VALIK-UURINGUD II

LOENGUTEEMAD

- Väikesed valimid ja *t*-jaotus
- Valimi mahu planeerimine
- Kaheväärtuselise tunnuse usaldusvahemik
- Kolme ja enama väärtusega kvalitatiivse tunnuse usalduspiirid
- Mediaani usaldusvahemik
- Valimi kaalumine
- Vea komponendid

KESKVÄÄRTUSE USALDUSVAHEMIK SUURE VALIMI KORRAL

Üldkogumi keskväärtuse usaldusvahemik

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

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

valimi standardhälve

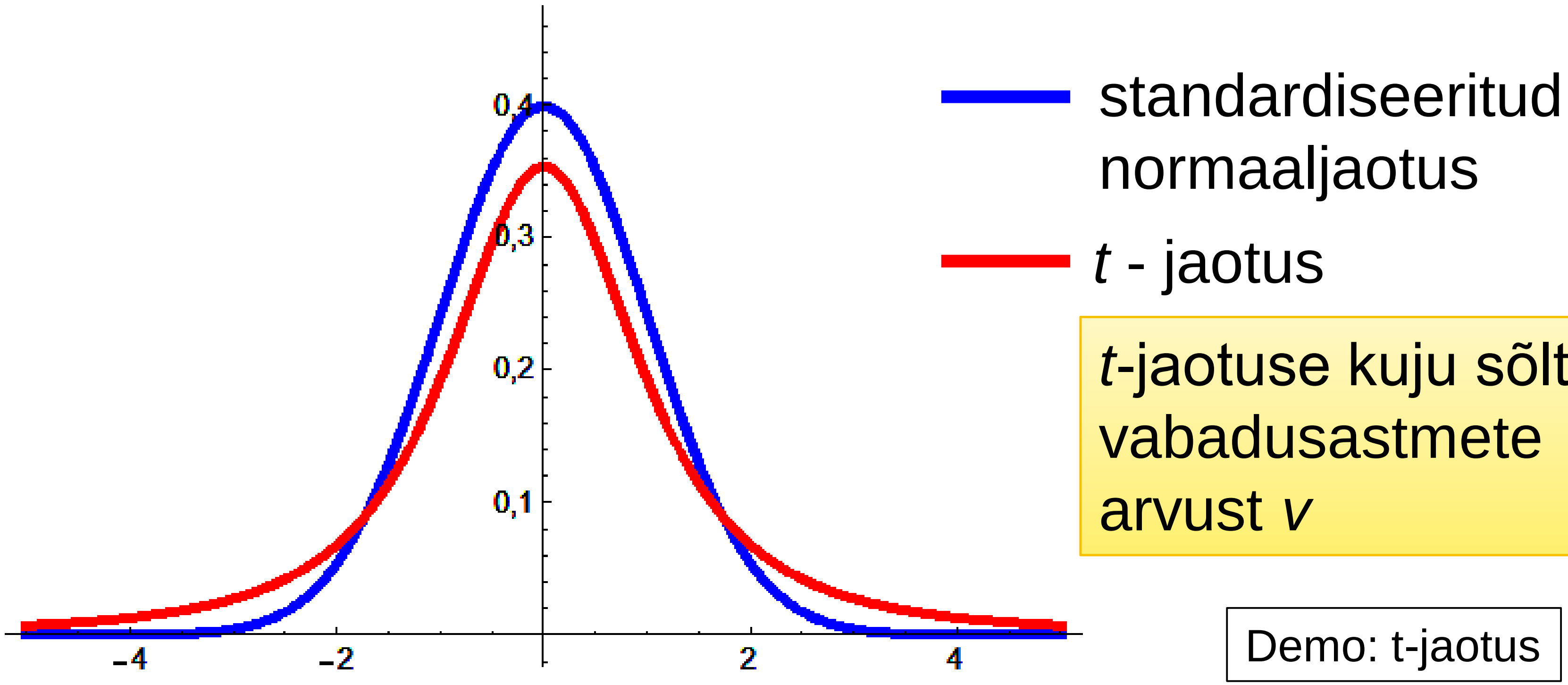
valimi maht

Standardiseeritud normaaljaotuse täiendkvantiil

VÄIKESED VALIMID
t-JAOTUS

Väikeste valimite korral valimite keskväärtuste jaotus erineb normaaljaotusest.

Kasutatakse *t*-jaotust ehk Studenti jaotust



t-jaotuse kuju sõltub vabadusastmete arvust *v*

Demo: t-jaotus

t-JAOTUS EHK STUDENTI JAOTUS

Miks Studenti jaotus?

William Gosset oli õppinud Oxfordis matemaatikat ja keemiat ning läks tööle Guinnessi pruulikotta. Seal rakendati tema statistikaalaseid teadmisi parima saagikusega odra valikul.

Katsetati erinevaid odrasorte, valimid olid väikesed. Gosset uuris väikeste valimite keskmiste jaotust. Oma uurimistulemusi ei tohtinud ta oma nime all avaldada tööandja keelu tõttu.

Avaldas pseudonüümi Student all.

William Sealy Gosset
(1876-1937)

VABADUSASTMETE ARV, I

Vabadusastmete arv on sõltumatute muutujate arv.

NÄIDE

Aktsiaportfellis on kolme ettevõtte aktsiaid. Portfelli struktuur:

ettevõtte A aktsiaid	20%
ettevõtte B aktsiaid	30%
ettevõtte C aktsiaid	?

On 3 muutujat. Aga sõltumatuid on ainult 2. Kui 2 on vabalt valitud, siis kolmas on määratud, seda ei saa enam vabalt valida.

Järelikult vabadusastmete arv on 2.

Sest esineb üks kitsendav tingimus: kolme osakaalu summa on 100%.

VABADUSASTMETE ARV, II

Summas $\sum_{i=1}^{n}(x_i - \bar{x})^2$ on sõltumatuid liidetavaid n -1

Miks? Sest kehtib tingimus

$$\sum_{i=1}^{n}(x_i - \bar{x}) = (x_1 - \bar{x}) + (x_2 - \bar{x}) + \ldots + (x_n - \bar{x}) = 0$$

Järelikult valimi standardhälbe leidmisel vabadusastmete arv

$$\boxed{\nu = n - 1}$$

USALDUSVAHEMIKU POOLLAIUS

Väikese valimi korral üldkogumi keskväärtuse usalduspiiride poollaius

$$\boxed{\Delta x = t_{\alpha/2}(\nu)\frac{s}{\sqrt{n}}}$$

$\bar{x}$ valimi keskmine

s valimi standardhälve

n valimi maht

$t_{\alpha/2}(\nu)$ $t(v)$-jaotuse täiendkvantiil, sõltub usaldatavusest β ($\alpha/2=1/2(1-\beta)$) ja vabadusastmete arvust $v = n - 1$.

VÕRDLUS STANDARDISEERITUD NORMAALJAOTUSEGA

Usaldatavus 0,95

Normaaljaotusele vastav kordaja 1,96

Valimi maht *n*	$t_{0,025}(v)$	Suhteline erinevus
3	4,30	120%
5	2,78	42%
10	2,26	15%
20	2,09	7%
30	2,05	4%
50	2,01	3%
100	1,98	1,2%
200	1,97	0,6%
500	1,965	0,2%
1000	1,962	0,1%
5000	1,9604	0,02%

Erinevus alla 5 %

Suureks loetakse valimid, kus $n > 30$. Siis võib rakendada normaaljaotust.

SUUR VÕI VÄIKE VALIM

Kogumi keskväärtuse usaldusvahemik

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

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

$$\Delta x = t_{\alpha/2}(\nu) \frac{s}{\sqrt{n}}$$

Suur valim (*n*>30)

Standardiseeritud normaaljaotuse täiendkvantiil

Väike valim (*n*<30)

t(*v*)-jaotuse täiendkvantiil

VALIMI MAHU PLANEERIMINE

NÄIDE: VALIMI MAHU PLANEERIMINE I

Viie juhuslikult valitud poe põhjal keskmise läbimüügi usaldusvahemik

$$42{,}6 \pm 32{,}5 \text{ tk nädalas.}$$

Valimi standardhälve *s* = 26,14, usaldatavus 0,95: tõenäosuskordaja *t*-jaotusest 2,78.

