VALIK-UURINGUD I

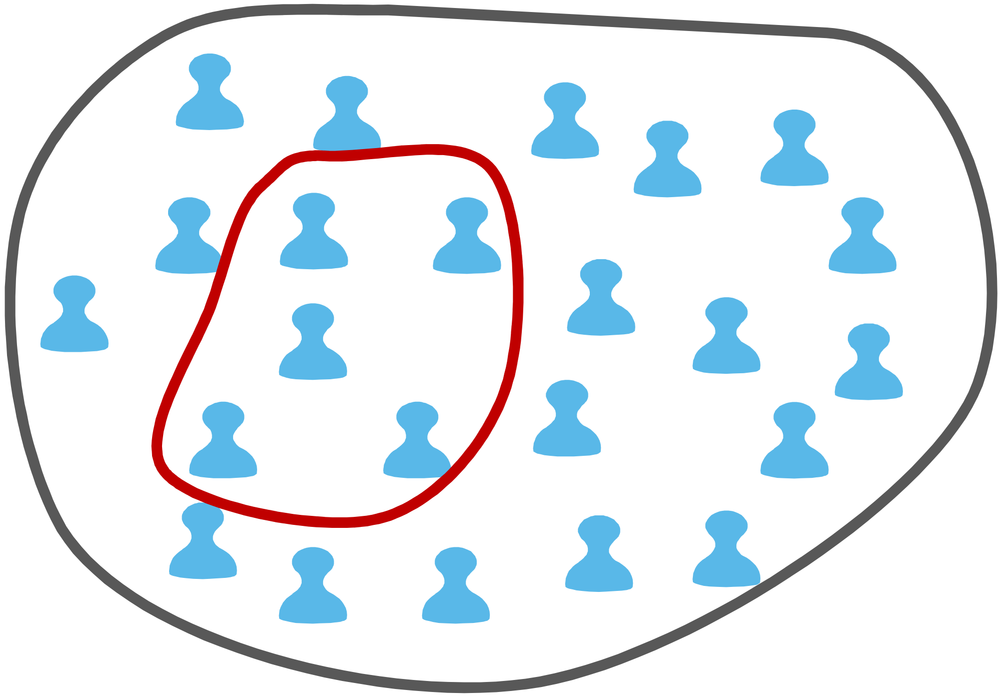
TEEMAD

- Kogum, valim ja valikumeetodid
- Punkthinnang
- Üldkogumi keskväärtuse, dispersiooni ja standardhälbe punkthinnangud
- Valimi keskmise valimjaotus
- Keskväärtuse usalduspiirid suure valimi korral

VALIM JA KOGUM

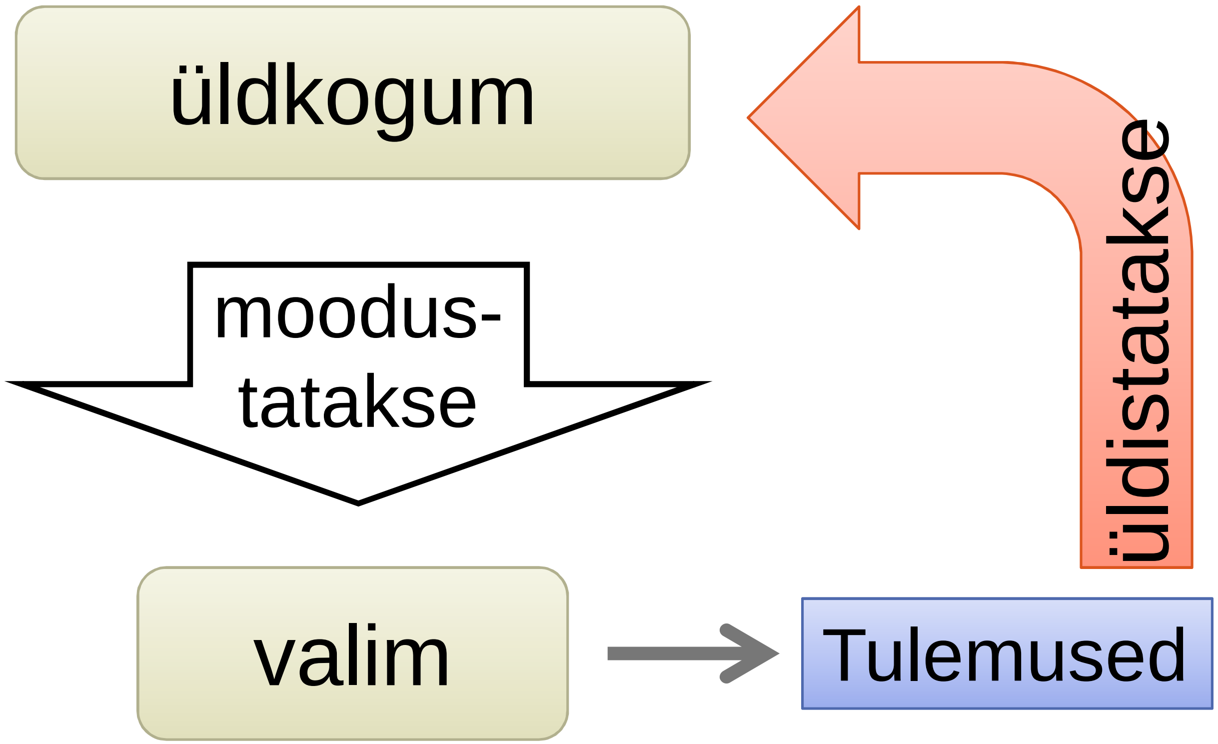
Statistiline uuring võib olla

- **kõikne** – uuritakse läbi terve üldkogum
- **valikuline** – uuritakse läbi üldkogumit esindav osa: valim



VALIKUURINGU EESMÄRK

Valikuuringu eesmärgiks on valimi põhjal järelduste tegemine üldkogumi kohta



VALIKUURINGUTE KASUTAMISE PÕHJUSED

- **Väiksem maksumus**
- **Suurem kiirus**
- **Suurem paindlikkus**
- **Laiem rakendatavus**
  - spetsiaalne aparatuur;
  - spetsiaalselt ettevalmistatud töötajad.
- **Suurem täpsus andmete kogumisel**
  - suurema kvalifikatsiooniga tööjõud;
  - võimalik paremini kontrollida töötlemisvigu.
- Mõnikord võib objekti testimine **rikkuda** objekti.

Video: The Original Mattress Factory Test center

NÄIDE: KULUD

- Kõikne statistika: rahvaloendus

Ca 65% e-loendusel, ülejäänute juurde küsitlejad

Värvati 2000 inimest

Planeeritud kulud 18,9 mln eurot

- Valikuuring TEAN ja OSKAN

Valim: 13 000 inimest

Intervjueerijate arv 120

Planeeritud kulud 1,6 mln eurot

LOEND

Objektide võtmine valimisse toimub loendi abil

Loend (freim) on vahend pääsemiseks üldkogumi objektide juurde.

Näited

- elanike register
- äriregister
- klientide andmebaas

Näiteks TNS Emor. Kodudes toimuvad küsitlused, loendiks on aadressibaasis olevad aadressid.

VASTAMISMÄÄR

Vastamismäär on vastanute osakaal valimis

1607 uuringu põhjal keskmine vastamismäär

- isikuküsitluste korral 52,7%
- organisatsioonide korral 35,7%

Eesti ettevõtete aastastatistika üldkogum, valim ja vastanud

Aasta	Kogum	Valim	Vastanud	Valimi osatähtsus kogumis, %	Vastamis-määr, %
2005	44 094	12 613	10 305	28,6	81,7
2010	61 293	11 643	9 231	19,0	79,3
2012	69 694	12 037	8 844	17,3	73,5

VALIKUMEETODID

Tõenäosuslikud valikumeetodid

- iga objekti korral teada selle valimisse kaasamise tõenäosus;
- võimalik hinnata valimi põhjal tehtud hinnangute **usaldusvahemikku** ja teame **tõenäosust**, et meid huvitava parameetri väärtus kogumis langeb sellesse vahemikku

Empiirilise valiku korral ei ole objektide valimisse sattumise tõenäosused teada

JUHUSLIKU VALIKU MEETODID

Tõmbeviisi valik

- Ühe juhusliku katse tulemus otsustab, millised üldkogumi objektid valimisse võetakse.
- Kasutatakse näiteks juhuarvude generaatorit.
- Valimi maht ette antud.

Loeteluviisi valik

- Korduvad juhuslikud katsed: üldkogumi iga objekti korral sooritatakse katse, mille tulemusel otsustatakse, kas see objekt võetakse valimisse või mitte
- Valimi maht pole täpselt ette teada.

