Statisztika elméleti összefoglaló

1. Tartalomjegyzék

2. Becsléselmélet

Alapfogalmak, jelölések

A sokaság elemszáma: N

A sokaság paramétereinek általános jele: Θ

A sokaság legfontosabb paraméterei:

- várható érték: μ
- szórás: σ
- arány: P

A minta elemszáma: n

A mintából számított paraméterek általános jele: $\widehat{\Theta}$.

A mintából számított legfontosabb paraméterek:

- várható érték: $\bar{y}$
- szórás: s
- arány: p

A becslések tulajdonságai

Torzítatlanság: a mintából számított becslések várható értéke megegyezik a sokasági paraméterrel, azaz

$$E(\hat{\theta}) = \theta$$

A torzítás mértéke: $Bs(\hat{\theta}) = \theta - E(\hat{\theta})$

A becslés varianciája: $Var(\hat{\theta})$

Becslések összehasonlítása

Két torzítatlan becslőfüggvény esetén a hatásosabbat kell választani.

Két torzítatlan becslőfüggvény közül azt tekintjük hatásosabbnak, amelyiknek kisebb a varianciája.

Egy torzított és egy torzítatlan becslőfüggvény közül az átlagos négyzetes hiba (MSE) alapján kell választani. Az a jobb becslés, amelynek kisebb az átlagos négyzetes hibája.

$$MSE(\hat{\theta}) = Bs^2(\hat{\theta}) + Var(\hat{\theta})$$

3. Intervallumbecslések

Mintavételi módok

- FAE: független azonos eloszlású (visszatevéses)
- EV: egyszerű véletlen (visszatevés nélküli)
- R: rétegzett
- TL: többlépcsős
- CS: csoportos

Általános képlet

$$Int_{1-\alpha}(\theta) = \hat{\theta} \pm \varphi_{1-\frac{\alpha}{2}} s_{\hat{\theta}}$$

A fenti képlet a konfidencia-intervallumok általános képlete.

A jelölések tartalma:

$1 - \alpha$: megbízhatósági szint

Θ: a sokaság paramétere

$\hat{\theta}$: a paraméter mintából becsült értéke (pontbecslés)

φ: valószínűségi változó

$s_{\hat{\theta}}$: a paraméterbecslés standard hibája

$\varphi_{1-\frac{\alpha}{2}} s_{\hat{\theta}} = \Delta_{\hat{\theta}}$: a konfidencia intervallum sugara, a becslés abszolút hibahatára

Konfidencia intervallumok konkrét esetekben

Sokasági átlag (várható érték, μ)

	kis minta ($n \leq 30$) ismert sokasági szórás (σ)	kis minta ($n \leq 30$) ismeretlen sokasági szórás (σ)	nagy minta (n>30)
FAE minta	$\bar{y} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	$\bar{y} \pm t_{1-\frac{\alpha}{2}}(v) \frac{s}{\sqrt{n}}$ szabadságfok: v=n-1	$\bar{y} \pm z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$
EV minta	$\bar{y} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \sqrt{1-\frac{n}{N}}$	$\bar{y} \pm t_{1-\frac{\alpha}{2}}(v) \frac{s}{\sqrt{n}} \sqrt{1-\frac{n}{N}}$ szabadságfok: v=n-1	$\bar{y} \pm z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \sqrt{1-\frac{n}{N}}$
R minta	-	-	$\overline{y_R} \pm z_{1-\frac{\alpha}{2}} \sqrt{\sum W^2 \frac{s_j^2}{n} \left(1-\frac{n}{N}\right)}$
AR minta	-	-	$\bar{y} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma_B}{\sqrt{n}}$

s: a mintából becsült szórás

$$s = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}}$$

Sokasági szórás (σ)

$$Int_{1-\alpha}(\sigma) = \left[\sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}}\right]$$

Sokasági arány (P)

	nagy minta (n>30)
FAE minta	$p \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$
EV minta	$p \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \sqrt{1-\frac{n}{N}}$

