Interval spolehlivosti pro rozdíl středních hodnot

Nechť náhodné výběry X_1 a X_2 jsou nezávislé a nechť rozptyly σ_1 a σ_2 jsou neznámé, přičemž platí $\sigma_1 \neq \sigma_2$. Přibližným $100(1-\alpha)$procentním intervalem spolehlivosti pro rozdíl $\mu_1 - \mu_2$ je interval

$$\bar{x}_1 - \bar{x}_2 - t_{1-\alpha/2}(v)\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + t_{1-\alpha/2}(v)\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},$$

přičemž

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{S_2^2}{n_2}\right)^2}$$

Testování korelace

Příklad 1

Jsou známy výsledky testů ze dvou předmětů zjištěné u 8 náhodně vybraných studentů určitého oboru.

č.	1	2	3	4	5	6	7	8
1. test	80	50	36	58	42	60	56	68
2. test	65	60	35	39	48	64	48	61

Za předpokladu, že uvedené výsledky jsou číselné realizace náhodného výběru z dvourozměrného normálního rozdělení, testujte na hladině významnosti 0,05 hypotézu o nezávislosti výsledků obou testů.

$$T = \frac{R}{\sqrt{1-R^2}}\sqrt{n-2} \sim t(n-2) \qquad H_0 : \rho = 0 \qquad H_1 : \rho \neq 0$$

$$R_{12} \doteq 0.6264 \qquad T_0 \doteq 1.18 \qquad t_{0.975}(6) \doteq 2.45$$

Hypotézu H_0 nezamítáme (způsobeno malým rozsahem výběru).

Příklad 2

U 600 vzorků rudy byl stanoven obsah železa dvěma metodami. Výběrový korelační koeficient výsledků těchto metod je 0,85. V literatuře se uvádí, že tento koeficient má být 0,9. Za předpokladu dvourozměrné normality naměřených dat testujte na hladině významnosti 0,05 hypotézu, že rozdíl mezi těmito koeficienty je náhodný.

$$Z = \frac{1}{2}\ln\frac{1+R}{1-R} \qquad U = (Z - Z_\rho)\sqrt{n-3} \sim N(0,1) \qquad H_0 : \rho = 0.9 \qquad H_1 : \rho \neq 0.9$$

$$Z \doteq 1.2562 \qquad Z_\rho \doteq 1.4722 \qquad U_0 \doteq -5.2777 \qquad u_{0.975} \doteq 1.9600$$

Hypotézu H_0 zamítáme.

Příklad 3

Lékařský výzkum se zabýval sledováním koncentrací látek A a B v moči pacientů trpících určitou ledvinovou chorobou. U 100 zdravých osob byl výběrový koeficient korelace 0,65 a u 142 osob trpících zmíněnou chorobou byl 0,37. Na hladině významnosti 0,05 testujte hypotézu, že se korelační koeficienty neliší. Předpokládejte dvourozměrnou normalitu.

$$Z = \frac{1}{2}\ln\frac{1+R}{1-R} \qquad U = \frac{(Z - Z_*)}{\sqrt{\frac{1}{n_-3} + \frac{1}{n_*-3}}} \sim N(0,1) \qquad H_0 : \rho = \rho_* \qquad H_1 : \rho \neq \rho_*$$

$$Z \doteq 0.7753 \qquad Z_* \doteq 0.3884 \qquad U_0 \doteq 2.9249 \qquad u_{0.975} \doteq 1.9600$$

Hypotézu H_0 zamítáme.

Test dobré shody pro normální rozdělení

Príklad 1. Zo základného súboru sme náhodne vybrali výberový súbor s rozsahom $n = 100$ a s výsledkami, ktoré sú uvedené v tab. 9.2.

Tab. 9.2

x	n_i	$n_i x_i'$	$(x_i' - \bar{x})^2$	$n_i(x_i' - \bar{x})^2$	p_i	$n_i^2/p_i n$
< 5	8	20	176,89	1 415,12	0,0537	11,92
5 – 10	10	75	68,89	688,90	0,1412	7,08
10 – 15	25	312,5	10,89	272,25	0,2573	24,29
15 – 20	32	560	2,89	92,48	0,2835	36,12
20 – 25	15	337,5	44,89	673,35	0,19215	11,71
> 30	10	275	136,89	1 368,90	0,07215	13,86
Spolu	$n = 100$	1 580,0		4 511,00	1,0000	104,98

Máme otestovať hypotézu, že základný súbor má normálne rozdelenie. Aby bolo možné určiť pravdepodobnosti p_i zo vzorca (9.51), resp. (9.53), musíme poznať strednú hodnotu m a štandardnú odchýlku σ, ktorými je definované normálne rozdelenie. Pretože tieto hodnoty neboli udané, odhadneme ich z výberového súboru metódou maximálnej vierohodnosti. Ako vieme z kapitoly 7, odhadmi parametrov m a σ, získanými metódou maximálnej vierohodnosti, sú výberový priemer $\bar{X}$, resp. výberová štandardná odchýlka S, kde

