

4 Tilastollinen päättely

4.1 Otoskeskiarvo ja luottamusväli

Kommentteja ratkaisuihin:

Keskiarvon luottamusväli voidaan ilmaista kahdella eri tavalla.

Tapa 1: Ilmoitetaan otoskeskiarvo $\bar{x}$ sekä virhemarginaali muodossa $\bar{x} \pm \text{virhemarginaali}$.

Tapa 2: Ilmoitetaan luottamusvälin ala- ja yläraja vähentämällä ja lisäämällä virhemarginaali otoskeskiarvoon. Tällöin luottamusväli ilmaistaan välinä *alaraja – yläraja*. Välin ilmaisemiseen voidaan käyttää myös hakasulkumerkintää: [*alaraja, yläraja*].

Näissä ratkaisuissa luottamusvälit ilmaistaan molemmilla tavoilla. Kumpi hyvänsä ilmaisutapa on yksinään riittävä.

Mikäli perusjoukon keskihajonta σ on tiedossa, luottamusväli muodostetaan näissä ratkaisuissa taulukkokirjoista löytyvän kaavan avulla.

Tässä tilanteessa luottamusväli voidaan muodostaa myös laskin- tai matematiikkaohjelmiston tilastot-toiminnolla (keskiarvon z-väli tai z-estimaatti).

231.

Satunnaismuuttuja $X \sim N(15, 6)$

$$\mu = 15, \sigma = 6$$

a) Kun otoskoko on $n = 9$, otoskeskiarvon $\bar{X}$

- odotusarvo on $\mu = 15$
- keskihajonta (eli keskiarvon keskivirhe) on $\frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{9}} = \frac{6}{3} = 2$

b) Perusjoukon jakauma on normaalijakauma $N(15, 6)$. Myös otoskeskiarvon jakauma on normaalijakauma $N(15, 2)$.

Molemmissa jakaumissa odotusarvo on $\mu = 15$.

Perusjoukon jakauman keskihajonta on $\sigma = 6$ kun taas otoskeskiarvon keskihajonta on 2, eli pienempi. Otoskeskiarvon jakauma on siis kapeampi.

Yhteistä: molemmat jakaumat ovat normaalijakaumia, joissa odotusarvo on 15.

Eroa: perusjoukossa hajonta on suurempaa.

232.

Satunnaismuuttuja $X = \text{"suomalaismiehen pituus (cm)"}.$

$$X \sim N(181, 6)$$

$$\mu = 181, \sigma = 6$$

$$n = 25$$

a) Keskiarvon keskivirhe: $\frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{25}} = \frac{6}{5} = 1,2 \text{ (cm)}.$

b) Otoskeskiarvon

- odotusarvo on $\mu = 181 \text{ (cm)}$
- keskihajonta on $\frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{25}} = 1,2 \text{ (cm)}.$

c) Satunnaismuuttuja $Y = \text{"25 miehen keskipituus (cm)"}.$

$$Y \sim N(181; 1,2)$$

$$\mu = 181, \sigma = 1,2$$

$$P(Y > 180) = 0,7976 \dots \approx 0,80$$

Laskinohjelmiston
normaalijakaumatoimintoon
syötetään: alaraja 180,
yläraja ∞ , $\mu = 181$ ja $\sigma = 1,2$.

d) Lasketaan keskipituuden $Y = 178$ (cm) normitettu arvo.

$$\frac{178 - \mu}{\sigma} = \frac{178 - 181}{1,2} = -2,5.$$

Normitetun arvon itseisarvo on $|-2,5| = 2,5 > 2$, joten arvo poikkeaa merkitsevästi odotusarvosta.

233.

a) Satunnaismuuttuja $X = \text{"hillomunkin paino (g)"}$.

$$X \sim N(80, 8)$$

$$\mu = 80, \sigma = 8$$

Lasketaan todennäköisyys, että $X > 85$.

$$P(Y > 85) = 0,2659 \dots \approx 0,27$$

Laskinohjelmiston
normaalijakaumatoimintoon
syötetään: alaraja 85, yläraja ∞ ,
 $\mu = 80$ ja $\sigma = 8$.

b) Satunnaismuuttuja $Y = \text{"12 munkin keskipaino (g)"}$

Muuttujan Y

- odotusarvo on $\mu = 80$ (g)
- keskihajonta on $\frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{12}} = 2,309 \dots$ (g).

Siis: $Y \sim N(80; 2,309\dots)$

$$\mu = 80, \sigma = 2,309\dots$$

Lasketaan todennäköisyys, että $Y > 85$.

$$P(Y > 85) = 0,01519 \dots \approx 0,015$$

Laskinohjelmiston
normaalijakaumatoimintoon
syötetään: alaraja 85, yläraja ∞ ,
 $\mu = 80$ ja $\sigma = 2,309\dots$

234.

a) Kirjoitetaan mittaustulokset laskentataulukkoon ja määritetään otoskeskiarvo.

- otoskeskiarvo on $\bar{x} = 99$ (mm)

Perusjoukon keskihajonta on $\sigma = 1$ (mm) ja otoskoko on $n = 10$.
Lasketaan keskiarvon keskivirhe.

$$\frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{10}} = 0,3162 \dots (\text{mm})$$

Luottamustasoksi on valittu $C = 0,99$. Tätä vastaava kriittinen arvo on $z^* = 2,58$.

Muodostetaan luottamusväli.

Keskiarvon 99 % :n
luottamusväli:

$$\bar{x} \pm 2,58 \cdot \frac{\sigma}{\sqrt{n}}$$

- Alaraja on $99 - 2,58 \cdot 0,3162 \dots = 98,184 \dots \approx 98,2$ (mm).
- Yläraja on $99 + 2,58 \cdot 0,3162 \dots = 99,815 \dots \approx 99,8$ (mm).

Luottamusväli ruuvien keskipituudelle on siis 98,2 mm – 99,8 mm.

b) Luottamusväli ilmaisee, että ruuvien keskipituus on 99 % varmuudella välillä 98,2 mm – 99,8 mm.

Tehtaan ilmoittama keskipituus 100,0 mm ei sisälly luottamusvälille, joten tämän otoksen perusteella tehtaan ilmoitus on virheellinen.

235.

a) Lasketaan otoskeskiarvo.

$$\bar{x} = \frac{5,8 + 5,7 + 4,8 + 5,9}{4} = \frac{22,2}{4} = 5,55$$

Perusjoukon keskihajonta on $\sigma = 0,5$.

Otoskoko on $n = 4$.

Keskiarvon keskivirhe on

$$\frac{\sigma}{\sqrt{n}} = \frac{0,5}{\sqrt{4}} = 0,25$$

b) Luottamustasoksi on valittu $C = 0,95$. Tätä vastaava kriittinen arvo on $z^* = 1,96$.

Muodostetaan luottamusväli.

- alaraja on $5,55 - 1,96 \cdot 0,25 = 5,06 \approx 5,1$.
- yläraja on $5,55 + 1,96 \cdot 0,25 = 6,04 \approx 6,0$.

Keskiarvon 95 %:n
luottamusväli:

$$\bar{x} \pm 1,96 \cdot \frac{\sigma}{\sqrt{n}}$$

Luottamusväli henkilön kolesteroliarvolle on siis 5,1–6,0.

c) Henkilön todellinen kolesteroliarvo on 95 % varmuudella välillä 5,1–6,0.

Mittausten perusteella tehtyyn johtopäätökseen jää 5 % virhemahdollisuus.

236.

a) Perusjoukon keskihajonta on $\sigma = 2,71$ (kg).

Otoskoko on $n = 20$.

Keskiarvon keskivirhe on

$$\frac{\sigma}{\sqrt{n}} = \frac{2,71}{\sqrt{20}} = 0,6059 \dots$$

Luottamustasoksi on valittu $C = 0,95$. Tätä vastaava kriittinen arvo on $z^* = 1,96$.

Otoskeskiarvo on $\bar{x} = 32,52$ (kg).

Muodostetaan luottamusväli.

Keskiarvon 95 %:n
luottamusväli:

$$\bar{x} \pm 1,96 \cdot \frac{\sigma}{\sqrt{n}}$$

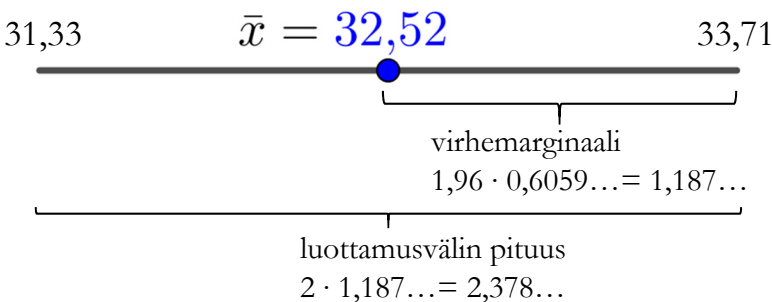
- Alaraja on $32,52 - 1,96 \cdot 0,6059 \dots = 31,332 \dots \approx 31,33$ (kg).
- Yläraja on $32,52 + 1,96 \cdot 0,6059 \dots = 33,707 \dots \approx 33,71$ (kg).

