
Johdatus tilastotieteeseen
Väliestimointi

Väliestimointi

Todennäköisyysjakaumien parametrien estimointi

Luottamusväli

Normaalijakauman odotusarvon luottamusväli

Normaalijakauman varianssin luottamusväli

Bernoulli-jakauman odotusarvon luottamusväli

Väliestimointi:

Mitä opimme? – 1/3

- **Tilastollisen tutkimuksen** tavoitteena on tehdä *johtopäätöksiä prosesseista tai mekanismeista*, jotka *generoivat* reaali maailman ilmiöitä koskevia **havaintoja**.
- Tavoitteeseen pyritään rakentamalla havainnot generoineille prosesseille tai mekanismeille **tilastollisia malleja**.
- Koska tilastollisten tutkimusasetelmien havaintoihin liittyy *aina satunnaisuutta* tai **epävarmuutta**, tilastollisina malleina käytetään **todennäköisyysmalleja**.
- Tilastollinen malli havainnot generoineelle prosessille tai mekanismille *on täysin määrätty*, jos havaintojen **todennäköisyysjakauma** tunnetaan.

Väliestimointi:

Mitä opimme? – 2/3

- Tilastollisena mallina käytettävän todennäköisyysjakauman **parametreja** *ei* sovellustilanteessa yleensä *tunneta*.
- Jakauman parametreille on löydettävä **estimaatit** eli *arviot*, jotta jakaumaa voisi hyödyntää mallina.
- *Tilastollisen tutkimuksen* tärkeimpiä osatehtäviä on **estimoida** eli *arvioida* mallina käytettävän todennäköisyysjakauman tuntemattomat parametrin *ilmiötä koskevista havainnoista*.
- *Havaintojen funktiota*, joka tuottaa ilmiötä koskeviin havaintoihin sovellettuna **estimaatteja** eli *arvioita* todennäköisyysjakauman tuntemattomalle parametrille, kutsutaan parametrin **estimaattoriksi**.

Väliestimointi: Mitä opimme? – 3/3

- *Tilastotieteen* tärkeimpiä osatehtäviä on *johtaa* estimaattoreita todennäköisyysjakauman parametreille.
- Estimaattoreiden johtamiseen käytetään tavallisesti joko **suurimman uskottavuuden menetelmää** tai **momenttimenetelmää**.
Ks. lukua **Estimointimenetelmät**.
- Todennäköisyysjakauman tuntemattomien parametrien *arvojen* määräämistä kutsutaan usein **piste-estimoinniksi** erotukseksi **väli-estimoinnista**, jossa parametreihin liitetään **luottamusväleiksi** kutsutut välit, *joka sisältävät parametrien todelliset arvot soveltajan valittavissa olevilla todennäköisyyksillä*.

Väliestimointi: Esitiedot

- Esitiedot: ks. seuraavia lukuja:
 - Tilastollisten aineistojen kerääminen ja mittaaminen**
 - Tilastollisten aineistojen kuvaaminen**
 - Otos ja otosjakaumat**
 - Estimointi**
 - Estimointimenetelmät**
- Tarvitset esitietoja myös seuraavista kalvokokoelman **Johdatus todennäköisyyslaskentaan** luvuista:
 - Satunnaismuuttujat ja todennäköisyysjakaumat**
 - Jakaumien tunnusluvut**
 - Diskreettejä jakaumia**
 - Jatkuvia jakaumia**
 - Normaalijakaumasta johdettuja jakaumia**

Väliestimointi: Lisätiedot

- Todennäköisyysjakaumien *parametreja koskevien tilastollisten hypoteesien testaamista* käsitellään luvuissa
 - Tilastolliset testit**
 - Testit suhdeasteikollisille muuttujille**
- *Jakaumaoletuksien testaamista* käsitellään luvussa
 - Yhteensopivuuden, homogeenisuuden ja riippumattomuuden testaaminen**

Väliestimointi

>> Todennäköisyysjakaumien parametrien estimointi

Luottamusväli

Normaalijakauman odotusarvon luottamusväli

Normaalijakauman varianssin luottamusväli

Bernoulli-jakauman odotusarvon luottamusväli

Parametrien estimointi

Avainsanat

Estimaatti

Estimaattori

Estimointi

Havainto

Havaintoarvo

Parametri

Otos

Otosjakauma

Parametri

Piste-estimointi

Tilastollinen aineisto

Tilastollinen malli

Todennäköisyysjakauma

Väliestimointi

Todennäköisyysjakaumat tilastollisten aineistojen kuvaajina

- **Tilastollinen aineisto** koostuu tutkimuksen kohteita kuvaavien muuttujien *havaituista arvoista*.
- Tilastollisissa tutkimusasetelmissä havaintoihin liittyy aina *epävarmuutta ja satunnaisuutta*.
- Tilastollisissa tutkimusasetelmissä tutkimuksen kohteita kuvaavat muuttujat tulkitaan *satunnaismuuttujiksi*, jotka *generoivat* muuttujien havaitut arvot.
- **Tilastollisella mallilla** tarkoitetaan *havaintoarvot generoineiden satunnaismuuttujien todennäköisyysjakaumaa*.

Todennäköisyysjakaumien parametrit 1/2

- Tarkastellaan jotakin tutkimuksen kaikkien mahdollisten kohteiden muodostaman perusjoukon S alkioiden ominaisuutta kuvaavaa *satunnaismuuttujaa* X .
- Oletetaan, että satunnaismuuttuja X noudattaa *todennäköisyysjakaumaa*, jonka *pistetodennäköisyys-* tai *tiheysfunktio*

$$f(x; \theta)$$

riippuu **parametrilla** θ .

- Merkintä:

$$X \sim f(x; \theta)$$

Todennäköisyysjakaumien parametrit 2/2

- Satunnaismuuttujan X pistetodennäköisyys- tai tiheysfunktio

$$f(x ; \theta)$$

kuvaa satunnaismuuttujan X *todennäköisyysjakaumaa* ja parametri θ kuvaa jotakin jakauman *karakteristista ominaisuutta*.

- Koska parametrin θ arvoa *ei* sovellustilanteessa *yleensä tunneta*, tilastollisen tutkimuksen tärkeimpiä osatehtäviä on **estimoida** eli *arvioida* tuntemattomalle parametrille θ sopiva arvo jakaumasta $f(x ; \theta)$ *poimitun otoksen perusteella*.

Todennäköisyysjakaumien parametrien estimointi

Yksinkertainen satunnaisotos

- Olkoon

$$X_1, X_2, \dots, X_n$$

yksinkertainen satunnaisotos jakaumasta, jonka pistetodennäköisyys- tai tiheysfunktio $f(x; \theta)$ riippuu parametrasta θ .

- Tällöin havainnot X_1, X_2, \dots, X_n ovat *riippumattomia, identtisesti jakautuneita satunnaismuuttujia*, joilla on *sama pistetodennäköisyys- tai tiheysfunktio $f(x; \theta)$* :

$$X_1, X_2, \dots, X_n \perp$$

$$X_i \sim f(x; \theta), i = 1, 2, \dots, n$$

Todennäköisyysjakaumien parametrien estimointi

Havainnot ja havaintoarvot

- Oletetaan, että satunnaismuuttujat (havainnot)

$$X_1, X_2, \dots, X_n$$

saavat *poimitussa otoksessa* **havaituiksi arvoikseen** luvut

$$x_1, x_2, \dots, x_n$$

- **Havaintoarvot**

$$x_1, x_2, \dots, x_n$$

vaihtelevat satunnaisesti otoksesta toiseen jakaumasta

$$f(x; \theta)$$

saatavin todennäköisyyksin.

Todennäköisyysjakaumien parametrien estimointi

Estimaattorit ja estimaatit 1/2

- Oletetaan, että todennäköisyysjakauman $f(x; \theta)$ parametrin θ estimointiin käytetään satunnaismuuttujien X_1, X_2, \dots, X_n funktiota eli *tunnuslukua*

$$T = g(X_1, X_2, \dots, X_n)$$

- Tällöin funktiota $T = g(X_1, X_2, \dots, X_n)$ kutsutaan parametrin θ **estimaattoriksi** ja *havaintoarvoista*

$$x_1, x_2, \dots, x_n$$

laskettua funktion g arvoa

$$t = g(x_1, x_2, \dots, x_n)$$

kutsutaan parametrin θ **estimaatiksi**.

Todennäköisyysjakaumien parametrien estimointi

Estimaattorit ja estimaatit 2/2

- Olkoon

$$T = g(X_1, X_2, \dots, X_n)$$

jakauman $f(x; \theta)$ parametrin θ *estimaattori*.

- Tällöin estimaattorin T havaintoarvoista

$$x_1, x_2, \dots, x_n$$

laskettu arvo eli *estimaatti*

$$t = g(x_1, x_2, \dots, x_n)$$

on satunnaismuuttujan T arvon realisaatio otoksessa.

Todennäköisyysjakaumien parametrien estimointi

Estimaattoreiden johtaminen

- *Hyvien estimaattoreiden johtaminen todennäköisyysjakaumien tuntemattomille parametreille on teoreettisen tilastotieteen keskeisiä ongelmia.*
 - Tärkeimmät estimaattoreiden johtamiseen käytettävät menetelmät:
 - **Momenttimenetelmä**
 - **Suurimman uskottavuuden menetelmä**
- Ks. lukua Estimointimenetelmät.

Todennäköisyysjakaumien parametrien estimointi

Piste-estimointi ja väliestimointi

- Todennäköisyysjakauman *parametrin arvon estimointia* kutsutaan usein **piste-estimoinniksi**.
- Parametrin estimaattiin on aina syytä liittää **luottamusväliksi** kutsuttu *väli, joka sisältää estimoidun parametrin todellisen, mutta tuntemattoman arvon tietyllä, soveltajan valittavissa olevalla todennäköisyydellä*.
- Luottamusvälin määräämistä kutsutaan **väliestimoinniksi**.