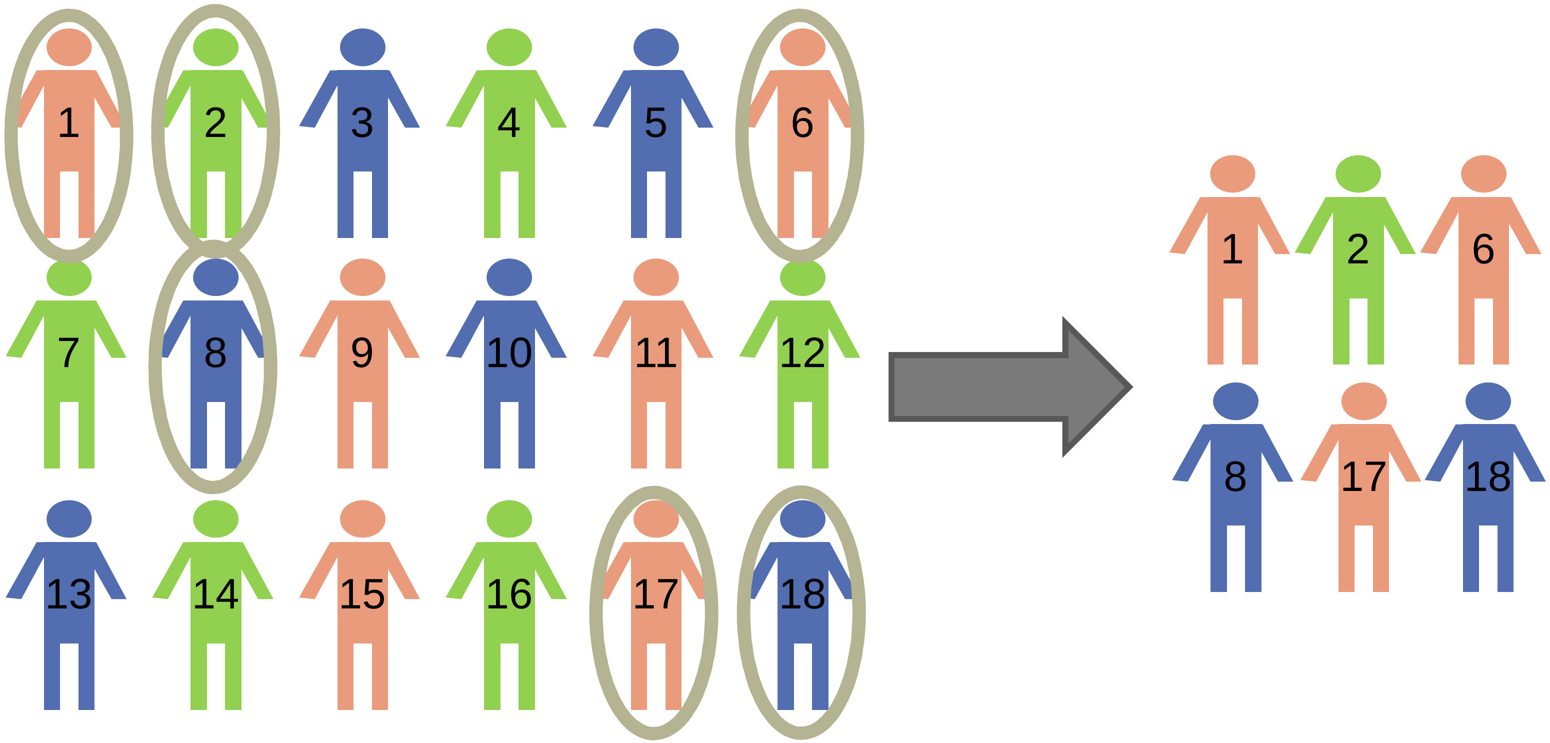
Demo: valikuviisid

TÕENÄOSUSLIKUD VALIKUMEETODID

- Lihtne juhuvalik
- Süstemaatiline valik
- Kihtvalik
- Klastervalik
- Kaheastmeline valik

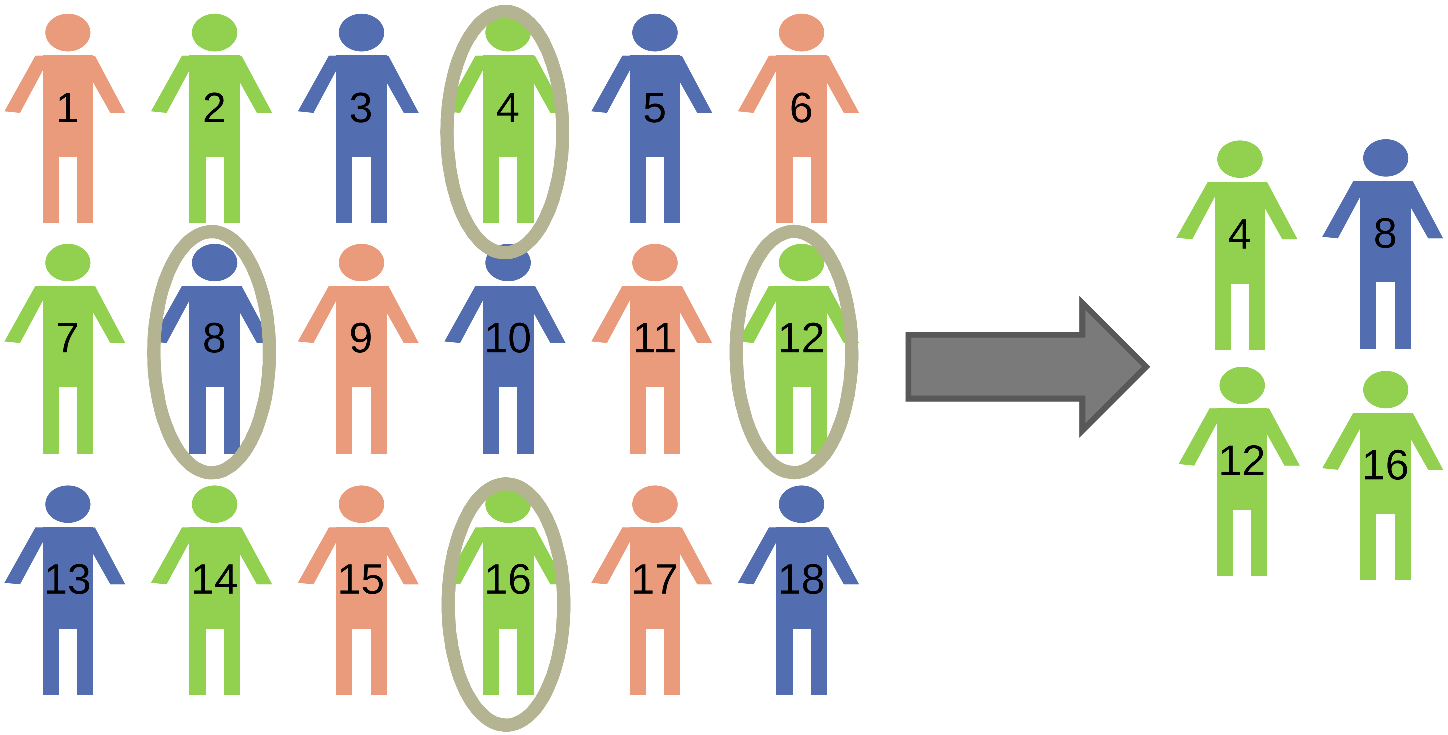
LIHTNE JUHUVALIK

- Kõigil objektidel ühesugune valimisse sattumine tõenäosus
- Ei taga piisavat representatiivsust



SÜSTEMAATILINE VALIK

- Objektide valik loendist toimub fikseeritud sammuga, mis määratakse juhuslikult.
- Ei taga piisavat representatiivsust.



NÄIDE: SÜSTEMAATILINE VALIK

Uuring „Vabatahtlikus tegevuses osalemine Eestis 2013“

Poliitikauuringute keskus Praxis.