Luottamusväli urosten keskipainolle on siis 31,33 kg – 33,71 kg.

b) Luottamusvälin keskikohta on otoskeskiarvo eli 32,52 kg.

Luottamusvälin pituus on kaksi kertaa virhemarginaali, eli

$$2 \cdot \underbrace{1,96 \cdot 0,6059 \dots}_{\text{virhemarginaali}} = 2 \cdot 1,187 \dots = 2,375 \dots \approx 2,38 \text{ (kg)}$$



c) Luottamusvälin keskikohta on otoskeskiarvo. Uudessa otoksessa otoskeskiarvo voi poiketa b-kohdan otoskeskiarvosta. Kahdesta eri otoksesta muodostetuilla luottamusväleillä ei siis välttämättä ole sama keskikohta.

Luottamusvälin pituus on kaksi kertaa virhemarginaali eli

$$\underbrace{2 \cdot 1,96 \cdot \frac{\sigma}{\sqrt{n}}}_{\text{virhe-
marginaali}}$$

Luottamusvälin pituuteen vaikuttavat siis perusjoukon keskihajonta $\sigma = 2,71$ (kg) sekä otoskoko $n = 20$. Samankokoisten otosten perusteella muodostetut luottamusvälit ovat siis keskenään yhtä pitkiä.

237.

Perusjoukon keskihajonta on $\sigma = 5$.

Otoskeskiarvo on $\bar{x} = 10$.

Valittu luottamustaso on 95 %, jota vastaava kriittinen arvo on $z^* = 1,96$.

a) Otoskoko on $n = 9$.

Keskiarvon keskivirhe on

$$\frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{9}} = \frac{5}{3}.$$

Virhemarginaali on $1,96 \cdot \frac{5}{3} = 3,266... \approx 3,27$.

Luottamusväli on $10 \pm 3,27$ eli $6,73 - 13,27$.

Luottamusvälin

- alaraja on $10 - 1,96 \cdot \frac{5}{3} = 6,733... \approx 6,73$
- yläraja on $10 + 1,96 \cdot \frac{5}{3} = 13,266... \approx 13,27$

b) Otoskoko on $n = 36$.

Keskiarvon keskivirhe on $\frac{5}{\sqrt{36}} = \frac{5}{6}$.

Virhemarginaali on $1,96 \cdot \frac{5}{6} = 1,633... \approx 1,63$.

Luottamusväli on $10 \pm 1,63$ eli $8,37 - 11,63$.

c) Otoskoko on $n = 144$.

Keskiarvon keskivirhe on $\frac{5}{\sqrt{144}} = \frac{5}{12}$.

Virhemarginaali on $1,96 \cdot \frac{5}{12} = 0,816... \approx 0,82$.

Luottamusväli on $10 \pm 0,82$ eli 9,18– 10,82.

d) Luottamusvälin pituus on kaksi kertaa virhemarginaali.

Otoskoko	Luottamusvälin pituus
$n = 9$	$2 \cdot 3,266... = 6,533...$
$n = 36$	$2 \cdot 1,633... = 3,266...$
$n = 144$	$2 \cdot 0,816... = 1,633...$

Voidaan tehdä seuraavat havainnot:

- Otoskoon kasvaessa luottamusväli lyhenee. Siis, mitä suurempi otos, sitä lyhyempi luottamusväli.
- Kun otoskoko nelinkertaistetaan ($36 = 4 \cdot 9$ ja $144 = 4 \cdot 36$), luottamusvälin pituus lyhenee puoleen.

e) Otokoko on $n = 9$, joten keskiarvon keskivirhe on

$$\frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{9}} = \frac{5}{3}.$$

Valittua luottamustasoa 99 % vastaava kriittinen arvo on $z^* = 2,58$.

Virhemarginaali on $2,58 \cdot \frac{5}{3} = 4,3$.

Luottamusväli on $10 \pm 4,3$ eli $5,7 - 14,3$ ja luottamusvälin pituus on $2 \cdot 4,3 = 8,6$.

a-kohdassa muodostetun luottamusvälin pituus on $6,533... \approx 6,5$, eli lyhyempi.

Kun luottamustaso kasvaa, luottamusväli pitenee (levenee).

Huomautus: Vastauksessa c-kohdassa ilmoitettu luottamusväli $10 \pm 4,29$ saadaan, jos kriittisenä arvona käytetään tarkempaa arvoa $z^* = 2,576$.

238.

a) Perusjoukon keskihajontaa ei tunneta, joten muodostetaan keskiarvolle t-väli.

Otoskeskiarvo on $\bar{x} = 182,0$

Otoskeskihajonta on $s = 6,82$

Otoskoko on $n = 34$

Luottamustaso on $C = 0,95$

Valitaan laskimessa tiedonsyöttömenetelmäksi "tilastot".

Syötetään tiedot esimerkiksi laskinohjelmiston luottamusväli-toimintoon.

Luottamusvälin

- alarajaksi saadaan $179,620... \approx 179,6$ (cm)
- ylärajaksi saadaan $184,379... \approx 184,4$ (cm)
- virhemarginaaliksi saadaan $ME = 2,379... \approx 2,4$ (cm).

Nuorten miesten keskipituuden luottamusväli on siis $182,0 \text{ cm} \pm 2,4 \text{ cm}$ eli $179,6 \text{ cm} - 184,4 \text{ cm}$.

b) Vuoden 2007 keskipituus 179 cm ei ole a-kohdassa muodostetulla luottamusvälillä $179,6 \text{ cm} - 184,4 \text{ cm}$.

Tämän perusteella voidaan siis sanoa, että nuorten miesten keskipituus on kasvanut.

Luottamusväli ilmaisee rajat, joiden välissä keskipituus vuonna 2017 on 95 %:n varmuudella.

239.

a) Perusjoukon keskihajontaa ei tunneta, joten muodostetaan keskiarvoille t-väli.

Riskitasoa 5 % vastaa luottamustaso $100 \% - 5 \% = 95 \% = 0,95$.

Otos 1:

Otoskeskiarvo on $\bar{x} = 24,5$

Otoskeskihajonta on $s = 4,2$

Otoskoko on $n = 150$

Luottamustaso on $C = 0,95$

Valitaan laskimessa
tiedonsyöttömenetelmäksi
"tilastot".

Syötetään tiedot esimerkiksi laskinohjelmiston luottamusväli-toimintoon.

Luottamusvälin

- alarajaksi saadaan $23,822... \approx 23,8$ (kertaa)
- ylärajaksi saadaan $25,177... \approx 25,2$ (kertaa)
- virhemarginaaliksi saadaan $ME = 0,677... \approx 0,7$ (kertaa).

Tekstiilin pesunkestävyyden keskiarvon luottamusväli on siis $24,5 \pm 0,7$ kertaa eli 23,8–25,2.

Otos 2:

Otoskeskiarvo on $\bar{x} = 27,3$

Otoskeskihajonta on $s = 4,0$

Otoskoko on $n = 150$

Luottamustaso on $C = 0,95$

Syötetään tiedot esimerkiksi laskinohjelmiston luottamusväli-toimintoon.

Luottamusvälin

- alarajaksi saadaan $26,654... \approx 26,7$ (kertaa)
- ylärajaksi saadaan $27,945... \approx 27,9$ (kertaa)
- virhemarginaaliksi saadaan $ME = 0,645... \approx 0,6$ (kertaa).

Tekstiilin pesunkestävyyden keskiarvon luottamusväli on siis $27,3 \pm 0,6$ kertaa eli 26,7–27,9.

Muodostetut luottamusvälit eivät ole päällekkäiset, vaan uudemman otoksen mukaisen luottamusvälin alaraja (26,7) on suurempi kuin edellisen otoksen mukaisen luottamusvälin yläraja (25,2).

Pesunkestävyyden voidaan tämän perusteella päätellä parantuneen.

240.

a) Merkintä $\alpha = 0,05$ tarkoittaa, että riskitasoksi on valittu 5 %.
Luottamustaso on tällöin $100 \% - 5 \% = 95 \% = 0,95$.

Perusjoukon keskihajontaa ei tunneta, joten muodostetaan keskiarvolle t-väli.

Otoskeskiarvo on $\bar{x} = 4,3$

Otoskeskihajonta on $s = 0,38$

Otoskoko on $n = 16$

Luottamustaso on $C = 0,95$



Valitaan laskimessa
tiedonsyöttömenetelmäksi
"tilastot".

