Preverjanje neparametričnih hipotez

Prilagoditveni test: hipotezi se nanašata na *tip porazdelitvene funkcije*. Ničelna hipoteza trdi, da je obravnavana naključna spremenljivka X porazdeljena z verjetnostno porazdelitvijo $f_0(x)$, alternativna pa, da to ne drži:

$$\begin{aligned} \mathrm{H}_0 &: f(x) = f_0(x)\,, \\ \mathrm{H}_1 &: f(x) \neq f_0(x)\,. \end{aligned} \tag{1}$$

H_0 preverimo tako, da razpon naključne spremenljivke X razdelimo na r intervalov oziroma razredov s širinami Δx_i. Za vzorec z n elementi določimo število meritev (frekvenco) n_i, ki pripada posameznem razredu. Za vsak razred lahko ocenimo verjetnost $p_i = n_i/n$, da X zavzame vrednost, ki pripada temu razredu. Ocenjene frekvence n_i nato primerjamo s frekvencami $n_{i_0} = p_{i_0} n$, ki jih za razrede izračunamo na podlagi v H_0 predpostavljene porazdelitve. Primerjavo izvedemo s testno statistiko, ki je utežena vsota kvadrata relativne razlike frekvenc:

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

Porazdelitev testne statistike χ^2 se asimptotično bliža porazdelitvi χ^2_{r-l-1}, kjer je l število parametrov predpostavljene porazdelitve, ki smo jih morali oceniti iz vzorca, da smo lahko izračunali frekvence n_{i_0}. Pri normalni porazdelitvi je $l = 2$, pri eksponentni in Poissonovi $l = 1$ in pri enakomerni $l = 0$. Če je n velik in če je $n_i \geq 5$ za vsak razred, je porazdelitev statistike χ^2 zelo podobna porazdelitvi χ^2_{r-l-1}. Če vrednost testne statistike presega kritično vrednost $\chi^2_{r-l-1;\alpha}$, H_0 zavrnemo. Za večjo preglednost računa navadno naredimo tabelo s stolpci x_i, n_i, n_i^2, $n_{i_0} = p_{i_0} n$ in n_i^2/n_{i_0}. Vrednost testne statistike je enaka vsoti zadnjega stolpca.

Test neodvisnosti: hipotezi se nanašata na *(ne)odvisnost dveh naključnih spremenljivk* oziroma 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 frekvence 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:

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

Pričakovano povezano verjetnost p_{ij_0} na podlagi H_0 izračunamo s produktom robnih relativnih frekvenc: $p_{ij_0} = p_{i_0} p_{j_0} = n_{i\star} n_{\star j}/n^2$. Ničelno hipotezo preverimo s testno statistiko:

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

Testna statistika χ^2 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.

Test homogenosti: hipotezi se nanašata na *(ne)homogenost v skupin glede na predpisani kriterij*, ki ima r možnih vrednosti. Za vsako skupino imamo vzorec z n_i vrednostmi, ki so glede na kriterij razdeljene v r razredov. Tako določimo frekvence n_{ij}, ki jih vpišemo v kontingenčno tabelo:

		Kriterij				
		y_1	y_2	$\cdots$	y_r	$n_{i\star} = n_i$
Skupine	x_1	n_{11}	n_{12}	$\cdots$	n_{1r}	$n_{1\star} = n_1$
	x_2	n_{21}	n_{22}	$\cdots$	n_{2r}	$n_{2\star} = n_2$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	x_v	n_{v1}	n_{v2}	$\cdots$	n_{vr}	$n_{r\star} = n_v$
	$n_{\star j}$	$n_{\star 1}$	$n_{\star 2}$	$\cdots$	$n_{\star r}$	$n_{\star\star} = n$

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

$$\begin{aligned} \mathrm{H}_0 : p_{1j} = p_{2j} = \cdots = p_{vj} = p_{\star j} \,, &\qquad \text{za vsak } j \\ \mathrm{H}_1 : p_{ij} \neq p_{kj} \,, &\qquad \text{za vsaj eno trojico } (i, j, k)\,. \end{aligned} \tag{6}$$

Pričakovano verjetnost za vsak razred kriterija ocenimo s $p_{\star j_0} = n_{\star j}/n$. Pričakovane vrednosti p_{ij_0} so torej $p_{ij_0} = n_i p_{\star j_0}/n = n_{i\star} n_{\star j}/n^2$. Cenilka za p_{ij_0} je enaka kot pri testu homogenosti, zato je tudi testna statistika enaka:

$$\chi^2 = n \sum_{i=1}^{v} \sum_{j=1}^{r} \frac{(p_{ij} - p_{ij_0})^2}{p_{ij_0}} = \sum_{i=1}^{v} \sum_{j=1}^{r} \frac{(n_{ij} - n_{i\star} \cdot n_{\star j}/n)^2}{n_{i\star} \cdot n_{\star j}/n} = n \left(\sum_{i=1}^{v} \sum_{j=1}^{r} \frac{n_{ij}^2}{n_{i\star} n_{\star j}} - 1 \right) , \tag{7}$$

ki je $\chi^2_{(v-1)(r-1)}$ porazdeljena. V kolikor je vrednost testne statistike večja od kritične $\chi^2_{(v-1)(r-1);\alpha}$, H_0 zavrnemo.