Matematikai statisztika képletgyüjtemeny

Legyen $X_1, X_2, \ldots, X_n$ egy n elemű minta.

Alapfogalmak

1. A minta átlaga: $\overline{X} = \frac{X_1+X_2+\ldots X_n}{n}$
2. A minta tapasztalati szórása $S = \sqrt{\frac{(X_1-\overline{X})^2+(X-X_2)^2+\cdots+(X_n-\overline{X})^2}{n}}$
3. A minta korrigált tapasztalati szórása $S^* = \sqrt{\frac{(X_1-\overline{X})^2+(X-X_2)^2+\cdots+(X_n-\overline{X})^2}{n-1}}$

Konfidenciaintervallum szerkesztése normális eloszlás várható értékére

1. Szórás ismert: $(\overline{X} - \frac{u_{\frac{\epsilon}{2}}\sigma_0}{\sqrt{n}}, \overline{X} + \frac{u_{\frac{\epsilon}{2}}\sigma_0}{\sqrt{n}})$
2. Szórás nem ismert: $\left(\overline{X} - \frac{t_{\epsilon/2}(n-1)S_n^*}{\sqrt{n}}, \overline{X} + \frac{t_{\epsilon/2}(n-1)S_n^*}{\sqrt{n}}\right)$

Próbák normális eloszlás váháto értékére

1. U-próba: $\Phi(u_{\epsilon/2}) = 1 - \frac{\epsilon}{2} \quad \Phi(u_\epsilon) = 1 - \epsilon$
   (a) Egymintás: $H_0 : \mu = \mu_0$: póbastatisztika: $u(\underline{X}) = \frac{\overline{x}-\mu_0}{\sigma_0}\sqrt{n}$,

H_1	konfidenciaintervallum
$\mu \neq \mu_0$	$(-u_{\epsilon/2}, u_{\epsilon/2})$
$\mu > \mu_0$	$(-\infty, u_\epsilon)$
$\mu < \mu_0$	$(-u_\epsilon, \infty)$

   (b) Kétmintás: $H_0 : \mu_1 = \mu_2$: próbastatisztika: $u(\underline{X}, \underline{Y}) = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$

H_1	konfidenciaintervallum
$\mu_1 \neq \mu_2$	$(-u_{\epsilon/2}, u_{\epsilon/2})$
$\mu_1 > \mu_2$	$(-\infty, u_\epsilon)$
$\mu_1 < \mu_2$	$(-u_\epsilon, \infty)$

2. t-próba:
   (a) Egymintás: $H_0 : \mu = \mu_0$: póbastatisztika: $u(\underline{X}) = \frac{\overline{x}-\mu_0}{S^*}\sqrt{n}$,

H_1	konfidenciaintervallum
$\mu \neq \mu_0$	$(-t_{\epsilon/2}(n-1), t_{\epsilon/2}(n-1))$
$\mu > \mu_0$	$(-\infty, t_\epsilon(n-1))$
$\mu < \mu_0$	$(-t_\epsilon(n-1), \infty)$

(b) : Kétmintás: $H_0 : \mu_1 = \mu_2$: próbastatisztika: $t(\underline{X}, \underline{Y}) = \frac{\overline{X}-\overline{Y}}{\sqrt{(n_1-1)(S_1^*)^2+(n_2-1)(S_2^*)^2}}\sqrt{\frac{n_1 n_2(n_1+n_2-2)}{n_1+n_2}}$,

H_1	konfidenciaintervallum
$\mu_1 \neq \mu_2$	$(-t_{\epsilon/2}(n_1+n_2-2), t_{\epsilon/2}(n_1+n_2-2))$
$\mu_1 > \mu_2$	$(-t_\epsilon(n_1+n_2-2), \infty)$
$\mu_1 < \mu_2$	$(-\infty, t_\epsilon(n_1+n_2-2))$

3. Illeszkedésvizsgálat χ^2-próbával:

   (a) Diszkrét eset: $H_0 : P(X = x_i) = p_i, \quad i = 1, 2, \ldots, r$,
   Próbastatisztika: $\chi^2 = \sum_{i=1}^r \frac{(v_i - np_i)^2}{np_i}$,
   Elutasítás: $\chi^2 \geq \chi^2_\epsilon(r-1)$.