Syötetään tiedot esimerkiksi laskinohjelmiston luottamusväli-toimintoon.

Luottamusvälin

- alarajaksi saadaan $4,097... \approx 4,1$
- ylärajaksi saadaan $4,502... \approx 4,5$
- virhemarginaaliksi saadaan $ME = 0,202... \approx 0,2$.

Asiakastyytyväisyyden keskiarvon luottamusväli on siis $4,3 \pm 0,2$ eli 4,1–4,5.

b) Ainakin jotain näyttöä tyytyväisyyden kohoamisesta on, sillä uusimmassa mittauksessa keskiarvo on korkeampi kuin pitkän aikavälin keskiarvo ($4,3 > 4,1$).

Luottamusvälin perusteella johtopäätökseen jää kuitenkin epävarmuutta, sillä pitkän aikavälin keskiarvo on juuri ja juuri a-kohdassa muodostetulla luottamusvälillä $4,1-4,5$.

Otoskoko $n = 16$ vaikuttaa pieneltä, mutta toisaalta perusjoukon koosta (kaikkien asiakkaiden lukumäärästä) ei ole tietoa.

241.

"Tarkkuus" tarkoittaa tässä sitä, kuinka paljon otoskeskiarvo keskimäärin poikkeaa perusjoukon keskiarvosta.

a) Keskiarvon keskivirhe ilmaisee otoskeskiarvon tarkkuuden.

Naisten otoksessa keskiarvon keskivirhe on 13,98.

Miesten otoksessa keskiarvon keskivirhe on 15,80 eli suurempi.

Naisten otoskeskiarvo on siis tarkempi.

Keskiarvon keskivirhe $\left(= \frac{\sigma}{\sqrt{n}}\right)$ on sitä pienempi, mitä suurempi otoskoko n on.

Naisten otoskoko on suurempi kuin miehillä ($68 > 52$).

Tästä syystä naisten keskiarvon keskivirhe on pienempi kuin vastaava luku miehillä.

b) Verrataan tutkimuksen perusteella saatuja keskivuokran luottamusvälejä toisiinsa.

Naisten maksaman keskivuokran (€/kk) 95 %:n luottamusväli on 257,24–313,04.

Miesten maksaman keskivuokran (€/kk) 95 %:n luottamusväli on 214,74–277,58.

Luottamusvälit ovat osittain päällekkäiset, joten ei voida sanoa, että naiset maksaisivat keskimäärin korkeampaa vuokraa kuin miehet.

242.

a) Kirjoitetaan mittaustulokset esimerkiksi laskinohjelmiston taulukkosovellukseen. Määritetään keskiarvo ja otoskeskihajonta tilastotoiminnoilla.

- keskiarvo on $\bar{x} = 4,5$
- otoskeskihajonta on $s = 0,583...$

Otoskoko on $n = 14$. Keskiarvon keskivirheeksi arvioidaan

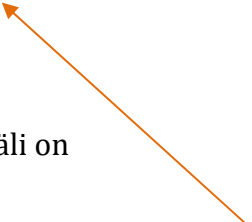
$$\frac{s}{\sqrt{n}} = \frac{0,583 \dots}{\sqrt{14}} = 0,155 \dots \approx 0,16.$$

b) Perusjoukon keskihajontaa ei tunneta, joten muodostetaan keskiarvolle t-väli esimerkiksi laskinohjelmiston taulukkosovelluksessa, luottamusväli-toiminnolla.

Luottamusvälin

- alarajaksi saadaan $4,163... \approx 4,2$
- ylärajaksi saadaan $4,836... \approx 4,8$
- virhemarginaaliksi saadaan $ME = 0,336... \approx 0,3$.

Sadeveden keskimääräisen pH-pitoisuuden luottamusväli on siis $4,5 \pm 0,3$ eli 4,2–4,8.



Valitaan laskimessa tiedonsyöttömenetelmäksi "tiedot". Valittu luottamustaso on $C = 0,95$.

243.

a) Kirjoitetaan punnitustulokset esimerkiksi laskinohjelmiston taulukkosovellukseen.

Perusjoukon keskihajontaa ei tunneta, joten muodostetaan keskiarvolle t-väli.

Luottamusvälin

- alarajaksi saadaan $91,317... \approx 91,3$ (g)
- ylärajaksi saadaan $104,882... \approx 104,9$ (g)
- otoskeskiarvo on $\bar{x} = 98,1$ (g)
- virhemarginaaliksi saadaan $ME = 6,782... \approx 6,8$ (g)



1) Valitaan tilastot-
toiminnoista
luottamusvälit: t-väli.
2) Valitaan
tiedonsyöttö-
menetelmäksi "tiedot".
3) Syötetään
havaintojen lista sekä
valittu luottamustaso
 $C = 0,95$.

Omenapuun omenien keskipainon luottamusväli on siis $98,1 \text{ g} \pm 6,8 \text{ g}$ eli $91,3 \text{ g} - 104,9 \text{ g}$.

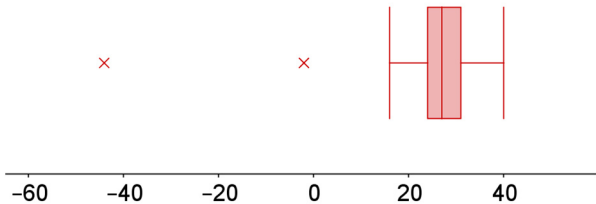
b) Otoksen perusteella muodostettu luottamusväli ilmaisee, millä välillä koko perusjoukon (omenapuun kaikkien omenien) tuntematon **keskipaino** on tietyllä varmuudella.

Luottamusväli ei anna tietoa yksittäisten omenien painosta.

Luottamusväliä ei siis voida tulkita tehtävänannossa ilmaistulla tavalla.

244.

a) Esimerkiksi matematiikkaohjelmiston tilastot-toiminnolla voidaan piirtää jakaumaa kuvaava laatikkokuvio, kun mittaustulokset on kopioitu laskentataulukkoon.



Poikkeavat havainnot on merkitty laatikkokuvioon rastilla (X).

Laatikkokuvion perusteella havainnosta kaksi pienintä poikkeaa merkitsevästi keskiarvosta. Kaksi pienintä arvoa ovat -44 ja -2. Lasketaan näiden arvojen normitetut arvot.

Määritetään mittaustulosten keskiarvo ja keskihajonta tilastotoiminnoilla.

- keskiarvo on $\bar{x} = 26,212 \dots$
- keskihajonta on $s = 10,745 \dots$

Arvon -44 normitettu arvo on

$$\frac{-44 - \bar{x}}{s} = \frac{-44 - 26,212 \dots}{1,745 \dots} = -6,53 \dots$$

Arvon -2 normitettu arvo on

$$\frac{-2 - \bar{x}}{s} = \frac{-2 - 26,212 \dots}{1,745 \dots} = -2,62 \dots$$

Normitettujen arvojen itseisarvot ovat suurempia kuin 2, joten kaksi pienintä arvoa poikkeaa merkitsevästi keskiarvosta.

Poistetaan kaksi pienintä arvoa, jolloin havaintojen lukumäärä on $N = 64$. Kun kaksi pienintä arvoa on poistettu, tunnusluvuiksi saadaan:

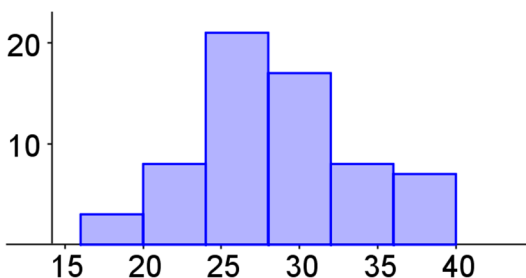
- keskiarvo on $\bar{x} = 27,75 \approx 27,8$
- otoskeskihajonta on $s = 5,083...$
- pienin arvo on $\min = 16$
- suurin arvo on $\max = 40$.

Vaihteluvälin pituus on $40 - 16 = 24$.

Havainnollistetaan jakaumaa histogrammilla. Valitaan ensimmäisen luokan todelliseksi alarajaksi pienin arvo 16 ja luokkien lukumääräksi esimerkiksi 6.

Luokkavälin leveydeksi tulee $\frac{24}{6} = 4$.

Matematiikkaohjelmisto tekee luokittelun ja kuvaa jakauman histogrammilla.