E-mail: matekportal@hotmail.com
Web: matekportal.ewk.hu
Tel.: 20/453-91-78

Értékösszegbecslés

$$Int_{1-\alpha}(Y') = N \cdot Int_{1-\alpha}(\mu)$$

Mintaelemszámok

	Átlagbecslés
FAE minta	$\dfrac{\left(z_{1-\frac{\alpha}{2}} \cdot \sigma\right)^2}{\Delta^2}$
EV minta	$\dfrac{\left(z_{1-\frac{\alpha}{2}} \cdot \sigma\right)^2}{\dfrac{\left(z_{1-\frac{\alpha}{2}} \cdot \sigma\right)^2}{N} + \Delta^2}$

Neyman-féle optimális mintaelosztás:

$$n_j = n\frac{N_j\sigma_j}{\sum N_j\sigma_j}$$

Független részminták módszere (FRM)

$$Int_{1-\alpha}(\theta) = \hat{\theta}_{FRM} \pm t_{1-\frac{\alpha}{2}}(n-1) \cdot s_{\hat{\theta}_{FRM}}$$

$$\hat{\theta}_{FRM} = \frac{\sum \hat{\theta}_i}{k}$$

$$s_{\hat{\theta}_{FRM}} = \sqrt{\frac{\sum(\hat{\theta}_i - \hat{\theta}_{FRM})^2}{k(k-1)}}$$

4. Hipotézisvizsgálat

Egymintás paraméteres próbák

Kétoldali próbák: egyenlőséget vizsgálnak egy sokasági paraméterre vonatkozóan

Egyoldali próbák: egyenlőtlenséget vizsgálnak egy sokasági paraméterre vonatkozóan

a) várható értékre vonatkozó próbák

- kis minta és ismert sokasági szórás, illetve nagy minta esetén

$$Z = \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

- kis minta és ismeretlen sokasági szórás esetén

$$T = \frac{\bar{y} - \mu_0}{\frac{s}{\sqrt{n}}}$$

b) szórásra vonatkozó próba

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

c) arányra vonatkozó próba

$$Z = \frac{p - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

Egymintás nemparaméteres próbák

Minden esetben egyoldali, felső kritikus értékkel végrehajtandó próbák.

a) Illeszkedésvizsgálat

$$\chi^2 = \sum \frac{(f_i - nP_i)^2}{nP_i} = n\left(\sum \frac{g_i^2}{P_i} - 1\right)$$

szabadságfok: k-b-1 (k: kategóriák száma; b: a becsült paraméterek száma)

b) Függetlenségvizsgálat

$$\chi^2 = \sum \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}$$

szabadságfok: (r-1)(c-1) (r: a táblázat sorainak; c: a táblázat oszlopainak száma)

p-érték (empirikus szignifikancia-szint)

Az a valószínűség, amely mellett a H_0 már éppen elvethető.

A próbafüggvény empirikus értékéhez, mint kritikus értékhez tartozó szignifikancia-szint.

Két-, és többmintás próbák

Két várható értékre vonatkozó próbák:

- nagy minta, vagy ismert szórások esetén:

$$Z = \frac{\bar{y} - \bar{x} - \delta_0}{\sqrt{\frac{\sigma_y^2}{n_y} + \frac{\sigma_x^2}{n_x}}}$$

- kis minta és ismeretlen szórások esetén:

$$T = \frac{\bar{y} - \bar{x} - \delta_0}{s_c \sqrt{\frac{1}{n_y} + \frac{1}{n_x}}}$$

$$s_c = \sqrt{\frac{s_y^2(n_y - 1) + s_x^2(n_x - 1)}{n_y + n_x - 2}}$$

szabadságfok: $n_y + n_x - 2$

Két arányra vonatkozó próbák:

- van feltételezett különbség (ε_0) a két arány közt

$$Z = \frac{p_y - p_x - \varepsilon_0}{\sqrt{\frac{p_y q_y}{n_y} + \frac{p_x q_x}{n_x}}}$$