Väliestimointi

Todennäköisyysjakaumien parametrien estimointi

>> Luottamusväli

Normaalijakauman odotusarvon luottamusväli

Normaalijakauman varianssin luottamusväli

Bernoulli-jakauman odotusarvon luottamusväli

Luottamusväli

Avainsanat

Estimaattori

Estimointi

Frekvenssitulkinta

Luottamuskerroin

Luottamustaso

Luottamusväli

Luottamusväli ja luottamustaso

- *Väliestimoinnissa todennäköisyysjakauman $f(x ; \theta)$ tuntemattomalle parametrille θ pyritään määräämään havainnoista riippuva väli, joka tietyllä, tutkijan valittavissa olevalla todennäköisyydellä, peittää parametrin todellisen arvon.*
- Konstruoitua väliä kutsutaan **luottamusväliksi** ja valittua todennäköisyyttä kutsutaan **luottamustasoksi**.
- Huomautus:
*Luottamustasolle voidaan antaa tulkinta **todennäköisyyden frekvenssitulkinnan** avulla.*

Luottamusvälin määrittäminen 1/3

- Oletukset:

(i) Olkoon $f(x ; \theta)$ satunnaismuuttujan X todennäköisyysjakauma, jonka määrää tuntematon parametri θ .

(ii) Olkoon X_1, X_2, \dots, X_n yksinkertainen satunnaisotos jakaumasta $f(x ; \theta)$.

(iii) Olkoon

$$\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$$

parametrin θ estimaattori.

Luottamusvälin määrittäminen 2/3

- Valitaan **luottamustaso**

$$1 - \alpha$$

ja määrätään sen jälkeen satunnaismuuttujat

$$L = L(X_1, X_2, \dots, X_n)$$

$$U = U(X_1, X_2, \dots, X_n)$$

siten, että

$$\Pr(\hat{\theta} - L \leq \theta) = \alpha / 2$$

$$\Pr(\hat{\theta} + U \geq \theta) = \alpha / 2$$

- Huomautus:

Satunnaismuuttujat L ja U riippuvat sekä havainnoista X_1, X_2, \dots, X_n että valitusta luottamustasosta $(1 - \alpha)$.

Luottamusvälin määrittäminen 3/3

- Tällöin väli

$$(\hat{\theta} - L, \hat{\theta} + U)$$

on parametrin θ **luottamusväli luottamustasolla** $(1 - \alpha)$.

- Luottamusvälin konstruktiosta seuraa, että väli

$$(\hat{\theta} - L, \hat{\theta} + U)$$

peittää tuntemattoman parametrin θ todellisen arvon todennäköisyydellä $(1 - \alpha)$:

$$\Pr(\hat{\theta} - L \leq \theta \leq \hat{\theta} + U) = 1 - \alpha$$

Luottamusvälin määrittäminen: Erikoistapaus

- Jos estimaattorin $\hat{\theta}$ jakauma on *symmetrinen*, parametrin θ **luottamusväli luottamustasolla** $(1 - \alpha)$ on muotoa

$$\hat{\theta} \pm A$$

jossa satunnaismuuttuja

$$A = A(X_1, X_2, \dots, X_n)$$

valitaan siten, että

$$\Pr(\hat{\theta} - A \leq \theta \leq \hat{\theta} + A) = 1 - \alpha$$

- Huomautus:

Satunnaismuuttuja A riippuu sekä havainnoista X_1, X_2, \dots, X_n että valitusta luottamustasosta $(1 - \alpha)$.

Luottamustason ja -välin frekvenssitulkinta

- Oletetaan, että luottamustasoksi on valittu $(1 - \alpha)$.
- Luottamustasolle ja siihen liittyvälle luottamusvälille voidaan antaa seuraava **frekvenssitulkinta**:
 - (i) Jos otantaa jakaumasta $f(x ; \theta)$ toistetaan, *keskimäärin* $100 \times (1 - \alpha) \%$ otoksista konstruoiduista luottamusväleistä *peittää* parametrin θ todellisen arvon.
 - (ii) Jos otantaa jakaumasta $f(x ; \theta)$ toistetaan, *keskimäärin* $100 \times \alpha \%$ otoksista konstruoiduista luottamusväleistä *ei peitä* parametrin θ todellista arvoa.

Johtopäätökset luottamusväleistä

- Oletetaan, että olemme tehneet *johtopäätöksen*, että konstruoitu luottamusväli peittää parametrin θ tuntemattoman todellisen arvon:
 - (i) Luottamusvälin konstruktiosta seuraa, että tehty johtopäätös *on oikea* $100 \times (1 - \alpha)$ %:ssa tapauksia.
 - (ii) Luottamusvälin konstruktiosta seuraa, että tehty johtopäätös *on väärä* $100 \times \alpha$ %:ssa tapauksia.
- Huomautus:

Virheellisen johtopäätöksen mahdollisuutta *ei saada häviämään*, ellei luottamusväliä tehdä *äärettömän leveäksi*, jolloin väli *ei enää sisällä informaatiota* parametrin θ oikeasta arvosta.

Luottamusvälit: Esimerkkejä

- Olkoon

$$X_1, X_2, \dots, X_n$$

yksinkertainen satunnaisotos jakaumasta $f(x; \theta)$.

- Tarkastellaan seuraavien jakaumien parametrien luottamusvälejä (ks. kalvokokoelman **Johdatus todennäköisyyslaskentaan lukuja Diskreettejä jakaumia ja Jatkuvia jakaumia**):
 - **Normaalijakauman odotusarvon luottamusväli**
 - **Normaalijakauman varianssin luottamusväli**
 - **Bernoulli-jakauman odotusarvoparametrin luottamusväli**

Väliestimointi

Todennäköisyysjakaumien parametrien estimointi

Luottamusväli

>> Normaalijakauman odotusarvon luottamusväli

Normaalijakauman varianssin luottamusväli

Bernoulli-jakauman odotusarvon luottamusväli

Normaalijakauman odotusarvon luottamusväli

Avainsanat

Frekvenssitulkinta

Luottamuskerroin

Luottamustaso

Luottamusväli

Normaalijakauma

Odotusarvo

Normaalijakauman odotusarvon luottamusväli

Esimerkki – 1/4

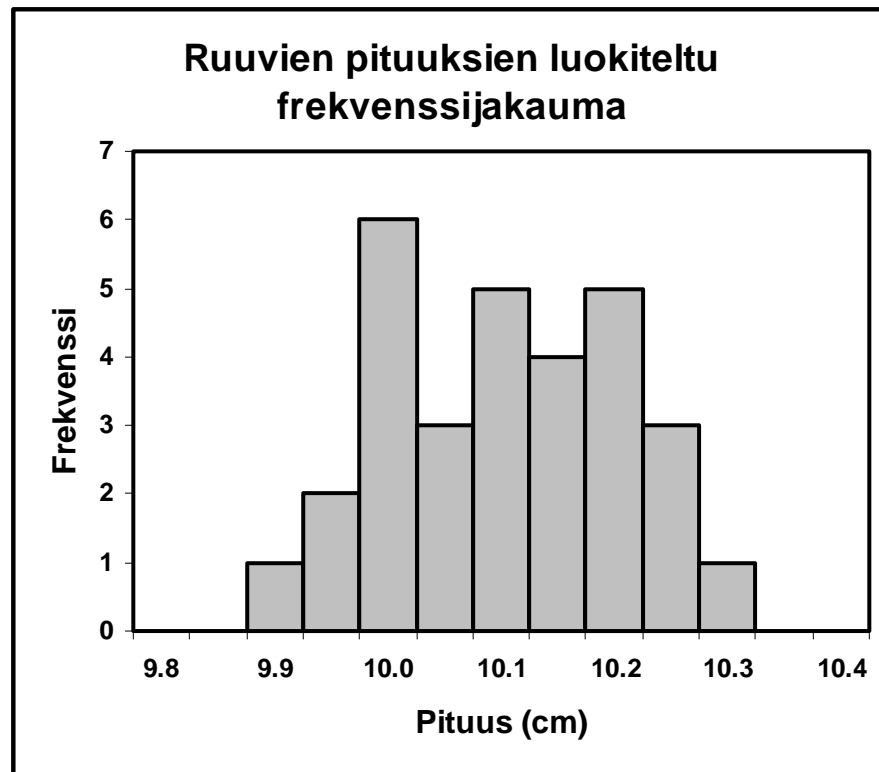
- Kone tekee *ruuveja*, joiden *pituudet vaihtelevat satunnaisesti noudattaen normaalijakaumaa*; ks. esimerkkiä luvussa **Tilastollisten aineistojen kuvaaminen**.
- Ruuvien joukosta poimitaan *yksinkertainen satunnaisotos*, jonka *koko* $n = 30$ ja otokseen poimittujen ruuvien pituudet mitataan.
- Taulukko oikealla esittää pituuksien *luokiteltua frekvenssi-jakaumaa*.

Luokkavälit	Luokkafrekvenssit
(9.85,9.90]	1
(9.90,9.95]	2
(9.95,10.00]	6
(10.00,10.05]	3
(10.05,10.10]	5
(10.10,10.15]	4
(10.15,10.20]	5
(10.20,10.25]	3
(10.25,10.30]	1

Normaalijakauman odotusarvon luottamusväli

Esimerkki – 2/4

- Kuva oikealla esittää otokseen poimittujen ruuvien pituuksien luokiteltua frekvenssijakaumaa vastaavaa *histogrammia*.
- *Luokkavälit* määräävät histogrammin suorakaiteiden *kannat*.
- Suorakaiteiden *korkeudet* on valittu niin, että suorakaiteiden *pinta-alat* suhtautuvat toisiinsa kuten vastaavat *luokka-frekvenssit*.