Suhteline viga on

$$\frac{32{,}5}{42{,}6} \approx 76\%$$

Soovime, et suhteline viga ei oleks suurem kui 40%:

$$\frac{\Delta x}{42{,}6} \leq 40\%$$

Kui suur peab olema valimi maht *n*?

NÄIDE: VALIMI MAHU PLANEERIMINE, II

$$\frac{\Delta x}{42,6} \leq 40\%$$

Leiame Δx maksimumväärtuse $\Delta x \leq 0,4 \cdot 42,6 = 17,04$

Kasutame Δx valemit $\Delta x = t_{\alpha/2}(\nu)\frac{s}{\sqrt{n}}$

Avaldame *n*

$$t_{\alpha/2}(\nu)\frac{s}{\sqrt{n}} \leq 17,04 \quad \Rightarrow \quad \left(t_{\alpha/2}(\nu)\frac{s}{17,04}\right)^2 \leq n$$

$$\left(2,78 \cdot \frac{26,14}{17,04}\right)^2 \leq n \quad \Rightarrow \quad 18,14 \leq n \quad \Rightarrow \quad \boxed{n = 19}$$

VALIMI MAHU PLANEERIMINE

Algul tuleb teha proovivalim.
Proovivalimi maht n_0 ja standardhälve $s_{0.}$

Kui soovime, et usaldusvahemiku poollaius oleks väiksem kui *d,* siis

$$\Delta x \le d \quad \Rightarrow \quad n \ge \left(t_{\alpha/2}(\nu) \frac{s_0}{d} \right)^2$$

Lisame proovivalimi objektidele veel objekte, kuni valimi maht on *n.*

d võib olla antud ka protsendina keskväärtusest.

KAHEVÄÄRTUSELINE TUNNUS

KAHEVÄÄRTUSELINE TUNNUS

Kas teate …?
Kas tarbite …?
Kas olete …. poolt?

Vastusevariandid kodeerime
"ei" 0
"jah“ 1

Kaheväärtuselise tunnuse keskväärtus on võrdne väärtuse "1" esinemise suhtelise sageduse ehk osakaaluga.

$$\bar{x} = p = \frac{m}{n}$$

n valimi maht
m ühtede arv

Standardhälve $$\sigma = \sqrt{p(1-p)}$$

KAHEVÄÄRTUSELISE TUNNUSE USALDUSPIIRID

Osakaal kogumis p (teadmata)

Osakaal valimis $\hat{p}$

Kehtib tsentraalne piirteoreem, p nihketa hinnang on $\hat{p}$

Osakaalu valimjaotuse standardhälve (standardviga):

$$s = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Usaldusvahemik

$$\boxed{\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

NÄIDE: VALIMISAKTIIVSUS

2019.a jaan küsitlus enne Riigikogu valimisi (Turu-uuringute AS). Valimi maht 1000. Valima lubas minna 80% küsitletutest.

Võtame usaldatavuseks β=0,95, kordaja $z_{\alpha/2}$=1,96

$$\Delta p = 1{,}96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1{,}96\sqrt{\frac{0{,}8\cdot(1-0{,}8)}{1000}} \approx 0{,}025$$

Usaldusvahemik $80{,}0\% \pm 2{,}5\%$ $(77{,}5\%;\ 82{,}5\%)_{0,95}$

Võtame usaldatavuseks β=0,99, kordaja $z_{\alpha/2}$=2,58

$$\Delta p = 2{,}58\sqrt{\frac{0{,}8\cdot(1-0{,}8)}{1000}} \approx 0{,}033$$

Usaldusvahemik $80{,}0\% \pm 3{,}3\%$ $(76{,}7\%;\ 83{,}3\%)_{0,99}$

Tegelikult osales 63,1%.

NÄIDE

Demo: osakaalu usalduspiirid

Mingi küsitlus. Valim maht 500.
Usaldatavuseks valime 0,95.

„Jah“ vastanuid 10% $\Delta p = 1{,}96\sqrt{\frac{0{,}1 \cdot 0{,}9}{500}} \approx 0{,}026$

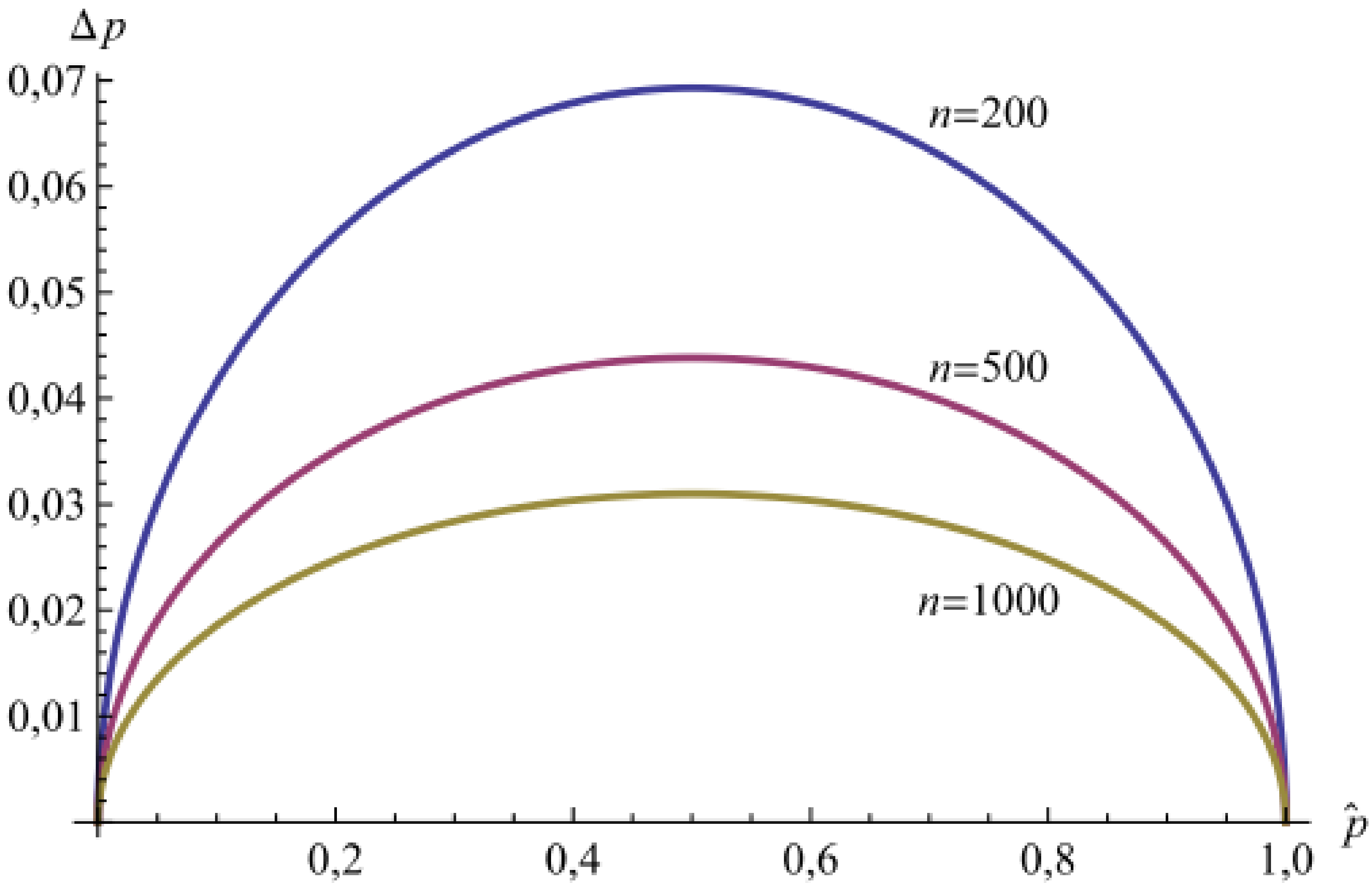
Usaldusvahemik $10{,}0\% \pm 2{,}6\%_{0,95}$

„Jah“ vastanuid 50% $\Delta p = 1{,}96\sqrt{\frac{0{,}5 \cdot 0{,}5}{500}} \approx 0{,}044$

Usaldusvahemik $50{,}0\% \pm 4{,}4\%_{0,95}$

Määramatus on suurem.