- nincs feltételezett különbség a két arány között

$$Z = \frac{p_y - p_x}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_y} + \frac{1}{n_x}\right)}}$$

Két szórásra vonatkozók próba

$$F = \frac{s_y^2}{s_x^2}$$

Szabadságfokok: $n_y - 1;\ n_x - 1$

Több várható érték azonosságára vonatkozó próba (varianciaanalízis)

$$F = \frac{s_k^2}{s_b^2}$$

$$s_k^2 = \frac{SSK}{M-1} \quad s_b^2 = \frac{SSB}{n-M}$$

$$SSK = \sum n_j\left(\bar{y}_j - \bar{y}\right)^2 \quad SSB = \sum n_j s_j^2$$

Szabadságfokok: $M - 1;\ n - M$

5. Regresszió-számítás

Kétváltozós regresszió

x: ok

y: okozat

Az általános egyenlet:

$$y = \underbrace{\hat{\beta}_0 + \hat{\beta}_1 x}_{\hat{y}} + e$$

A paraméterek becslése

$$\hat{\beta}_1 = \frac{\sum d_x d_y}{\sum d_x^2} = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$d_x = x_i - \bar{x}$$

Illeszkedés:

$$s_e = \sqrt{\frac{\sum e^2}{n-2}} \qquad \sum e^2 = SSE = \sum (y_i - \hat{y})^2$$

Kapcsolatszorosság

$$r = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}} = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{(\sum x^2 - n\bar{x}^2)(\sum y^2 - n\bar{y}^2)}} \qquad r^2 = \frac{SSR}{SST}$$

$$SSR = \hat{\beta}_1^2 \sum d_x^2 = \sum (\hat{y} - \bar{y})^2 \qquad SST = \sum d_y^2 = \sum (y_i - \bar{y})^2$$

Rugalmasság

$$El(\hat{y}, x) = \frac{\hat{\beta}_1 x}{\hat{y}}$$

E-mail: matekportal@hotmail.com
Web: matekportal.ewk.hu
Tel.: 20/453-91-78

Intervallumbecslések

- paraméterekre

$$Int_{1-\alpha}(\beta_1) = \hat{\beta}_1 \pm t_{1-\frac{\alpha}{2}}(n-2)s_{\hat{\beta}_1}$$

$$s_{\hat{\beta}_1} = \frac{s_e}{\sqrt{\sum d_x^2}}$$

$$Int_{1-\alpha}(\beta_0) = \hat{\beta}_0 \pm t_{1-\frac{\alpha}{2}}(n-2)s_{\hat{\beta}_0}$$

$$s_{\hat{\beta}_0} = s_e\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum d_x^2}}$$

- várható értékre

$$Int_{1-\alpha}(E(Y_*)) = \hat{y}_* \pm t_{1-\frac{\alpha}{2}}(n-2)s_{\hat{y}_*}$$

$$s_{\hat{y}_*} = s_e\sqrt{\frac{1}{n} + \frac{(x_* - \bar{x})^2}{\sum d_x^2}}$$

Hipotézisek

- paraméter szignifikanciája

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$T = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}$

szabadságfok: n-2

$F = \frac{SSR}{\frac{SSE}{n-2}}$

Szabadságfokok: 1, n-2

Idősoros regresszió

Durbin-Watson próba (reziduális autokorreláció)

$H_0: \rho = 0$

$H_1: \rho \neq 0$

Durbin-Watson segédregresszió:

$$\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t$$

Próbafüggvény:

$$d = \frac{\sum (e_t - e_{t-1})^2}{\sum e_t^2}$$

Nemlineáris regresszió

- exponenciális regresszió

$$\hat{y} = \hat{\beta}_0 \hat{\beta}_1^x$$

- hatványkitevős regresszió

$$\hat{y} = \hat{\beta}_0 x^{\hat{\beta}_1}$$

Kapcsolatszorosság

$$I = \sqrt{1 - \frac{SSE}{SST}}$$

Többváltozós regresszió

Általánosságban

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k$$

Paraméterek becslése

$$\vec{\beta} = (X^T X)^{-1} X^T y$$

Illeszkedés:

$$s_e = \sqrt{\frac{\sum e^2}{n-k-1}} \quad \sum e^2 = SSE = \sum (y_i - \hat{y})^2 = e^T e$$