Normaalijakauman odotusarvon luottamusväli

Esimerkki – 3/4

- Yhteenveto otostiedoista:

Pituuksien *aritmeettinen keskiarvo*:

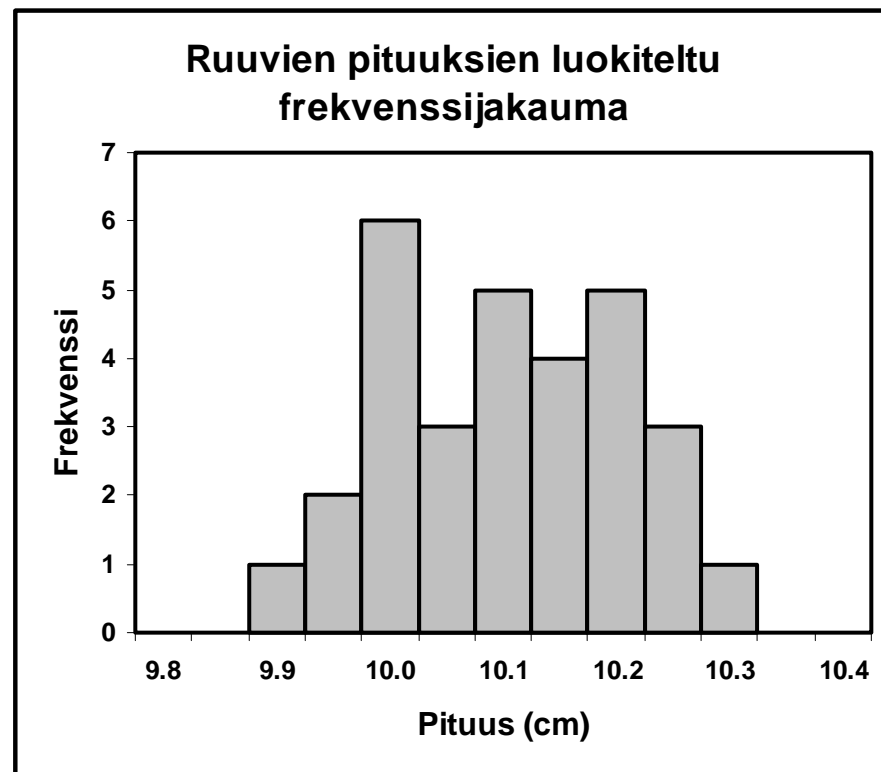
$$\bar{X} = 10.09 \text{ cm}$$

Pituuksien *keskihajonta*:

$$s = 0.1038 \text{ cm}$$

- Huomautus:

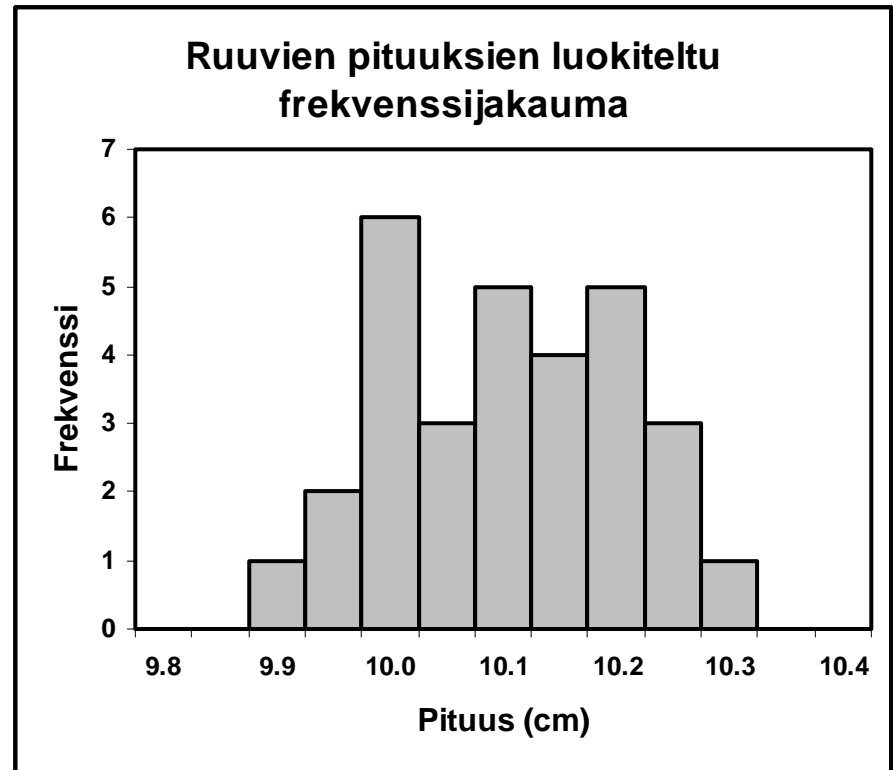
Jos otantaa toistetaan, kaikki otosta koskevat tiedot (sekä havaintoarvot että niistä lasketut otossuureet) *vaihtelevat satunnaisesti* otoksesta toiseen.



Normaalijakauman odotusarvon luottamusväli

Esimerkki – 4/4

- Ongelma:
Mitä koneen tekemien ruuvien *todellisesta keskipituudesta* voidaan tietää *yhdestä otoksesta* saatujen tietojen perusteella?
- Ratkaisu:
Konstruoidaan *väli, joka valitulla todennäköisyydellä sisältää ruuvien todellisen keskipituuden.*



Normaalijakauma ja sen parametointi

- Satunnaismuuttuja X noudattaa **normaalijakaumaa** $N(\mu, \sigma^2)$, jos sen *tiheysfunktio* on

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

$$-\infty < \mu < +\infty, \sigma > 0$$

- Normaalijakauman *parametreina* ovat jakauman *odotusarvo*

$$E(X) = \mu$$

ja *varianssi*

$$\text{Var}(X) = \sigma^2$$

Normaalijakauman odotusarvon luottamusväli

Otos normaalijakaumasta

- Olkoon

$$X_1, X_2, \dots, X_n$$

yksinkertainen satunnaisotos normaalijakaumasta

$$N(\mu, \sigma^2)$$

- Tällöin satunnaismuuttujat X_1, X_2, \dots, X_n ovat *riippumattomia* ja noudattavat *samaa normaalijakaumaa*

$$N(\mu, \sigma^2)$$

Normaalijakauman parametrien estimointi

- *Estimoidaan* normaalijakauman $N(\mu, \sigma^2)$ parametrit μ ja σ^2 niiden *harhattomilla estimaattoreilla*:

(i) *Odotusarvoparametrin* μ harhaton estimaattori:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

(ii) *Varianssiparametrin* σ^2 harhaton estimaattori:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Normaalijakauman odotusarvon luottamusväli

Luottamustaso

- Määrätään **luottamusväli normaalijakauman odotusarvoparametrille μ** .
- Valitaan **luottamustasoksi**
 $1 - \alpha$
- Luottamustaso kiinnittää todennäköisyyden, jolla konstruoitava luottamusväli peittää normaalijakauman odotusarvon μ todellisen arvon.

Normaalijakauman odotusarvon luottamusväli

Luottamuskertoimet

- Olkoon valittu luottamustaso $(1 - \alpha)$.
- Määrätään **luottamuskertoimet** $-t_{\alpha/2}$ ja $+t_{\alpha/2}$ siten, että

$$\Pr(t \leq -t_{\alpha/2}) = \frac{\alpha}{2}$$

$$\Pr(t \geq +t_{\alpha/2}) = \frac{\alpha}{2}$$

jossa satunnaismuuttuja t noudattaa *Studentin t-jakaumaa* vapausastein $(n - 1)$:

$$t \sim t(n - 1)$$

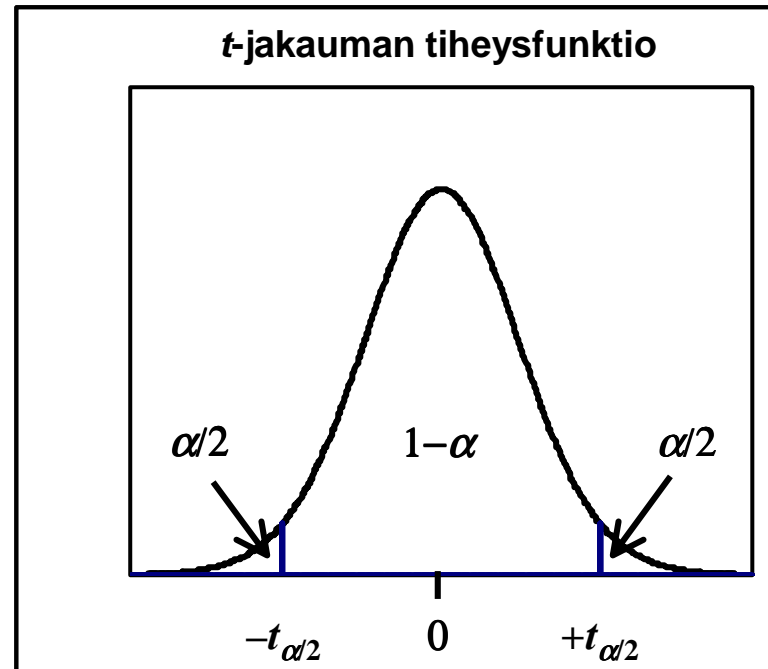
- Luottamuskertoimet $-t_{\alpha/2}$ ja $+t_{\alpha/2}$ toteuttavat ehdon

$$\Pr(-t_{\alpha/2} \leq t \leq +t_{\alpha/2}) = 1 - \alpha$$

Normaalijakauman odotusarvon luottamusväli

Luottamuskertoimien määrittäminen: Havainnollistus

- Luottamuskertoimet $-t_{\alpha/2}$ ja $+t_{\alpha/2}$ jakavat t -jakauman tiheysfunktion kuvaajan alle jäävän todennäköisyysmassan (= 1) kolmeen osaan:
 - (1) Pisteeseen $-t_{\alpha/2}$ vasemmalle puolelle jää $\alpha/2$ % massasta.
 - (2) Pisteeseen $+t_{\alpha/2}$ oikealle puolelle jää $\alpha/2$ % massasta.
 - (3) Pisteiden $-t_{\alpha/2}$ ja $+t_{\alpha/2}$ väliin jää $(1 - \alpha)$ % massasta.
- Huomaa, että
$$\alpha/2 + \alpha/2 + (1 - \alpha) = 1$$