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Pretože tu disponujeme výberovými súbormi zoskupenými do intervalov, na výpočet $\bar{x}$ a s použijeme približné vzorce:

$$\bar{x} = \frac{\sum_{l=1}^{r} n_l x'_l}{n}$$

$$s = \sqrt{\frac{1}{n} \sum_{l=1}^{r} n_l (x'_l - \bar{x})^2}$$

kde x'_l je stredom intervalu (x_l, x_{l+1}) pre $l = 1, 2, \ldots, r$. Tab. 9.2 udáva hodnoty, ktoré umožňujú vypočítať $\bar{x}$ a s. Podľa toho

$$\bar{x} = 15{,}8$$
$$s = 6{,}72$$

Keď sme určili hodnoty $\bar{x}$ a s, môžeme vypočítať p_l za predpokladu, že hypotéza je správna, a že súbor má normálne rozdelenie. Tieto pravdepodobnosti môžeme vypočítať dvojakým spôsobom: Buď použijeme distribučnú funkciu normálneho rozdelenia

$$p_1 = F_0(x_{l+1}) - F_0(x_l) \tag{9.54}$$

buď tak, že za východisko zvolíme hodnotu normálneho rozdelenia

$$p_l \doteq f_0(x'_l) \frac{(x_{l+1} - x_l)}{s} \tag{9.55}$$

Je zrejmé, že ak použijeme vzorec (9.55), dostaneme len približné hodnoty pravdepodobností p_l. Použijeme preto presnejší vzorec (9.54). Aby bolo možné v tabuľkách odčítať požadované hodnoty normálnej distribučnej funkcie $F_0(x_l)$, musíme od každého x_l odpočítať strednú hodnotu 15,8 a výsledok vydeliť štandardnou odchýlkou 6,72.

Z tabuliek distribučnej funkcie normálneho rozdelenia

pre $\dfrac{5 - 15{,}8}{6{,}72} = -1{,}61$ dostaneme 0,0537

pre $\dfrac{10 - 15{,}8}{6{,}72} = -0{,}86$ dostaneme 0,1949

$$\text{pre} \quad \frac{15 - 15{,}8}{6{,}72} = -0{,}12 \quad \text{dostaneme} \quad 0{,}4522$$

$$\text{pre} \quad \frac{20 - 15{,}8}{6{,}72} = 0{,}63 \quad \text{dostaneme} \quad 0{,}7357$$

$$\text{pre} \quad \frac{25 - 15{,}8}{6{,}72} = 1{,}46 \quad \text{dostaneme} \quad 0{,}92785$$

Pre prvý interval platí:

$$p_1 = F(-1{,}61) - F(-\infty) = 0{,}0537 - 0 = 0{,}0537$$

Za dolnú hranicu prvého intervalu berieme $-\infty$, lebo predpokladáme, že premenná má normálne rozdelenie. Pre druhý interval platí:

$$p_2 = F(-0{,}86) - F(-1{,}61) = 0{,}1949 - 0{,}0537 = 0{,}1412$$

Podobne dostaneme:

$$p_3 = F(-0{,}12) - F(-0{,}86) = 0{,}4522 - 0{,}1949 = 0{,}2573$$
$$p_4 = F(0{,}63) - F(-0{,}12) = 0{,}7357 - 0{,}4522 = 0{,}2835$$
$$p_5 = F(1{,}46) - F(0{,}63) = 0{,}92785 - 0{,}7357 = 0{,}19215$$
$$p_6 = F(+\infty) - F(1{,}46) = 1 - 0{,}92785 = 0{,}07215$$

Ak sme našli hodnoty p_l, môžeme vypočítať štatistiku χ^2 podľa vzorca (9.53). Pre tento cieľ v prípade každého intervalu vypočítame $\frac{n_l^2}{p_l n}$. Tieto hodnoty sú uvedené v tab. 9.2. Po sčítaní dostaneme:

$$\sum_{l=1}^{r} \frac{n_l^2}{p_l n} = 104{,}98$$

Podľa vzorca (9,53) napokon dostaneme hodnotu štatistiky χ^2

$$\chi^2 = \sum_{l=1}^{r} \frac{n_l^2}{p_l n} - n = 104{,}98 - 100 = 4{,}98$$

Ako vieme, premenná χ^2 bude mať $r - m - 1$ stupňov voľnosti. Keďže $r = 6$ (toľko máme intervalov) a $m = 2$ (toľko bolo odhadovaných parametrov), preto

$$r - m - 1 = 6 - 2 - 1 = 3$$

Zvolíme hladinu významnosti $\alpha = 0{,}05$ a v tabuľkách rozdelenia χ^2 nájdeme hodnotu χ^2_α, ktorá spĺňa podmienku (9.52). Dostaneme $\chi^2_\alpha = 7{,}815$. Pretože pozorované $\chi^2_\alpha = 4{,}98$ neprevyšuje kritickú hodnotu χ^2_α, nie je dôvod na zamietnutie postavenej hypotézy s normálnym rozdelením základného súboru.

Použitá literatura

- Budíková M., Lerch T., Mikoláš Š.: Základní statistické metody. MU Brno 2006
- Hátle J., Likeš J.: Základy počtu pravděpodobnosti a matematické statistiky. SNTL/ALFA 1974
- Sadowski W.: Matematická štatistika. ALFA 1975