   (b) Folytonos eset: $H_0 : P(X < x) = F(x), \quad x \in \mathbb{R}$
   Osszuk fel a számegyenest

$$-\infty = x_0 < x_1 < \cdots < x_{r-1} < x_r = \infty$$

   osztópontokkal r részre. Jelölje v_i azon kisérletek számát, amikor a kisérlet eredménye az $[x_{i-1}, x_i[$ intervallumba esik.
   Próbastatisztika: $\chi^2 = \sum_{i=1}^r \frac{(v_i - np_i)^2}{np_i}$,
   Elutasítáa: $\chi^2 \geq \chi^2_\epsilon(r-1)$.

   Ha menet közben b darab paramétert becslünk, akkor $\chi^2_\epsilon(r-1-b)$-vel kell számolni!

4. Homogenitaásvizsgálat: Legyen $X_1, X_2, \ldots, X_n$ az X háttérváltozóból származó és $Y_1, Y_2, \ldots, Y_m$ az Y háttérváltozóból származó minta. $H_0 : P(X < x) = P(Y < x)$, minden $x \in \mathbb{R}$

   (a) Diszkrét eset: Legyen $v_i = |\{k : X_k = x_i\}$, $u_i = |\{k : Y_k = x_i\}$.
   Próbastatisztika: $\chi^2 = nm \sum_{i=1}^r \frac{(\frac{v_i}{n} - \frac{u_i}{m})^2}{u_i + v_i}$
   Elutasítás: $\chi^2 \geq \chi^2_\epsilon(r-1)$

   (b) Folytonos eset: Osszuk fel a számegyenest

$$-\infty = x_0 < x_1 < \cdots < x_{r-1} < x_r = \infty$$

   osztópontokkal r részre. Legyen $v_i = |\{k : X_k \in [[x_{i-1}, x_i\}|$ és $u_i = |\{k : y_k \in [[x_{i-1}, x_i\}|$
   Próbastatisztika: $\chi^2 = nm \sum_{i=1}^r \frac{(\frac{v_i}{n} - \frac{u_i}{m})^2}{u_i + v_i}$
   Elutasítás: $\chi^2 \geq \chi^2_\epsilon(r-1)$

5. Becsléses függetlenségvizsgálat: $H_0 : P(A_i B_j) = P(A_i)P(B_j), \quad$ minden $i = 1, \ldots, r, j = 1, \ldots, s$

   (a) Diszkrét eset: Ha X és Y diszkrét eloszlásúak, akkor jelölje $x_1, \ldots, x_r$ illetve $y_1, \ldots, y_s$ az értékkészletüket és legyen $v_{i.} = |\{k : X_k = x_i\}|, \quad i = 1, \ldots, r$, $v_{.j} = |\{k : Y_k = y_j\}|, \quad j = 1, \ldots, s$, $v_{ij} = |\{k : X_k = x_i, Y_k = y_j\}|, \quad i = 1, \ldots, r, j = 1, \ldots s$.
   Próbastatisztika: $\chi^2 = n \sum_{i=1}^r \sum_{j=1}^s \frac{(v_{ij} - \frac{v_{i.} v_{.j}}{n})^2}{v_{i.} v_{.j}}$,
   Elutasítás: $\chi^2 \geq \chi^2_\epsilon((r-1)(s-1))$

   (b) Folytonos eset: Vesszük a szokásos felosztását a valós számegyenesnek és legyen $v_{i.} = |\{k : X_k \in [x_{i-1}, x_i[\}|, \quad i = 1, \ldots, r$ $v_{.j} = |\{k : Y_k \in [y_{j-1}, y_j[\}|, \quad j = 1, \ldots, s$ A többi megegyezik a diszkrét esettel.

Regrssziós egyenes

Az (x_1, y_1), $(x_2, y_2), \ldots, (x_n, y_n)$ mintához írjuk fel az $y = ax + b$ regressziós egyenes paraméterei:

$$a = \frac{n \sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$b = \frac{(\sum_{i=1}^{n} x_i^2)(\sum_{i=1}^{n} y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} x_i y_i)}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$