

Matemaatiline statistika

Üldkogum

Definitsioon 1. Üldkogumiks nimetatakse objektide, mille kohta soovitakse teha statistilisi järeldusi, hulka.

Üldkogumi objekte kirjeldatakse ühe või mitme **tunnusega**.

Definitsioon 2. Valim on üldkogumi objektide hulga lõplik alamhulk, mille põhjal tehakse järeldusi üldkogumi objektide tunnuse või tunnuste kohta.

Eeldame järgnevalt, et 1. valim on juhuslik, st. iga valimisse sattuv objekt on üldkogumist valitud juhuslikult, 2. objektid võetakse valimisse üksteisest sõltumatult, 3. iga üldkogumi objekt võib esineda valimis ühekordselt (kordusteta valim).

Valimi maht - objektide arv valimis.

Punkthinnangud

Olgu X -üldkogumi objektide tunnus, mis on juhuslik suurus m parameetrist $\alpha_1, \dots, \alpha_m$. Ehk suuruse X jaotustihedus on kujul $f(x, \alpha_1, \dots, \alpha_m)$. Valimi $\{x_1, \dots, x_n\}$ põhjal saame leida α_k empiirilise väärtuse $\alpha_k^* = \alpha_k^*(x_1, \dots, x_n)$ ehk nn **punkthinnangu**. Parameetri α_k punkthinnangut α_k^* nimetatakse 1. **nihketa hinnanguks**, kui $E\alpha_k^* = \alpha_k$; 2. **mõjusaks**, kui iga $\varepsilon > 0$ korral

$$\lim_{n \rightarrow +\infty} P(|\alpha_k - \alpha_k^*| < \varepsilon) = 1;$$

3. **efektiivseks**, kui sellel on parameetri α_k teiste punkthinnangutega vähim dispersioon.

Üldkogumi keskmise hinnang on valimkeskmine

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

kus n on valimi maht ja x_1, \dots, x_n on tunnuse x valimil mõõdetud väärtused. \bar{x} on juhuslik suurus, sest valim on juhuslik. Kehtib $E\bar{x} = EX$.

Leiame $D\bar{x}$:

$$D\bar{x} = \frac{DX}{n}$$

Kui X on normaaljaotusega, siis \bar{x} on ka efektiivne.

Üldkogumi dispersiooni punkthinnangud I Kui üldkogumi keskvärtus on teada

$$\mu_2^* = \frac{1}{n} \sum_{i=1}^n (x_i - EX)^2$$

Nihketa hinnang. Kui eksisteerib μ_4 , siis on see hinnang ka mõjus, sest $D\mu_2^* = \frac{1}{n}(\mu_4 - \mu_2^2)$.

II Kui üldkogumi keskvärtus pole teada Valimi dispersioon

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

on dispersiooni DX nihketa punkthinnang. Kehtib $s^2 = \frac{n}{n-1}(\overline{x^2} - \bar{x}^2)$.

X on normaaljaotusega

Lause 1. Kui $X \sim N(a, \sigma)$, siis 1. $\bar{x} \sim N(a, \sigma/\sqrt{n})$ ja $\sqrt{n}(\bar{x} - a)/\sigma \sim N(0, 1)$; 2. \bar{x} ja s^2 on sõltumatud juhuslikud suurused; 3. $(n-1)s^2/\sigma^2$ on juhuslik suurus, millel on χ^2 -jaotus vabadusastmete arvuga $n-1$.

Juhusliku vektori (X, Y) arvkarakteristikute punkthinnangud I

Kui EX ja EY on teada, siis

$$K_{x,y}^{***} = \frac{1}{n} \sum_{i=1}^n (x_i - EX)(y_i - EY)$$

on $cov(X, Y)$ nihketa punkthinnang.

