PARAMETRIČNI PREIZKUSI ZNAČILNOSTI

MAJHNI VZORCI velikosti n

Test	Predpostavke	Testna statistika	Porazdelitev	Namen testa
$H_0\,(\mu = \mu_0)$	$X \sim N\,(\mu, \sigma)$, σ znana	$Z = \frac{\bar{X}-\mu_0}{\sigma}\sqrt{n}$	$N\,(0,1)$	populacijsko povprečje
$H_0\,(\mu = \mu_0)$	$X \sim N\,(\mu, \sigma)$, σ neznana	$T = \frac{\bar{X}-\mu_0}{S}\sqrt{n}$	$S\,(n-1)$	populacijsko povprečje
$H_0\,(\sigma = \sigma_0)$	$X \sim N\,(\mu, \sigma)$	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$	$\chi^2\,(n-1)$	populacijski standardni odklon
$H_0\,(\mu = \nu)$	$X \sim N\,(\mu, \sigma)$, $Y \sim N\,(\nu, \tau)$, σ, τ znana	$T = \frac{\bar{X}-\bar{Y}}{S}$	$N\,(0,1)$	enakost populacijskih povprečij
	m velikost vzorca X, n vel. vzorca Y	$S = \sqrt{\frac{\sigma^2}{m} + \frac{\tau^2}{n}}$		
$H_0\,(\mu = \nu)$	$X \sim N\,(\mu, \sigma)$, $Y \sim N\,(\nu, \sigma)$, σ neznana	$T = \frac{\bar{X}-\bar{Y}}{S}\sqrt{\frac{nm}{n+m}}$	$S\,(m+n-2)$	enakost populacijskih povprečij
	m velikost vzorca X, n vel. vzorca Y	$S = \sqrt{\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}}$		
$H_0\,(\sigma = \tau)$	$X \sim N\,(\mu, \sigma)$, $Y \sim N\,(\nu, \tau)$	$F = \frac{S_X^2}{S_Y^2}$	$F\,(m-1, n-1)$	homogenost varianc

VELIKI VZORCI velikosti n

Test	Predpostavke	Testna statistika	Porazdelitev	Namen testa
$H_0\,(E\,(X) = \mu_0)$	X kakorkoli, $\sigma = \sigma\,(X)$ znana	$Z = \frac{\bar{X}-\mu_0}{\sigma}\sqrt{n}$	$\simeq N\,(0,1)$	populacijsko povprečje
$H_0\,(E\,(X) = \mu_0)$	X kakorkoli, $\sigma = \sigma\,(X)$ neznana	$Z = \frac{\bar{X}-\mu_0}{S}\sqrt{n}$	$\simeq N\,(0,1)$	populacijsko povprečje
$H_0\,(\sigma = \sigma_0)$	$X \sim N\,(\mu, \sigma)$	$Z = \frac{S}{\sigma_0}\sqrt{2\,(n-1)} - \sqrt{2n-3}$	$\simeq N\,(0,1)$	populacijski standardni odklon
$H_0\,(p = p_0)$	$\bar{p}$ vzorčni delež, n poskusov	$Z = \frac{\bar{p}-p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$	$\simeq N\,(0,1)$	populacijski delež
$H_0\,(p = q)$	$\bar{p}$ vzorčni delež, m poskusov	$Z = \frac{\bar{p}-\bar{q}}{\sqrt{\hat{p}(1-\hat{p})}}\sqrt{\frac{nm}{n+m}}$	$\simeq N\,(0,1)$	enakost populacijskih deležev
	$\bar{q}$ vzorčni delež, n poskusov	$\hat{p} = \frac{m\bar{p}+n\bar{q}}{m+n}$		
$H_0\,(\mu = \nu)$	$X \sim N\,(\mu, \sigma)$, $Y \sim N\,(\nu, \sigma)$, σ neznana	$Z = \frac{\bar{X}-\bar{Y}}{S}$	$\simeq N\,(0,1)$	enakost populacijskih povprečij
	m velikost vzorca X, n vel. vzorca Y	$S = \sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}$		
$H_0\,(\sigma = \tau)$	$X \sim N\,(\mu, \sigma)$, $Y \sim N\,(\nu, \tau)$	$F = \frac{S_X^2}{S_Y^2}$	$F\,(m-1, n-1)$	homogenost varianc

NEPARAMETRIČNI PREIZKUSI ZNAČILNOSTI

1. **Pearsonov hi kvadrat (tip porazdelitve)**

   Zalogo vrednosti X razdelimo na r razredov S_k, ki jim pripadajo hipotetične verjetnosti $p_k = P\left(X \in S_k | H_0\right)$. Naj bodo n_k vzorčne frekvence razredov S_k in n velikost vzorca. Testna statistika:

$$\chi^2 = \sum_{k=1}^{r} \frac{\left(n_k - np_k\right)^2}{np_k} \approx \chi^2\left(r-1\right).$$

   Če je potrebno predhodno oceniti m parametrov je $\chi^2 \approx \chi^2\left(r-m-1\right)$. Pogoj: vsi $np_k \geq 5$!