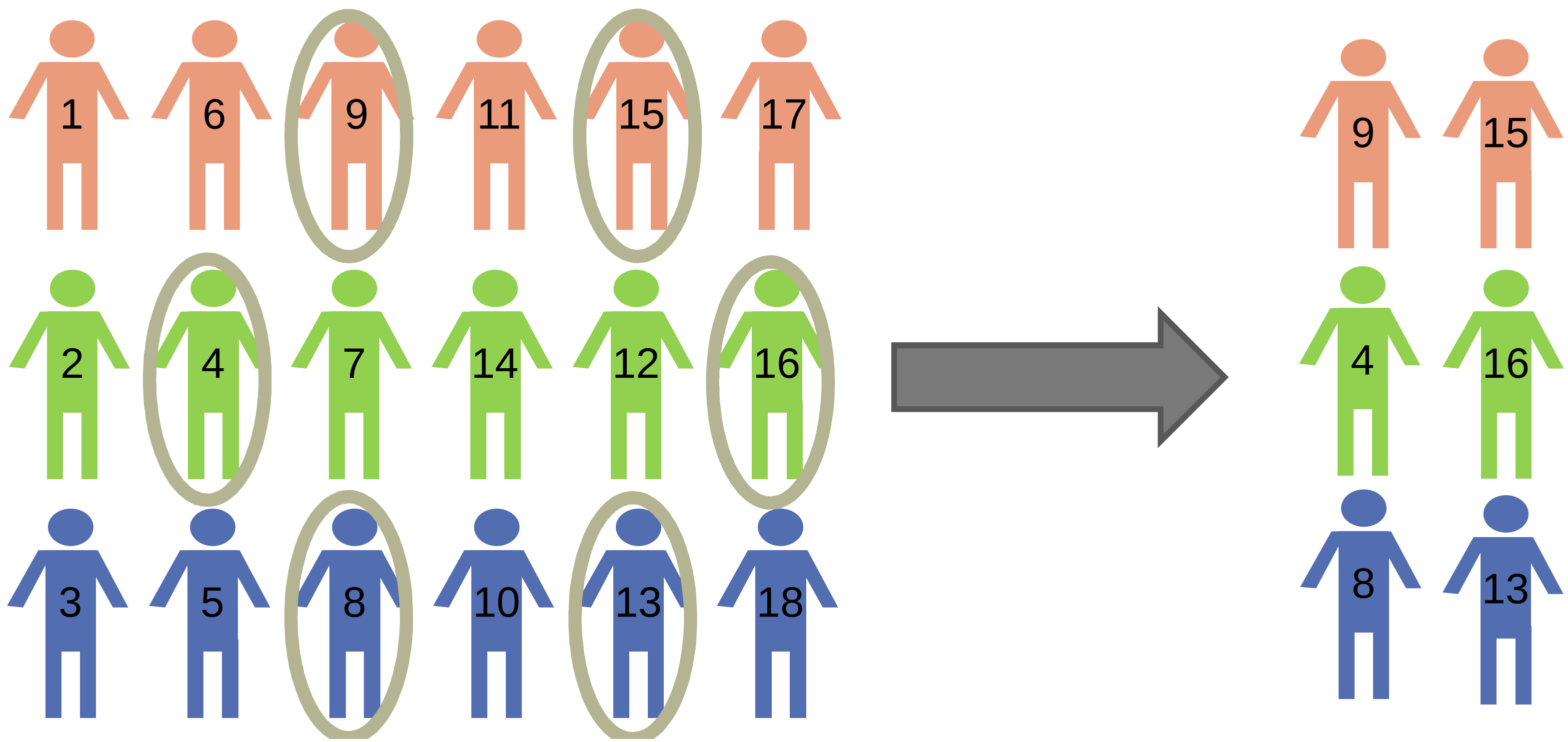
Valimi suurus 1000 inimest.

Aadressi valikul rakendati lähte-aadressi meetodit, mille puhul antakse igale küsitlejale ette juhuslikult valitud aadress esimese intervjuu läbiviimiseks.

Edasi liiguti kindla sammu alusel – iga kolmas korter või iga teine eramaja, et tagada valikusse sattunud elupaikade juhuslikkus.

KIHTVALIK

- Üldkogum jaotatakse tausttunnuste järgi kihtideks.
- Igas kihis rakendatakse mingit tõenäosuslikku valikumeetodit.



NÄIDE: KIHTVALIK

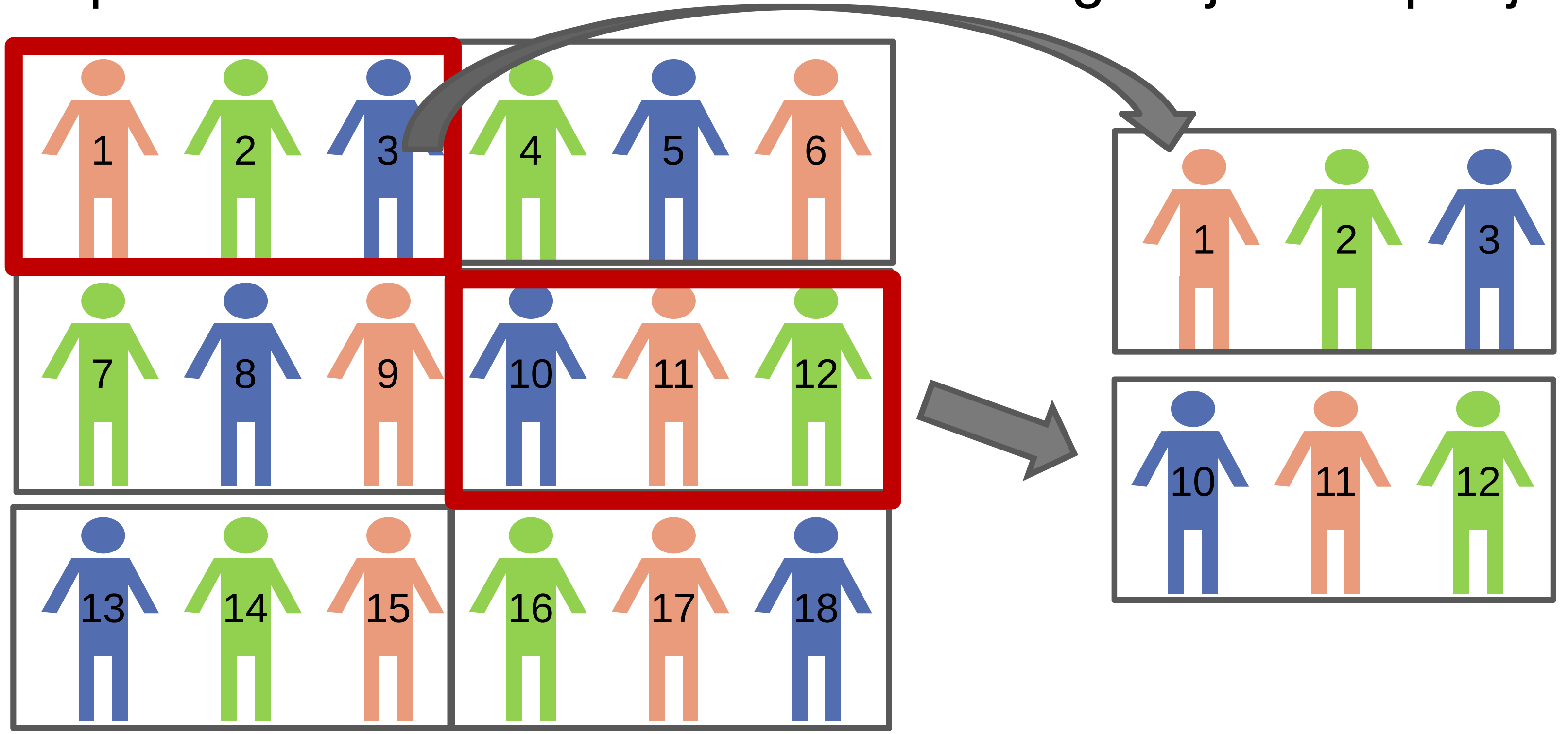
Avaliku teenistuse personalijuhtimise uuring 2010

Saar Poll OÜ

Sihtrühm	Valimi suurus
Personalijuhid	57
Asutuste juhid	57

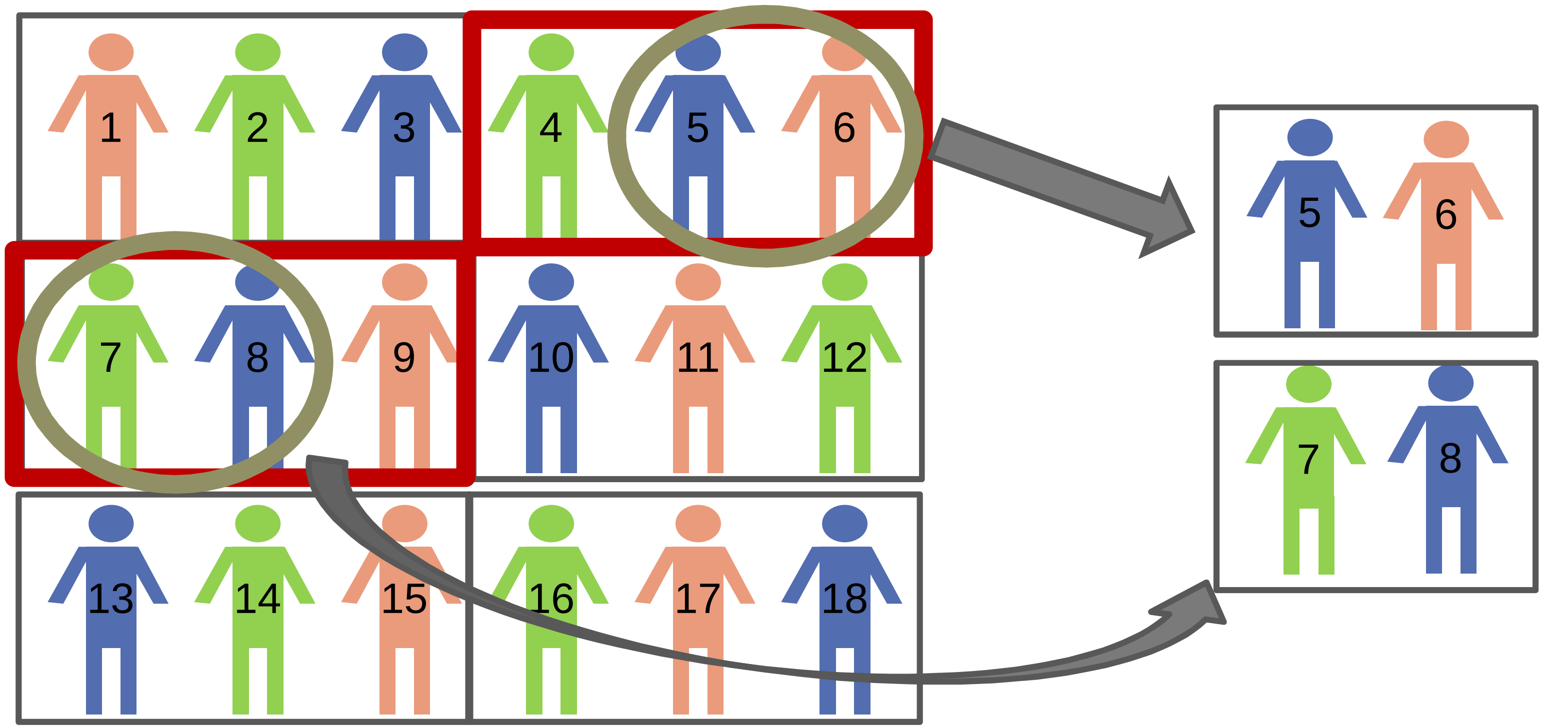
KLASTERVALIK

- Üldkogum koosneb objektigruppidest ehk klastritest (üldhariduskoolid, vallad).
- Toimub juhuslik klastrite valik ning iga klastri parameetrid leitakse selle klastri kõigi objektide põhjal.



