1 Empirikus közép (átlag)

Legyen adott $\xi_1, \xi_2, ..., \xi_n$ minta. A mintákat felhasználva a következőképpen számolható ki az empirikus közép:

$$\overline{\xi} = \frac{\xi_1 + \xi_2 + ... + \xi_n}{n} \tag{1}$$

2 Empirikus és korrigált empirikus szórásnégyzet

Legyen adott $\xi_1, \xi_2, ..., \xi_n$ minta. A mintákat felhasználva a következőképpen számolható ki az empirikus szórásnégyzet:

$$s_n^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\xi_i - \overline{\xi}\right)^2 \tag{2}$$

valamint a korrigált empirikus szórásnégyzet:

$$s_n^{2*} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(\xi_i - \overline{\xi}\right)^2 \tag{3}$$

3 Illeszkedésvizsgálat

Legyen:

- $\xi_1, \xi_2, ..., \xi_n$ események egy tetszőleges A eseménytéren
- k_i: ξ_i esemény bekövetkezéseinek száma,
- N: $k_1 + k_2 + ... + k_n$; összes bekövetkezés száma,
- p_i: ξ_i esemény valószínűsége,
- n: események száma,

ekkor az illeszkedés vizsgálat az alábbi módon hajtható végre:

$$\chi^2 = \sum_{i=1}^{n} \frac{(k_i - N \cdot p_i)^2}{N \cdot p_i} \tag{4}$$

Ha kiszámoltuk χ^2 értékét, össze kell azt vetnünk a χ^2 táblázattal. A pontosság általában 95%, ebben az esetben a 0.05-ös oszlopból kell kinézni a megoldást $(1 - 0.95 = 0.05)$. A feladatban a szabadsági fok $n - 1$, ebben a sorban kell vizsgálni a táblázatot. Ha a számított χ^2 kisebb mint a táblázatban található, akkor a becslést elfogadjuk.

Pro tipp: R-ben a χ^2 táblázatos ellenőrzés végrehajtható a `qchisq(p=.95, df=n-1)` paranccsal, ami visszaadja a számot amihez egyeztetni kell a számolt χ^2 értékünket. Értelemszerűen a `p` paraméter a pontossági fokot, míg a df paraméter a szabadsági fokot adja meg.

4 Homogenitásvizsgálat

Legyen:

- $\xi_1, \xi_2, ..., \xi_r$ és $\eta_1, \eta_2, ..., \eta_r$ események (Figyelem, r darab van mind a két mintasorban),
- ξ_i egy gyakoriság a ξ mintában (pl $\xi_3 = 10$),
- η_i egy gyakoriság az η mintában (pl $\eta_5 = 3$),
- $n = \xi_1 + \xi_2 + ... + \xi_r$,
- $m = \eta_1 + \eta_2 + ... + \eta_r$.

ekkor a homogenitás vizsgálat a következőképpen írható le:

$$\chi^2 = \sum_{i=1}^{r} \frac{\left(\frac{\xi_i}{n} - \frac{\eta_i}{m}\right)^2}{\xi_i + \eta_i} \cdot n \cdot m \tag{5}$$

Ha kiszámoltuk χ^2 értékét, össze kell azt vetnünk a χ^2 táblázattal. A pontosság általában 95%, ebben az esetben a 0.05-ös oszlopból kell kinézni a megoldást $(1 - 0.95 = 0.05)$. A feladatban a szabadsági fok $r - 1$, ebben a sorban kell vizsgálni a táblázatot. Ha a számított χ^2 kisebb mint a táblázatban található, akkor a becslést elfogadjuk.