Luottamusväli normaalijakauman odotusarvolle

- Normaalijakauman **odotusarvoparametrin μ luottamusväli luottamustasolla $(1 - \alpha)$** on muotoa

$$\left(\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right)$$

jossa

\bar{X} = havaintojen *aritmeettinen keskiarvo*

s^2 = havaintojen harhaton *otosvarianssi*

n = havaintojen *lukumäärä*

$-t_{\alpha/2}, +t_{\alpha/2}$

= luottamustasoon $(1 - \alpha)$ liittyvät *luottamuskertoimet t -jakaumasta vapausastein $(n - 1)$*

Normaalijakauman odotusarvon luottamusväli

Luottamusväli ja sen pituus 1/2

- Normaalijakauman odotusarvon μ luottamusväli luottamustasolla $(1 - \alpha)$ esitetään usein muodossa

$$\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

koska väli on symmetrinen keskipisteensä \bar{X} suhteen.

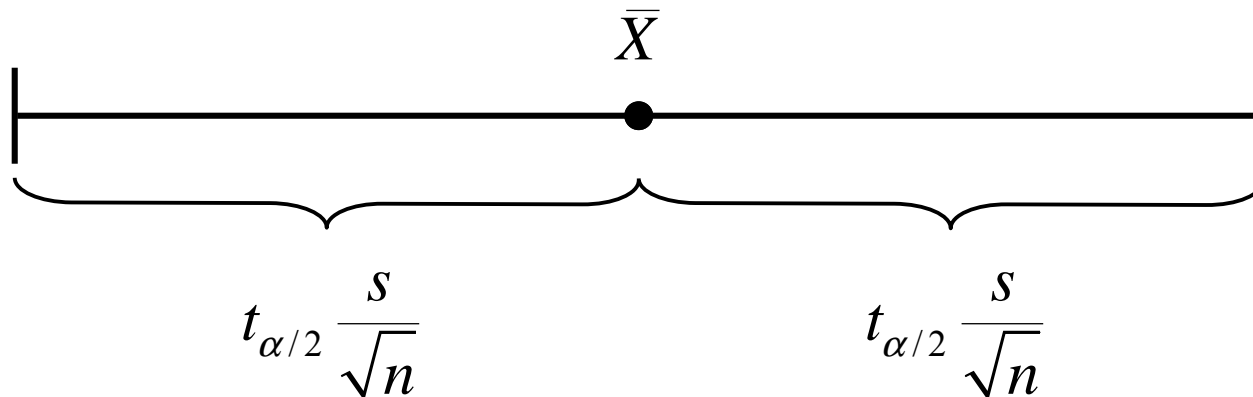
- Normaalijakauman odotusarvon μ **luottamusvälin pituus** on

$$2 \times t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Normaalijakauman odotusarvon luottamusväli

Luottamusväli ja sen pituus 2/2

- Normaalijakauman odotusarvon μ luottamusväli luottamustasolla $(1 - \alpha)$:



Normaalijakauman odotusarvon luottamusväli

Luottamusvälin johto 1/8

- Olkoon

$$X_1, X_2, \dots, X_n$$

yksinkertainen satunnaisotos normaalijakaumasta

$$N(\mu, \sigma^2)$$

- Olkoon

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

havaintojen X_1, X_2, \dots, X_n *aritmeettinen keskiarvo* ja

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

havaintojen X_1, X_2, \dots, X_n (harhaton) *otosvarianssi*.

Normaalijakauman odotusarvon luottamusväli

Luottamusvälin johto 2/8

- Määritellään satunnaismuuttuja

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

- Satunnaismuuttuja t voidaan kirjoittaa muotoon

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma}{\sqrt{s^2}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s^2/\sigma^2}{n-1}}}$$

- Määrätään satunnaismuuttujan t jakauma tarkastelemalla erikseen satunnaismuuttujan t osoittajaa ja nimittäjää.

Normaalijakauman odotusarvon luottamusväli

Luottamusvälin johto 3/8

- Satunnaismuuttujan t osoittaja määrittelee satunnaismuuttujan

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

- Satunnaismuuttuja Z noudattaa normaalijakautuneen otoksen *aritmeettisen keskiarvon otosjakaumaa koskevan tuloksen perusteella standardoitua normaalijakaumaa* $N(0,1)$:

$$Z \sim N(0,1)$$

Todistus: ks. lukua **Otos ja otosjakaumat**.

Normaalijakauman odotusarvon luottamusväli

Luottamusvälin johto 4/8

- Satunnaismuuttuja t nimittäjä määrittelee satunnaismuuttujan

$$V = (n - 1) \frac{s^2}{\sigma^2}$$

- Satunnaismuuttuja V noudattaa normaalijakautuneen otoksen *varianssin otosjakaumaa koskevan tuloksen perusteella* χ^2 -jakaumaa *vapausastein* $(n - 1)$:

$$V \sim \chi^2(n - 1)$$

Todistus: ks. lukua **Otos ja otosjakaumat**.

Normaalijakauman odotusarvon luottamusväli

Luottamusvälin johto 5/8

- Edellä on todettu, että

$$Z \sim N(0,1)$$

$$V \sim \chi^2(n-1)$$

- Lisäksi voidaan osoittaa, että satunnaismuuttujat Z ja V ovat *riippumattomia*; todistus: ks. lukua **Otos ja otosjakaumat**.
- Siten satunnaismuuttuja

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{Z}{\sqrt{\frac{V}{n-1}}}$$

noudattaa *Studentin t-jakaumaa* vapausastein $(n - 1)$ suoraan jakauman *määritelmän mukaan*:

$$t \sim t(n-1)$$

Normaalijakauman odotusarvon luottamusväli

Luottamusvälin johto 6/8

- Määrätään t -jakaumasta vapausastein $(n - 1)$ piste $+t_{\alpha/2}$ siten, että

$$\Pr(t \geq +t_{\alpha/2}) = \frac{\alpha}{2}$$

jolloin (t -jakauman symmetrian perusteella)

$$\Pr(t \leq -t_{\alpha/2}) = \frac{\alpha}{2}$$

ja edelleen

$$\Pr(-t_{\alpha/2} \leq t \leq +t_{\alpha/2}) = 1 - \alpha$$

Normaalijakauman odotusarvon luottamusväli

Luottamusvälin johto 7/8

- Tarkastellaan epäyhtälöketjua

$$-t_{\alpha/2} \leq t \leq +t_{\alpha/2}$$

- Sijoittamalla tähän epäyhtälöketjuun satunnaismuuttujan t lauseke, saadaan epäyhtälöketju

$$-t_{\alpha/2} \leq \frac{\bar{X} - \mu}{s/\sqrt{n}} \leq +t_{\alpha/2}$$

- Tästä epäyhtälöketjusta saadaan sen kanssa *yhtäpitävä* epäyhtälöketju

$$\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Normaalijakauman odotusarvon luottamusväli

Luottamusvälin johto 8/8

- Yhdistämällä saatu epäyhtälö siihen, että

$$\Pr(-t_{\alpha/2} \leq t \leq +t_{\alpha/2}) = 1 - \alpha$$

saadaan vihdoin

$$\Pr\left(\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

Luottamusvälin frekvenssitulkinta 1/2

- Normaalijakauman odotusarvon μ luottamusvälin

$$\left(\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right)$$

konstruktiosta seuraa, että

$$\Pr \left(\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right) = 1 - \alpha$$

- Siten konstruoitu luottamusväli *peittää* parametrin μ todellisen arvon todennäköisyydellä $(1 - \alpha)$ ja se *ei peitä* parametrin μ todellista arvoa todennäköisyydellä α .

Luottamusvälin frekvenssitulkinta 2/2

- Normaalijakauman odotusarvon μ luottamusvälille voidaan antaa seuraava **frekvenssitulkinta**:
 - (i) Jos otantaa jakaumasta $N(\mu, \sigma^2)$ toistetaan, *keskimäärin*
 $100 \times (1 - \alpha) \%$
otoksista konstruoiduista luottamusväleistä *peittää* parametrin μ todellisen arvon.
 - (ii) Jos otantaa jakaumasta $N(\mu, \sigma^2)$ toistetaan, *keskimäärin*
 $100 \times \alpha \%$
otoksista konstruoiduista luottamusväleistä *ei peitä* parametrin μ todellista arvoa.

Normaalijakauman odotusarvon luottamusväli

Luottamusvälin ominaisuudet 1/2

- Normaalijakauman odotusarvon μ luottamusvälin *keskipiste* \bar{X} vaihtelee otoksesta toiseen.
- Luottamusvälin *pituus*

$$2 \times t_{\alpha/2} \frac{s}{\sqrt{n}}$$

vaihtelee otoksesta toiseen.

- Luottamusvälin *pituus* riippuu valitusta luottamustasosta $(1 - \alpha)$, havaintojen lukumäärästä n ja otosvarianssista s^2 .

Normaalijakauman odotusarvon luottamusväli

Luottamusvälin ominaisuudet 2/2

- Luottamusväli *lyhenee (pitenee)*, jos *luottamustasoa* $(1 - \alpha)$ *pienennetään (kasvatetaan)*.
- Luottamusväli *lyhenee (pitenee)*, jos *havaintojen lukumäärää n kasvatetaan (pienennetään)*.
- Luottamusväli *lyhenee (pitenee)*, jos *otosvarianssi s^2 pienenee (kasvaa)*.