KAHEASTMELINE VALIK

- Esimesel astmel klastrite juhuslik valik.
- Teisel astmel klastritest objektide juhuslik valik.



NÄIDE: NOORSOOUURING KISS1999

Uuriti noorte seksuaalset küpsemist.

Loend: eesti ja vene õppekeelega üldhariduskoolide 9. kl.

Kihtvalim: eesti ja vene koolid, tausttunnus emakeel

Kaheastmeline valik:

1. **Klastervalik** õpilaste arvu järgi: väikesed, keskmised, suured klassid.

2. **Juhuvalik** klastrites.

EMPIIRILINE VALIK I

Empiirilise valiku korral ei ole objektide valimisse sattumise tõenäosused teada.

- **Kvootide meetod.**

Ette antakse valim struktuur, st antakse ette kvoodid, kui palju objekte tuleb vaadeldavatest tunnusrühmadest valida. Kvoodid määratakse vastavalt üldkogumi struktuurile.

- **Tasakaalustatud valik.**

Analoogne kvootide meetodiga, kuid üldkogumi ja valimi võrdlemiseks ei kasutata mitte taustunnuse väärtuse sagedusi vaid näiteks keskmisi.

EMPIIRILINE VALIK II

- **Ekspertvalik ehk subjektiivne valik**

Näiteks nimetab ekspert kümme tänavat, mis on tema arvates uuritava linna jaoks tüüpilised ning andmed korjatakse nendelt tänavatelt.

- **Sobivusvalim**

Valitakse välja objektid, mis mingil põhjusel on sobivad. Näiteks küsitletakse inimesi, keda isiklikult tuntakse.

KUI SUUR PEAB OLEMA VALIM?

- Sõltub tunnuse varieerumisest üldkogumis.
- Kui tunnus ei varieeru, siis valimi maht n=1.
- Suurema varieerumise korral suurem valim.
- Täpsemalt vt valimi mahu planeerimine

NÄITED ERINEVATEST VALIMITEST

Uuring	Üldkogum	Kogumi maht	Valimi maht	Valimi osakaal kogumist
Erakondade reiting (Emor)	18 a ja vanemad	1068 tuh	1000	0,094%
Tean ja oskan	16-65-a	860 tuh	13 000	1,5%
Sotsiaaluuring (Eesti Statistikaamet)	Leibkonnad	600 tuh	7000	1,2%
Ettevõtted (Eesti Statistikakaamet)		70 tuh	12 000	17%

VALIMI ESINDUSLIKKUS

Esinduslik ehk representatiivne valim:

- piisavalt suur (suurus sõltub varieerumisest);
- kõigil üldkogumi objektidel võimalus sattuda valimisse;
- kajastab üldkogumi seesmist struktuuri, taustunnuste jaotumist
  - meeste ja naiste osakaal valimis sama, mis kogumis
  - vanuseline struktuur valimi sama, mis kogumis.

  Olulised on need taustunnused, mis mõjutavad uuritavaid tunnuseid.

NÄIDE: VALIMISED USA-s 1936

Alates aastast 1916 oli ajakiri Literary Digest edukalt prognoosinud USA presidendivalimiste võitjaid.

Aastal 1936 kasutati esmakordselt telefoniküsitlust. Valimi maht ca 2,4 mln.

Prognoos: Alf Landon

Tegelikult Franklin Roosevelt

Gallupi valim 50 tuhat, prognoos Roosevelt oli õige.

VALIKUNIHE

Valikunihe tekib siis, kui valimi moodustamisel on teatud tüüpi objektidel suurem võimalus valimisse sattuda kui teistel.

Näiteks

- küsitlused veebis
- küsitlused telesaadetes (helistage numbril …, kui vastate "ei" ja numbril …, kui vastate "jaa")

NÄIDE: KÜSITLUS TALLINNAS

Tallinnas läbiviidud küsitlus 19.-25. märts 2012.
Kas toetate tasuta ühistransporti?
Küsitlus viidi läbi kaubanduskeskustes, linnaosade küsitluspunktides.

Kogum 416 000
Valim 68 059

„Jah“ 51 242
„Ei“ 16 939

Linnapea sõnul oli 51 242 inimese antud mandaat piisav.

Valimi suurusest on tähtsam valimi **esinduslikkus**

Valesti koostatud suur valim annab halvema tulemuse kui õigesti koostatud väike valim.

HINNANGUD

Uuringu eesmärk: teha järeldusi üldkogumi kohta.

Näiteks: Eesti elanike sissetulekute aritmeetiline keskmine on …… eurot.

Kõikne statistika: saame selle aritmeetilise keskmise leida.

Valikuuring: saame seda aritmeetilist keskmist **hinnata.**

Valimi põhjal leitakse üldkogumi parameetrite **hinnangud.** Kasutatakse spetsiaalseid hindamisreegleid.

PUNKTHINNANG

Punkthinnang: parameetri hindamise tulemuseks on üks arv

Hinnang leitakse valimi põhjal
Valim on juhuvalim

$\Downarrow$

Punkthinnang on juhuslik suurus

Demo: punkthinnang

HINNANGU NIHE

Hinnangu nihe on hinnangute keskväärtuse $E[\hat{a}]$ ja

hinnatava parameetri tegeliku väärtuse a vahe:

$$b = E[\hat{a}] - a$$

Näiteks 5 valimit

Valim	Keskmine
1	1216
2	902
3	964
4	996
5	904

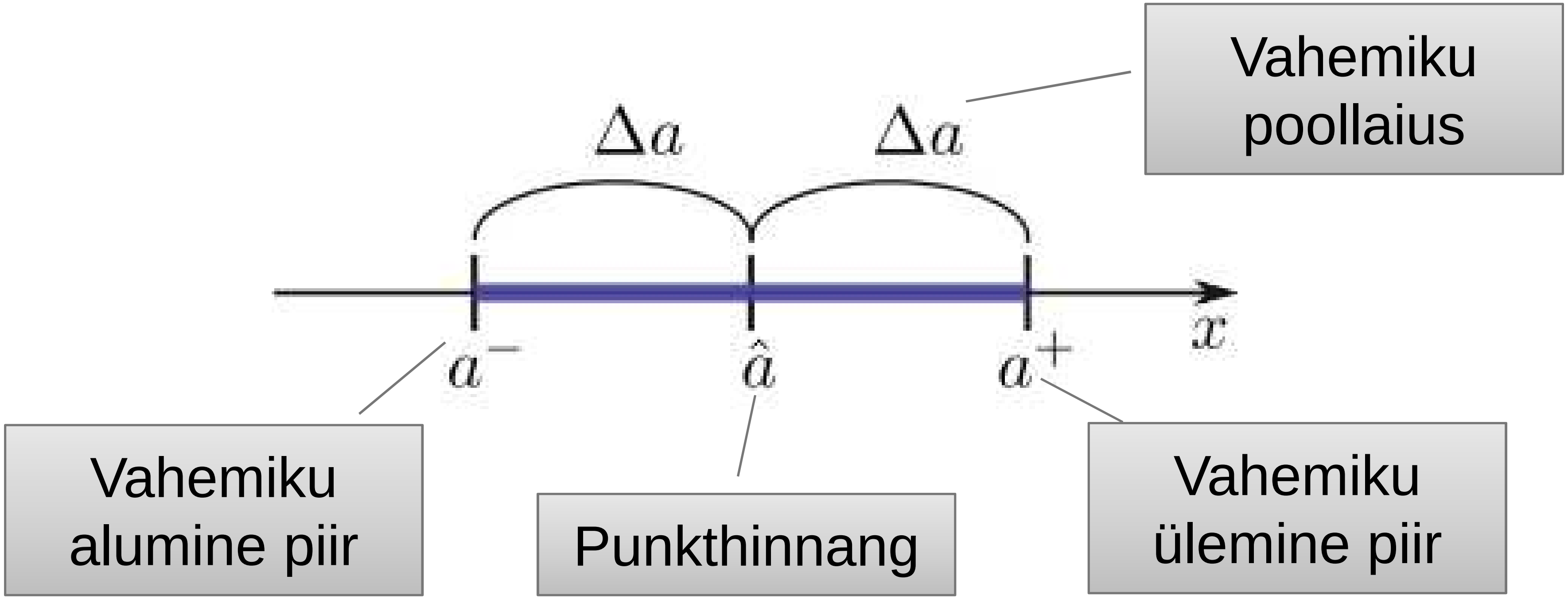
Hinnangute keskväärtus $E[\hat{a}] = 996,4$