Mittaustulos	<i>f</i>
16–19	3
20–23	8
24–27	21
28–31	17
32–35	8
36–40	7

Frekvenssijakauman perusteella mittaustulokset jakautuvat melko symmetrisesti keskiarvon 27,8 ympärille ja suurin osa mittaustuloksista on lähellä keskiarvoa.

b) Kun kaksi pienintä arvoa on poistettu, tunnusluvut ovat:

- keskiarvo on $\bar{x} = 27,75 \approx 27,8$
- otoskeskihajonta on $s = 5,083...$

Keskivirheeksi saadaan

$$\frac{s}{\sqrt{n}} = \frac{5,083 \dots}{\sqrt{64}} = 0,6354 \dots \approx 0,64.$$

c) Määritetään luottamusväli (t-väli) laskin- tai matematiikkaohjelmiston tilastot-toiminnoilla. Luottamusväliä varten tarvitaan:

- otoskeskiarvo $\bar{x} = 27,75$
- otoskeskihajonta $s = 5,083...$
- otoskoko $n = 64$
- luottamustaso $C = 0,99$

Luottamusväliksi saadaan $27,75 \pm 1,687 \dots \approx 27,8 \pm 1,7$ eli 26,1–29,4.

245.

Keskiarvon luottamusväli on

$$\bar{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}}.$$

Luottamusvälin pituus lasketaan lausekkeella

$$2 \cdot z^* \cdot \frac{\sigma}{\sqrt{n}}$$

I Kun otoskokoa kasvatetaan, luottamusväli lyhenee.

Oikein. Otoskoko n on lausekkeessa nimittäjässä, joten otoskoon kasvaessa lausekkeen arvo pienenee eli luottamusväli lyhenee.

II Kun luottamustasoa kasvatetaan, luottamusvälin pituus lyhenee.

Väärin. Luottamustaso määrää kriittisen kertoimen z^* arvon. Kun luottamustaso kasvaa, myös kriittinen arvo kasvaa eli luottamusväli pitenee.

III Kun otoskokoa kasvatetaan, luottamustaso kasvaa.

Väärin. Tutkimuksen tekijä (tai muu ihminen) päättää luottamustason.

246.

Keskiarvon 95 %:n luottamusväli on

$$\bar{x} \pm 1,96 \cdot \frac{\sigma}{\sqrt{n}},$$

jossa

- $1,96 \cdot \frac{\sigma}{\sqrt{n}} = \text{virhemarginaali}$
- $\frac{\sigma}{\sqrt{n}} = \text{keskiarvon keskivirhe.}$

Otoskoko on $n = 25$.

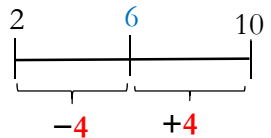
a) Otoskeskiarvo on luottamusvälin keskipiste. Lasketaan luottamusvälin päätepisteiden keskiarvo.

$$\frac{2 + 10}{2} = \frac{12}{2} = 6$$

Siis otoskeskiarvo on $\bar{x} = 6$.

b) Virhemarginaali on $10 - 6 = 4$ (tai: $6 - 2 = 4$).

Muodostetaan yhtälö ja ratkaistaan keskivirhe.



$$1,96 \cdot \frac{\sigma}{\sqrt{n}} = 4$$

$$\frac{\sigma}{\sqrt{n}} = \frac{4}{1,96} = 2,040 \dots \approx 2$$

c) Sijoitetaan otoskoko $n = 25$ keskivirheen lausekkeeseen ja ratkaistaan perusjoukon keskihajonta σ .

$$\frac{\sigma}{\sqrt{25}} = 2,040 \dots$$

$$\frac{\sigma}{5} = 2,040 \dots$$

$$\sigma = 5 \cdot 2,040 \dots = 10,204 \dots \approx 10$$

247.

Perusjoukon keskihajontaa ei tunneta, joten muodostetaan keskiarvolle t-väli.

Otoskeskiarvo on $\bar{x} = 10,2$ (%)

Otoskeskihajonta on $s = 3,1$ (%)

Otoskoko on $n = 25$

Luottamustaso on $C = 0,95$

Valitaan laskimessa tiedonsyöttömenetelmäksi "tilastot".

Syötetään tiedot esimerkiksi laskinohjelmiston luottamusväli-toimintoon.

Luottamusvälin

- alarajaksi saadaan $8,920... \approx 8,9$ (%)
- ylärajaksi saadaan $11,479... \approx 11,5$ (%)
- virhemarginaaliksi saadaan $ME = 1,279... \approx 1,3$ (%).

Nikkelipitoisuuden keskiarvon luottamusväli on siis $10,2 \% \pm 1,3$ %-yksikköä eli $8,9 \% - 11,5 \%$.

Tulos tarkoittaa: voidaan olla 95 %:n varmoja, että kyseisen alueen kallioperän keskimääräinen nikkelipitoisuus on välillä $8,9 \% - 11,5 \%$.

248.

a) Kirjoitetaan mittaustulokset esimerkiksi laskinohjelmiston taulukkosovellukseen. Määritetään keskiarvo ja otoskeskihajonta tilastotoiminnoilla.

- keskiarvo on $\bar{x} = 117,65 \approx 117,7$ (°C)
- otoskeskihajonta on $s = 1,820 \dots \approx 1,8$ (°C)

b) Perusjoukon keskihajonta on tiedossa, joten luottamusväli voidaan muodostaa taulukkokirjoista löytyvällä laskukaavalla.

Luottamustasolla 95 % luottamusväli on muotoa

$$\bar{x} \pm 1,96 \cdot \frac{\sigma}{\sqrt{n}}$$

Perusjoukon keskihajonta on $\sigma = 2,0$ (°C) ja otoskoko on $n = 10$. Virhemarginaaliksi saadaan

$$1,96 \cdot \frac{2,0}{\sqrt{10}} = 1,239 \dots \approx 1,2 \text{ (°C)}.$$

Luottamusväli on siis $117,7 \text{ °C} \pm 1,2 \text{ °C}$.

Luottamusvälin

- alaraja on $117,65 - 1,239 \dots = 116,410 \dots \approx 116,4$ (°C)
- yläraja on $117,65 + 1,239 \dots = 118,889 \dots \approx 118,9$ (°C).

Luottamusväli on siis $116,4 \text{ °C} - 118,9 \text{ °C}$.

c) Perusjoukon keskihajonta on $\sigma = 2,0$ (°C) ja otoskoko on $n = 50$.

Virhemarginaaliksi saadaan

$$1,96 \cdot \frac{2,0}{\sqrt{50}} = 0,554 \dots \approx 0,6 \text{ (°C)}.$$

Otoskeskiarvo on $\bar{x} = 117,37 \approx 117,4$ (°C).

Sulamispisteen luottamusväli on siis $117,4^\circ\text{C} \pm 0,6^\circ\text{C}$.

Luottamusvälin

- alaraja on $117,37 - 1,239 \dots = 116,130 \dots \approx 116,1$ (°C)
- yläraja on $117,37 + 1,239 \dots = 118,609 \dots \approx 118,6$ (°C).

Sulamispisteen luottamusväli on siis $116,1^\circ\text{C} - 118,6^\circ\text{C}$.

d) Perusjoukon keskihajonta ei ole tiedossa, joten muodostetaan keskiarvolle t-väli.

Luottamusvälin

- alarajaksi saadaan $116,347... \approx 116,3$ (°C)
- ylärajaksi saadaan $118,952... \approx 119,0$ (°C)
- otoskeskiarvo on $\bar{x} = 117,65 \approx 117,7$ (°C)
- virhemarginaaliksi saadaan $ME = 1,302... \approx 1,3$ (°C).

Sulamispisteen luottamusväli on siis $117,7\text{ °C} \pm 1,3\text{ °C}$ eli $116,3\text{ °C} - 119,0\text{ °C}$.

Kun perusjoukon keskihajontaa ei tunneta, se arvioidaan otoksen perusteella otoskeskihajonnalla. Arvioon tulee tällöin lisää epävarmuutta, mikä kasvattaa luottamusvälin pituutta.

Jos hajonta σ olisi tuntematon, b-kohdan luottamusvälin virhemarginaali kasvaisi eli luottamusväli pitenisi.

249.

a) Perusjoukon keskihajontaa ei tunneta, joten muodostetaan keskiarvolle t-väli.

Otoskeskiarvo on $\bar{x} = 0,60$ ($\mu\text{g/l}$)

Otoskeskihajonta on $s = 0,44$ ($\mu\text{g/l}$)

Otoskoko on $n = 28$

Luottamustaso on $C = 0,99$

Valitaan laskimessa tiedonsyöttömenetelmäksi "tilastot".

Syötetään tiedot esimerkiksi laskinohjelmiston luottamusväli-toimintoon.

Luottamusvälin

- alarajaksi saadaan $0,369... \approx 0,37$ ($\mu\text{g/l}$)
- ylärajaksi saadaan $0,830... \approx 0,83$ ($\mu\text{g/l}$)
- virhemarginaaliksi saadaan $ME = 0,230... \approx 0,23$ ($\mu\text{g/l}$)

Lyijypitoisuuden keskiarvon luottamusväli on siis $0,60 \mu\text{g/l} \pm 0,23 \mu\text{g/l}$ eli $0,37 \mu\text{g/l} - 0,83 \mu\text{g/l}$.