Normaalijakauman odotusarvon luottamusväli

Vaatimukset luottamusvälille

- Olisi toivottavaa pystyä konstruoimaan odotusarvolle μ mahdollisimman *lyhyt* luottamusväli, johon liittyvä luottamustaso olisi mahdollisimman *korkea*.
- Vaatimusten samanaikainen täyttäminen *ei ole* kuitenkaan *mahdollista, jos otoskoko pidetään kiinteänä*:
 - (i) *Luottamustason kasvattaminen pidentää luottamusväliä, jolloin tieto parametrin μ todellisen arvon sijainnista tulee epätarkemmaksi.*
 - (ii) *Luottamusvälin lyhentäminen pienentää luottamustasoa, jolloin tieto parametrin μ todellisen arvon sijainnista tulee epävarmemmaksi.*

Normaalijakauman odotusarvon luottamusväli

Johtopäätökset luottamusvälistä

- Oletetaan, että olemme tehneet *johtopäätöksen*, että konstruoitu luottamusväli peittää odotusarvoparametrin μ todellisen arvon:
 - (i) Luottamusvälin konstruktiosta seuraa, että tehty johtopäätös *on oikea* $100 \times (1 - \alpha)$ %:ssa tapauksia.
 - (ii) Luottamusvälin konstruktiosta seuraa, että tehty johtopäätös *on väärä* $100 \times \alpha$ %:ssa tapauksia.
- Huomautus:

Virheellisen johtopäätöksen mahdollisuutta *ei saada häviämään*, ellei luottamusväliä tehdä *äärettömän leveäksi*, jolloin väli *ei enää sisällä informaatiota* odotusarvoparametrin μ todellisesta arvosta.

Normaalijakauman odotusarvon luottamusväli

Esimerkki (jatkuu) – 1/4

- Erään koneen tekemien ruuvien joukosta poimittiin *yksinkertainen satunnaisotos* ja ruuvien pituudet mitattiin.

Otoskoko: $n = 30$

Pituuksien aritmeettinen keskiarvo: $\bar{X} = 10.09$ cm

Pituuksien otoskeskihajonta: $s = 0.1038$ cm

- Konstruoidaan ruuvien *todelliselle keskipituudelle μ luottamusväli luottamustasolla 0.95*.
- Valitaan *luottamuskertoimet $-t_{0.025}$ ja $+t_{0.025}$ siten, että*

$$\Pr(t \leq -t_{0.025}) = \Pr(t \geq +t_{0.025}) = 0.025$$

jossa satunnaismuuttuja t noudattaa *Studentin t -jakaumaa* vapausastein $n - 1 = 29$.

- Luottamuskertoimet $-t_{0.025}$ ja $+t_{0.025}$ toteuttavat ehdon

$$\Pr(-t_{0.025} \leq t \leq +t_{0.025}) = 0.95$$

Normaalijakauman odotusarvon luottamusväli

Esimerkki (jatkuu) – 2/4

- t -jakauman *taulukoista* nähdään, että

$$\Pr(t \geq +2.045) = 0.025$$

$$\Pr(t \leq -2.045) = 0.025$$

kun vapausasteiden lukumäärä

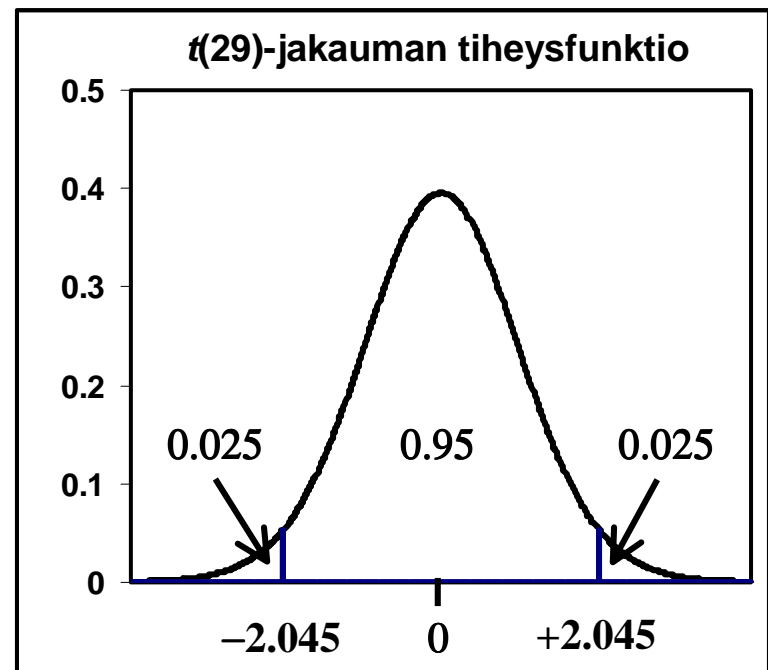
$$n - 1 = 29$$

- Siten *luottamuskertoimet* ovat:

$$+t_{0.025} = +2.045$$

$$-t_{0.025} = -2.045$$

- Kuvio oikealla havainnollistaa luottamuskertoimien valintaa.



Normaalijakauman odotusarvon luottamusväli Esimerkki (jatkuu) – 3/4

- *Luottamusväliksi* saadaan:

$$\begin{aligned}\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} &= 10.09 \pm 2.045 \times \frac{0.1038}{\sqrt{30}} \\ &= 10.09 \pm 0.04 \\ &= (10.05, 10.13)\end{aligned}$$

- Siten tiedämme, että *ruuvien todellinen keskipituus on todennäköisyydellä 0.95 välillä*
(10.05, 10.13)

Normaalijakauman odotusarvon luottamusväli

Esimerkki (jatkuu) – 4/4

- *Luottamustason 0.95 tulkinta:*

Oletetaan, että poimimme koneen tekemien ruuvien joukosta *toistuvasti* yksinkertaisia satunnaisotoksia, joiden koko on 30 ja konstruoinme *jokaisesta* otoksesta 95 %:n luottamusvälin edellä esitetyllä menetelmällä.

Tällöin:

- (i) *Konstruoidusta väleistä keskimäärin 95 % peittää ruuvien todellisen, mutta tuntemattoman keskipituuden.*
- (ii) *Konstruoidusta väleistä keskimäärin 5 % ei peitä ruuvien todellista, mutta tuntematonta keskipituutta.*

Normaalijakauman odotusarvon luottamusväli

Otoskoon määrääminen 1/4

- Monissa tutkimustilanteissa toivotaan, että parametrille voitaisiin muodostaa *mahdollisimman lyhyt* luottamusväli.
- Normaalijakauman odotusarvon μ luottamusvälin pituus riippuu otoskoosta siten, että *väli lyhenee, jos havaintojen lukumäärää n kasvatetaan*.
- Tämä normaalijakauman odotusarvon μ luottamusvälin ominaisuus mahdollistaa tietyin ehdoin otoskoon valitsemisen sellaisella tavalla, että *luottamusväliksi saadaan* (suunnilleen) *halutun mittainen väli*.
- Oletetaan siis, että normaalijakauman odotusarvo-parametrille μ halutaan konstruoida luottamusväli, jonka *toivottu pituus* on $2A$.

Normaalijakauman odotusarvon luottamusväli

Otoskoon määrääminen 2/4

- *Jos normaalijakauman varianssi σ^2 tunnetaan, normaalijakauman odotusarvoparametrin μ luottamusväliksi luottamustasolla $(1 - \alpha)$ saadaan*

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

jossa

\bar{X} = havaintojen *aritmeettinen keskiarvo*

σ^2 = havaintojen *oletettu varianssi*

n = *otoskoko*

$-z_{\alpha/2}$, $+z_{\alpha/2}$

= luottamustasoon $(1 - \alpha)$ liittyvät *luottamuskertoimet normaalijakaumasta* $N(0,1)$

Normaalijakauman odotusarvon luottamusväli

Otoskoon määrääminen 3/4

- Jos siis käytettävissä on *ennakkotietoa perusjoukon varianssista* σ^2 , voidaan otoskoko n ratkaista yhtälöstä:

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = A$$

jossa

$z_{\alpha/2}$ = luottamustasoon $(1 - \alpha)$ liittyvä *luottamuskerroin* normaalijakaumasta $N(0,1)$

σ^2 = havaintojen *oletettu varianssi*

n = *otoskoko*

$2A$ = *toivottu pituus* luottamusvälille

Normaalijakauman odotusarvon luottamusväli

Otoskoon määrääminen 4/4

- Siten *tarvittava otoskoko* on

$$n = \left(\frac{z_{\alpha/2} \sigma}{A} \right)^2$$

- Huomautus:

Tarvittavan otoskoon kaavasta nähdään, että mitä lyhyempää luottamusväliä toivotaan, sitä suurempi otos on poimittava:

Esimerkiksi, jos luottamusvälin pituus halutaan puolittaa, pitää havaintoja kerätä 4 kertaa enemmän.

