

Usalduspiirid

Kvantitatiivsed meetodid

Indrek Soidla

Usaldusvahemik

- Valimi alusel arvutatud statistik (nt aritmeetiline keskmine) kehtib täpselt valimi kohta
- Kui täpselt kirjeldab see populatsiooni?
- Usaldusvahemik – vahemik, mis katab statistiku tegeliku väärtuse populatsioonis teatud tõenäosusega (täpsemalt: teatud läve ületava tõenäosusega)
 - NB! Eeldab, et valim on juhuslik kogum populatsioonist
 - = tõenäosuslik valim
- Punkthinnang vs vahemikhinnang
- Erinevate statistikute puhul arvutatakse mõnevõrra erinevalt, aga mõte sama

Usaldusvahemik

- Usalduspiirid aritmeetilise keskmise puhul

$$\mu \geq m - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\mu \leq m + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

- μ – keskmine populatsioonis
- m – keskmine valimis
- $z_{1-\frac{\alpha}{2}}$ – standardiseeritud normaaljaotuse $1 - \frac{\alpha}{2}$ kvantiil, kus α on vea tõenäosuse piir
 - nt usaldusnivool 90% $\alpha = 10\%$, vastav $z_{1-\frac{\alpha}{2}} = 1,64$
 - nt usaldusnivool 95% $\alpha = 5\%$, vastav $z_{1-\frac{\alpha}{2}} = 1,96$
 - nt usaldusnivool 99% $\alpha = 1\%$, vastav $z_{1-\frac{\alpha}{2}} = 2,58$
- σ – standardhälve populatsioonis, suure valimi korral ligilähedane standardhälbele valimis
- n – valimimaht
- Näide R-s

Millest sõltub usaldusvahemiku laius?

- Valimimaht
 - Suurem valimimaht => kitsam usaldusvahemik
- Tunnuse hajuvus
 - Väiksem hajuvus => kitsam usaldusvahemik
- Usaldusnivoo
 - Madalam usaldusnivoo => kitsam usaldusvahemik
- Kas selleks siis, et tulemuses maksimaalselt kindel olla, on vaja suuremat valimimahtu, tunnuse väiksemat hajuvust ja madalamat usaldusnivood?
- Madalam usaldusnivoo tähendab ka suuremat eksimisvõimalust

Usaldusvahemik

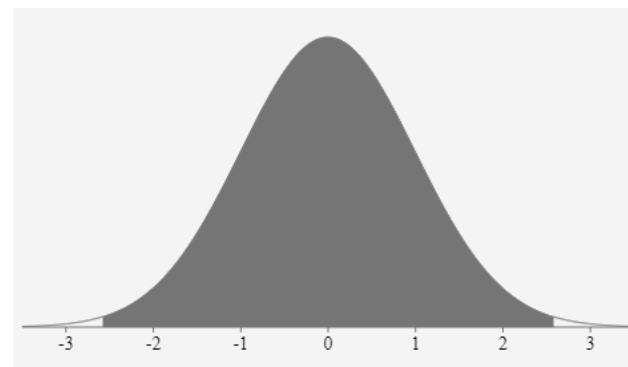
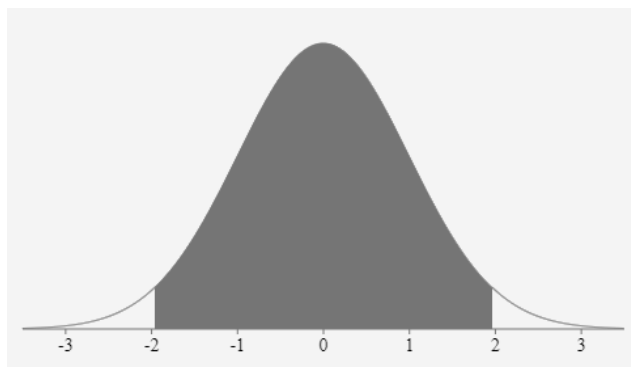
- Mida saame väita usalduspiiride põhjal?
- Usaldusnivool 95% saame väita, et...
 - nt Kesk-Eestis on hinnangute tegelik keskmine vahemikus 4,14 ja 4,92

Mida näitab usaldusnivoo?

