Formule za 7. vajo pri predmetu Naključni pojavi

V tej vaji obravnavamo testiranje neparametričnih hipotez in sicer pri treh tipih testov.
Pri prvem testu, ki se imenuje prilagoditveni, se hipotezi nanašata na obliko verjetnostne porazdelitve. Ničelna hipoteza npr. trdi, da je obravnavana naključna spremenljivka porazdeljena normalno, alternativna pa, da to ne drži:

$$\begin{aligned} \mathrm{H}_0 &: f(x) = \mathcal{N}(m, \sigma)\,, \\ \mathrm{H}_1 &: f(x) \neq \mathcal{N}(m, \sigma)\,. \end{aligned} \tag{1}$$

H_0 preverimo tako, da razpon naključne spremenljivke X razdelimo na r intervalov oziroma razredov in nato za vsak razred na podlagi vzorca n vrednosti ocenimo verjetnost p_i, da naključna spremenljivka zavzame vrednost iz tega razreda. Ocenjene verjetnosti nato primerjamo z verjetnostmi p_{i_0}, ki jih za razrede izračunamo na podlagi v H_0 predpostavljene porazdelitve. Primerjavo izvedemo s statistiko:

$$\chi^2 = n \sum_{i=1}^{r} \frac{(p_i - p_{i_0})^2}{p_{i_0}} = \sum_{i=1}^{r} \frac{(n_i - n_{i_0})^2}{n_{i_0}}\,, \tag{2}$$

v kateri sta n_i in n_{i_0} izmerjena in predpostavljena pogostost za i-ti razred. Statistika je približno χ^2_{r-l-1} porazdeljena, če je n velik in $n_i \geq 5$ za vsak razred. l v številu prostostnih stopenj statistike pomeni število parametrov predpostavljene porazdelitve, ki smo jih morali oceniti iz vzorca, da smo lahko izračunali predpostavljene verjetnosti p_{i_0} oziroma pogostosti n_{i_0}. Pri normalni porazdelitvi je $l = 2$, pri eksponentni $l = 1$ in pri enakomerni $l = 0$. V kolikor vrednost testne statistike presega kritično $\chi^2_{r-l-1;\alpha}$, H_0 zavrnemo. Za večjo preglednost računa, vmesne rezultate vpisujemo v tabelo, v kateri sta prva dva stolpca običajno podana z vzorcem, preostale tri pa moramo izračunati:

x_i	n_i	p_{i_0}	$n_{i_0} = p_{i_0} \cdot n$	$(n_i - n_{i_0})^2 / n_{i_0}$
x_1	n_1	p_{1_0}	n_{1_0}	$(n_1 - n_{1_0})^2 / n_{1_0}$
x_2	n_2	p_{2_0}	n_{2_0}	$(n_2 - n_{2_0})^2 / n_{2_0}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_r	n_r	p_{r_0}	n_{r_0}	$(n_r - n_{r_0})^2 / n_{r_0}$
$\sum_i^r$	n	1.0	n	χ^2

Drugi tip testa se imenuje test neodvisnosti. Hipotezi se nanašata na (ne)odvisnost dveh naključnih spremenljivk oz. vplivov X in Y, katerih vrednosti lahko razdelimo na r oziroma c razredov. Na vzorcu n vrednosti za vsak par razredov (x_i, y_j) določimo pogostosti n_{ij} in jih vpišemo v kontingenčno tabelo:

		Y				
		y_1	y_2	$\cdots$	y_c	$n_{i\star}$
X	x_1	n_{11}	n_{12}	$\cdots$	n_{1c}	$n_{1\star}$
	x_2	n_{21}	n_{22}	$\cdots$	n_{2c}	$n_{2\star}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	x_r	n_{r1}	n_{r2}	$\cdots$	n_{rc}	$n_{r\star}$
	$n_{\star j}$	$n_{\star 1}$	$n_{\star 2}$	$\cdots$	$n_{\star c}$	$n_{\star\star} = n$

V tabeli znak $\star$ na mestu indeksa pomeni vsoto po tem indeksu:

$$n_{i\star} = \sum_{j=1}^{c} n_{ij} \qquad \text{oziroma} \qquad n_{\star j} = \sum_{i=1}^{r} n_{ij}\,. \tag{3}$$