Tegelik väärtus $a = 942,8$

Hinnangu nihe $b = 53,6$

VAHEMIKHINNANG

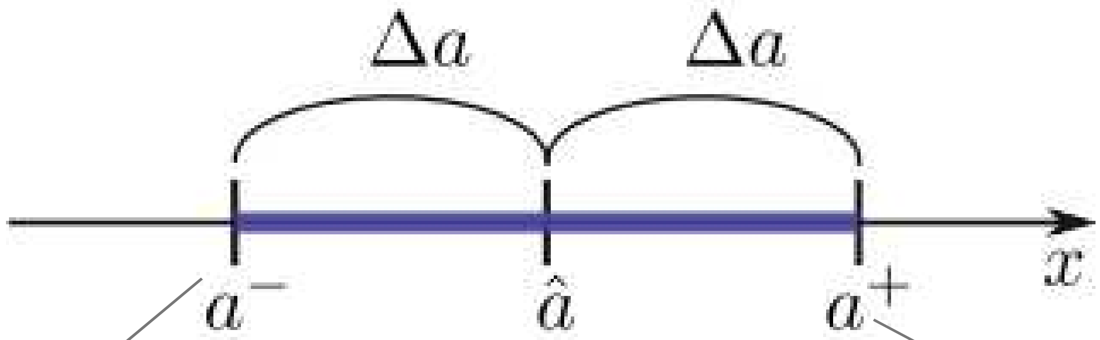
Vahemikhinnang on valimi põhjal määratud vahemik, mis katab parameetri tegeliku väärtuse etteantud (küllalt suure) tõenäosusega.



USALDUSVAHEMIK

Parameetri *a* **usaldusvahemikuks** usaldatavusega *β* nimetatakse vahemikku, mis katab parameetri *a* väärtuse tõenäosusega *β*:

$$\beta = P(|\hat{a} - a| < \Delta a)$$



Alumine usalduspiir

Ülemine usalduspiir

ÜLDKOGUMI KESKVÄÄRTUSE PUNKTHINNANG

Üldkogumi keskväärtuse μ punkthinnanguks on valimi aritmeetiline keskmine:

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

Summeerimine üle valimisse kuuluvate objektide

n valimi maht

ÜLDKOGUMI DISPERSIOONI PUNKTHINNANG

Üldkogumi dispersiooni σ^2 punkthinnang

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

Summeerimine üle valimisse kuuluvate objektide

See on **valimi dispersioon**

MIKS NIMETAJAS *n* -1?

Kui kasutame valemit $$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

saame nihkega hinnangu. Hinnangu keskväärtus

$$E\left[S^2\right] = \frac{n-1}{n}\sigma^2$$ Tõestus õpikus lisa A.7

Nihketa hinnangu jaoks

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

VALIMI KESKMISE VALIMJAOTUS

Eeldused

- lihtne juhuvalik
- suured valimid

Valimite keskmised alluvad normaaljaotusele

Ei sõltu sellest, millisele jaotusele allub vaadeldav tunnus kogumis

Demo: keskmiste jaotus

TSENTRAALNE PIIRTEOREEM

Küllalt suure valimi mahu *n* korral alluvad valimite

keskmised $\overline{X}$ normaaljaotusele keskväärtusega μ ja

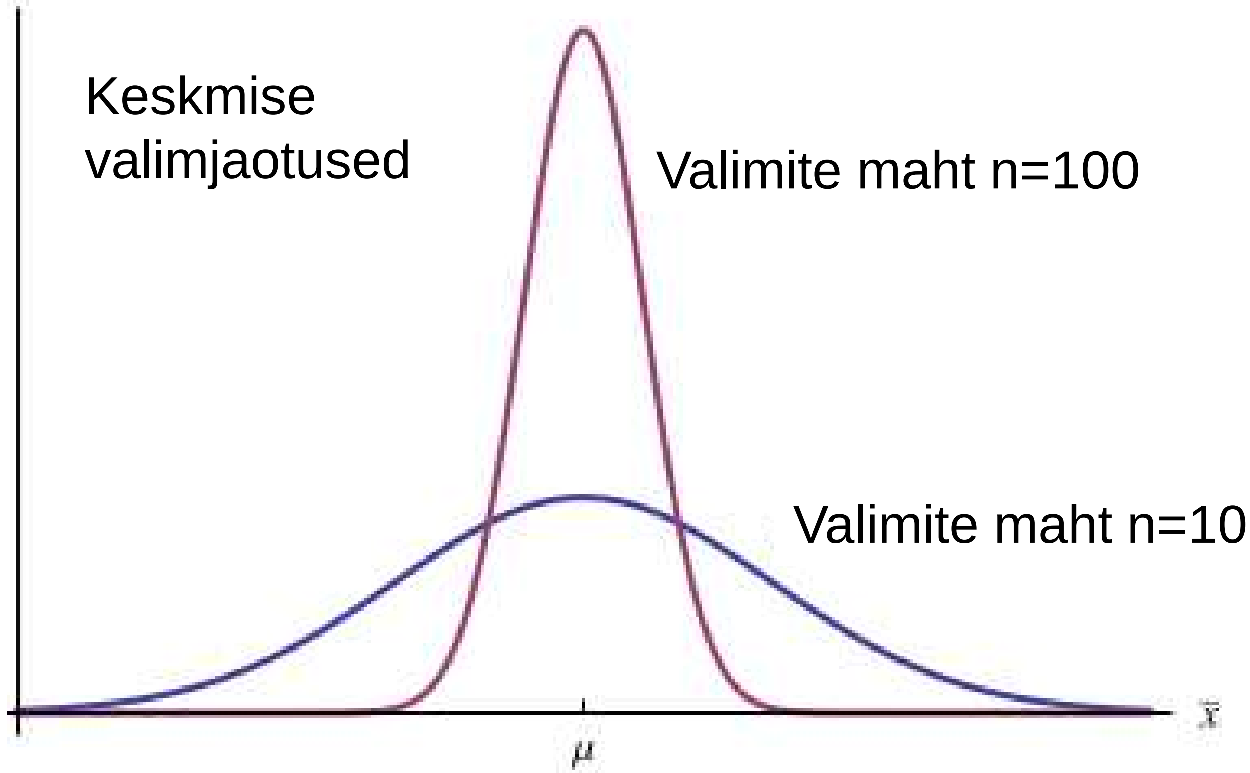
standardhälbega $\frac{\sigma}{\sqrt{n}}$, kus σ on kogumi standardhälve.

$$\overline{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

VALIMJAOTUSE STANDARDHÄLVE

σ iseloomustab üksikväärtuste hajumist.

$\dfrac{\sigma}{\sqrt{n}}$ iseloomustab valimite (maht *n*) keskmiste hajuvust, see on valimi keskmiste valimjaotuse standardhälve



VALIMITE KESKMISTE HAJUVUS

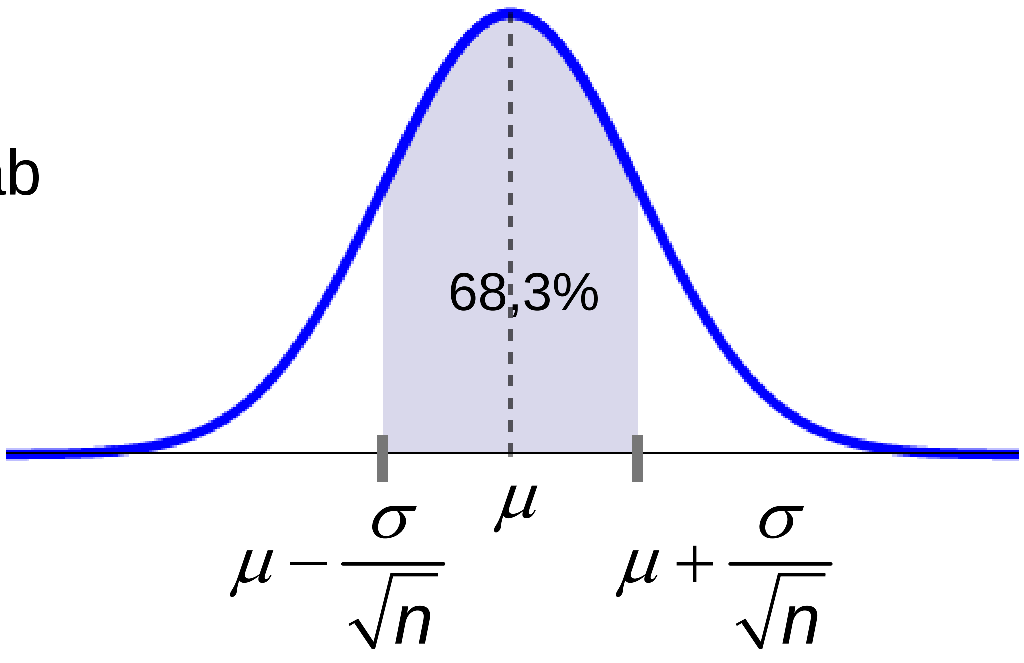
Normaaljaotuse korral jääb vahemikku

- keskväärtus ± 1 standardhälve 68,3%
- keskväärtus ± 2 standardhälvet 95,4%
- keskväärtus ± 3 standardhälvet 99,7%