SÕLTUVUS OSAKAALUST VALIMIS



NÄIDE: VARIANDI "JAH" VALIJATE ARV VÄIKE

Pettuseriskide alane uuring Eestis 2014. Valimi maht 112.

Küsimus: "Kas järgnev tegevus tundub Teile õigustatud, et majandussurutises toime tulla?„:
"Rahalised maksed (majandus)tegevuse jätkamiseks ja elavdamiseks„
Vastusevariandid „jah, „ei“.

"Jah" valis 2 vastajat. Osakaal $\frac{2}{112} \approx 0,0179 = 1,79\%$

Usaldusvahemiku poollaius

$$\Delta p = 1,96 \cdot \sqrt{\frac{0,0179 \cdot (1-0,0179)}{112}} \approx 0,025 = 2,5\%$$

Alumine piir $p^{-} = 1,79\% - 2,5\% = -0,71\%$

PROBLEEM!

KORRIGEERITUD USALDUSPIIRID

Korrigeerime

"Jah" valinute arv $m \to m+2$

Valimi maht $n \to n+4$

Nelja lisamise reegel

Korrigeeritud osakaal $\tilde{p} = \dfrac{m+2}{n+4}$

Korrigeeritud usalduspiirid

$$\tilde{p} \pm \Delta\tilde{p},$$

$$\Delta\tilde{p} = z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$$

Demo: osakaalu usalduspiirid, korrigeerimine

SUURE VALIMI TINGIMUS

Valemit $\hat{p} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

võib kasutada juhul, kui kumbagi väärtust omavate elementide arv on valimis ≥ 5, st kehtib tingimus

$$\boxed{n \cdot \min(\hat{p}, 1-\hat{p}) \geq 5}$$

Kui see tingimus pole täidetud, tuleb kasutada korrigeeritud usalduspiire (nelja lisamise reegel).

NÄIDE: VARIANDI "JAH" VALIJATE ARV VÄIKE

Pettuseriskide alane uuring Eestis 2014
Valimi maht 112

"Jah" valis 2 vastajat.

Korrigeeritud osakaal $\tilde{p} = \frac{m+2}{n+4} = \frac{2+2}{112+4} \approx 0,0345 = 3,45\%$

Usaldusvahemiku poollaius, korrigeeritud

$$\Delta\tilde{p} = z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} = 1,96 \cdot \sqrt{\frac{0,0345 \cdot (1-0,0345)}{112+4}} \approx 0,0332 = 3,32\%$$

Seda, et rahalised maksed (majandus)tegevuse jätkamiseks ja elavdamiseks on majandussurutise korral õigustatud, arvab

$3,45\% \pm 3,32\%$ juhtidest

KOLME JA ENAMA VÄÄRTUSEGA KVALITATIIVSE TUNNUSE OSAKAALU USALDUSPIIRID

ERAKONDADE REITINGUD 2019-2023

Uuringufirma Norstat

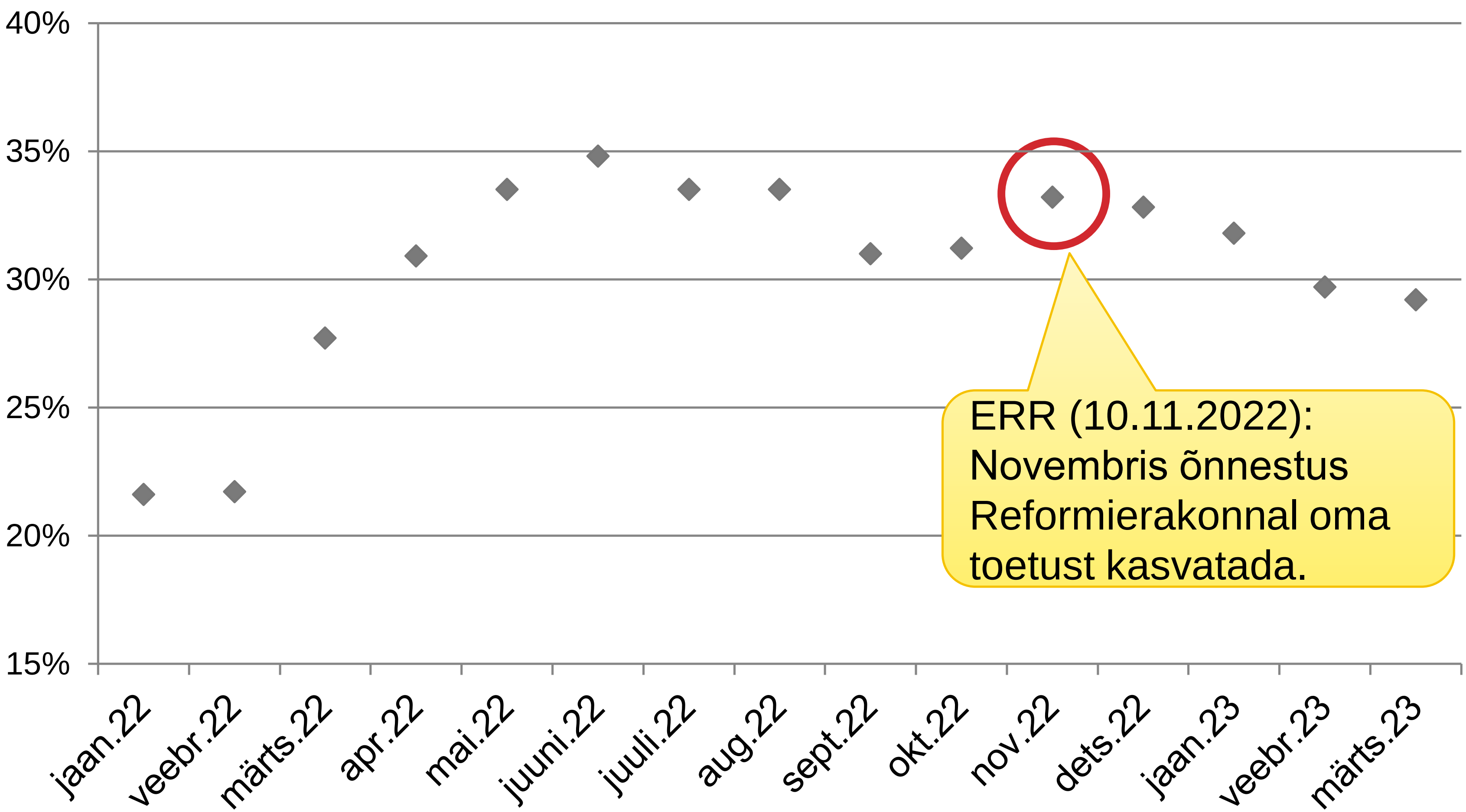
Valimi maht 4000 isikut

Valikuvariante 8

Keskerakond
Reformierakond
Eesti Konservatiivne Rahvaerakond
Sotsiaaldemokraatlik Erakond
Isamaa
Eesti 200
Parempoolsed
Eestimaa Rohelised

