_0}{S}\sqrt{n}$$

má rozdělení $\mathrm{t}(n-1)$ a tedy platí

$$P\left(t_{\frac{\alpha}{2}}(n-1) < \frac{\overline{X} - \mu_0}{S}\sqrt{n} < t_{1-\frac{\alpha}{2}}(n-1)\right) = 1 - \alpha.$$

$$H_0\colon \mu = \mu_0, \qquad H_1\colon \mu \neq \mu_0.$$

H_0 *zamítáme*, pokud

$\dfrac{\overline{x} - \mu_0}{s}\sqrt{n}$ neleží v intervalu $\big(t_{\frac{\alpha}{2}}(n-1),\ t_{1-\frac{\alpha}{2}}(n-1)\big)$.

χ^2 – test dobré shody (goodness of fit test).

Předpokládaná *hladina významnosti* α.

a) Určíme předpokládaný typ rozdělení (hypotéza H_0) a stanovíme na kolika parametrech závisí jeho distribuční funkce. Počet parametrů označme s.

b) Vypočteme bodové odhady všech parametrů.

c) Ve výběrovém souboru nalezneme největší hodnotu $x_{\max}$ a nejmenší hodnotu $x_{\min}$.

d) Interval $\langle x_{\min}, x_{\max}\rangle$ rozdělíme na r podintervalů $\langle t_{i-1}, t_i)$, $i = 1, \dots, r$, délka intervalů nemusí být stejná. Ke každému z těchto intervalů vypočteme (teoretickou) pravděpodobnost

$$p_i = P(t_{i-1} \le x < t_i) = F(t_i) - F(t_{i-1}).$$

Při výpočtu těchto pravděpodobností používáme hodnoty parametrů stanovené v bodě b).
Není-li pro některé i splněna podmínka $np_i \ge 5$, sloučíme příslušné intervaly a přepočítáme pravděpodobnosti p_i.

e) Hodnoty výběrového souboru roztřídíme do jednotlivých intervalů a určíme absolutní četnosti n_i těchto intervalů (t.j. počet hodnot v každém intervalu).

f) Vypočteme hodnotu statistiky

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i}.$$

g) Porovnáme ji s kvantilem

$$\chi^2_{1-\alpha}(r - s - 1).$$

Je-li

$$\chi^2 \leq \chi^2_{1-\alpha}(r - s - 1),$$

pak H_0 přijímáme, je-li

$$\chi^2 > \chi^2_{1-\alpha}(r - s - 1),$$

pak H_0 zamítáme.

Spolehlivost 90% $\implies$ $\alpha = 0.1$

	$\langle 171, 176)$	$\langle 176, 181)$	$\langle 181, 186)$	$\langle 186, 191)$
p_i	0.127	0.377	0.300	0.092
n_i	7	12	11	4
np_i	4.64	13.19	10.50	3.22

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i} = \frac{2.36^2}{4.64} + \frac{1.19^2}{13.19} + \frac{0.5^2}{10.5} + \frac{0.78^2}{3.33} = 1.51$$

$$\chi^2_{1-\alpha}(r - s - 1) = \chi^2_{0.90}(1) = 2.71 > 1.51.$$

Přijímáme H_0, soubor má normální rozdělení $N(179.8, 24.1)$.

Spolehlivost 90% $\implies$ $\alpha = 0.1$

	$\langle 156, 171)$	$\langle 171, 177)$	$\langle 177, 184)$	$\langle 184, 191)$
p_i	0.217	0.202	0.371	0.175
n_i	3	11	17	9
np_i	4.48	12.32	14.84	7.00

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i} = \frac{1.48^2}{4.48} + \frac{1.32^2}{12.32} + \frac{2.16^2}{14.84} + \frac{2^2}{7} = 1.52$$

$$\chi^2_{1-\alpha}(r-s-1) = \chi^2_{0.90}(1) = 2.71 > 1.52.$$

Přijímáme H_0, soubor má normální rozdělení $N(178.4, 47.8)$.

Alternativní rozdělení

$X = 0, 1; \qquad p(1) = p, \quad p(0) = 1 - p.$

$E(X) = p, \qquad \text{var}(X) = p - p^2 = p\,(1 - p).$

Binomické rozdělení

$$X = 0, 1, \ldots, n; \qquad p(k) = \binom{n}{k} p^k\,(1 - p)^{n-k}$$

Značení: $X \sim \text{Bi}\,(n,\, p).$

Jsou-li $X_1, \ldots, X_n$ nezávislé veličiny s alternativním rozdělením, pak $X = X_1 + \cdots + X_n$ má binomické rozdělení $\text{Bi}\,(n,\, p)$.

$E(X) = E(X_1) + \cdots + E(X_n) = np$

$\text{var}(X) = \text{var}(X_1) + \cdots + \text{var}(X_n) = np(1 - p)$

$$P(X \leq k) = \sum_{i=0}^{k} \binom{n}{i} p^i\,(1 - p)^{n-i}$$

pro $p = \frac{1}{2}$: $\qquad P(X \leq k) = \left(\frac{1}{2}\right)^n \sum_{i=0}^{k} \binom{n}{i}$

Naparametrické testy – znaménkový test.

Medián $q_{0.5}$: $\quad P(X < q_{0.5}) = P(X > q_{0.5}) = 0.5.$

$X_1, \dots, X_n$ – náhodný výběr ze spojitého (neznámého) rozdělení, x_0 libovolné číslo, dána hladina významnosti α.

$$H_0: \quad q_{0.5} = x_0, \qquad H_1: \quad q_{0.5} < x_0.$$

Definujme

$$Z_i = X_i - x_0, \quad i = 1, \dots, n$$

a nechť Z udává počet kladných hodnot mezi $Z_1, \dots, Z_n$;

$$Z \sim \mathrm{Bi}\left(n, \tfrac{1}{2}\right).$$

H_0 *zamítáme*, pokud $Z < c$, kde

$$\left(\frac{1}{2}\right)^n \sum_{i=0}^{c} \binom{n}{i} \le \alpha < \left(\frac{1}{2}\right)^n \sum_{i=0}^{c+1} \binom{n}{i}.$$

Příklad

i	1	2	3	4	5	6	7	8	9	10	11	12
X_i	32	21	28	15	8	22	17	35	20	31	17	11
$X_i - 25$	$+$	$-$	$+$	$-$	$-$	$-$	$-$	$+$	$-$	$+$	$-$	$-$
Z_i	1	0	1	0	0	0	0	1	0	1	0	0

$$Z = 4$$

$$p(Z = 0) = \left(\tfrac{1}{2}\right)^{12} = 0.00024 < 0.05$$

$$p(Z \leq 1) = \left(\tfrac{1}{2}\right)^{12} \left(1 + \tbinom{12}{1}\right) = 0.00318 < 0.05$$

$$p(Z \leq 2) = \left(\tfrac{1}{2}\right)^{12} \left(1 + \tbinom{12}{1} + \tbinom{12}{2}\right) = 0.0193 < 0.05$$

$$p(Z \leq 3) = \left(\tfrac{1}{2}\right)^{12} \left(1 + \tbinom{12}{1} + \tbinom{12}{2} + \tbinom{12}{3}\right) = 0.0730 > 0.05$$

$c = 2 < 4 \Longrightarrow H_0$ **nezamítáme**.