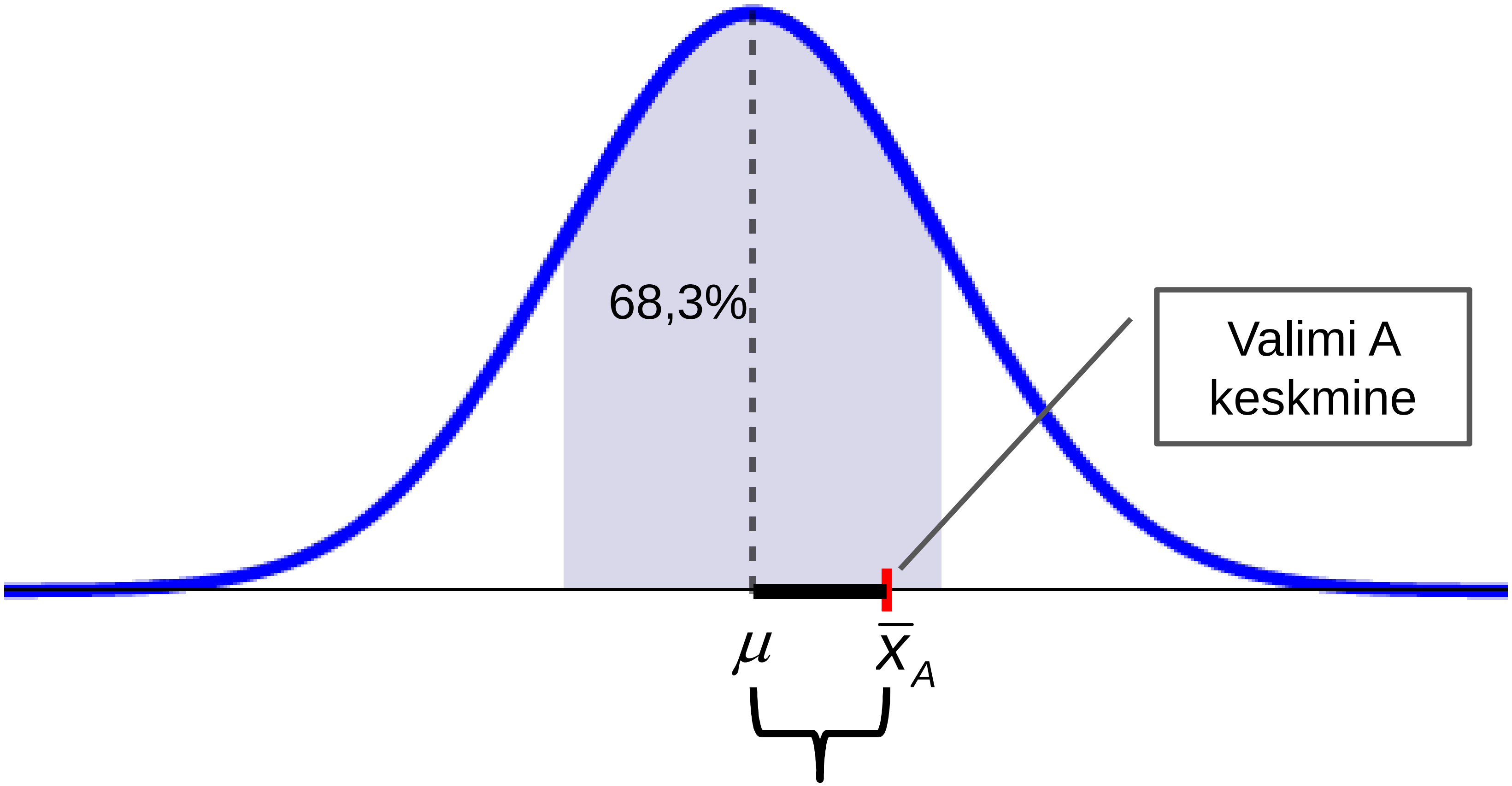
kõikidest väärtustest

Tõenäosusega 68,3% jääb
ühe konkreetse valimi
keskmine vahemikku

$$\mu \pm \frac{\sigma}{\sqrt{n}}$$



ÜHE VALIMI KESKMINE



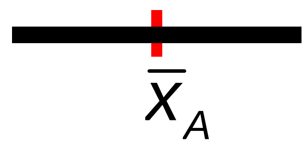
Kogumi keskväärtuse μ kaugus valimi keskmisest $\bar{x}_A - \mu$

Tõenäosusega 68,3% $$\left|\bar{x}_A - \mu\right| < \frac{\sigma}{\sqrt{n}}$$

VALIMI KESKMINE TEADA

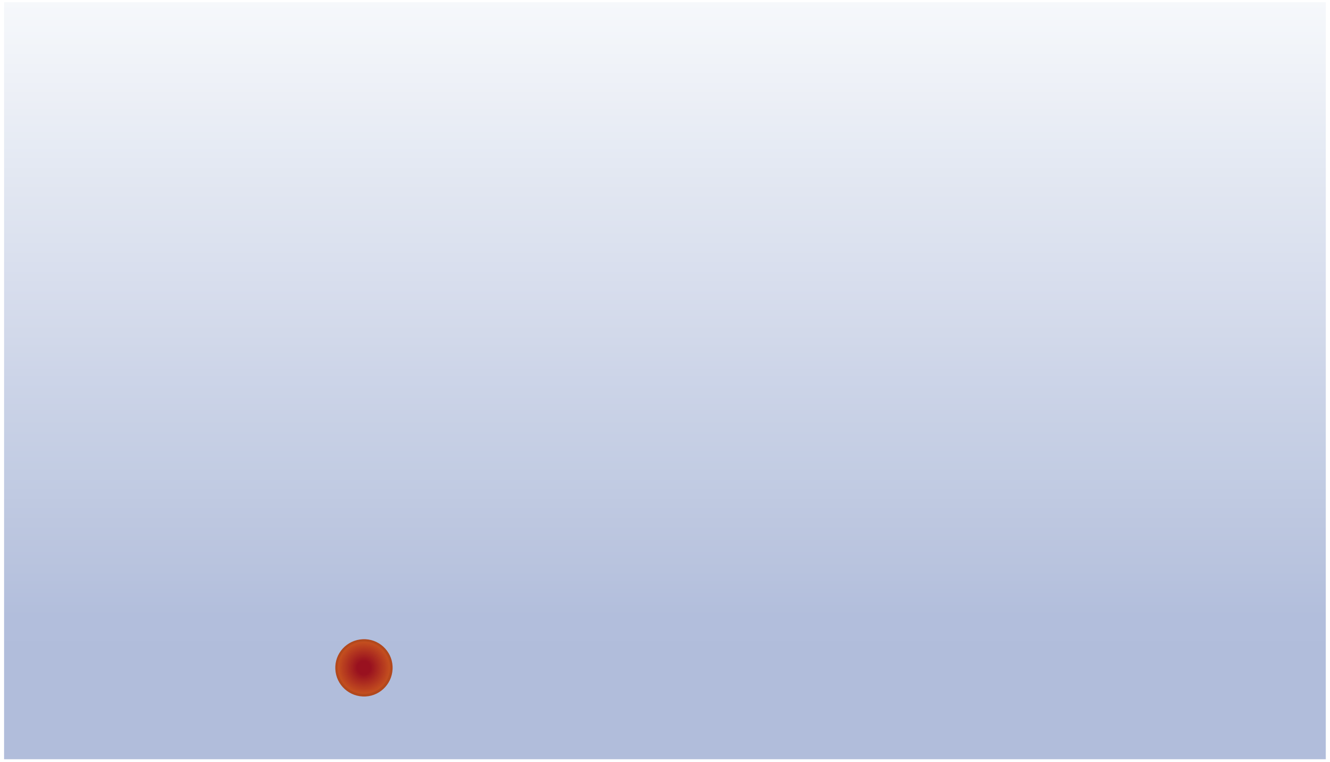
Valimvaatlusest valimi keskmine $\bar{x}_A$ teada

Kus on μ?



Tõenäosusega 68,3% $$\left|\bar{x}_A - \mu\right| < \frac{\sigma}{\sqrt{n}}$$

Kui me teame, kuidas mahakukkunud õunad hajuvad ümber puu,



siis leides ühe õuna, teame, kui kaugelt puud otsida.