NÄIDE: TOETUS REFORMIREAKONNALE 2022-2023

Allikas: https://reitingud.ee/reitingud/



Demo: viie valiku valimjaotus

VARIANDI *i* OSAKAALU USALDUSVAHEMIKU POOLLAIUS

Iga valikuvariandi *i* jaoks

$$\hat{p}_i \pm \Delta p_i$$

$$\Delta p_i = \sqrt{\chi^2_{\alpha/k}(1)}\sqrt{\frac{\hat{p}_i(1-\hat{p}_i)}{n}}$$

$\hat{p}_i$ *i*-nda variandi valinute osakaal valimis

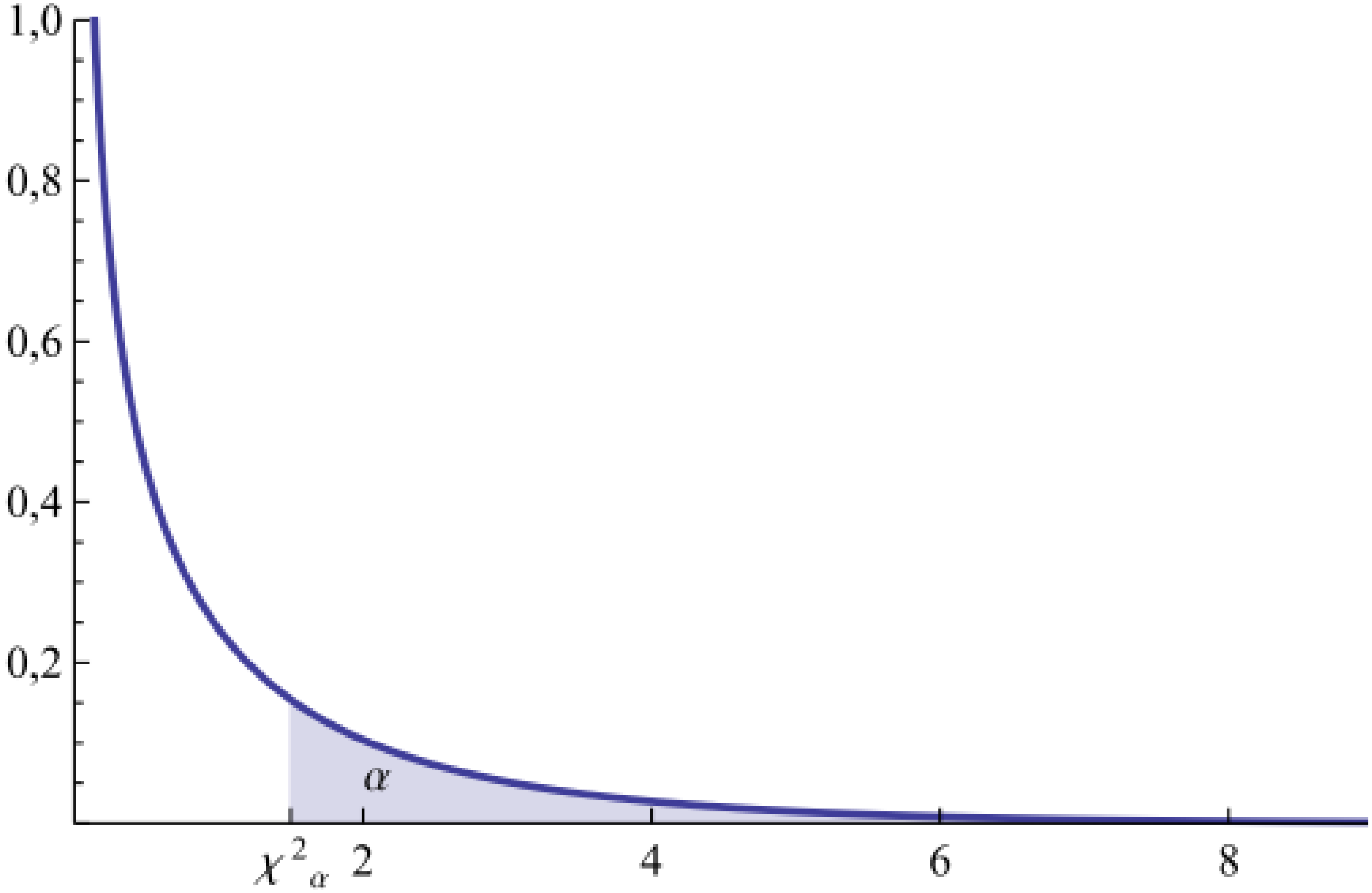
n valimi maht

$\chi^2_{\alpha/k}(1)$ χ^2 (1) jaotuse täiendkvantiil

α vea tõenäosus, $\alpha = 1 - \beta$

k valikute arv

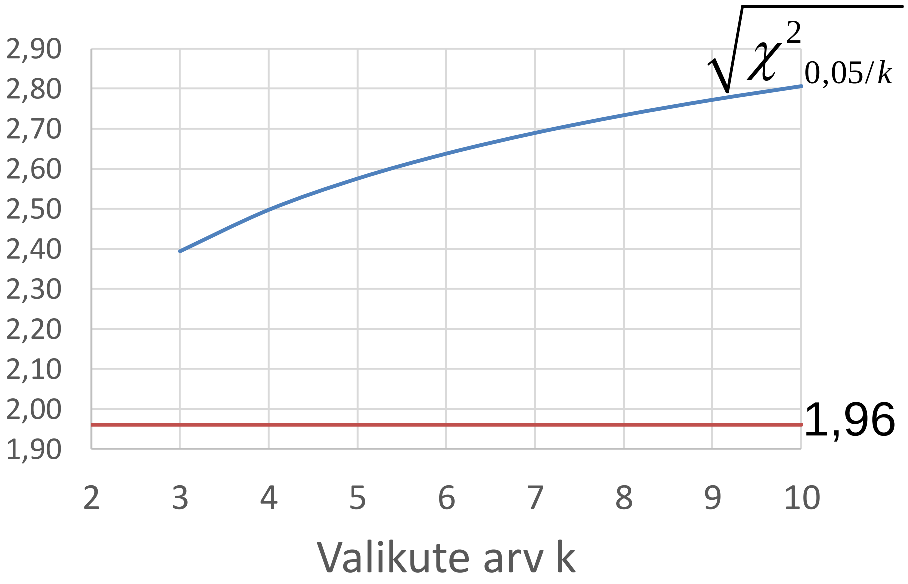
χ^2 - JAOTUS



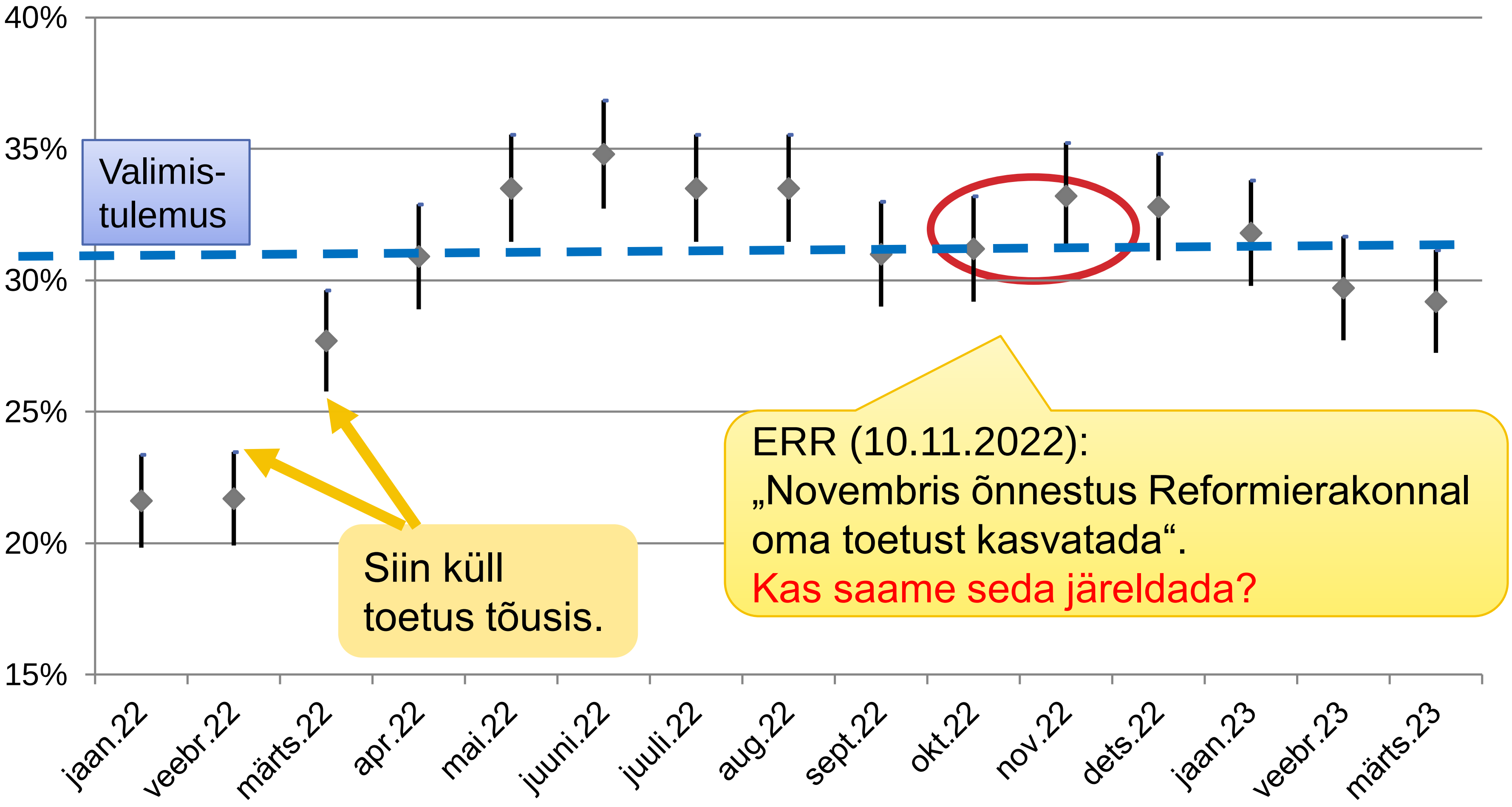
χ^2 – JAOTUSE TÄIENDKVANTIILID

	β=0,95 $z_{0,025}$=1,96	β=0,9 $z_{0,05}$=1,64
k	$\sqrt{\chi^2_{0,05/k}}$	$\sqrt{\chi^2_{0,1/k}}$
3	2,39	2,13
4	2,50	2,24
5	2,58	2,33
6	2,64	2,39
7	2,69	2,45
8	2,73	2,50
9	2,77	2,54
10	2,81	2,58

Tõenäosuskordaja kahe valiku ning rohkem kui kahe valiku puhul.
Usaldatavus 0,95



NÄIDE: TOETUS REFORMIERAKONNALE 2022-2023



NÄIDE: ROHELISTE TOETAJAID VALIMITES

Valimi maht 4000

	okt. 23	nov. 23	dets. 23	jaan. 23	veebr. 23
Osakaal	1,4%	1,1%	2,2%	1,9%	1,5%
n_i	56	44	88	76	60

n_i on väike, alla 100

Ka siin on väikese n_i korral vaja teist lähenemist.

Allikas: uuringufirma Norstat

TÄPNE VALEM

Kasutada juhul, kui $n_i < 100$

osakaalu alumine piir $$p_i^- = \frac{\chi^2 + 2n_i - \sqrt{D}}{2(n + \chi^2)},$$

osakaalu ülemine piir $$p_i^+ = \frac{\chi^2 + 2n_i + \sqrt{D}}{2(n + \chi^2)}$$

$$D = \chi^2 \left(\chi^2 + 4(n - n_i)\frac{n_i}{n} \right)$$

n_i *i*-nda variandi valinute arv valimis
n valimi maht