- Valikuviga – juhuslik viga, mille põhjus on see, et üldkogumi asemel käsitletakse juhuvalimit
 - Üldkogumi asemel käsitletakse juhuvalimit – valim ei pruugi olla täpne koopia populatsioonist
 - St tunnuste jaotused valimis ei pruugi olla identsed tunnuste jaotustega populatsioonis
 - Jahuvaliku tõttu ei saa jaotused valimis ja üldkogumis erineda väga palju, aga teatud (väikeses) ulatuses saavad
 - St meil võib juhuslikult olla valim, mis (meie analüüsitava tunnuse jaotuse seisukohalt) erineb populatsioonist vähe, aga võib juhuslikult olla ka valim, mis erineb populatsioonist rohkem
- Usaldusnivood võib tõlgendada kui indikaatorit selle kohta, kas usaldusvahemiku arvutamisel arvestatakse suurema või väiksema valikuveaga (valitakse vastavalt kas rangem või leebem usaldusnivoo)

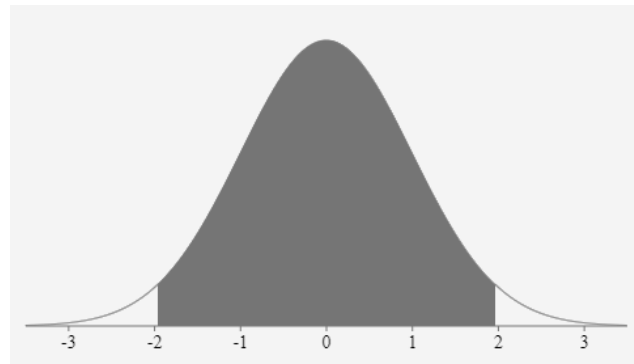
Mida näitab usaldusnivoo?

- Teisiti öeldes,
 - kui võtame usaldusnivoo 95%,
 - arvestame võimalusega, et
 - meie valim on (meie analüüsitava tunnuse jaotuse seisukohalt) 95 protsendi populatsiooniga kõige sarnasemate valimite seas;
 - me ei arvesta võimalusega, et meie valim võib olla 5 protsendi populatsioonist kõige rohkem erineva valimi seas
- Samalaadselt,
 - kui võtame rangema usaldusnivoo, 95% asemel 99%,
 - arvestame usaldusvahemiku arvutamisel võimalusega, et
 - meie valim ei pruugi olla isegi 95 protsendi populatsiooniga kõige sarnasemate valimite seas,
 - st võib populatsioonist veel rohkem erineda;
 - samas ei arvesta me võimalusega, et meie valim võib olla 1 protsendi populatsioonist kõige rohkem erineva valimi seas



Mida näitab usaldusnivoo?

- Mida rangema usaldusnivoo valime,
 - seda suuremat valimi erinevust populatsioonist peame võimalikuks ja
 - seda laiem on usaldusvahemik (sest arvestame võimalusega, et meie valim erineb populatsioonist rohkem)



Kuidas tõlgendada usaldusvahemikku?

- Millised tõlgendused alljärgnevaist on õiged?
- *Keskmise usaldusvahemik sisaldab 95% indiviidide väärtused*
- *95%-lise tõenäosusega sisaldab usaldusvahemik keskmise tegelikku väärtust*
- *Usaldusvahemik katab keskmise tegeliku väärtuse 95 juhul 100-st*
 - Mida siin tähendab „95 juhul 100-st“?
- *Tõenäosus, et keskmise tegelik väärtus on usaldusvahemikus, on 95%*
 - Miks on selline väide problemaatiline? Vaatame lähemalt

Usaldusvahemik ja tõenäosus

- Ei ole täpne öelda, et saame usaldusvahemikku 95% usaldada, või usaldusvahemikus 95% kindel olla
- Tõsi, meie ei tea täpselt, kas arvutatud usaldusvahemik katab tegeliku väärtuse või mitte, aga
- mingi valimi alusel juba leitud usaldusvahemik kas katab tegeliku väärtuse või ei kata
- Kui ikkagi tahame rääkida tõenäosusest, et juba arvutatud usaldusvahemik katab tegeliku väärtuse, siis see tõenäosus on kas 100% või 0%, mitte 95%
 - St usaldusvahemik kas katab tegeliku väärtuse või ei kata

