Uporabna statistika

Gregor Dolinar

Fakulteta za elektrotehniko
Univerza v Ljubljani

4. december 2013

Diskretna porazdelitev

Primer

Velikost vzorca $n = 100$.

n_k število elementov vzorca s k napakami

Velja (k, n_k):

$(0, 24), (1, 30), (2, 31), (3, 11), (4, 2), (5, 1), (6, 1)$.

H_0 : slučajna spremenljivka X je Poissonovo porazdeljena

$\alpha = 0.05$

$$\lambda = E(X) = \frac{\sum_{i=0}^{6} k \cdot n_i}{100} =$$

$$p_i = P(X = i) = \frac{e^{-\lambda}\lambda^i}{i!}, \quad E_i = np_i$$

$$X_0^2 = \sum_{i=0}^{6} \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2_{0.05,5} = 11.07, \quad X_0^2 \overset{?}{<} \chi^2_{0.05,5}$$

Zvezna porazdelitev

Primer

Dolžine intervalov običajno določimo tako, da so pričakovane frekvence $E_i = np_i$ za vsak interval enake (ploščina pod krivuljo je na vsakem intervalu enaka).
Če preverjamo, če je porazdelitev normalna, potem si pri določanju mej lahko pomagam tako, da normalno porazdelitev standardiziramo $\frac{x-\overline{x}}{s}$.

Kontingenčna tabela

Podatke razdelimo v podskupine glede na 2 različna kriterija.
Na primer, študente razdelimo v skupine glede na oceno pri predmetu Statistika in pri predmetu Matematika.
Vsak podatek pripada 2 skupinama glede na 2 kriterija, torej jih lahko razporedimo v dvodimenzionalno tabelo.
Zanima nas, ali sta kriterija med sabo neodvisna.

Podatke razporedimo v dvodimenzionalno tabelo.

	1	2	...	c
1	O_{11}	O_{12}	...	O_{1c}
2	O_{21}	O_{22}	...	O_{2c}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r	O_{r1}	O_{r2}	...	O_{rc}

Hipoteza H_0 : razporeditev elementov po vrsticah tabele je neodvisna od razporeditve po stolpcih.

Vsota elementov i-te vrstice, deljena z n (število elementov vzorca):

$$\hat{u}_i = \frac{1}{n} \sum_{j=1}^{c} O_{ij}.$$

Dobimo verjetnost, da je element vzorca v i-ti vrstici (neodvisno od tega v katerem stolpcu je).

Vsota elementov j-tega stolpca, deljena z n:

$$\hat{v}_j = \frac{1}{n} \sum_{i=1}^{r} O_{ij}.$$

Dobimo verjetnost, da je element vzorca v r-tem stolpcu (neodvisno od tega v kateri vrstici je).

Če je razporeditev n elementov vzorca po stolpcih neodvisna od razporeditve po vrsticah, je število elementov v i-ti vrstici in j-tem stolpcu enako

$$E_{ij} = n\hat{u}_i\hat{v}_j.$$

Za velike n je

$$X_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

približno χ^2 porazdeljena slučajna spremenljivka z $(r-1)(c-1)$ prostostnimi stopnjami.
Hipotezo o neodvisnosti zavrnemo, če je izračunana vrednost testne statistike večja od $\chi^2_{\alpha,(r-1)(c-1)}$.

Primer

	10	9	8	$o < 8$
10	25	6	17	13
9	17	16	15	6
8	18	4	18	10
$o < 8$	10	8	11	20