Väliestimointi

Todennäköisyysjakaumien parametrien estimointi

Luottamusväli

Normaalijakauman odotusarvon luottamusväli

>> Normaalijakauman varianssin luottamusväli

Bernoulli-jakauman odotusarvon luottamusväli

Normaalijakauman varianssin luottamusväli

Avainsanat

Estimointi

Frekvenssitulkinta

Luottamuskerroin

Luottamustaso

Luottamusväli

Normaalijakauma

Odotusarvo

Varianssi

Normaalijakauma ja sen parametointi

- Satunnaismuuttuja X noudattaa **normaalijakaumaa** $N(\mu, \sigma^2)$, jos sen *tiheysfunktio* on

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

$$-\infty < \mu < +\infty, \sigma > 0$$

- Normaalijakauman *parametreina* ovat jakauman *odotusarvo*

$$E(X) = \mu$$

ja *varianssi*

$$\text{Var}(X) = \sigma^2$$

Normaalijakauman varianssin luottamusväli

Otos normaalijakaumasta

- Olkoon

$$X_1, X_2, \dots, X_n$$

yksinkertainen satunnaisotos normaalijakaumasta

$$N(\mu, \sigma^2)$$

- Tällöin satunnaismuuttujat X_1, X_2, \dots, X_n ovat *riippumattomia* ja noudattavat *samaa normaalijakaumaa*

$$N(\mu, \sigma^2)$$

Normaalijakauman parametrien estimointi

- *Estimoidaan* normaalijakauman $N(\mu, \sigma^2)$ parametrit μ ja σ^2 niiden *harhattomilla estimaattoreilla*:

- (i) *Odotusarvoparametrin* μ harhaton estimaattori:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

- (ii) *Varianssiparametrin* σ^2 harhaton estimaattori:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Normaalijakauman varianssin luottamusväli

Luottamustaso

- Määrätään **luottamusväli normaalijakauman varianssiparametrille σ^2** .
- Valitaan **luottamustasoksi**
 $1 - \alpha$
- Luottamustaso kiinnittää todennäköisyyden, jolla konstruoitava luottamusväli peittää normaalijakauman varianssin σ^2 todellisen arvon.

Normaalijakauman varianssin luottamusväli

Luottamuskertoimet

- Olkoon valittu luottamustaso $(1 - \alpha)$.
- Määrätään **luottamuskertoimet** $\chi^2_{1-\alpha/2}$ ja $\chi^2_{\alpha/2}$ siten, että

$$\Pr(\chi^2 \leq \chi^2_{1-\alpha/2}) = \frac{\alpha}{2}$$

$$\Pr(\chi^2 \geq \chi^2_{\alpha/2}) = \frac{\alpha}{2}$$

jossa satunnaismuuttuja χ^2 noudattaa χ^2 -jakaumaa vapausastein $(n - 1)$:

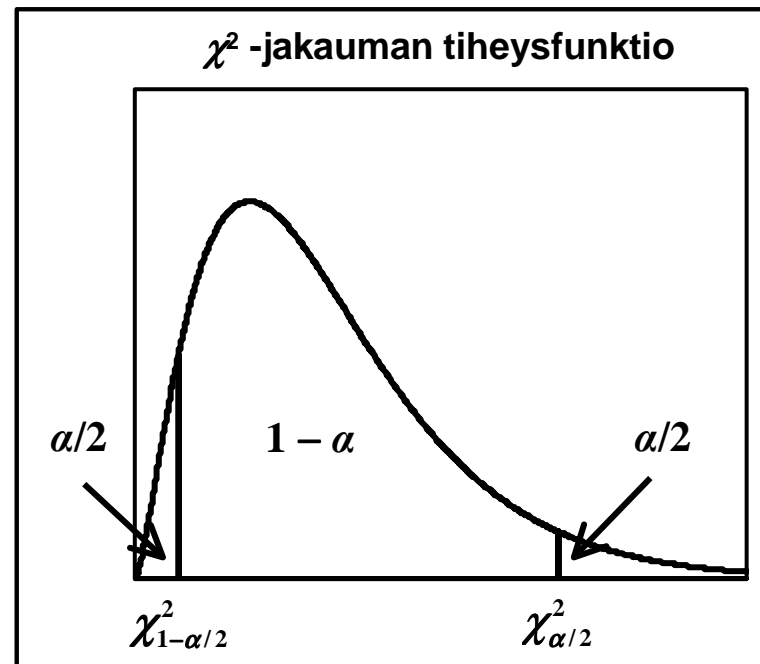
$$\chi^2 \sim \chi^2(n - 1)$$

- Luottamuskertoimet $\chi^2_{1-\alpha/2}$ ja $\chi^2_{\alpha/2}$ toteuttavat ehdon

$$\Pr(\chi^2_{1-\alpha/2} \leq \chi^2 \leq \chi^2_{\alpha/2}) = 1 - \alpha$$

Luottamuskertoimien määrittäminen: Havainnollistus

- Luottamuskertoimet $\chi^2_{1-\alpha/2}$ ja $\chi^2_{\alpha/2}$ jakavat χ^2 -jakauman tiheysfunktion kuvaajan alle jäävän todennäköisyysmassan (= 1) kolmeen osaan:
 - (1) Pisteiden $\chi^2_{1-\alpha/2}$ vasemmalle puolelle jää $\alpha/2$ % massasta.
 - (2) Pisteiden $\chi^2_{\alpha/2}$ oikealle puolelle jää $\alpha/2$ % massasta.
 - (3) Pisteiden $\chi^2_{1-\alpha/2}$ ja $\chi^2_{\alpha/2}$ väliin jää $(1 - \alpha)$ % massasta.
- Huomaa, että
$$\alpha/2 + \alpha/2 + (1 - \alpha) = 1$$



Luottamusväli normaalijakauman varianssille

- Normaalijakauman **varianssiparametrin σ^2 luottamusväli luottamustasolla $(1 - \alpha)$** on muotoa

$$\left(\frac{(n-1)s^2}{\chi_{\alpha/2}^2}, \frac{(n-1)s^2}{\chi_{1-\alpha/2}^2} \right)$$

jossa

s^2 = havaintojen harhaton *otosvarianssi*

n = havaintojen *lukumäärä*

$$\chi_{1-\alpha/2}^2, \chi_{\alpha/2}^2$$

= luottamustasoon $(1 - \alpha)$ liittyvät *luottamuskertoimet χ^2 -jakaumasta vapausastein*

$$(n - 1)$$

Normaalijakauman varianssin luottamusväli

Luottamusväli ja sen pituus

- Normaalijakauman varianssin σ^2 **luottamusvälin pituus** on

$$(n-1)s^2 \left(\frac{1}{\chi_{1-\alpha/2}^2} - \frac{1}{\chi_{\alpha/2}^2} \right)$$

Normaalijakauman varianssin luottamusväli

Luottamusvälin johto 1/5

- Olkoon

$$X_1, X_2, \dots, X_n$$

yksinkertainen satunnaisotos normaalijakaumasta

$$N(\mu, \sigma^2)$$

- Olkoon

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

havaintojen X_1, X_2, \dots, X_n *aritmeettinen keskiarvo* ja

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

havaintojen X_1, X_2, \dots, X_n (harhaton) *otosvarianssi*.

Normaalijakauman varianssin luottamusväli

Luottamusvälin johto 2/5

- Määritellään satunnaismuuttuja

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

- Satunnaismuuttuja χ^2 noudattaa normaalijakautuneen otoksen *varianssin otosjakaumaa koskevan tuloksen perusteella* χ^2 -jakaumaa vapausastein $(n-1)$:

$$\chi^2 \sim \chi^2(n-1)$$

Todistus: ks. lukua **Otos ja otosjakaumat**.

Normaalijakauman varianssin luottamusväli

Luottamusvälin johto 3/5

- Määrätään χ^2 -jakaumasta vapausastein $(n - 1)$ piste $\chi^2_{1-\alpha/2}$ siten, että

$$\Pr(\chi^2 \leq \chi^2_{1-\alpha/2}) = \frac{\alpha}{2}$$

ja piste $\chi^2_{\alpha/2}$ siten, että

$$\Pr(\chi^2 \geq \chi^2_{\alpha/2}) = \frac{\alpha}{2}$$

jolloin

$$\Pr(\chi^2_{1-\alpha/2} \leq \chi^2 \leq \chi^2_{\alpha/2}) = 1 - \alpha$$

Normaalijakauman varianssin luottamusväli

Luottamusvälin johto 4/5

- Tarkastellaan epäyhtälöketjua

$$\chi_{1-\alpha/2}^2 \leq \chi^2 \leq \chi_{\alpha/2}^2$$

- Sijoittamalla tähän epäyhtälöketjuun satunnaismuuttujan χ^2 lauseke, saadaan epäyhtälöketju

$$\chi_{1-\alpha/2}^2 \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi_{\alpha/2}^2$$

- Tästä epäyhtälöketjusta saadaan sen kanssa *yhtäpitävä* epäyhtälöketju

$$\frac{(n-1)s^2}{\chi_{\alpha/2}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{1-\alpha/2}^2}$$

Normaalijakauman varianssin luottamusväli

Luottamusvälin johto 5/5

- Yhdistämällä saatu epäyhtälö siihen, että

$$\Pr(\chi_{1-\alpha/2}^2 \leq \chi^2 \leq \chi_{\alpha/2}^2) = 1 - \alpha$$

saadaan vihdoin

$$\Pr\left(\frac{(n-1)s^2}{\chi_{\alpha/2}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{1-\alpha/2}^2}\right) = 1 - \alpha$$

Luottamusvälin frekvenssitulkinta 1/2

- Normaalijakauman varianssin σ^2 luottamusvälin

$$\left(\frac{(n-1)s^2}{\chi_{\alpha/2}^2}, \frac{(n-1)s^2}{\chi_{1-\alpha/2}^2} \right)$$

konstruktiosta seuraa, että

$$\Pr \left(\frac{(n-1)s^2}{\chi_{\alpha/2}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{1-\alpha/2}^2} \right) = 1 - \alpha$$

- Siten konstruoitu luottamusväli *peittää* parametrin σ^2 todellisen arvon todennäköisyydellä $(1 - \alpha)$ ja se *ei peitä* parametrin σ^2 todellista arvoa todennäköisyydellä α .