Tulos tarkoittaa: voidaan olla 99 % varmoja, että kyseisen järven veden keskimääräinen lyijypitoisuus on välillä $0,37 \mu\text{g/l} - 0,83 \mu\text{g/l}$.

b) Luottamusväli ilmaisee, mille välille perusjoukon tuntematon keskiarvo sijoittuu tietyllä luottamustason ilmaisemalla varmuudella. Luottamusväli ei anna tietoa yksittäisistä havainnoista (mittauksista). Tehtävänannon luonnehdinta luottamusvälille on siis virheellinen.

250.

Perusjoukoksi tulkitaan tässä Kimmon kilpailukauden kaikki heittotulokset. Kilpailuissa saadut tulokset tulkitaan satunnaisotokseksi tästä perusjoukosta.

Riskitasoa 5 % vastaa luottamustaso $100 \% - 5 \% = 95 \% = 0,95$.

Kirjoitetaan heittotulokset esimerkiksi laskinohjelmiston taulukkosovellukseen.

Perusjoukon keskihajontaa ei tunneta, joten muodostetaan keskiarvolle t-väli.

Luottamusvälin

- alarajaksi saadaan $20,744... \approx 20,74$ (m)
- ylärajaksi saadaan $20,915... \approx 20,92$ (m)
- otoskeskiarvo on $\bar{x} = 20,83$ (m)
- virhemarginaaliksi saadaan $ME = 0,085... \approx 0,09$ (m)

Heittotulosten keskiarvon luottamusväli on siis $20,83 \text{ m} \pm 0,09 \text{ m}$ eli $20,74 \text{ m} - 20,92 \text{ m}$.

Viime kesän keskiarvo $20,80 \text{ m}$ kuuluu muodostetulle luottamusvälille. Tämän perusteella ei siis voida sanoa, että heittotulos olisi parantunut edellisvuodesta.

251.

Kopioidaan mittaustulokset matematiikkaohjelmiston taulukkosovellukseen.

Määritetään tunnusluvut tilastotoiminnoilla.

- keskiarvo on $\bar{x} = 4,751 \dots$
- otoskeskihajonta on $s = 0,777 \dots$
- pienin arvo on $\min = 2,87$
- suurin arvo on $\max = 6,95$.

Vaihteluvälin pituus on
 $6,95 - 2,87 = 4,08$.

Matematiikkaohjelmisto tekee automaattisesti luokittelun, jossa

- ensimmäisen luokan todellinen alaraja on aineiston pienin arvo 2,87
- viimeisen luokan todellinen yläraja on aineiston suurin arvo 6,95
- luokkien lukumäärä on 5.

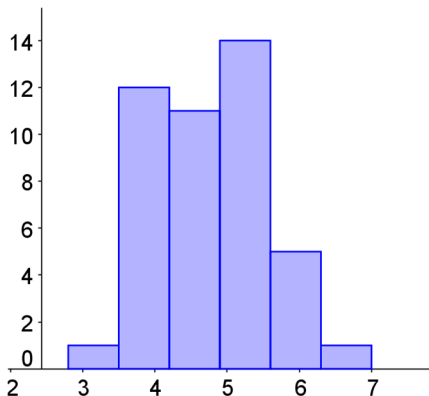
Histogrammia varten aineisto on luokiteltava ja histogrammi muoto riippuu valitusta luokittelusta. Luokitteluun ei ole olemassa yhtä oikeaa tapaa, mutta luokittelussa voidaan edetä kappaleessa 2.1 esitetyllä tavalla.

Valitaan ensimmäisen luokan todelliseksi alarajaksi esimerkiksi arvo 2,8 ja luokkien lukumääräksi esimerkiksi 6.

Luokkavälin leveydeksi tulee $\frac{4,08}{6} = 0,68 \approx 0,7$.

Matematiikkaohjelmisto tekee luokittelun ja kuvaa jakauman histogrammilla, kun luokat asetetaan käsin ja ohjelmistoon syötetään:

- luokittelun alku: 2,8
- luokkavälin leveys: 0,7



Frekvenssitaulukossa luokkarajat esitetään pyöristettyinä lukuina. Edellisen luokan yläraja ja seuraavan luokan alaraja ovat eri lukuja.

Pituus (m)	<i>f</i>
2,8–3,4	1
3,5–4,1	12
4,2–4,8	11
4,9–5,5	14
5,6–6,2	5
6,3–7,0	1

Kuvaajan perusteella jakauma vaikuttaa symmetriseltä: suurin osa havainnoista on lähellä keskiarvoa $\bar{x} = 4,751 \dots$ (m) ja havaintojen määrä pienenee keskiarvosta alas- ja ylöspäin siirryttäessä.

Normitetaan pienin arvo 2,87 (m) ja suurin arvo 6,95 (m).

Arvon 2,87 normittu arvo on

$$\frac{2,87 - \bar{x}}{s} = \frac{2,87 - 4,75 \dots}{0,77 \dots} = -2,420 \dots$$

Normitetun arvon itseisarvo on $|-2,420 \dots| = 2,420 > 2$, joten pienin arvo poikkeaa merkitsevästi odotusarvosta.

Arvon 6,95 normittu arvo on

$$\frac{6,95 - \bar{x}}{s} = \frac{6,95 - 4,75 \dots}{0,77 \dots} = 2,829 \dots$$

Normitetun arvon itseisarvo on $|2,829 \dots| = 2,829 > 2$, joten myös suurin arvo poikkeaa merkitsevästi odotusarvosta.

b) Perusjoukon keskihajontaa ei tunneta, joten muodostetaan keskipituudelle t-väli.

Luottamusvälin

- alarajaksi saadaan $4,515... \approx 4,52$ (m)
- ylärajaksi saadaan $4,987... \approx 4,99$ (m)
- otoskeskiarvo on $\bar{x} = 4,751 ... \approx 4,75$ (m)
- virhemarginaaliksi saadaan $ME = 0,236... \approx 0,24$ (m).

Valkohain keskipituudelle saadaan siis luottamusväli $4,75 \text{ m} \pm 0,24 \text{ m}$ eli $4,52 \text{ m} - 4,99 \text{ m}$.

c) Arvo $5,0 \text{ m}$ ei kuulu b-kohdassa muodostetulle luottamusvälille $4,52 \text{ m} - 4,99 \text{ m}$, kun luottamusvälin rajat ilmaistaan kahden desimaalin tarkkuudella. Jos käytetään yksidesimaalisia likiarvoja, luottamusväliksi saadaan $4,5 \text{ m} - 5,0 \text{ m}$, jolloin arvo $5,0 \text{ m}$ kuuluu luottamusvälille.

Tämän perusteella voidaan siis sanoa, että keskipituus voi olla jopa $5,0 \text{ m}$.

4 Tilastollinen päättely

4.2 Prosenttiosuuden virhemarginaali ja luottamusväli

Kommentteja ratkaisuihin:

Prosenttiosuuden luottamusväli voidaan ilmaista kahdella eri tavalla.

Tapa 1: Ilmoitetaan otoksesta laskettu prosenttiosuus $\hat{p}$ sekä tutkimuksen virhemarginaali muodossa $\hat{p} \pm \text{virhemarginaali}$.

Tapa 2: Ilmoitetaan luottamusvälin ala- ja yläraja vähentämällä ja lisäämällä virhemarginaali otoksesta laskettuun prosenttiosuuteen. Tällöin luottamusväli ilmaistaan välinä *alaraja* – *yläraja*. Välin ilmaisemiseen voidaan käyttää myös hakasulkumerkintää: [*alaraja*, *yläraja*].

Näissä ratkaisuissa luottamusvälit ilmaistaan molemmilla tavoilla. Kumpi hyvänsä ilmaisutapa yksinään on riittävä.

Prosenttiosuuden luottamusväli voidaan muodostaa laskin- tai matematiikkaohjelmiston tilastot-toiminnolla (1 väittämän z-väli tai suhteen z-estimaatti)

Näissä ratkaisuissa luottamusvälit muodostetaan kuitenkin taulukkokirjoista löytyvän kaavan avulla.

252.

Todennäköisyys, että poimittu pallo on vihreä, on
 $P(\text{vihreä}) = 25 \% = 0,25$.

Olkoon $X = \text{"vihreiden pallojen lukumäärä 100 pallon otoksessa"}$.

Pallojen poimiminen voidaan tulkita toistokokeeksi. Toistojen lukumäärä on $n = 100$ ja "onnistumisen" eli vihreän pallon saamisen todennäköisyys on joka toistolla sama: $p = 0,25$.

