8. Neparametrični testi

Doslej smo preizkušali hipoteze o neznanih parametrih (običajno smo predpostavljali normalno porazdelitev); to so parametrični testi.

Če na porazdelitev sl. spremenljivke X ne moremo nič privzeti, potem uporabimo neparametrične teste.

1) test z znaki; neparametrični analog testa T

Na populaciji dana sl. sprem. X s porazd. funkcijo F_X

in Y s porazd. f. F_Y. Obravnavajmo dva sl. vektorja $(X_1, .., X_n)$ in $(Y_1, .., Y_n)$, kjer X_i in Y_i dobimo na istem elementu populacije.
Preizkušamo hipotezo $H_0(F_X = F_Y)$
Def. razlike $D_i = X_i - Y_i$, $i = 1, 2, .., n$
Pr. vzemimo, da so vrednosti za $(D_1, D_2, .., D_n)$ različne od 0; sicer jih izpustimo in zmanjšamo n.
Če velja H_0, potem je $P(D_i > 0) = \frac{1}{2} = P(D_i < 0)$.
Naj bo S^+ število pozitivnih vrednosti v $(D_1, .., D_n)$.
Potem je $S^+ \sim Bin(n, \frac{1}{2})$, tj. $p_k = P(S^+ = k) = \binom{n}{k} \cdot \frac{1}{2^n}$

Pri dani stopnji značilnosti α je kritično območje $\{k : k \leq k_\alpha \text{ ali } k \geq n - k_\alpha\}$, kjer je k_α določen z zahtevama:

$$P(S^+ \leq k_\alpha) = \sum_{k=0}^{k_\alpha} p_k \leq \frac{\alpha}{2} \quad \text{in} \quad P(S^+ \leq k_\alpha + 1) > \frac{\alpha}{2}$$

Pri velikih n ima S^+ pribl. porazd. $N\left(\frac{n}{2}, \frac{\sqrt{n}}{2}\right)$, kar pomeni

$$Z = \frac{S^+ - \frac{n}{2}}{\frac{\sqrt{n}}{2}} \approx N(0,1)$$

<u>Zgled</u> Zniževanje krvnega tlaka pri 10 bolnikih

X_i... sistolični krvni tlak pred zdravljenjem, Y_i pa po njem.

$D_i = X_i - Y_i$ razlika tlakov

$-8 \quad \cancel{0} \quad 2 \quad 4 \quad 9 \quad 14 \quad 19 \quad 22 \quad 32 \quad 35$

$n = 9$, vrednost za S^+ je 8, $\alpha = 0.05$

$$P(S^+ \le 1) = p_0 + p_1 = \binom{9}{0}\cdot\frac{1}{2^9} + \binom{9}{1}\cdot\frac{1}{2^9} = \frac{10}{2^9} \doteq 0.020 < 0.025 = \frac{\alpha}{2}$$

$$P(S^+ \le 2) = \frac{10}{2^9} + \binom{9}{2}\cdot\frac{1}{2^9} = \frac{46}{2^9} \doteq 0.090 > \frac{\alpha}{2}.$$

Torej je $k_\alpha = 1$ in kritično območje $\{0, 1, 8, 9\}$.

H_0 (zdravilo ne učinkuje) zavrnemo.

3) <u>Inverzijski test</u> (<u>Wilcoxon-Mann-Whitney</u>)

Vzorca $(X_1, \dots, X_m)$ in $(Y_1, \dots, Y_n)$ neodvisna in $m \le n$. Vzorčne vrednosti $(x_1, \dots, x_m)$ in $(y_1, \dots, y_n)$ razvrstimo po velikosti:

$$z_1 \le z_2 \le z_3 \le \dots \le z_{m+n}$$

Pripišemo <u>mesta</u> (<u>range</u>) $R_i = \text{rang}\, X_i = k$, če je $x_i = z_k$, $i = 1, 2, \dots, m$

Sl. sprem. $V = R_1 + R_2 + \dots + R_m$ zavzame vrednosti med $1 + 2 + \dots + m = \frac{m(m+1)}{2}$ (vsi x_i so na začetku) in

$$(n+1)+(n+2)+\dots+(n+m) = n\cdot m + \frac{m(m+1)}{2}$$

Če velja $H_0\ (F_X = F_Y)$ in $m+n \geq 20$, $m, n \geq 4$, potem ima V pribl. normalno porazd.

$$N\left(\frac{(m+n+1)\cdot m}{2}, \sqrt{\frac{m\cdot n(m+n+1)}{12}}\right),$$ torej

$$Z = \sqrt{\frac{3}{mn(m+n+1)}}\cdot\left(2\cdot V - (m+n+1)\cdot m\right) \sim N(0,1)$$

Ta test je neparametrični analog primerjalnemu testu.

<u>Zgled</u> zdravili proti nespečnosti, $m = n = 10$

Vzorčne vrednosti razvrstimo po velikosti.

				×			×		×	
z_i:	-1.6	-1.2	-0.2	-0.1	-0.1	0.0	0.1	0.7	0.8	0.8
rang:	1	2	3	4.5	4.5	6	7	8	9.5	9.5
	×	×	×		×			×	×	×
z_i:	1.1	1.6	1.9	2.0	3.4	3.4	3.7	4.4	4.6	5.5
rang:	11	12	13	14	15.5	15.5	17	18	19	20

vrednost za V: $4.5+7+9.5+11+12+13+15.5+18+19+20=$

$=129.5$, za Z: $\sqrt{\dfrac{3}{10\cdot 10\cdot 21}}\left(2\cdot 129.5-21\cdot 10\right)=1.85$

Pri $\alpha=0.05$ je $z_{\alpha/2}=1.96$, torej $K_\alpha=(-\infty,-1.96]\cup[1.96,\infty)$.

H_0 ne moremo zavrniti.

Inverzija med x_i in y_j se pojavi, če ima y_j manjši rang kot x_i.

Če sl. sprem. U meri število inverzij v zap. $z_1 \leq z_2 \leq \ldots \leq z_{m+n}$, potem je

$$V = U + \frac{m(m+1)}{2}$$