USALDUSVAHEMIK

Tõenäosusega 68,3% $$\bar{x} - \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + \frac{\sigma}{\sqrt{n}}$$

$$\mu = \bar{x} \pm \frac{\sigma}{\sqrt{n}}$$ usaldatavusega β=68,3%

TÕENÄOSUSKORDAJA

$$\mu = \bar{x} \pm \frac{\sigma}{\sqrt{n}}$$ usaldatavusega β=68,3%

$$\mu = \bar{x} \pm 2 \cdot \frac{\sigma}{\sqrt{n}}$$ usaldatavusega β=95,4%

$$\mu = \bar{x} \pm 3 \cdot \frac{\sigma}{\sqrt{n}}$$ usaldatavusega β=99,7%

$$\mu = \bar{x} \pm k_\beta \cdot \frac{\sigma}{\sqrt{n}}$$ k_β on **tõenäosuskordaja**

TÕENÄOSUSKORDAJA LEIDMINE

k_β =1, β=68,3%
k_β =2, β=95,4%
k_β =3, β=99,7%

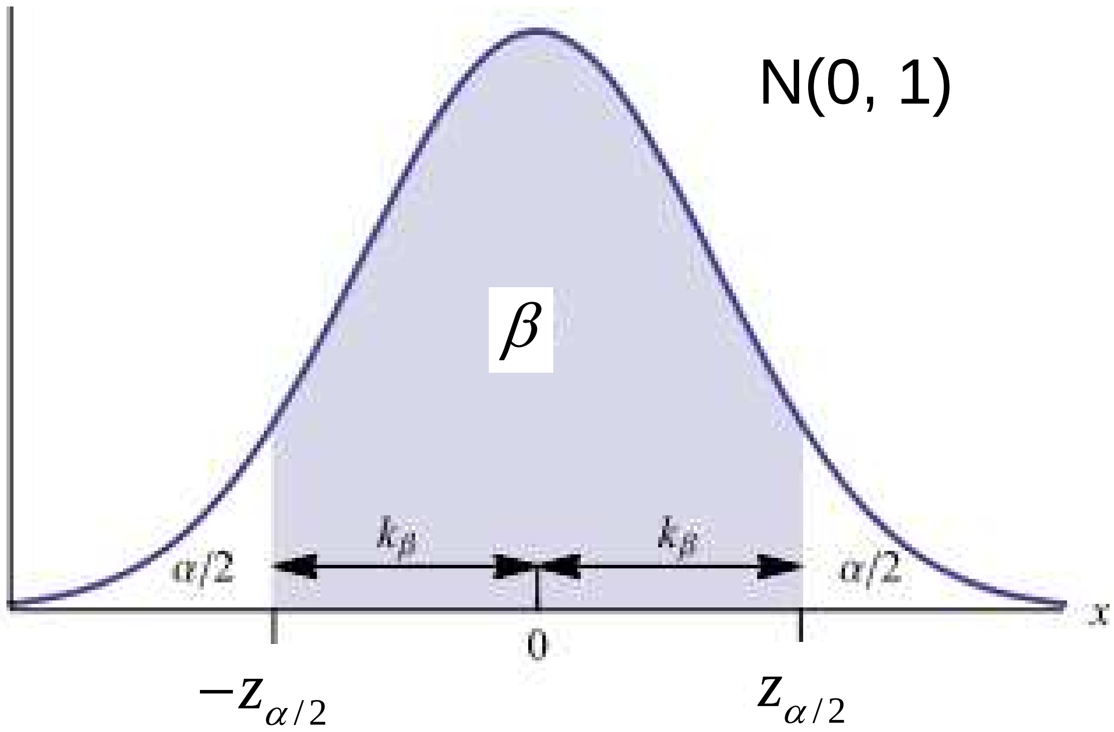
k_β täpselt, β ligikaudu

Aga kui β täpselt?
Näiteks 90%?

k_β on standardiseeritud normaaljaotuse täiendkvantiil

$$k_\beta = z_{\alpha/2}$$

$$\alpha = 1 - \beta$$



TÕENÄOSUSKORDAJA VÄÄRTUSED

- keskväärtus ± 1 standardhälve 68,3%
- keskväärtus ± 2 standardhälvet 95,4%
- keskväärtus ± 3 standardhälvet 99,7%

Täpsed väärtused

Ligikaudsed väärtused

Ligikaudsed väärtused

Täpsed väärtused

- keskväärtus ± 1,15 standardhälvet 75%
- keskväärtus ± 1,64 standardhälvet 90%
- keskväärtus ± 1,96 standardhälvet 95%

k_β β

STANDARDVIGA

Praktikas pole meil üldkogumi standardhälbe tegelik väärtus σ teada ning kasutatakse selle hinnangut, valimi standardhälvet *s*.

$$\frac{\sigma}{\sqrt{n}} \quad \rightarrow \quad \boxed{se = \frac{s}{\sqrt{n}}} \quad \text{standardviga}$$

kus valimi standardhälve

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

STANDARDHÄLVE JA STANDARDVIGA

Valimi standardhälve

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

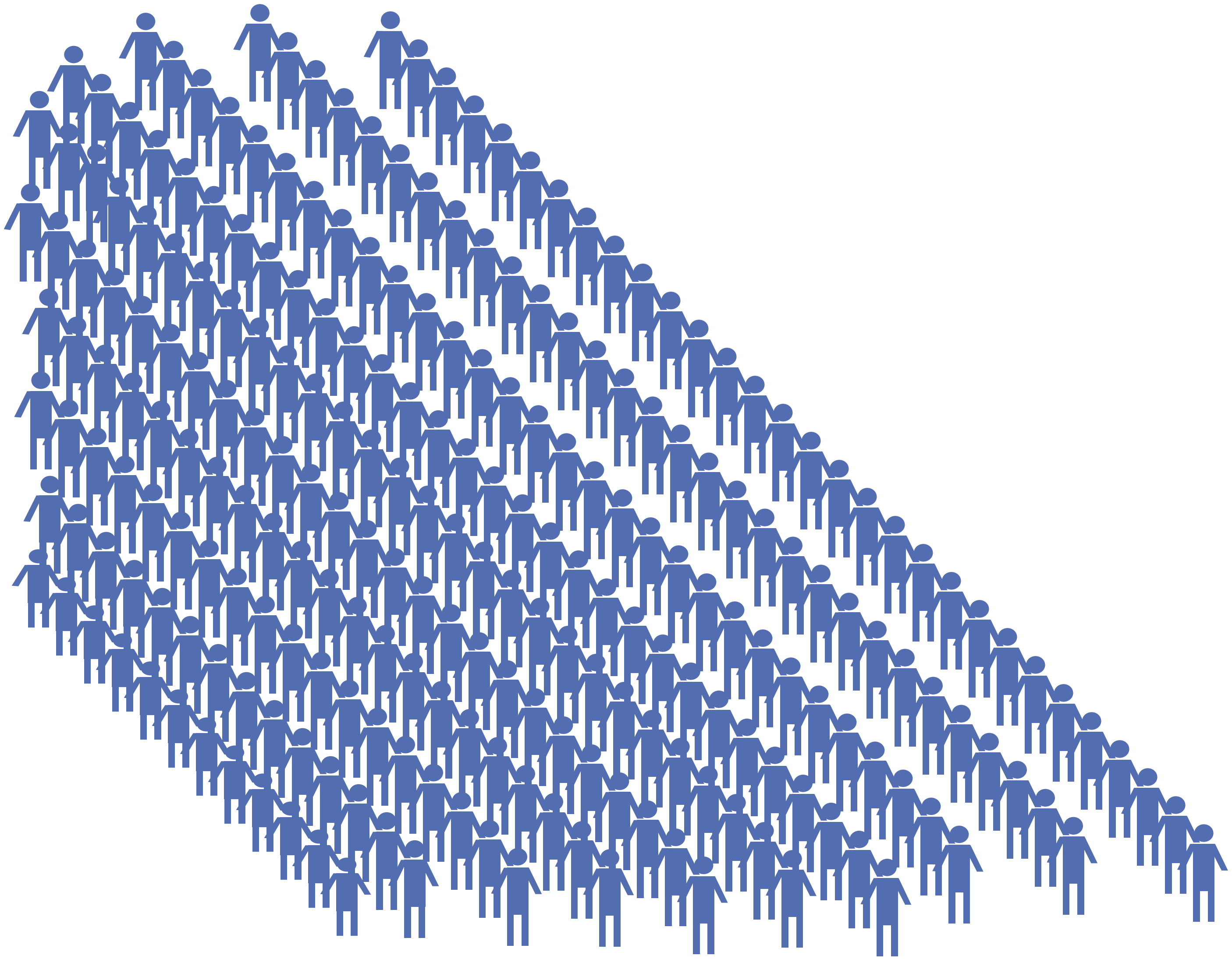
Kogumi standardhälbe hinnang. Iseloomustab üksikute objektide hajumist

Standardviga

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

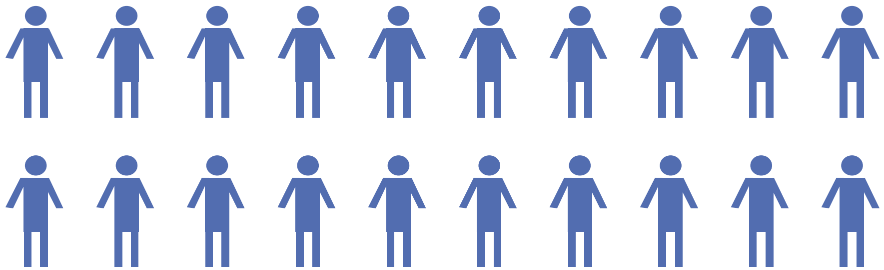
Keskväärtuse valimjaotuse standardhälbe hinnang. Iseloomustab valimite keskmiste hajumist

LÕPMATULT SUUR KOGUM



Valimi maht n=3
Kogumi maht
enne N=200.
Valimi võtmisel
väheneb 197-ni,
so 1,5%.