$$\chi^2 = \chi^2_{\alpha/k}(1)$$ $\chi^2(1)$ jaotuse täiendkvantiil

NÄITEID KVALITATIIVSETEST TUNNUSTEST, KUS 3 JA ENAM VÄÄRTUST

Nimiskaalas

- **Sotsiaalmajanduslik seisund:**
  - töötav,
  - töötu,
  - õpilane,
  - pensionär.
- **Eluruumi tüüp:**
  - eramaja,
  - mitmepereelamu,
  - korter.

Järjestusskaalas

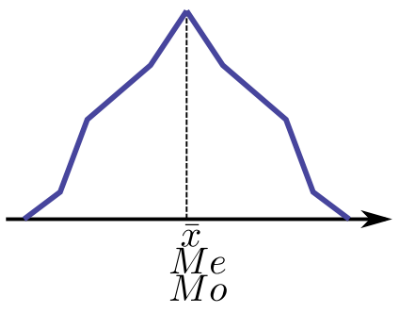
- **Haridustase:**
  - I taseme haridus;
  - II taseme haridus;
  - III taseme haridus.
- **Toimetulek:**
  - suurte raskustega;
  - mõningate raskustega;
  - tuleb toime.

Valida võib ainult ühe variandi!

MEDIAANI USALDUS-VAHEMIK

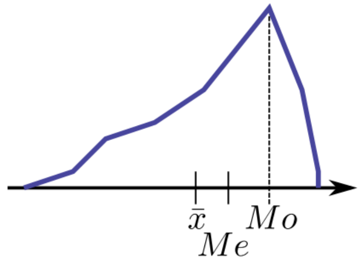
SÜMMEETRILINE JA ASÜMMEETRILINE JAOTUS

- Kui juhusliku suuruse jaotus on sümmeetriline, langevad keskväärtus ja mediaan kokku.
- Mediaani punkthinnanguks on siis valimi aritmeetiline keskmine.
- Mediaani usalduspiirideks keskväärtuse usalduspiirid.
- Järelikult tuleb leida keskväärtuse usaldusvahemik.



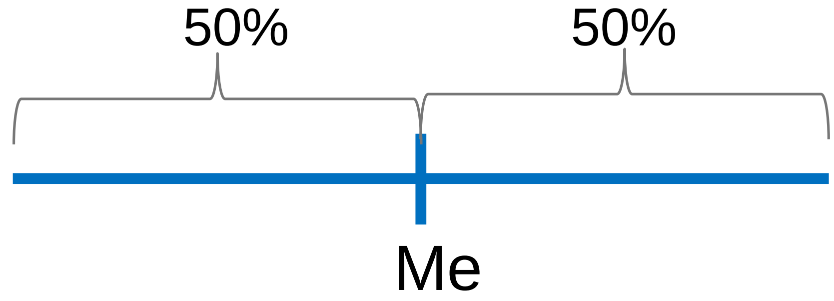
Asümmeetrilise jaotuse korral see ei kehti.

Järgnevalt Thomson-Savuri protseduur, mida saab kasutada suvalise jaotuse korral.



ÜLDKOGUMI MEDIAAN JA VALIM

Üldkogum



Teeme juhuvalimi mahuga n

Tunnuse väärtus x_i valimis sisalduva objekti i jaoks

tõenäosusega 0,5 x_i < Me Selliseid objekte n^- tk

tõenäosusega 0,5 x_i > Me

Objektide arv n^- on juhuslik suurus, mis allub binoomjaotusele B(n, 0,5).

THOMSON-SAVURI PROTSEDUUR, 1

Juhuvalim 35, 57, 36, 30, 57, 58, 60, 59, 63. Valimi maht $n = 9$.

Järjestame valimi elemendid

30	35	36	57	57	58	59	60	63

Leiame mediaani usalduspiirid usaldatavusega 0,95.

