Formule za 8. vajo pri predmetu Naključni pojavi

Enakost povprečij $r \geq 2$ normalno porazdeljenih populacij s podobno varianco lahko preverimo s testom, ki se imenuje analiza variance. Pri tem testu ničelna hipoteza trdi, da so vsa povprečja enaka, alternativna pa, da se vsaj eno povprečje razlikuje od drugih:

$$\begin{aligned} \mathrm{H}_0 &: m_1 = m_2 = \cdots = m_r\,, \\ \mathrm{H}_1 &: m_i \neq m_j\,, \quad \text{za vsaj en par } (i,j)\,. \end{aligned} \tag{1}$$

Za preverjanje enakosti potrebujemo po en vzorec s po n_i elementi iz vsake populacije, skupaj torej r vzorcev z $N = \sum_i n_i$ elementi:

vzorec 1	x_{11}	x_{12}	$\ldots$	x_{1n_1}
vzorec 2	x_{21}	x_{22}	$\ldots$	x_{2n_2}
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
vzorec r	x_{r1}	x_{r2}	$\ldots$	x_{rn_r}

Test temelji na primerjavi povprečnega odstopanja izmerjene spremenljivke med vzorci s povprečnim odstopanjem spremenljivke znotraj vzorcev. V kolikor so odstopanja med vzorci značilno večja kot odstopanja znotraj vzorcev, H_0 zavrnemo. Uporabimo F-porazdeljeno testno statistiko:

$$F = \frac{S_1^2}{S_2^2}\,, \tag{2}$$

ki ima $(r-1, N-r)$ prostostnih stopenj. S_1^2 predstavlja povprečno kvadratično odstopanje med posameznimi vzorci, medtem ko S_2^2 predstavlja povprečno kvadratično odstopanje znotraj vzorcev. Veličine, potrebne za izračun testne statistike, vpisujemo v tabelo:

Odstopanje	Vsota kvadratičnih odstopanj	Število prostostnih stopenj	Povprečje kvadratičnih odstopanj	Statistika
Med skupinami	Q_1	$r-1$	$S_1^2 = \dfrac{Q_1}{r-1}$	$F = \dfrac{S_1^2}{S_2^2}$
Znotraj skupin	Q_2	$N-r$	$S_2^2 = \dfrac{Q_2}{N-r}$	
Celotno	Q	$N-1$		

Pri izračunu si lahko pomagamo z naslednjimi formulami:

$$Q = \sum_{i=1}^{r}\sum_{j=1}^{n_i} x_{ij}^2 - \frac{x_{\bullet\bullet}^2}{N} \qquad x_{\bullet\bullet} = \sum_{i=1}^{r}\sum_{j=1}^{n_i} x_{ij} \tag{3}$$

$$Q_1 = \sum_{i=1}^{r} \frac{x_{i\bullet}^2}{n_i} - \frac{x_{\bullet\bullet}^2}{N} \qquad x_{i\bullet} = \sum_{j=1}^{n_i} x_{ij} \tag{4}$$

$$Q_2 = Q - Q_1 \qquad N = \sum_{i=1}^{r} n_i \tag{5}$$

Linearnost odvisnosti med dvema spremenljivkama X in Y lahko s korelacijskim koeficientom:

$$r_{XY} = \frac{\mathrm{Cov}\left[X, Y\right]}{\sqrt{\mathrm{Var}\left[X\right] \mathrm{Var}\left[Y\right]}} = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left(n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2 \right) \left(n \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i \right)^2 \right)}}\,, \tag{6}$$

katerega zaloga vrednosti je $[-1, 1]$. Vrednosti $|r_{XY}| \approx 1$ kažejo na izrazito linearno odvisnost med X in Y, medtem ko vrednosti $|r_{XY}| < 0.5$ kažejo, da spremenljivki nista linearno odvisni. V slednjem primeru sta spremenljivki lahko neodvisni ali pa je njuna odvisnost nelinearna.
Matematični model linearne odvisnosti je:

$$Y = aX + b\,, \tag{7}$$

graf te enačbe pa je premica. Koeficienta a in b izračunamo:

$$\begin{aligned} a &= \frac{\mathrm{Cov}\left[X, Y\right]}{\mathrm{Var}\left[X\right]} = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2}\,, \\ b &= \mathrm{E}\left[Y\right] - a\mathrm{E}\left[X\right] = \frac{1}{n} \sum_{i=1}^{n} y_i - \frac{a}{n} \sum_{i=1}^{n} x_i\,. \end{aligned} \tag{8}$$

S tema koeficientoma določena premica se najbolje prilega izmerjenim točkam (x_i, y_i), saj je povprečno kvadratično odstopanje izmerjenih točk od premice minimalno.
Linearni model lahko uporabimo tudi, kadar zveza med spremenljivkama ni linearna. V takih primerih si lahko pomagamo z logaritmiranjem zveze, v kolikor jo s tem lineariziramo, ali pa uvedemo nove spremenljivke, kot npr.:

$$Y = aX^2 + b = aZ + b\,, \quad \text{kjer je } X^2 = Z\,. \tag{9}$$