Ničelna hipoteza pri tem testu trdi, da sta X in Y neodvisni, alternativna hipoteza pa, da nista:

$$\begin{aligned} &\mathrm{H}_0 : p_{ij} = p_i \cdot p_j \,, && \text{za vsak par } (i,j) \\ &\mathrm{H}_1 : p_{ij} \neq p_i \cdot p_j \,, && \text{za vsaj en par } (i,j)\,. \end{aligned} \tag{4}$$

Ničelno hipotezo preverimo s testno statistiko:

$$\chi^2 = n \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(p_{ij} - p_{ij_0})^2}{p_{ij_0}} = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - n_{i\star} \cdot n_{\star j}/n)^2}{n_{i\star} \cdot n_{\star j}/n}\,, \tag{5}$$

ki je $\chi^2_{(r-1)(c-1)}$ porazdeljena. V kolikor je vrednost testne statistike večja od kritične $\chi^2_{(r-1)(c-1);\alpha}$, H_0 zavrnemo. V testni statistiki uporabimo naslednjo cenilko za povezano verjetnost p_{ij_0}:

$$p_{ij_0} = p_{i_0} \cdot p_{j_0} = \frac{n_{i\star} \cdot n_{\star j}}{n^2}\,. \tag{6}$$

Tretji tip testa se imenuje test homogenosti. Hipotezi se nanašata na (ne)homogenost r skupin glede na predpisani kriterij, ki ima c možnih vrednosti. Za vsako od r skupin imamo vzorec z n_i vrednostmi, ki jih razdelimo glede na kriterij v c razredov. Tako določimo pogostosti n_{ij}, ki jih vpišemo v kontingenčno tabelo:

		Kriterij				
		y_1	y_2	$\cdots$	y_c	$n_{i\star} = n_i$
Skupine	x_1	n_{11}	n_{12}	$\cdots$	n_{1c}	$n_{1\star} = n_1$
	x_2	n_{21}	n_{22}	$\cdots$	n_{2c}	$n_{2\star} = n_2$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	x_r	n_{r1}	n_{r2}	$\cdots$	n_{rc}	$n_{r\star} = n_r$
	$n_{\star j}$	$n_{\star 1}$	$n_{\star 2}$	$\cdots$	$n_{\star c}$	$n_{\star\star} = n$

Tabela je podobna kot pri testu neodvisnosti, le da so v tem primeru pogostosti n_i določene že pred testom in niso odvisne od razvrščanja. Ničelna hipoteza pri tem testu trdi, da so skupine homogene glede na kriterij, alternativna hipoteza pa, da niso:

$$\begin{aligned} \mathrm{H}_0 : p_{1j} = p_{2j} = \cdots = p_{rj} = p_{\star j}\,, &\qquad \text{za vsak } j \\ \mathrm{H}_1 : p_{ij} \neq p_{kj}\,, &\qquad \text{za vsaj eno trojico } (i,j,k)\,. \end{aligned} \tag{7}$$

Ničelno hipotezo preverimo s testno statistiko:

$$\chi^2 = n \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(p_{ij} - p_{ij_0})^2}{p_{ij_0}} = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - n_{i\star} \cdot n_{\star j}/n)^2}{n_{i\star} \cdot n_{\star j}/n}\,, \tag{8}$$

ki je $\chi^2_{(r-1)(c-1)}$ porazdeljena. V kolikor je vrednost testne statistike večja od kritične $\chi^2_{(r-1)(c-1);\alpha}$, H_0 zavrnemo. V tem primeru za oceno predpostavljenih verjetnosti $p_{\star j_0}$ in p_{ij_0} uporabimo cenilki:

$$p_{\star j_0} = \frac{n_{\star j}}{n}\,, \qquad p_{ij_0} = \frac{n_i \cdot p_{\star j_0}}{n} = \frac{n_{i\star} \cdot n_{\star j}}{n^2}\,. \tag{9}$$

Cenilka za p_{ij_0} je enaka kot pri testu homogenosti, zato je tudi testna statistika pri obeh testih enaka.