Luottamusvälin frekvenssitulkinta 2/2

- Normaalijakauman odotusarvon σ^2 luottamusvälille voidaan antaa seuraava **frekvenssitulkinta**:
 - (i) Jos otantaa jakaumasta $N(\mu, \sigma^2)$ toistetaan, *keskimäärin*
 $100 \times (1 - \alpha) \%$
otoksista konstruoiduista luottamusväleistä *peittää* parametrin σ^2 todellisen arvon.
 - (ii) Jos otantaa jakaumasta $N(\mu, \sigma^2)$ toistetaan, *keskimäärin*
 $100 \times \alpha \%$
otoksista konstruoiduista luottamusväleistä *ei peitä* parametrin σ^2 todellista arvoa.

Normaalijakauman varianssin luottamusväli

Luottamusvälin ominaisuudet 1/2

- Normaalijakauman varianssin σ^2 luottamusvälin *pituus*

$$(n-1)s^2 \left(\frac{1}{\chi_{1-\alpha/2}^2} - \frac{1}{\chi_{\alpha/2}^2} \right)$$

vaihtelee otoksesta toiseen.

- Luottamusvälin *pituus* riippuu valitusta luottamustasosta $(1 - \alpha)$ ja otosvariانسsista s^2 sekä (jonkin verran) havaintojen lukumäärästä n .

Luottamusvälin ominaisuudet 2/2

- Luottamusväli *lyhenee (pitenee)*, jos *luottamustasoa* $(1 - \alpha)$ *pienennetään (kasvatetaan)*.
- Luottamusväli *lyhenee (pitenee)*, jos *otosvarianssi* s^2 *pienenee (kasvaa)*.
- Luottamusvälin pituus *pysyy suunnilleen samana*, jos *havaintojen lukumäärää* n *muuttuu*.

Normaalijakauman varianssin luottamusväli

Vaatimukset luottamusvälille

- Olisi toivottavaa pystyä konstruoimaan varianssi-parametrille σ^2 mahdollisimman *lyhyt* luottamusväli, johon liittyvä luottamustaso olisi samanaikaisesti mahdollisimman *korkea*.
- Vaatimusten samanaikainen täyttäminen *ei ole* kuitenkaan mahdollista:
 - (i) *Luottamustason kasvattaminen pidentää luottamusväliä*, jolloin tieto parametrin σ^2 todellisen arvon sijainnista tulee *epätarkemmaksi*.
 - (ii) *Luottamusvälin lyhentäminen pienentää luottamustasoa*, jolloin tieto parametrin σ^2 todellisen arvon sijainnista tulee *epävarmemmaksi*.

Johtopäätökset luottamusvälistä

- Oletetaan, että olemme tehneet *johtopäätöksen*, että konstruoitu luottamusväli peittää varianssiparametrin σ^2 todellisen arvon:
 - (i) Luottamusvälin konstruktiosta seuraa, että tehty johtopäätös *on oikea* $100 \times (1 - \alpha)$ %:ssa tapauksia.
 - (ii) Luottamusvälin konstruktiosta seuraa, että tehty johtopäätös *on väärä* $100 \times \alpha$ %:ssa tapauksia.
- Huomautus:

Virheellisen johtopäätöksen mahdollisuutta *ei saada häviämään*, ellei luottamusväliä tehdä *äärettömän leveäksi*, jolloin väli *ei enää sisällä informaatiota* varianssiparametrin σ^2 todellisesta arvosta.

Väliestimointi

Todennäköisyysjakaumien parametrien estimointi

Luottamusväli

Normaalijakauman odotusarvon luottamusväli

Normaalijakauman varianssin luottamusväli

>> Bernoulli-jakauman odotusarvon luottamusväli

Bernoulli-jakauman odotusarvon luottamusväli

Avainsanat

Bernoulli-jakauma
Estimointi
Frekvenssitulkinta
Keskeinen raja-arvolause
Luottamuskerroin
Luottamustaso
Luottamusväli
Normaalijakauma
Odotusarvo
Todennäköisyysparametri

Bernoulli-jakauma ja sen parametointi 1/2

- Olkoon A on kiinnostuksen kohteena oleva *tapahtuma* ja

$$\Pr(A) = p$$

$$\Pr(A^c) = 1 - p = q$$

- Määritellään satunnaismuuttuja

$$X = \begin{cases} 1, & \text{jos } A \text{ tapahtuu} \\ 0, & \text{jos } A \text{ ei tapahdu} \end{cases}$$

- Tällöin X noudattaa **Bernoulli-jakaumaa** parametrinaan

$$p = \Pr(A) = E(X)$$

- Merkitään:

$$X \sim \text{Ber}(p)$$

Bernoulli-jakauma ja sen parametointi 2/2

- *Bernoulli-jakauman* $\text{Ber}(p)$ *pistetodennäköisyysfunktio* on

$$f(x; p) = p^x q^{1-x}, \quad x = 0, 1$$

$$0 < p < 1$$

$$q = 1 - p$$

Bernoulli-jakauman odotusarvon luottamusväli

Otos Bernoulli-jakaumasta

- Olkoon

$$X_1, X_2, \dots, X_n$$

yksinkertainen satunnaisotos Bernoulli-jakaumasta

$$\text{Ber}(p)$$

- Tällöin satunnaismuuttujat X_1, X_2, \dots, X_n ovat *riippumattomia* ja noudattavat *samaa Bernoulli-jakaumaa*

$$\text{Ber}(p)$$

Bernoulli-jakauman odotusarvoparametrin estimointi 1/2

- *Estimoidaan* Bernoulli-jakauman $\text{Ber}(p)$ odotusarvoparametri p sen *harhattomalla estimaattorilla*:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$$

Bernoulli-jakauman odotusarvoparametrin estimointi 2/2

- Koska

$$X_i = \begin{cases} 1, & \text{jos } A \text{ tapahtuu} \\ 0, & \text{jos } A \text{ ei tapahdu} \end{cases}$$

niin

$$\sum_{i=1}^n X_i = f$$

jossa f on tapahtuman A *frekvenssi* otoksessa.

- Siten Bernoulli-jakauman parametrin p *estimaattori*

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{f}{n}$$

on tapahtuman A *suhteellinen frekvenssi* otoksessa.

Bernoulli-jakauman odotusarvon luottamusväli

Luottamustaso

- Määrätään **luottamusväli Bernoulli-jakauman odotusarvoparametrille p** .
- Valitaan **luottamustasoksi**
 $1 - \alpha$
- Luottamustaso kiinnittää todennäköisyyden, jolla konstruoitava luottamusväli peittää Bernoulli-jakauman parametrin p todellisen arvon.

Bernoulli-jakauman odotusarvon luottamusväli

Luottamuskertoimet

- Olkoon valittu luottamustaso $(1 - \alpha)$.
- Määrätään **luottamuskertoimet** $-z_{\alpha/2}$ ja $+z_{\alpha/2}$ siten, että

$$\Pr(z \leq -z_{\alpha/2}) = \frac{\alpha}{2}$$

$$\Pr(z \geq +z_{\alpha/2}) = \frac{\alpha}{2}$$

jossa satunnaismuuttuja z noudattaa *standardoitua normaalijakaumaa*:

$$z \sim N(0,1)$$

- Luottamuskertoimet $-z_{\alpha/2}$ ja $+z_{\alpha/2}$ toteuttavat ehdon

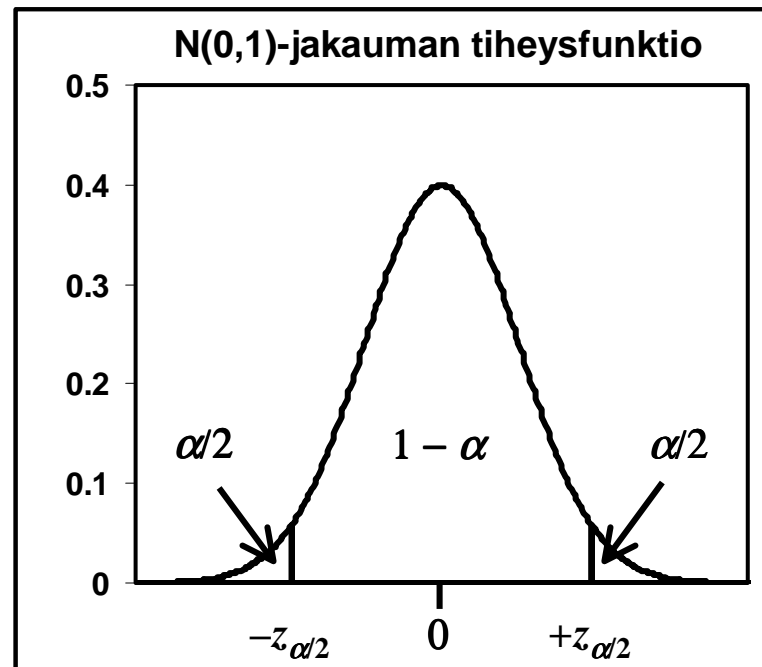
$$\Pr(-z_{\alpha/2} \leq z \leq +z_{\alpha/2}) = 1 - \alpha$$

Bernoulli-jakauman odotusarvon luottamusväli

Luottamuskertoimien määrittäminen:

Havainnollistus

- Luottamuskertoimet $-z_{\alpha/2}$ ja $+z_{\alpha/2}$ jakavat normaalijakauman tiheysfunktion kuvaajan alle jäävän todennäköisyysmassan (= 1) kolmeen osaan:
 - (1) Pisteeseen $-z_{\alpha/2}$ vasemmalle puolelle jää $\alpha/2$ % massasta.
 - (2) Pisteeseen $+z_{\alpha/2}$ oikealle puolelle jää $\alpha/2$ % massasta.
 - (3) Pisteiden $-z_{\alpha/2}$ ja $+z_{\alpha/2}$ väliin jää $(1 - \alpha)$ % massasta.
- Huomaa, että
$$\alpha/2 + \alpha/2 + (1 - \alpha) = 1$$



Luottamusväli Bernoulli-jakauman odotusarvolle

- Bernoulli-jakauman **odotusarvoparametrin** p **approksimatiivinen luottamusväli luottamustasolla** $(1 - \alpha)$ on muotoa

$$\left(\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right)$$

jossa

\hat{p} = parametrin p *harhaton estimaattori*

n = havaintojen *lukumäärä*

$-z_{\alpha/2}, +z_{\alpha/2}$

= luottamustasoon $(1 - \alpha)$ liittyvät *luottamuskertoimet normaalijakaumasta* $N(0,1)$

Bernoulli-jakauman odotusarvon luottamusväli

Luottamusväli ja sen pituus 1/2

- Bernoulli-jakauman odotusarvoparametrin p approksimatiivinen luottamusväli luottamustasolla $(1 - \alpha)$ esitetään usein muodossa

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

koska väli on symmetrinen keskipisteensä \hat{p} suhteen.

