

# JÄRELDAV STATISTIKA

## Väljavõttelise vaatluse vead

Valim<>valim, valim<>populatsioon

1) Tõenäosusteooria tsentraalne piirteoreem:

kui mistahes populatsioonist tuletada suuri ühesuguste mahtudega valimeid, siis on nende valimite keskväärtused peaaegu normaaljaotusega.

2) Valimite keskväärtuste aritmeetiline keskmine  $\approx$  populatsiooni keskväärtus.

Valimi keskväärtus paikneb:

$$95\% \quad M_{\text{pop}} \pm 1,96 \sigma_{\text{pop}}$$

$$99\% \quad M_{\text{pop}} \pm 2,58 \sigma_{\text{pop}}$$

$$99,73\% \quad M_{\text{pop}} \pm 3,0 \sigma_{\text{pop}}$$

$$\sigma_{\text{pop}} = ?$$

Valimi keskväärtuste teoreetilise jaotuse standardhälve on väljavõtu vea mõõduks. Teda nimetatakse keskväärtuse standard e esindusveaks ( $SE_{M,m}$ ).

$$SE_M = \sigma_{\text{pop}}/\sqrt{N}, \text{ kus}$$

$$SE_M = \sigma_{\text{valim}}/\sqrt{N}$$

Näide.

Kui kaugel eeldame valimi keskväärtust  $M_V=75$  asuvat populatsiooni keskväärtusest  $M_{\text{pop}}$ , kui  $\sigma_V=18$ ,  $N=36$ ?

$$SE_M = 18/\sqrt{36} = 3$$

$$M_V - Z \cdot SE_M \leq M_{\text{pop}} \leq M_V + Z \cdot SE_M$$

$$0,3174 \quad 68,26\% \quad 75 \pm 1,0 \cdot 3 \quad 72,0 \leq M_{\text{pop}} \leq 78,0$$

$$0,05 \quad 95\% \quad 75 \pm 1,96 \cdot 3 \quad 69,1 \leq M_{\text{pop}} \leq 80,9$$

$$0,01 \quad 99\% \quad 75 \pm 2,58 \cdot 3 \quad 67,3 \leq M_{\text{pop}} \leq 82,7$$

$$0,0027 \quad 99,73\% \quad 75 \pm 3,0 \cdot 3 \quad 66,0 \leq M_{\text{pop}} \leq 84,0$$

**Usaldusnivoo** - tõenäosus jääda usaldusvahemikku.

**Olulisuse nivoo** - tõenäosus jääda väljaspoole usaldusvahemikku ( $p, \alpha$ ).

### ***Student'i jaotus (t-väärtused)***

Valimi väikeste mahtude korral pole tegu mitte normaal-, vaid Student'i jaotusega (t-jaotusega).

Et saada t-jaotuse korral usaldatavusnivood, mille piires asub 95% teoreetilisi valimi kesk-väärtusi, tuleb meil standardhälbe ühikutes liikuda  $M_{pop}$  suhtes kaugemale.

t-jaotus sõltub valimi suurusest.

| t-väärtused       |          |          |
|-------------------|----------|----------|
| Vabadus<br>astmed | $p=0,05$ | $p=0,01$ |
| 1                 | 12,706   | 63,657   |
| 2                 | 4,303    | 9,925    |
| 5                 | 2,571    | 4,032    |
| 8                 | 2,306    | 3,355    |
| 10                | 2,228    | 3,196    |
| 14                | 2,145    | 2,977    |
| 16                | 2,120    | 2,921    |
| 20                | 2,086    | 2,845    |
| 30                | 2,042    | 2,75     |
| 60                | 2,0      | 2,66     |
| 120               | 1,98     | 2,617    |
| $\infty$          | 1,96     | 2,576    |

### Näide.

Määrata  $M_{pop}$  usaldusvahemik, kui  $M_V=50$ ,  $\sigma_V=10$ ,  $N=17$  ( $df=N-1=16$ ),  $p=0,05$ .

$$SE_M = 10/\sqrt{17} = 10/4,12 = 2,42$$

$$M_V - t_p * SE_M \leq M_{pop} \leq M_V + t_p * SE_M$$

$$t_{(0,05;df=16)} = 2,12$$

$$50 \pm 2,12 * 2,42 = 50 \pm 5,13; \quad 44,87 \leq M_{pop} \leq 55,13$$

**Vastus:**  $M_{pop}$  omab 95% tõenäosust omada väärtust 44,87 ja 55,13 vahel.

### *Statistilised hüpoteesid*

Hüpotees - oletus, mida uurija omab tunnuste vaheliste seoste või erinevuste kohta.

Hüpoteeside kontrollimine - seose/erinevuse heakskiitmine või kummutamine teatud usaldusnivool. Näitab, kas erinevus/seos on statistiliselt oluline või ta on põhjustatud väljavõtuveast.

Nullhüpotees ( $H_0$ ) - statistiliselt oluliste erinevuste puudumine.

Alternatiivne e. sisukas hüpotees ( $H_1$ ) - statistiliselt oluliste erinevuste olemasolu.

### Näide.

2 rühma: 11 poissi, keskmine hinne  $M_V=14$ ,  $\sigma_V= 3$   
11 tüdrukut, keskmine hinne  $M_V=10$ ,  $\sigma_V= 2$ . Kas rühmade keskmised hinded erinevad oluliselt?

$H_0$ : keskmised hinded ei erine oluliselt;

$H_1$ : keskmised hinded erinevad oluliselt.

$$t_{0,05} = 2,09; \quad t_{arvutuslik} = 3,67$$

**Kriitiline väärtus** - kriteeriumi väärtus, mille korral veel kehtib  $H_0$ .

### ***I ja II liiki viga***

**I liiki viga** - tõese  $H_0$  kummutamine;  $\alpha=0,05$

**II liiki viga** - väär  $H_0$  vastuvõtmine;  $\beta$

I liiki viga vähendamiseks vähendada olulisuse nivood  $p$  või suurendada valimi mahtu. See aga suurendab teist liiki viga tõenäosust.

**Kriteeriumi võimsus**  $= 1 - \beta$ , so. tõenäosus kummutada  $H_0$ , mis ongi tegelikult vale.

**Kriteeriumi efektiivsus**  $= (N_{\text{param}}/N_{\text{mitteparam}})*100\%$ , so. kuivõrd peame suurendama valimi mahtu, et mitteparameetiline kriteerium muutuks sama võimsaks kui parameetiline kriteerium.

**Näide.**

**Kriteeriumi efektiivsus = 70%,  $N_{\text{param}}=21$**

**$N_{\text{mitteparam}}= 21*100/70= 30$ .**