5 Függetlenségvizsgálat

Legyen:

- $A_1, A_2, ..., A_r$ szempontok,
- $B_1, B_2, ..., B_s$ szempontok,
- $K_{1,1}, K_{1,2}, ..., K_{r,s}$ megfigyelések számai, amikor A és B teljesül egyszerre (pl $K_{3,2} = A_3$ és B_2 is teljesül),
- $N_{i,.} = K_{i,1} + K_{i,2} + ... + K_{i,s}$, az A gyakorisága,
- $N_{.,j} = K_{1,j} + K_{2,j} + ... + K_{r,j}$, a B gyakorisága,
- $n = N_{1,.} + N_{2,.} + ... + N_{r,.} = N_{.,1} + N_{.,2} + ... + N_{.,s}$, peremértékek összege az oszlopokban és a sorokban megegyezik.

Szemléltetés:

	B_1	B_2	B_3	$N_{i,.}$
A_1	$K_{1,1}$	$K_{1,2}$	$K_{1,3}$	$N_{1,.}$
A_2	$K_{2,1}$	$K_{2,2}$	$K_{2,3}$	$N_{2,.}$
$N_{.,j}$	$N_{.,1}$	$N_{.,2}$	$N_{.,3}$	n

A jelölések bevezetése után a következő képletet használjuk:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(N_{i,j} - \frac{N_{i,.} \cdot N_{.,j}}{n}\right)^2}{\frac{N_{i,.} \cdot N_{.,j}}{n}} \tag{6}$$

Ha kiszámoltuk χ^2 értékét, össze kell azt vetnünk a χ^2 táblázattal. A pontosság általában 95%, ebben az esetben a 0.05-ös oszlopból kell kinézni a megoldást $(1 - 0.95 = 0.05)$. A feladatban a szabadsági fok $(r - 1) \cdot (s - 1)$, ebben a sorban kell vizsgálni a táblázatot. Ha a számított χ^2 kisebb mint a táblázatban található, akkor a becslést elfogadjuk.

6 Lineáris regresszió

Legyenek adottak az $X = x_1, x_2, ..., x_n$ és a hozzá tartozó $Y = y_1, y_2, ..., y_n$ koordináták. Az ezen pontokat közelítő egyenes egyenlete $f(x) = \hat{a} + \hat{b}x$, ahol

$$\hat{b} = \frac{\sum_{i=1}^{n} \left(x_i - \overline{X}\right) \left(y_i - \overline{Y}\right)}{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2} \tag{7}$$

$$\hat{a} = \overline{Y} - \hat{a} \cdot \overline{X} \tag{8}$$

7 Exponenciális regresszió

Az exponenciális regresszió akkor használatos, ha a minta jobban hasonlít egy exponenciális függvény görbéjére mint egy egyenesre. Hasonlóan a lineáris regresszióhoz, itt is egy $X = x_1, x_2, ..., x_n$ és $Y = y_1, y_2, ..., y_n$ mintapárral dolgozunk.

Maga az exponenciális görbe a következő képlettel írható le:

$$f(x) = a \cdot e^{bx} \tag{9}$$

Hogy megkapjuk a képletben szereplő a és b paramétereket, a mintát először vissza kell vezetnünk lineáris regresszióra. Ennek a képlete a következő:

$$f(x) = a_0 + a_1 x \tag{10}$$

Az így előállított egyenes segítségével fogjuk megmondani az a és b paramétereket:

$$a = e^{a_0}, \qquad b = a_1 \tag{11}$$

A regressziós egyenes megszerkesztéséhez ki kell számolnunk az a_0 és a_1 paramétereket. A számolás valójában itt kezdődik, az eredményeket felhasználva eljutunk a 11-es, majd végül a 9-es képlethez. A számolás előtt vezessük be a következő jelölést:

- $Z = ln(Y)$, exponenciális alapú, logaritmikusan transzformált Y koordináták

A számolás elkezdhető:

$$a_1 = \frac{n \cdot \sum z_i x_i - \sum z_i \sum x_i}{n \cdot \sum x_i^2 - (\sum x_i)^2} \tag{12}$$

$$a_0 = \overline{Z} - a_i \overline{X} \tag{13}$$

Az így kapott a_0 és a_1 paraméterek behelyettesíthetőek a 11-es képletbe. Az így megkapott a és b paraméterek behelyettesíthetőek a 9-es képletbe, ezzel megkapva a közelítő görbét.