II Kui EX ja EY ei ole teada, siis

$$K_{x,y}^* = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

on $cov(X, Y)$ nihketa punkthinnang. Kehtib $K_{x,y}^* = \frac{n}{n-1}(\overline{xy} - \bar{x}\bar{y})$, kus $\overline{xy} =$

$$\frac{1}{n} \sum_{i=1}^n x_i y_i.$$

Korrelatsioonikordaja punkthinnanguks

$$r_{x,y}^* = \frac{K_{x,y}^*}{s_x s_y}$$

Vahemikhinnangud

Olgu α juhusliku suuruse X parameeter ja $\alpha^* = \alpha^*(x_1, \dots, x_n)$ parameetri α hinnang. Kui $\varepsilon > 0$ on kindel suurus, siis vahemiku $(\alpha^* - \varepsilon, \alpha^* + \varepsilon)$ otspunktid on samuti juhuslikud suurused.

$$P(\alpha \in (\alpha^* - \varepsilon, \alpha^* + \varepsilon)) = \beta.$$

β – usaldusnivoo $l_\beta = (\alpha^* - \varepsilon, \alpha^* + \varepsilon)$ – usaldusvahemik $\alpha^* - \varepsilon$ ja $\alpha^* + \varepsilon$ – usalduspiirid Kui

$$P(\alpha \leq \alpha^* - \varepsilon) = P(\alpha \geq \alpha^* + \varepsilon) = \frac{1 - \beta}{2},$$

siis on tegemist **sümmeetrilise usaldusvahemikuga**.

Üldkogumi keskvärtuse usaldusvahemik

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad E\bar{x} = EX, \quad D\bar{x} = \frac{\sigma^2}{n}$$

Tsentraalse piirteoreemi kohaselt on \bar{x} asümptootiliselt normaalne, st kui n on küllalt suur, siis \bar{x} on ligikaudu normaaljaotusega parameetritega EX ja $\sqrt{D\bar{x}} = \sigma/\sqrt{n}$.

$$P(EX \in (\bar{x} - \varepsilon_\beta, \bar{x} + \varepsilon_\beta)) = \beta$$

$$l_\beta = \left(\bar{x} - \frac{s}{\sqrt{n}} \Phi^{-1} \left(\frac{\beta}{2} \right), \bar{x} + \frac{s}{\sqrt{n}} \Phi^{-1} \left(\frac{\beta}{2} \right) \right).$$

Põhilisi jaotusi

Öeldakse, et juhuslik suurus

$$Y_n = X_1^2 + X_2^2 + \dots + X_n^2$$

allub χ^2 -**jaotusele**, kui $X_k \sim N(0, 1)$ on sõltumatud juhuslikud suurused. Arvu n nimetatakse χ^2 -jaotusega juhusliku suuruse Y_n vabadusastmete arvuks. Saab näidata, et

$$f_{Y_n}(y) = \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} y^{n/2-1} e^{-y/2}, & \text{kui } y > 0, \\ 0, & \text{kui } y \leq 0, \end{cases}$$

kus $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt$ ($\alpha > 0$) on Euleri gammafunktsioon.

Põhilisi jaotusi

Õeldakse, et juhuslik suurus

$$T_n = \frac{X}{Y_n} \sqrt{n}$$

allub **Studenti jaotusele** ehk t -jaotusele vabadusastmete arvuga n , kui $X \sim N(0, 1)$ ja Y_n on sõltumatud juhuslikud suurused ning Y_n on χ^2 -jaotusega vabadusastmete arvuga n .

$$f_{T_n}(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, \quad t \in R, \quad n = 1, 2, \dots$$

Põhilisi jaotusi

Õeldakse, et juhuslik suurus

$$Z_{n,m} = \frac{X_n/n}{Y_m/m}$$

allub **Fisher'i jaotusele** vabadusastmetega n ja m , kui sõltumatud juhuslikud suurused X_n ja Y_m on χ^2 -jaotusega vastavalt vabadusastmete arvuga n ja m .

$$f_{Z_{n,m}}(z) = \begin{cases} \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{n}{m}\right)^{n/2} z^{(n-2)/2} \left(1 + \frac{n}{m}z\right)^{-(n+m)/2}, & \text{kui } z > 0, \\ 0, & \text{kui } z \leq 0. \end{cases}$$