Usaldusvahemik ja tõenäosus

- Arvutades mitmete valimite põhjal keskmiste usaldusvahemikud usaldusnivool 95%, katavad neist vähemalt 95% keskmise tegelikku väärtust
- Usaldusnivoo – tegelikku väärtust katvate usaldusvahemike minimaalne sagedus hüpoteetiliste valimite korral
- NB! Enamik valetõlgendusi tulevad sisse seal, kus kasutatakse tõenäosuse terminit
- Kui pole kindel, et tõlgendad õigesti, ära tõenäosuse (ega muud sarnast) mõistet sisse too
- Oluline: kui kõneleme tõenäosusest, siis tähendab see
 - tõenäosust saada andmeid (mille puhul keskmise usaldusvahemik katab tegeliku väärtuse)
 - mitte tõenäosust, et statistiku (nt keskmise) tegelik väärtus on nii või nii suur

Miks on oluline terminoloogias täpne olla?

- Et potentsiaalset lugejat mitte eksitada :)
- Et ise kompetentseid järeldusi teha

Mida saame väita usalduspiiride põhjal?

- Usaldusnivool 95% saame väita, et...
 - Kesk-Eestis on hinnangute tegelik keskmine vahemikus 4,14 ja 4,92
 - Kirde-Eestis on hinnangute tegelik keskmine vahemikus 3,73 ja 4,43
 - Lääne-Eestis on hinnangute tegelik keskmine vahemikus 4,26 ja 4,88
 - keskmised võivad neis regioonides kattuda
- Kas saame väita, et usaldusnivool 95% keskmised neis regioonides ei erine?
- Kui statistiliselt täpne olla, siis ei saa
- Miks?
- Täpsema hinnangu annab keskmiste erinevuse (vahe) usaldusvahemik

Keskmiste vahe usalduspiirid

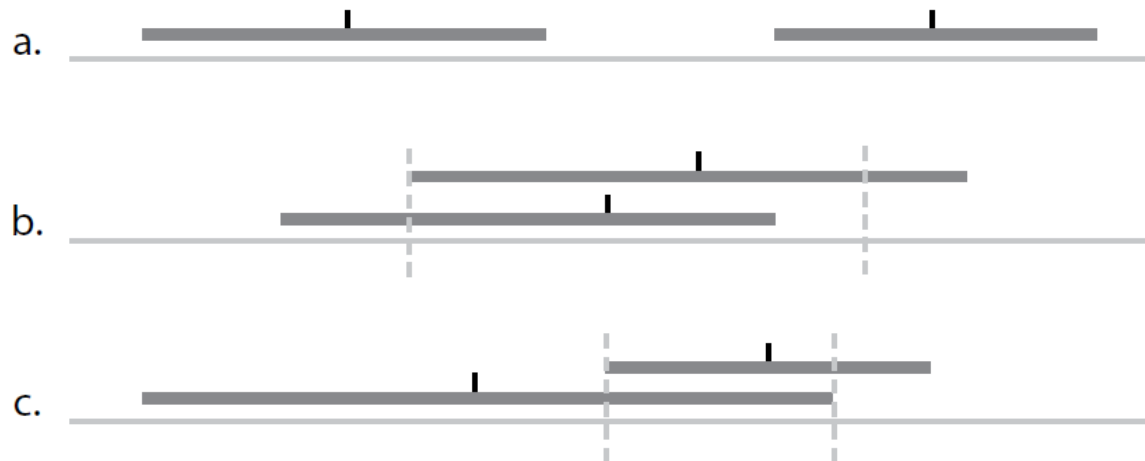
- Keskmiste vahe usalduspiiride arvutamine

$$\mu_1 - \mu_2 = m_1 - m_2 \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- R-s võimalik arvutada, lihtsamalt saab t-testiga
 - Nt sõltumatute kogumite t-test hindabki, kas kahe grupi keskmiste vahe erineb nullist
- Meie näites Kirde- ja Lääne-Eesti keskmiste vahe usaldusvahemik ei sisalda nulli, seega
 - hoolimata keskmise rahulolu usaldusvahemike kattumisest saame väita, et
 - usaldusnivool 95% keskmine hinnang Kirde- ja Lääne-Eestis erineb

Usaldusvahemik

- Lihtne viis keskmiste usaldusvahemike alusel võrdlemise „usaldamiseks“



- a. Saab erinevust väita
- b. Erinevust ei saa väita
- c. Kontrolli keskmiste erinevuse usaldusvahemikku või mõne statistilise testi alusel