Selleks otsime usaldusvahemikku, nii et vahemikust välja jäämise tõenäosus

$$\alpha \leq 0,05$$

THOMSON-SAVURI PROTSEDUUR, 2

Alumine piir

30	35	36	57	57	58	59	60	63

Tõenäosus, et kõik 9 arvu on mediaanist suuremad, st 0 arvu on mediaanist väiksemad

$$P(n^{-}=0)=0,002$$

Arvutamiseks kasutame binoomjaotust
n=9, *p*=0,5

THOMSON-SAVURI PROTSEDUUR, 3

Alumine piir

Ülemine piir

30	35	36	57	57	58	59	60	63

Tõenäosus, et kõik 9 arvu on mediaanist väiksemad

$$P(n^{-}=0)=0,002$$

$$P(n^{-}=9)=0,002$$

$$\alpha = 0,002+0,002=0,004$$

Kehtib tingimus $\alpha < 0,05$ Liigume edasi

THOMSON-SAVURI PROTSEDUUR, 4

Alumine piir

30	35	36	57	57	58	59	60	63

Tõenäosus, et 8 arvu on mediaanist suuremad, st 0 või 1 arvu on mediaanist väiksemad

$$P(n^{-} \leq 1) = 0,02$$

THOMSON-SAVURI PROTSEDUUR, 5

Alumine piir

Ülemine piir

30	35	36	57	57	58	59	60	63

Tõenäosus, et mediaan on siin

Tõenäosus, et mediaan on siin

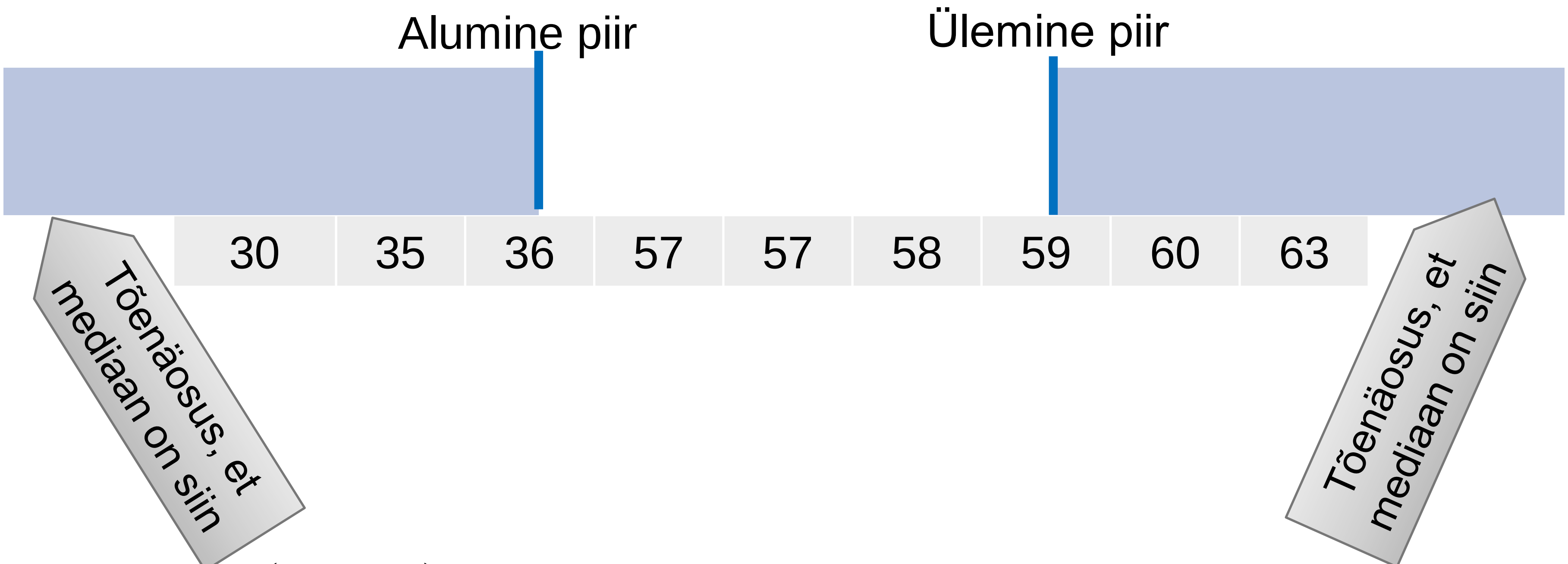
$P(n^{-} \leq 1) = 0,02$

$P(n^{-} \geq 8) = 0,02$

$$\alpha = 0,02 + 0,02 = 0,04$$

Kehtib tingimus $\alpha < 0,05$ Liigume edasi

THOMSON-SAVURI PROTSEDUUR, 6



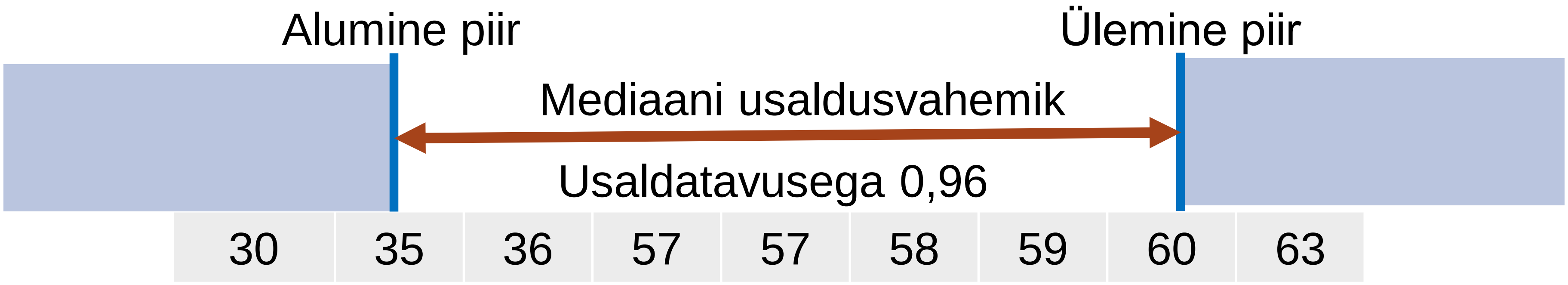
$$P(n^{-} \leq 2) = 0,09$$

$$P(n^{-} \geq 7) = 0,09$$

$$\alpha = 0,09 + 0,09 = 0,18$$

Liiga suur, lähme 1 samm tagasi

THOMSON-SAVURI PROTSEDUUR, 7

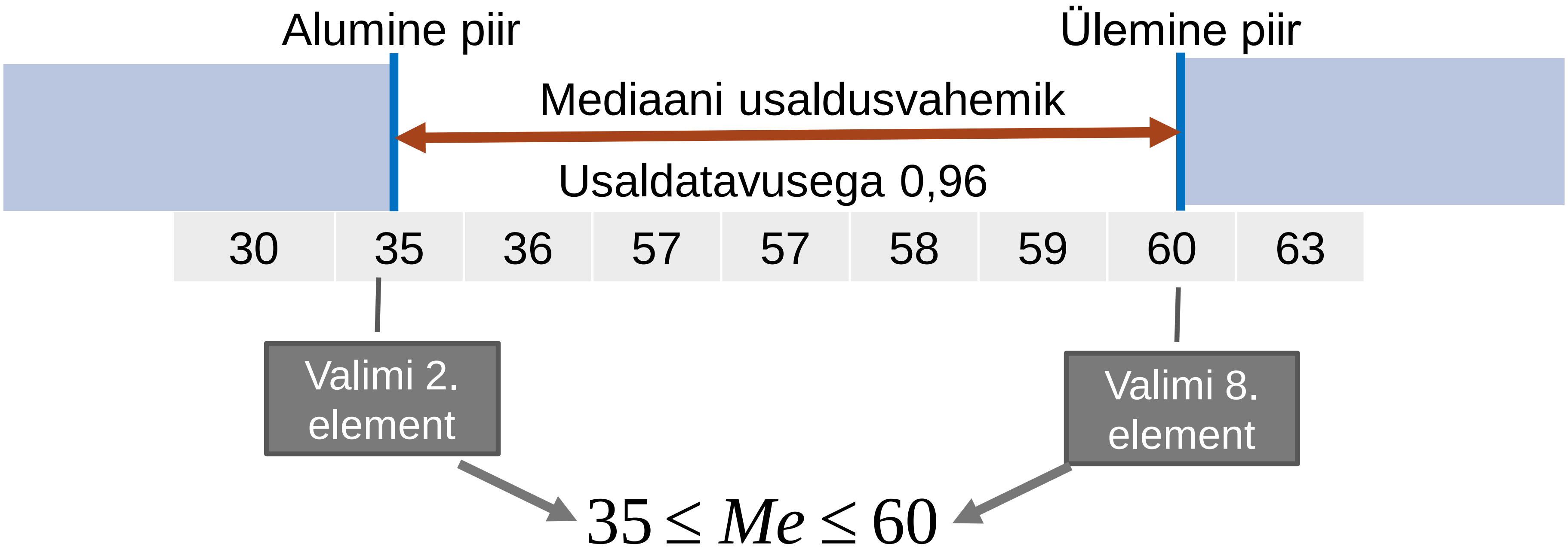


$$P(n^{-} \leq 1) = 0,02$$

$$P(n^{-} \geq 8) = 0,02$$

$$\alpha = 0,02 + 0,02 = 0,04 < 0,05$$ Suuremaks ei saa

THOMSON-SAVURI PROTSEDUUR, 8



Mediaani punkthinnangut asümmeetrilisel juhul leida ei saa.

Demo: mediaani usadusvahemik

MEDIAANI USALDUSVAHEMIKU LEIDMINE, KOKKUVÕTE

Otsime sellist usaldusvahemikku, mis
1) oleks võimalikult kitsas;
2) vahemikust välja jäämise tõenäosus ei ületaks α.

Järjestatud valimist leiame,
1) mitmes valimi element vastab alumisele piirile (järjenumber k);
2) ja mitmes ülemisele piirile (järjenumber n-k+1).

Vastavate elementide väärtused annavad mediaani usaldusvahemiku alumise ja ülemise piiri.

NÄIDE: KULUTUSED TOIDUKAUPADELE, 1

Leibkonnauuring 2012. Valimi maht 9080.