Siis: $X \sim \text{Bin}(100; 0,25)$

a) Lasketaan tapahtuman $X = 25$ todennäköisyys.

$$P(X = 25) = 0,0917... \approx 0,09$$

Laskinohjelmiston
binomijakaumatoimintoon syötetään:
 $n = 100, p = 0,25$ ja $x = 25$.

b) Otoksessa on 100 palloa, joten yksi pallo vastaa yhtä prosenttia.

$$25 - 3 = 22 \quad 25 + 3 = 28$$

Vihreiden pallojen lukumäärä otoksessa poikkeaa arvosta 25 korkeintaan kolme prosenttiyksikköä, kun $22 \leq X \leq 28$.

$$P(22 \leq X \leq 28) = 0,581... \approx 0,58$$

Laskinohjelmiston binomijakaumatoimintoon syötetään: $n = 100$,
 $p = 0,25$, alaraja 22 ja yläraja 28.

253.

a) Tutkimuksessa tutkitaan erään kaupungin lukiolaisia, joten perusjoukkona on tämän kaupungin lukiolaiset.

b) Otoksen koko on $n = 150$. Näistä 63 vastasi käyttävänsä Firefoxia, joten käyttäjien prosenttiosuus on otoksessa

$$\hat{p} = \frac{63}{150} = 0,42 = 42 \%$$

c) Prosenttiosuuden 95 % virhemarginaali on

$$\pm 1,96 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}.$$

$$n = 150$$

$$\hat{p} = 0,42$$

$$1 - \hat{p} = 1 - 0,42 = 0,58$$

Virhemarginaaliksi saadaan

$$\pm 1,96 \cdot \sqrt{\frac{0,42 \cdot 0,58}{150}} = \pm 0,0789 \dots \approx \pm 7,9 \%$$

eli $\pm 7,9$ prosenttiyksikköä.

254.

a) Otoksen perusteella saatu prosenttiosuus antaa arvion eli estimoi prosenttiosuutta koko perusjoukossa.

Työharjoittelua tai TET-jaksoa kannattavien lukiolaisten prosenttiosuus on noin 53 %.

b) Virhemarginaali määrittelee kannatusosuuden 53 % ympärille rajat, joiden sisällä todellinen kannatusosuus on luottamustason ilmaisemalla varmuudella.

- $53 - 4,4 = 48,6$ (%)
- $53 + 4,4 = 57,4$ (%)

Työharjoittelua tai TET-jaksoa kannattavien lukiolaisten prosenttiosuus on välillä 48,6 % – 57,4 %.

c) Kun johtopäätöksiä tehdään otoksen perusteella (ilman, että asiaa kysytään jokaiselta perusjoukkoon kuuluvalta), tuloksiin jää aina epävarmuutta. Tämä johtuu siitä, että otokseen osuu sattumanvarainen joukko vastaajia.

Tämä tilastollinen epävarmuus ilmaistaan ilmoittamalla tutkimuksen riskitaso 5 % tai luottamustaso $100 \% - 5 \% = 95 \%$.

Voidaan siis olla 95 % varmoja siitä, että työharjoittelua tai TET-jaksoa kannattavien lukiolaisten prosenttiosuus on välillä 48,6 % – 57,4 %.

255.

a) Prosenttiosuuden 95 %:n virhemarginaali on

$$\pm 1,96 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

$$n = 550$$

$$\hat{p} = 0,192$$

$$1 - \hat{p} = 1 - 0,192 = 0,808$$

Virhemarginaaliksi saadaan

$$\pm 1,96 \cdot \sqrt{\frac{0,192 \cdot 0,808}{550}} = \pm 0,0329 \dots \approx \pm 3,3 \%$$

eli $\pm 3,3$ prosenttiyksikköä.

b) Luottamusväli on $19,2 \% \pm 3,3 \%$ -yksikköä. Luottamusvälin

- alaraja on $19,2 - 3,3 = 15,9 \%$
- yläraja on $19,2 + 3,3 = 22,5 \%$.

Luottamusväli on siis $15,9 \% - 22,5 \%$.

c) Luottamusväli ilmaisee rajat, joiden sisällä kaupunkipyöriä käyttävien asukkaiden prosenttiosuus on 95 %:n varmuudella. Päättelyyn jäävä riskitaso on $100\% - 95\% = 5\%$ eli on 5 %:n mahdollisuus, että todellinen prosenttiosuus on luottamusvälin ulkopuolella.

d) Ei tarkoita. Riskitaso ilmaisee otoksen perusteella tehtyyn johtopäätökseen, eli luottamusväliin, liittyvän epävarmuuden. Ei voida olla täysin varmoja, että kaupunkipyöriä käyttävien prosenttiosuus olisi b-kohdassa muodostetulla välillä.

256.

a) Prosenttiosuuden 95 %:n virhemarginaali on

$$\pm 1,96 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

$$n = 1003$$

$$\hat{p} = 0,44$$

$$1 - \hat{p} = 1 - 0,44 = 0,56$$

Virhemarginaaliksi saadaan

$$\pm 1,96 \cdot \sqrt{\frac{0,44 \cdot 0,56}{1003}} = \pm 0,0307 \dots \approx \pm 3 \%$$

eli ± 3 prosenttiyksikköä.

Luottamusväli on 44 % ± 3 %-yksikköä.

Luottamusvälin

- alaraja on $44 - 3 = 41$ (%)
- yläraja on $44 + 3 = 47$ (%).

Luottamusväliksi saadaan siis 41 % – 47 %.

b) Luottamusväli ilmaisee, millä välillä prosenttiosuus on tutkimuksen perusjoukossa. Luottamustaso ilmaisee, kuinka varma tutkimuksessa saatu johtopäätös on.

Voidaan olla 95 %:n varmoja siitä, että tutkimuksen perusjoukossa niiden ihmisten prosenttiosuus, jotka pitävät ulkomaille muuttoa itselleen todennäköisenä vaihtoehtona, on välillä 41 % – 47 %.

257.

Otoksen koko on $n = 5000$.

EU-jäsenyyttä kannattavien prosenttiosuus otoksessa oli

$$\hat{p} = \frac{2880}{5000} = 0,576 = 57,6 \, \%.$$

a) Prosenttiosuuden 99 %:n virhemarginaali on

$$\pm 2,58 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

$$n = 5000$$

$$\hat{p} = 0,576$$

$$1 - \hat{p} = 1 - 0,576 = 0,424$$

Virhemarginaaliksi saadaan

$$\pm 2,58 \cdot \sqrt{\frac{0,576 \cdot 0,424}{5000}} = \pm 0,0180 \dots \approx \pm 1,8 \, \%$$

eli $\pm 1,8$ prosenttiyksikköä.

b) Luottamusväli on $57,6 \% \pm 1,8 \%$ -yksikköä.

Luottamusvälin

- alaraja on $57,6 - 1,8 = 55,8 \%$
- yläraja on $57,6 + 1,8 = 59,4 \%$.

Luottamusväliksi saadaan siis $55,8 \% - 59,4 \%$.

Vaaleissa EU-jäsenyyden kannattajia on $57,6 \% \pm 1,8 \%$ -yksikköä eli kannattajien prosenttiosuus on välillä $55,8 \% - 59,4 \%$.

c) Lopullinen äänestystulos $52,2 \%$ ei kuulu b-kohdassa kyselyn perusteella muodostettuun luottamusväliin. Ovensuukysely ei siis onnistunut ennustamaan vaalitulosta.

258.

a) Tutkimukseen vastasi 2260 nuorta, joten otoskoko on $n = 2260$.
Prosenttiosuuden 95 %:n virhemarginaali on

$$\pm 1,96 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

WhatsApp:

$\hat{p} = 91 \% = 0,91$, jolloin $1 - \hat{p} = 1 - 0,91 = 0,09$.

Virhemarginaaliksi saadaan

$$\pm 1,96 \cdot \sqrt{\frac{0,91 \cdot 0,09}{2260}} = \pm 0,0117 \dots \approx \pm 1,2 \%$$

Luottamusväli on $91 \% \pm 1,2 \%$ eli $89,8 \% - 92,2 \%$.

YouTube:

$\hat{p} = 90 \% = 0,90$, jolloin $1 - \hat{p} = 1 - 0,90 = 0,10$.

Virhemarginaaliksi saadaan

$$\pm 1,96 \cdot \sqrt{\frac{0,90 \cdot 0,10}{2260}} = \pm 0,0123 \dots \approx \pm 1,2 \%$$

Luottamusväli on $90 \% \pm 1,2 \%$ eli $88,8 \% - 91,2 \%$.

Luottamusvälin

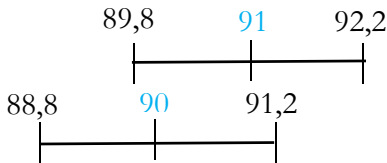
alaraja on

$91 \% - 1,2 \% = 89,8 \%$,

yläraja on

$91 \% + 1,2 \% = 92,2 \%$

b) Tutkimuksen perusteella WhatsAppia käyttävien nuorten prosenttiosuus on välillä 89,8 % – 92,2 % nuorista ja YouTubea käyttävien nuorten prosenttiosuus välillä 88,8 % – 91,2 %.