LÕPLIK KOGUM



Valimi maht n=3.
Kogumi maht enne *N*=20.
Valimi võtmisel väheneb 17-ni,
so 15%.

LÕPLIK JA LÕPMATU KOGUM

Tsentraalne piirteoreem kehtib, kui kogumi maht $N \to \infty$.

Siis valimjaotuse standardhälbe hinnang $se = \dfrac{s}{\sqrt{n}}$

Kui kogumi maht N on lõplik ja väheneb valimi tegemisel

siis $$se = \frac{s}{\sqrt{n}}\sqrt{1-\frac{n}{N}}$$

Kasutada siis, kui $$\frac{n}{N} > 5\%$$

KOGUMI KESKVÄÄRTUSE USALDUSPIIRID

Suure (n>30) valimi korral on üldkogumi keskväärtuse usalduspiirid usaldatavusega β

$$\boxed{\begin{array}{l} \bar{x} \pm \Delta x \\ \Delta x = z_{\alpha/2} \dfrac{s}{\sqrt{n}} \end{array}}$$

$\bar{x}$ valimi keskmine

s valimi standardhälve

n valimi maht

$z_{\alpha/2}$ standardiseeritud normaaljaotuse täiendkvantiil, α=1-β

Demo: usalduspiirid

KOGUMI KESKVÄÄRTUSE USALDUSPIIRID LÕPLIKU KOGUMI KORRAL

Lõpliku kogumi mahu *N* korral

$$\bar{x} \pm \Delta x$$

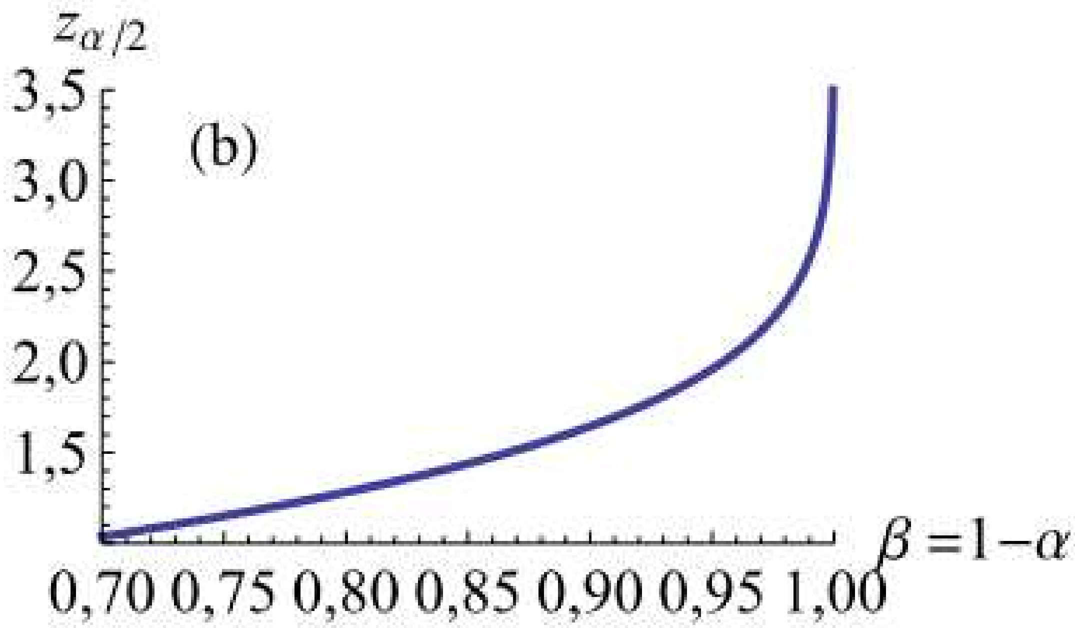
$$\Delta x = z_{\alpha/2} \frac{s}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

USALDATAVUSE VALIK

Kõige sagedamini kasutatakse usaldatavust **0,95.**
Mõnikord ka 0,90 või 0,99.

β	0,9	0,95	0,99
$z_{\alpha/2}$	1,64	1,96	2,59

Tõenäosuskordaja sõltuvus usaldatavusest

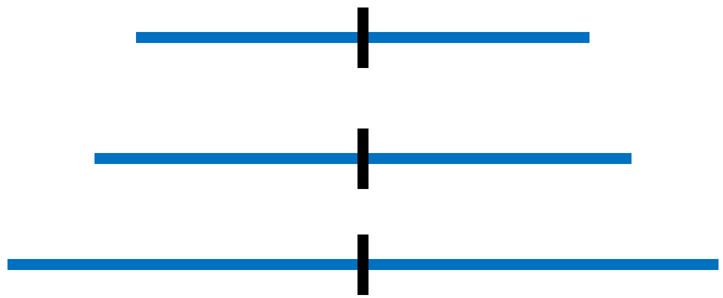


Demo: usaldatavus

USALDATAVUS JA USALDUSVAHEMIKU LAIUS

Ühe ja sama valimi korral
suurem usaldatavus = laiem usaldusvahemik (suurem määramatus)

Usaldatavus *β*	Kordaja
0,90	1,64
0,95	1,96
0,99	2,58

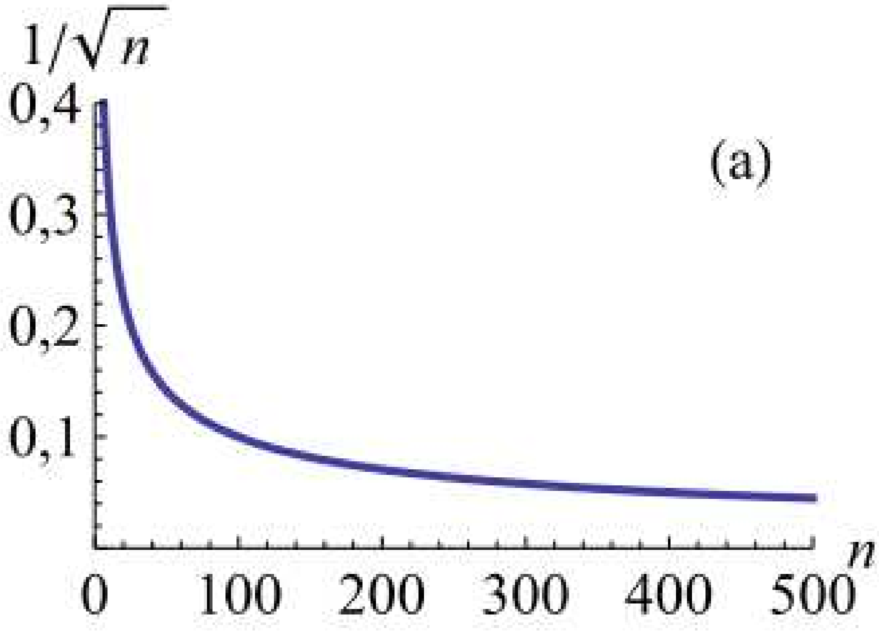


SÕLTUVUS VALIMI MAHUST

$$\Delta x = z_{\alpha/2} \frac{s}{\sqrt{n}}$$

$$\alpha = 1 - \beta$$

Muuta saame usaldatavust ja valimi mahtu



MILLEST SÕLTUB USALDUSVAHEMIKU POOLLAIUS

standardhälvet muuta ei saa

$$z_{(1-\beta)/2} \frac{s}{\sqrt{n}}$$

usaldatavust saame valida

valimi mahtu saame valida enne uuringut

NÄIDE: KULUD TOIDULE I

Leibkonna eelarve uuring 2012.

Valimi maht $n = 9080$

Valimi keskmine $\overline{x} = 915{,}30$ € leibkonnaliikme kohta aastas

Valimi standardhälve $s = 541{,}72$ €

Usaldatavuseks võtame $\beta = 0{,}95$

Vastav tõenäosuskordaja $z_{0{,}025} = 1{,}96$

$$\Delta x = z_{\alpha/2} \frac{s}{\sqrt{n}} = 1{,}96 \cdot \frac{541{,}72}{\sqrt{9080}} \approx 11{,}1$$

NÄIDE: KULUD TOIDULE II

Usaldatavusega 0,95 oli keskmine kulu toiduainetele ja mittealkohoolsetele jookidele

$$915{,}3 \pm 11{,}1$$

eurot aastas pereliikme kohta.

Teine võimalus usaldusvahemiku esitamiseks:

(904,2, 926,4) eurot aastas

USALDUSPIIRID: ESITUSVIIS 1

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

Esitatud on keskväärtus ja usaldusvahemiku poollaius

Millistesse piiridesse jääb: vaja arvutada

Näiteks

25 ± 2 usaldatavusega 0,95

$25 \pm 2_{\ 0,95}$

NB! Alati tuleb märkida usaldatavus!

USALDUSPIIRID: ESITUSVIIS 2

alumine piir $\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}$

ülemine piir $\bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}$

Esitatud on alumine ja ülemine usalduspiir

Keskväärtus ja usaldusvahemiku laius vaja arvutada

Näiteks

Vahemik (23, 27) usaldatavusega 0,95