Kulutused toidukaupadele leibkonnaliikme kohta aastas?

Alumise usalduspiiri järjenumber k=4447 (leitud binoomjaotusest).
Ülemise usalduspiiri järjenumber n-k+1=9080-4447+1=4634.

Vastavate järjenumbritega elemendid järjestatud valimis

Järjenumber	Kulud toidule, eurot
…..	……
4447	814,82
…..	……
4634	838,83
…..	…..

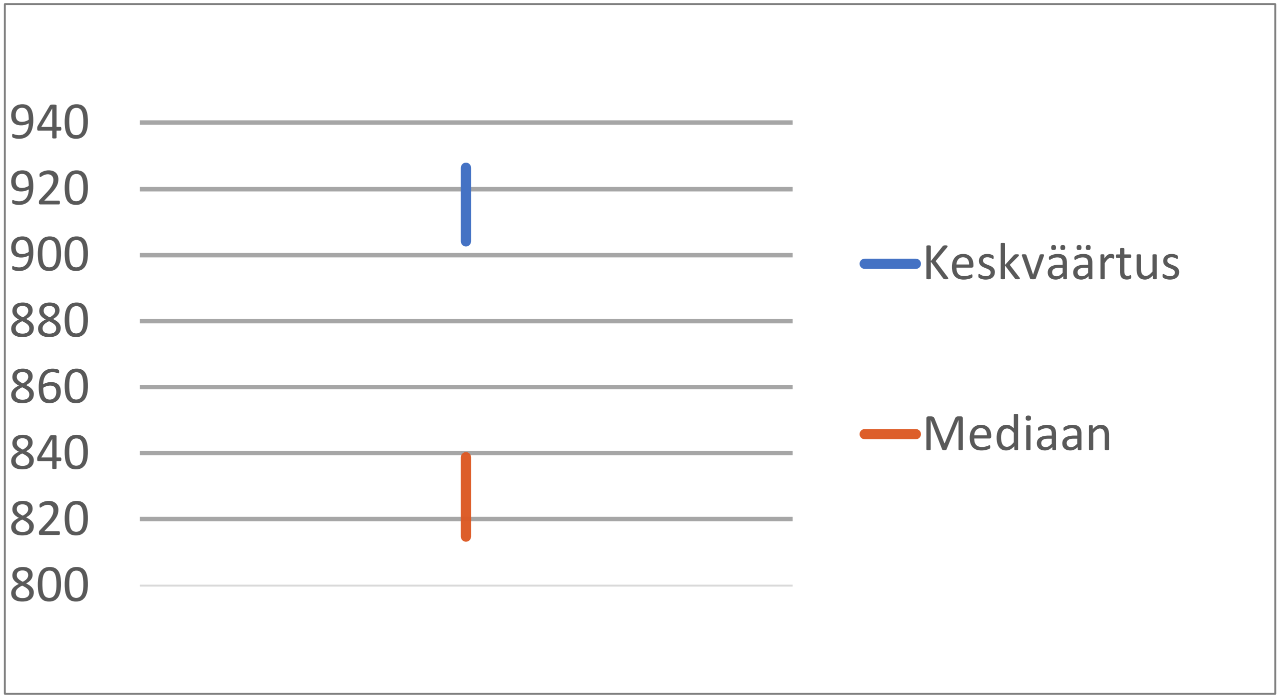
Mediaani usaldusvahemik (814,82; 838,83) eurot pereliikme kohta aastas, usaldatavus 0,95.

NÄIDE: KULUTUSED TOIDUKAUPADELE, 2

Leibkonnauuring 2012. Valimi maht 9080.
Kulutused toidukaupadele leibkonnaliikme kohta aastas?

Mediaani usaldusvahemik (814,82, 838,83) eurot pereliikme kohta aastas, usaldatavus 0,95.

Võrdluseks keskväärtuse usaldusvahemik (904,2, 926,4) eurot aastas.



MEDIAANI USALDUSVAHEMIK SUURTE VALIMITE KORRAL

Suure valimi korral binoomjaotus läheneb normaaljaotusele.
Mediaani usaldusvahemiku piirid suure valimi korral.

Alumine piir: järjestatud valimi element järjenumbriga

$$k = 0,5(n - z_{\alpha/2}\sqrt{n})$$

Ülemine piir: järjestatud valimi element järjenumbriga

$$n - k + 1$$

Näiteks kulud toidule leibkonnauuringust. n=9080

$$k = 0,5 \cdot \left(9080 - 1,96 \cdot \sqrt{9080}\right) \approx 4447$$

MEDIAANI USALDUSPIIRID, KOKKUVÕTE

Sümmeetrilise jaotuse korral kasuta keskväärtuse usalduspiire.

Asümmeetrilise jaotuse korral reasta valimi elemendid kasvavalt. Seejärel leia alumise piiri järjenumber *k*

suur valim: $k = 0{,}5(n - z_{\alpha/2}\sqrt{n})$

väike valim: Thomson-Savuri protseduur.

Ülemise piiri järjenumber $n - k + 1$

Mediaani usalduspiirideks on vastavate järjenumbritega elemendid valimis.

VALIMI KAALUMINE

NÄIDE: VALIMI KAALUMINE I

Ülikooli lõpetanute hulgas viidi läbi küsitlus. Vastanute hulgas oli kolme teaduskonna lõpetanuid.

1. etapp: kaalude leidmine

Teadus-kond	Lõpetanute osakaal kogumis	Osakaal vastanute hulgas
A	50%	30%
B	20%	50%
C	30%	20%

$$\frac{0{,}5}{0{,}3} = 1{,}667, \quad \frac{0{,}2}{0{,}5} = 0{,}4, \quad \frac{0{,}3}{0{,}2} = 1{,}5$$

NÄIDE: VALIMI KAALUMINE II

2. etapp: kaalude omistamine
Igale vastajale omistatakse kaal vastavalt sellele, millise teaduskonna lõpetaja.