Luottamusvälit ovat osin päällekkäiset.

Tutkimuksen perusteella ei siis voida olla varmoja, että WhatsApp olisi käytetyin sosiaalisen median palvelu 13–17-vuotiaiden suomalaisnuorten keskuudessa.

259.

a)

SDP:

$$n = 2431$$

$$\hat{p} = 21,2 \% = 0,212$$

$$1 - \hat{p} = 1 - 0,212 = 0,788$$

Virhemarginaaliksi saadaan

Prosenttiosuuden 95 %:n
virhemarginaali on

$$\pm 1,96 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

$$\pm 1,96 \cdot \sqrt{\frac{0,212 \cdot 0,788}{2431}} = \pm 0,0162 \dots \approx \pm 1,6 \%$$

Luottamusväli on $21,2 \% \pm 1,6 \%$ eli $19,6 \% - 22,8 \%$.

KOK:

$$n = 2431$$

$$\hat{p} = 20,5 \% = 0,205$$

$$1 - \hat{p} = 1 - 0,205 = 0,795$$

Virhemarginaaliksi saadaan

Luottamusvälin

alaraja on

$$21,2 \% - 1,6 \% = 19,6 \%$$

yläraja on

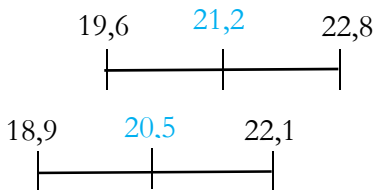
$$21,2 \% + 1,6 \% = 22,8 \%$$

$$\pm 1,96 \cdot \sqrt{\frac{0,205 \cdot 0,795}{2431}} = \pm 0,0160 \dots \approx \pm 1,6 \%$$

Luottamusväli on $20,5 \% \pm 1,6 \%$ eli $18,9 \% - 22,1 \%$.

b) Tutkimuksen virhemarginaali on suurten puolueiden (SDP ja KOK) osalta $\pm 1,6\% \approx \pm 2\%$ eli noin kaksi prosenttiyksikköä suuntaansa. Maininta virhemarginaalista on siis oikea.

Kyselyyn vastanneiden joukossa SDP:n kannatus on Kokoomuksen kannatusta suurempaa ($21,5\% > 20,5\%$).



a-kohdan luottamusvälit ovat kuitenkin osin päällekkäiset. Luottamusvälien perusteella ei siis voida päätellä, että SDP:n kannatus olisi suurinta kaikkien äänioikeutettujen joukossa.

Luottamusvälien perusteella uutisen tekemään johtopäätökseen jää siis epävarmuutta.

260.

Prosenttiosuuden 95 %:n virhemarginaali on

$$\pm 1,96 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

$$n = 1613$$

$$\hat{p} = \frac{470}{1613} = 0,2913 \dots \approx 0,291 = 29,1 \%$$

$$1 - \hat{p} = 1 - 0,291 = 0,709$$

Virhemarginaaliksi saadaan

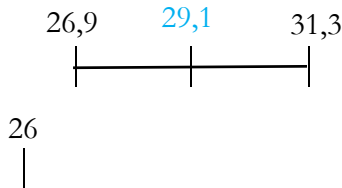
$$\pm 1,96 \cdot \sqrt{\frac{0,291 \cdot 0,709}{1613}} = \pm 0,0221 \dots \approx \pm 2,2 \%$$

Luottamusväli on 29,1 % $\pm$ 2,2 % eli 26,9 % – 31,3 %.

Verrataan aiemmin saatua prosenttiosuutta nyt saatuun prosenttiosuuden luottamusväliin.

Tutkimuksen perusteella saatu luottamusväli ilmaisee, että 26,9 % – 31,3 % suomalaisista uskoo oman taloudellisen tilanteen kohenemiseen.

Aiempi prosenttiosuus 26 % ei kuulu tälle luottamusvälille.



Voidaan siis perustellusti todeta, että suomalaisten usko oman taloudellisen tilanteen kohenemiseen on muuttunut (kasvanut).

261.

a) Prosenttiosuuden 95 %:n virhemarginaali on

$$\pm 1,96 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

Kysely 1:

$$n = 1500$$

$$\hat{p} = 0,58$$

$$1 - \hat{p} = 1 - 0,58 = 0,42$$

Virhemarginaaliksi saadaan

$$\pm 1,96 \cdot \sqrt{\frac{0,58 \cdot 0,42}{1500}} = \pm 0,0249 \dots \approx \pm 2,5 \, \%.$$

Kysely 2:

$$n = 1000$$

$$\hat{p} = 0,68$$

$$1 - \hat{p} = 1 - 0,68 = 0,32$$

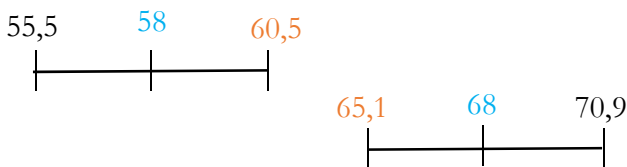
Virhemarginaaliksi saadaan

$$\pm 1,96 \cdot \sqrt{\frac{0,68 \cdot 0,32}{1000}} = \pm 0,0289 \dots \approx \pm 2,9 \, \%.$$

b) Kyselyn 1 mukaan Niinistön kannatusosuus on $58\% \pm 2,5\%$ eli kannatusosuus on välillä $55,5\% - 60,5\%$.

Kyselyn 2 mukaan Niinistön kannatusosuus on $68\% \pm 2,9\%$ eli välillä $65,1\% - 70,9\%$.

Kahdessa eri kyselyssä saadut arviot Niinistön kannatusosuudelle eivät ole päällekkäiset. Virhemarginaalit eivät siis selitä eroa kannatusmittauksissa.



c) Eri otoksiin osuu eri joukko ihmisiä, joten otosten perusteella lasketut kannatusosuudet (58% ja 68%) eroavat sekä toisistaan että kannatusosuudesta koko perusjoukossa.

Tutkimuksen virhemarginaali kuvaa ainoastaan *tilastollista* virhettä, eli otannasta johtuvaa vaihtelua. Kun puhutaan virhemarginaalista, oletetaan, että sattuma on ainut tekijä, joka vaikuttaa siihen, kenet perusjoukosta tulevat vastaajaksi. Todellisuudessa näin harvemmin on, vaan mielipidekyselyjen tuloksiin vaikuttavat monet muutkin tekijät. Esimerkiksi:

- Otanta voi epäonnistua, ja vastaajiksi valikoituu jokin tietty ihmisryhmä (tätä ilmiötä kutsutaan valikoitumisharhaksi).
- Kaikki vastaajat eivät anna sitä vastausta, miten he todellisuudessa ajattelevat tai toimivat (tätä ilmiötä kutsutaan vastausharhaksi).

262.

a) Viallisten tuotteiden osuus otoksessa oli $\hat{p} = 0,0630 = 6,3 \%$.

Tutkimuksen virhemarginaali oli $ME = 0,01399 \approx 1,4 \%$

Tutkimuksen mukaan virheellisten tuotteiden osuus tuote-erässä on siis $6,3 \% \pm 1,4 \%$ -yksikköä eli $4,9 \% - 7,7 \%$.

Luottamusvälin

alaraja on $6,3 \% - 1,4 \% = 4,9 \%$,

yläraja on $6,3 \% + 1,4 \% = 7,7 \%$.

b) Tutkimuksen luottamustaso on 99 %, joten 99 %:n varmuudella voidaan sanoa, että viallisia tuotteita on vähintään 4,9 %, mutta korkeintaan 7,7 %.

- $0,049 \cdot 100\,000 = 4900$ (kappaletta)
- $0,077 \cdot 100\,000 = 7700$ (kappaletta)

Koko tuote-erässä viallisten lukumäärä on siis vähintään 4900 mutta korkeintaan 7700.

263.

Tutkitaan asiaa luottamusvälien avulla.

Miehet:

$$\hat{p} = 0,6440 = 64,4 \%$$

$$ME = 0,0316 \approx 3,2 \%$$

Luottamusvälin

- alaraja on $0,6124 \approx 61,2 \%$
- yläraja on $0,6756 \approx 67,6 \%$

Luottamusväli Spotify-palvelua käyttävien miesten prosenttiosuudelle on siis $64,4 \% \pm 3,2 \%$ eli $61,2 \% - 67,6 \%$.