2. **Testiranje neodvisnosti s kontingenčno tabelo**

   Vrednosti spremenljivke X razdelimo na r razredov A_i in vrednosti spremenljivke Y razdelimo na s razredov B_j. Naj bodo n_{ij} celične frekvence od $A_i B_j$ in $n_{i.}$ robne frekvence od A_i ter $n_{.j}$ robne frekvence od B_j. Testna statistika:

$$\chi^2 = n\left(\sum_{i,j} \frac{n_{ij}^2}{n_{i.} n_{.j}} - 1\right) \approx \chi^2\left(\left(r-1\right)\left(s-1\right)\right)$$

   Yatesova korektura v primeru $r = s = 2$:

$$\chi_Y^2 = \frac{n\left(\left|n_{11}n_{22} - n_{21}n_{12}\right| - \frac{n}{2}\right)^2}{n_{.1}n_{.2}n_{2.}n_{1.}} \approx \chi^2\left(1\right).$$

3. **Wilcoxonov test predznačnih rangov (enakost porazdelitev pri dveh odvisnih vzorcih)**

   Imamo vzorec velikosti n parnih podatkov (x_1, y_1), (x_2, y_2),...,(x_n, y_n). Rangiramo absolutne razlike $|x_i - y_i|$ in naj bo W^+ vsota rangov pozitivnih razlik in W^- vsota rangov negativnih razlik. Testna statistika:

$$W = \min\{W^+, W^-\},$$

   katere kritične vrednosti so tabelirane (Tabela H). Za $n \geq 30$ lahko uporabimo aproksimacijo:

$$Z = \frac{4W - n(n+1)}{\sqrt{n(n+1)\left(2n+1\right)}}\sqrt{\frac{3}{2}} \approx N(0,1).$$

4. **Mann-Whitneyev test (enakost porazdelitev pri dveh neodvisnih vzorcih)**

   Imamo vzorec $x_1, x_2, ..., x_m$ spremenljivke X in vzorec $y_1, y_2, ..., y_n$ spremenljivke Y. Podatke združimo in rangiramo; naj bo R_X vsota rangov spremenljivke X in R_Y vsota rangov Y. Število inverzij glede na X oziroma število inverzij glede na Y:

   $$U_X = R_X - \frac{m(m-1)}{2}\,, \quad U_Y = R_Y - \frac{n(n-1)}{2}.$$

   Testna statistika:

   $$U = \min\{U_X, U_Y\},$$

   katere kritične vrednosti so tabelirane (Tabela G). Za $m + n \geq 20$ in $n, m \geq 5$ lahko uporabimo aproksimacijo:

   $$Z = \frac{2U - mn}{\sqrt{mn(m+n+1)}}\sqrt{3} \approx N(0,1).$$

5. **Friedmanov test (enakost porazdelitev pri k odvisnih vzorcih)**

   Imamo vzorec podatkov velikosti n za vektor k spremenljivk $(X_1, X_2, ..., X_k)$: $(x_{11}, x_{21}, ..., x_{k1})$, $(x_{12}, x_{22}, ..., x_{k2})$,...,$(x_{1n}, x_{2n}, ..., x_{kn})$. Podatke pri vsaki m-ti "meritvi" $(x_{1m}, x_{2m}, ..., x_{km})$ rangiramo z rangi od 1 do k. Naj bo R_i vsota rangov, ki jo imajo podatki $x_{i1}, x_{i2}, ..., x_{in}$ spremenljivke X_i. Testna statistika:

   $$\chi^2_F = \frac{12}{nk(k+1)}\sum_{i=1}^{k} R_i^2 - 3n(k+1) \approx \chi^2\,(k-1)\,.$$

   Pogoj: $k = 3$ in $n > 13$; $k = 4$ in $n > 8$; $k = 5$ in $n > 5$.

6. **Kruskal-Wallisov test (enakost porazdelitev pri k neodvisnih vzorcih)**

   Imamo k vzorcev podatkov velikosti $n_1 + n_2 + ... + n_k = n$: X_1 - $x_{11}, x_{12}, ..., x_{1n_1}$; X_2 - $x_{21}, x_{22}, ..., x_{2n_2}$;...;$X_k$ - $x_{k1}, x_{k2}, ..., x_{kn_k}$. Vse podatke združimo in rangiramo. Naj bo R_i vsota rangov i-tega vzorca (spremenljivka X_i). Testna statistika:

   $$H = \frac{12}{n(n+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(n+1) \approx \chi^2\,(k-1)\,.$$

   Pogoj: $k \geq 3$ in vsak $n_i \geq 5$.