Isik	Teaduskond	Kaal
1	A	1,667
2	A	1,667
3	A	1,667
4	B	0,400
5	B	0,400
6	B	0,400
7	B	0,400
8	B	0,400
9	C	1,500
10	C	1,500
		10

NÄIDE: VALIMI KAALUMINE III

3. etapp: vastus küsimusele brutokuupalga kohta, kaalutud keskmise leidmine

Isik	Teaduskond	Vastus	Kaal	Vastus × kaal
1	A	760	1,667	1266,7
2	A	800	1,667	1333,3
3	A	930	1,667	1550,0
4	B	850	0,400	340,0
5	B	1200	0,400	480,0
6	B	1100	0,400	440,0
7	B	950	0,400	380,0
8	B	800	0,400	320,0
9	C	1050	1,500	1575,0
10	C	890	1,500	1335,0
		Kokku	**10**	9020

$$\overline{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{9020}{10} = 902$$

VEA KOMPONENDID

HINNANGU VIGA

$$u = \hat{a} - a$$

$\hat{a}$ valimi põhjal leitud parameetri punkthinnang

a parameetri tegelik väärtus

Tegelikku väärtust me ei tea => vea suurust pole võimalik leida.

Teatud juhtudel on võimalik hinnata vea ülempiiri.

VEA KOMPONENDID

- Valikuviga
- Loendiviga
- Kaoviga
- Objektide asendamise viga
- Mõõtmisviga
- Töötlusviga

VALIKUVIGA

- Põhjustatud valimi kasutamisest.
- Mõõdab hinnangu varieeruvust üle erinevate valimite, mida antud disainiga on võimalik saada.
- Tõenäosusliku valikuuringu korral saab valikuvea ülemmäära hinnata (standardviga, usaldusvahemik).

LOENDIVIGA

Põhjustatud ebakorrektsest loendist

Ülekaetus - loend sisaldab üldkogumisse mittekuuluvaid objekte.

Alakaetus – loend ei sisalda kõiki kogumisse kuuluvaid objekte. Tihti erinevad need loendisse kuuluvatest objektidest, siis saame nihkega hinnangu.

Loendivea suurust pole tavaliselt võimalik hinnata.

NÄIDE: LOENDIVIGA

Leibkonna eelarve uuring 2010

Üldkogum: tavaleibkondades elavad kõik 2010.a 1. jaan seisuga vähemalt 15-aastased Eesti alalised elanikud, v.a pikka aega (vähemalt aasta) institutsioonides viibijad.
Loend: rahvastikuregister
Valim 7803 isikut

Loendiviga	Arv	% valimist	% loendiveast
Küsitletav surnud	85	1,1	21,8
Küsitletav viidud institutsiooni (hooldekodusse, vanglasse)	41	0,5	10,5
Küsitletav viibib välismaal vähemalt aasta	264	3,4	67,7
KOKKU	**390**	**5,0**	**100,0**

KAOVIGA

Mingil põhjusel ei saada andmeid kõigi valimisse sattunud objektide kohta.

- **Objekti kadu**

Nt inimene ei ela märgitud aadressil.

- **Tunnuse väärtuse kadu**

Nt keeldub vastamast, uuritava tunnuse väärtus puudub.

Kadu iseloomustab nii küsitluse korraldust kui ka vastajate hoiakuid.

NÄIDE: KAO PÕHJUSED JA NENDE OSATÄHTSUS

Leibkonna eelarve uuring 2010. Valim 7803 isikut.

Kao põhjus	Arv	Osatähtsus kaos, %
Kontakti puudumine	**1951**	**51,6**
küsitletavat pole kodus	896	23,7
ei ela antud aadressil	508	13,4
…………………………………………………………		
Keeldumine	**1391**	**36,8**
keeldub (kategooriliselt) vastamast	882	23,3
keeldub ajanappuse tõttu	259	6,9
………………………………………………………..		
Muud põhjused	**439**	**11,6**
kõrge vanus, ei ole võimeline uuringus osalema	103	2,7

NÄIDE VASTAMISMÄÄR

Ettevõtete aastastatistika üldkogum, valim ja vastanud

Aasta	Kogum	Valim	Vastanud	Valimi osatähtsus kogumis, %	Vastamis-määr, %
2007	54 958	12494	10 414	22,7	83,4
2008	58 219	12939	10 258	22,2	79,3
2009	59 254	11703	11 703	19,8	77,3

Allikas: Eesti Statistikaamet, www.stat.ee Tabel EM026, Ettevõtete aastastatistika üldkogum, valim ja vastanud

KAO KOMPENSEERIMINE

- **Kaalumine**

Vastanute tunnuste väärtused korrutatakse läbi eelnevalt leitud kaaludega.

- **Puuduvate väärtuste asendamine (imputeerimine)**

Kasutatakse sarnaste tunnustega objekte (isikuid).

MÕÕTMISVIGA

- **Mõõtmisvahendi viga**
  - Aja mõõtmisel kella viga
  - Küsimustiku korral: valesti sõnastatud küsimus
- **Mõõtmisolukorra viga**
  - Eri isikute küsitlemisel võib mõõtmisolukord olla erinev ja see võib avaldada mõju tulemusele
- **Intervjueerija viga**
  - Küsimus esitatakse valesti
  - Vastus märgitakse üles valesti

TÖÖTLUSVIGA

Tekivad andmete kodeerimisel, sisestamisel, analüüsimisel

Vähendamiseks:

- sisestamisel väärtusvahemike kontroll;
- eraldi kodeerida tunnuse väärtuse puudumine.

NÄIDE: UURING TÄISKASVANUTE RASVUMISEST

Postimees 24. märts 2014:

Kolmandik täiskasvanud eestlastest on rasvunud

Aastal 2012 Tervise Arengu instituudi uuring: rasvunud on 19% täiskasvanutest. Valimi suurus 5000.

2014. a TÜ-s kaitstud doktoritöö, autor Triin Eglit: rasvunud on 32% täiskasvanutest. Valimi suurus 495.

Miks nii suur erinevus?

NÄIDE: UURING TÄISKASVANUTE RASVUMISEST

Mõlemad valimid juhuvalimid.

Tervise Arengu instituudi uuring: küsitlus postituse teel. Paluti leida ja kirja panna oma kehamassiindeks.

Võimalikud vead.

1. Ülekaalulised on vähem altid sellele vastama.
   Kaoviga
2. Inimesed ise kalduvad näitama väiksemat kehamassiindeksit, kui see tegelikult on.
   Mõõtmisviga (tahtlik)
3. Vanemad inimesed ei pruugi oma kaalu ja pikkust täpselt teada.
   Mõõtmisviga

Triin Eglit külastas inimesi ise ja mõõtis nende kehamassiindeksit.

???

11.05.-31.07. 2020 viidi Eestis läbi koroonaviiruse antikehade seroepidemioloogiline uuring KOROSERO-EST-1. Valimisse kaasati juhuvaliku alusel isikud kahest perearstikeskusest, kokku 3608 isikut. Uuringus keeldus osalemast 1650 valimisse sattunut, so 46% kutsututest.

Allikas https://www.ctm.ee/et/covid-19/koroonaviirus-sars-cov-2-antikehade-serolevimusuuring-eestis-kahe-piirkonna-testuuring/

Millise veaga on tegemist?

A. Loendiviga
V B. Kaoviga
C. Mõõtmisviga
D. Viga puudub

???

Isikuküsitlusel paluti vastajatel märkida, millisesse vahemikku jäi nende netosissetulek eelmisel kuul. 25 inimest jätsid sellele küsimusele vastamata. Vastuste sisestamisel andmebaasi tekkis nende isikute sissetuleku lahtrisse arv 0.

Millise veaga on tegemist?

A. Mõõtmisviga
V B. Töötlusviga
C. Viga ei ole