Kapcsolatszorosság

$$R^2 = \frac{SSR}{SST} = 1 - \frac{1}{q_{yy}} = \frac{r_{y1}^2 + r_{y2}^2 - 2r_{y1}r_{y2}r_{12}}{1 - r_{12}^2}$$

Parciális korreláció

$$r_{yj.12\ldots k} = \frac{-q_{yj}}{\sqrt{q_{yy}q_{jj}}}$$

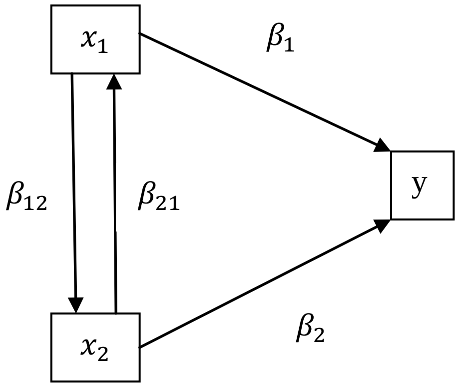
Háromváltozós esetben

$$r_{y1.2} = \frac{r_{y1} - r_{y2}r_{12}}{\sqrt{\left(1 - r_{y2}^2\right)\left(1 - r_{12}^2\right)}}$$

Rugalmasság

$$El\left(\hat{y}; x_j\right) = \frac{\hat{\beta}_j x_j}{\hat{y}}$$

Útelemzés



Közvetlen hatás: β_1 és β_2 regresszió: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$

Közvetett hatások: $\beta_{12}\beta_2$ és $\beta_{21}\beta_1$ regressziók: $\hat{x}_2 = \hat{\beta}_0 + \hat{\beta}_{12} x_1$ és $\hat{x}_1 = \hat{\beta}_0 + \hat{\beta}_{21} x_2$

Teljes hatások: $\beta_1 + \beta_{12}\beta_2$ és $\beta_2 + \beta_{21}\beta_1$ regressziók: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$ és $\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2$

Multikollinearitás

$$VIF_j = \frac{1}{1 - R_j^2}$$

$1 < VIF_j < 2, \quad gyenge$

$2 < VIF_j < 5, \quad erős$

$$5 < VIF_j, \quad káros$$

Hipotézisek

- parciális T-próba

  $H_0: \beta_j = 0$
  $H_1: \beta_j \neq 0$

  $T = \frac{\hat{\beta}_j}{s_{\hat{\beta}_j}}$
  szabadságfok: n-k-1

- globális F-próba (varianciaanalízis)

$$F = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} = \frac{n-k-1}{k}\frac{R^2}{1-R^2}$$

szabadságfokok: k, n-k-1

ANOVA-tábla

A variancia eredete	Négyzetösszeg	Szabadságfok	Átlagos négyzetösszeg	F-érték
Regresszió	SSR	k	$\frac{SSR}{k}$	$\frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}}$
Maradék	SSE	n-k-1	$\frac{SSE}{n-k-1}$	
Teljes	SST	n-1		

Intervallum-becslések

- paraméterekre

$$Int_{1-\alpha}(\beta_j) = \hat{\beta}_j \pm t_{1-\frac{\alpha}{2}}(n-k-1)s_{\hat{\beta}_j}$$

$$s_{\hat{\beta}_j} = s_e\sqrt{(X^TX)^{-1}_{jj}}$$

- várható értékre

$$Int_{1-\alpha}(E(Y_*)) = \hat{y}_* \pm t_{1-\frac{\alpha}{2}}(n-k-1)s_{\hat{y}_*}$$

$s_{\hat{y}_*} = s_e$