Naiset:

$$\hat{p} = 0,5959 \approx 59,6 \%$$

$$ME = 0,0333 \approx 3,3 \%$$

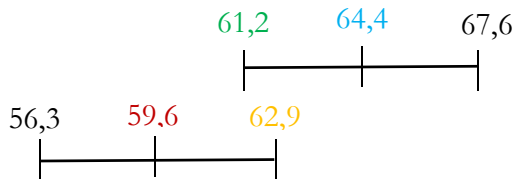
Luottamusvälin

- alaraja on $0,5626 \approx 56,3 \%$
- yläraja on $0,6292 \approx 62,9 \%$

Luottamusväli Spotify-palvelua käyttävien naisten prosenttiosuudelle on siis $59,6 \% \pm 3,3 \%$ eli $56,3 \% - 62,9 \%$.

Otoksessa miehet käyttivät naisia enemmän Spotify-palvelua ($64,4 \% > 59,6 \%$).

Tutkimuksen perusteella saadut luottamusvälit ovat kuitenkin osin päällekkäiset: miesten luottamusvälin alaraja (61,2 %) on pienempi kuin naisten luottamusvälin yläraja (62,9 %). Luottamusvälien perusteella ei siis voida sanoa, että miehet käyttäisivät palvelua naisia enemmän.



264.

Edellisen vuoden tietoja käytetään ratkaisussa ”otoksena”, jossa $n = 612$.

Takuuaikana korjattiin 92 autoa, joten korjattujen autojen prosenttiosuus on

$$\hat{p} = \frac{92}{612} = 0,15032 \dots \approx 15,03 \%,$$

$$\text{jolloin } 1 - \hat{p} = 1 - 0,1503 \dots = 0,8496 \dots$$

Prosenttiosuuden 95 %:n virhemarginaali on

$$\pm 1,96 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

eli virhemarginaaliksi saadaan

$$\pm 1,96 \cdot \sqrt{\frac{0,1503 \dots \cdot 0,8496 \dots}{612}} = 0,02831 \dots \approx 2,83 \%.$$

Edellisvuoden tietojen perusteella takuuaikana korjaukseen tulevien autojen prosenttiosuus on $15,03 \% \pm 2,83 \%$ eli prosenttiosuus on välillä **12,20 %** – **17,86 %**.

Seuraavana vuonna myytävien autojen lukumäärä kasvaa 25 %, joten autoja myydään

$$1,25 \cdot 612 = 765 \text{ (kappaletta).}$$

Takuukorjaukseen tulevien autojen lukumäärälle saadaan:

- alaraja $0,122 \cdot 765 = 93,33 \approx 90$ (autoa)
- yläraja $0,1786 \cdot 765 = 136,629 \approx 140$ (autoa).

Seuraavana vuonna takuukorjaukseen tulevien autojen määrä on siis välillä 90–140 autoa.

265.

Tuote A:

$$n = 600$$

$$\hat{p} = \frac{215}{600} = 0,3583 \dots$$

$$1 - \hat{p} = 1 - 0,3583 \dots = 0,6417 \dots$$

Virhemarginaaliksi saadaan

$$\pm 1,96 \cdot \sqrt{\frac{0,3583 \dots \cdot 0,6417 \dots}{600}} = \pm 0,0383 \dots \approx \pm 3,8 \%,$$

joten tuotteen A käyttäjien prosenttiosuus on välillä 35,8 % $\pm$ 3,8 % eli
32,0 % – 39,6 %.

Tuote B:

$$n = 600$$

$$\hat{p} = \frac{198}{600} = 0,33$$

$$1 - \hat{p} = 1 - 0,33 = 0,67$$

Virhemarginaaliksi saadaan

$$\pm 1,96 \cdot \sqrt{\frac{0,33 \cdot 0,67}{600}} = \pm 0,0376 \dots \approx \pm 3,8 \%,$$

joten tuotteen B käyttäjien prosenttiosuus on välillä 33,0 % $\pm$ 3,8 % eli
29,2 % – 36,8 %.

Saadut luottamusväli ovat osittain päällekkäiset. Tämän perusteella ei siis voida sanoa, että tuotteen A markkinaosuus (eli tuotetta käyttävien kuluttajien prosenttiosuus) olisi suurempi kuin tuotteella B.

266.

Prosenttiosuuden 95 %:n virhemarginaali on

$$\pm 1,96 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

$$n = 1000$$

$$\hat{p} = 0,81$$

$$1 - \hat{p} = 1 - 0,81 = 0,19$$

Virhemarginaaliksi saadaan

$$\pm 1,96 \cdot \sqrt{\frac{0,81 \cdot 0,19}{1000}} = \pm 0,0243 \dots \approx \pm 2 \%$$

eli ± 2 prosenttiyksikköä.

Luottamusväli on 81 % ± 2 %-yksikköä.

Luottamusvälin

- alaraja on $81 - 2 = 79$ (%)
- yläraja on $81 + 2 = 83$ (%).

Nykyistä asevelvollisuusarmeijaa kannattavien suomalaisten prosenttiosuus on siis kyselyn perusteella välillä 79 % – 83 %.

267.

a)

$$n = 1000$$

$$\hat{p} = \frac{644}{1000} = 0,644 = 64,4 \%$$

$$1 - \hat{p} = 1 - 0,644 = 0,356$$

Tutkimuksen virhemarginaali on

$$\pm 1,96 \cdot \sqrt{\frac{0,644 \cdot 0,356}{1000}} = \pm 0,0296 \dots \approx \pm 3,0 \%$$

eli $\pm 3,0$ prosenttiyksikköä.

Lisäydinvoiman rakentamista vastustavien luottamusväli on $64,4 \% \pm 3,0 \%$ -yksikköä.

Luottamusvälin

- alaraja on $64,4 - 3,0 = 61,4 \%$
- yläraja on $64,4 + 3,0 = 67,4 \%$.

Luottamusväliksi saadaan siis $61,4 \% - 67,4 \%$.

b) Aikaisemmassa kyselyssä lisäydinvoiman rakentamista vastusti 51 % vastaajista. Prosenttiosuus ei ole a-kohdassa muodostetulla luottamusvälillä $61,4 \% - 67,4 \%$. Tämän perusteella on siis perusteita väittää, että lisäydinvoiman vastustus on lisääntynyt.

268.

a) Prosenttiosuuden 95 %:n virhemarginaali on

$$\pm 1,96 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

Tapa 1: Päätellään vastaus virhemarginaalin laskukaavasta.

Otoskoko n on kaavassa nimittäjässä (jakajana). Kun lausekkeen nimittäjä kasvaa, lausekkeen arvo pienenee.

Kun otoskoko on $n = 1060$, on virhemarginaali ± 3 prosenttiyksikköä. Isommalla otoskoolla $n = 2000$ (> 1060) virhemarginaali on pienempi kuin ± 3 prosenttiyksikköä.

Tapa 2: Lasketaan virhemarginaali.

$$n = 2000$$

$$\hat{p} = 46 \% = 0,46$$

$$1 - \hat{p} = 1 - 0,46 = 0,54$$

Virhemarginaaliksi saadaan

$$\pm 1,96 \cdot \sqrt{\frac{0,46 \cdot 0,54}{2000}} = 0,0218 \dots \approx 2,18 \%$$

Virhemarginaali olisi siis 2,18 prosenttiyksikköä, eli pienempi kuin ± 3 prosenttiyksikköä.

b) Luottamustaso määrää virhemarginaalin laskukaavassa

$$\pm z^* \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

olevan kertoimen z^* arvon. Luottamustason kasvaessa kerroin z^* kasvaa.

Kun luottamustaso on 95 %, virhemarginaali on ± 3 prosenttiyksikköä. Matalammalla luottamustasolla 90 % (< 95 %) kerroin on pienempi, eli virhemarginaali on pienempi kuin ± 3 prosenttiyksikköä.

c) Perusjoukon koko ei vaikuta tutkimuksen virhemarginaaliin. Virhemarginaali olisi siis edelleen ± 3 prosenttiyksikköä.

269.

a)

$$n = 2000$$

$$\hat{p} = 16,6 \% = 0,166$$

$$1 - \hat{p} = 1 - 0,166 = 0,834$$

Virhemarginaaliksi saadaan

$$\pm 1,96 \cdot \sqrt{\frac{0,166 \cdot 0,834}{2000}} = \pm 0,0163 \dots \approx \pm 1,6 \%$$

Virhemarginaali on siis 1,6 prosenttiyksikköä suuntaansa.

Puolueen A kannatus on 16,6 % ja sen mukaan laskettu virhemarginaali (1,6 pros.yks.) on pienempi kuin suurimman puolueen perusteella laskettu virhemarginaali (1,7 pros.yks.)

b) Ei päde. Erisuuruisilla kannatusprosentteilla on erisuuret virhemarginaalit. Mielipidemittauksissa ilmoitetaan yleensä suurimman puolueen mukaan laskettu virhemarginaali.