- Bernoulli-jakauman parametrin p approksimatiivisen **luottamusvälin pituus** on

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

Bernoulli-jakauman odotusarvon luottamusväli

Luottamusväli ja sen pituus 2/2

- Bernoulli-jakauman odotusarvoparametrin p approksimatiivinen luottamusväli luottamustasolla $(1 - \alpha)$:

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

Bernoulli-jakauman odotusarvon luottamusväli

Luottamusvälin johto 1/5

- Olkoon

$$X_1, X_2, \dots, X_n$$

yksinkertainen satunnaisotos Bernoulli-jakaumasta

$$\text{Ber}(p)$$

- Olkoon

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$$

harhaton estimaattori parametrille p .

- Huomaa, että \hat{p} on havaintojen X_1, X_2, \dots, X_n *aritmeettinen keskiarvo.*

Bernoulli-jakauman odotusarvon luottamusväli

Luottamusvälin johto 2/5

- Määritellään satunnaismuuttuja

$$z = \frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})/n}}$$

- Satunnaismuuttuja z noudattaa Bernoulli-jakautuneen otoksen *suhteellisen osuuden otosjakaumaa koskevan tuloksen perusteella suurissa otoksissa approksimatiivisesti standardoitua normaalijakaumaa* $N(0,1)$:

$$z \sim_a N(0,1)$$

Ks. kalvokokoelman **Johdatus todennäköisyyslaskentaan** lukua **Konvergenssikäsitteet ja raja-arvolauseet**.

Bernoulli-jakauman odotusarvon luottamusväli

Luottamusvälin johto 3/5

- Määrätään normaalijakaumasta $N(0,1)$ piste $+z_{\alpha/2}$ siten, että

$$\Pr(z \geq +z_{\alpha/2}) = \frac{\alpha}{2}$$

jolloin (normaalijakauman symmetrian perusteella)

$$\Pr(z \leq -z_{\alpha/2}) = \frac{\alpha}{2}$$

ja edelleen

$$\Pr(-z_{\alpha/2} \leq z \leq +z_{\alpha/2}) = 1 - \alpha$$

Bernoulli-jakauman odotusarvon luottamusväli

Luottamusvälin johto 4/5

- Tarkastellaan epäyhtälöketjua

$$-z_{\alpha/2} \leq z \leq +z_{\alpha/2}$$

- Sijoittamalla tähän epäyhtälöketjuun satunnaismuuttujan z lauseke, saadaan epäyhtälöketju

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

- Tästä epäyhtälöketjusta saadaan sen kanssa *yhtäpitävä* epäyhtälöketju

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

Bernoulli-jakauman odotusarvon luottamusväli

Luottamusvälin johto 5/5

- Yhdistämällä saatu epäyhtälö siihen, että

$$\Pr(-z_{\alpha/2} \leq z \leq +z_{\alpha/2}) =_a 1 - \alpha$$

saadaan vihdoin

$$\Pr\left(\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) =_a 1 - \alpha$$

Luottamusvälin frekvenssitulkinta 1/2

- Bernoulli-jakauman parametrin p luottamusvälin

$$\left(\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

konstruktiosta seuraa, että

$$\Pr \left(\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right) =_a 1 - \alpha$$

- Siten konstruoitu luottamusväli *peittää* parametrin p todellisen arvon approksimatiivisesti todennäköisyydellä $(1 - \alpha)$ ja se *ei peitä* parametrin p todellista arvoa approksimatiivisesti todennäköisyydellä α .

Luottamusvälin frekvenssitulkinta 2/2

- Bernoulli-jakauman parametrin p approksimatiiviselle luottamusvälille voidaan antaa seuraava **frekvenssitulkinta**:
 - (i) Jos otantaa jakaumasta $\text{Ber}(p)$ toistetaan, *keskimäärin* $100 \times (1 - \alpha) \%$ otoksista konstruoiduista luottamusväleistä *peittää* parametrin p todellisen arvon.
 - (ii) Jos otantaa jakaumasta $\text{Ber}(p)$ toistetaan, *keskimäärin* $100 \times \alpha \%$ otoksista konstruoiduista luottamusväleistä *ei peitä* parametrin p todellista arvoa.

Bernoulli-jakauman odotusarvon luottamusväli

Luottamusvälin ominaisuudet 1/2

- Bernoulli-jakauman parametrin p luottamusvälin *keskipiste* \hat{p} vaihtelee otoksesta toiseen.
- Luottamusvälin *pituus*

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

vaihtelee otoksesta toiseen.

- Luottamusvälin *pituus* riippuu valitusta luottamustasosta $(1 - \alpha)$, havaintojen lukumäärästä n ja estimaattorista \hat{p} .

Bernoulli-jakauman odotusarvon luottamusväli

Luottamusvälin ominaisuudet 2/2

- Luottamusväli *lyhenee (pitenee)*, jos *luottamustasoa* $(1 - \alpha)$ *pienennetään (kasvatetaan)*.
- Luottamusväli *lyhenee (pitenee)*, jos *havaintojen lukumäärää n kasvatetaan (pienennetään)*.
- Luottamusväli on *lyhimmillään*, kun

$$\hat{p} \approx 0 \text{ tai } \hat{p} \approx 1$$

- Luottamusväli on *pisimmillään*, kun

$$\hat{p} = \frac{1}{2}$$

Bernoulli-jakauman odotusarvon luottamusväli

Vaatimukset luottamusvälille

- Olisi toivottavaa pystyä konstruoimaan parametrille p mahdollisimman *lyhyt* luottamusväli, johon liittyvä luottamustaso olisi samanaikaisesti mahdollisimman *korkea*.
- Vaatimusten samanaikainen täyttäminen *ei ole* kuitenkaan mahdollista, *jos otoskoko pidetään kiinteänä*:
 - (i) *Luottamustason kasvattaminen pidentää luottamusväliä*, jolloin tieto parametrin p todellisen arvon sijainnista tulee *epätarkemmaksi*.
 - (ii) *Luottamusvälin lyhentäminen pienentää luottamustasoa*, jolloin tieto parametrin p todellisen arvon sijainnista tulee *epävarmemmaksi*.

Bernoulli-jakauman odotusarvon luottamusväli

Johtopäätökset luottamusvälistä

- Oletetaan, että olemme tehneet *johtopäätöksen*, että konstruoitu luottamusväli peittää odotusarvoparametrin p todellisen arvon:
 - (i) Luottamusvälin konstruktiosta seuraa, että tehty johtopäätös *on oikea* $100 \times (1 - \alpha)$ %:ssa tapauksia.
 - (ii) Luottamusvälin konstruktiosta seuraa, että tehty johtopäätös *on väärä* $100 \times \alpha$ %:ssa tapauksia.
- Huomautus:

Virheellisen johtopäätöksen mahdollisuutta *ei saada häviämään*, ellei luottamusväliä tehdä *äärettömän leveäksi*, jolloin väli *ei enää sisällä informaatiota* odotusarvoparametrin p todellisesta arvosta.

Bernoulli-jakauman odotusarvon luottamusväli

Otoskoon määrääminen 1/5

- Monissa tutkimustilanteissa toivotaan, että parametrin luottamusväli saataisiin *mahdollisimman lyhyeksi*.
- Bernoulli-jakauman odotusarvoparametrin p luottamusvälin pituus riippuu otoskoosta siten, että *väli lyhenee, jos havaintojen havaintojen lukumäärää n kasvatetaan*.
- Tämä odotusarvoparametrin p luottamusvälin ominaisuus mahdollistaa tietyin ehdoin otoskoon valinnan niin, että *luottamusvälistä tulee (suurin piirtein) toivotun mittainen*.
- Oletetaan siis, että Bernoulli-jakauman odotusarvoparametrille p halutaan konstruoida luottamusväli, jonka *toivottu pituus on*

2A

Bernoulli-jakauman odotusarvon luottamusväli

Otoskoon määrääminen 2/5

- *Bernoulli-jakauma odotusarvoparametrin p luottamusväli luottamustasolla $(1 - \alpha)$ on muotoa*

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

jossa

\hat{p} = parametrin p *harhaton estimaattori*

n = havaintojen *lukumäärä*

$-z_{\alpha/2}$, $+z_{\alpha/2}$

= luottamustasoon $(1 - \alpha)$ liittyvät *luottamuskertoimet normaalijakaumasta* $N(0,1)$

Bernoulli-jakauman odotusarvon luottamusväli

Otoskoon määrääminen 3/5

- Jos siis käytettävissä on *ennakkotietoa parametrusta* p , voidaan otoskoko n ratkaista yhtälöstä:

$$z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} = A$$

jossa

$z_{\alpha/2}$ = luottamustasoon $(1 - \alpha)$ liittyvä *luottamuskerroin normaalijakaumasta* $N(0,1)$

p = odotusarvon *oletettu arvo*

n = *otoskoko*

$2A$ = *toivottu pituus* luottamusvälille

Bernoulli-jakauman odotusarvon luottamusväli

Otoskoon määrääminen 4/5

- Siten *tarvittava otoskoko* on

$$n = \left(\frac{z_{\alpha/2} \sqrt{p(1-p)}}{A} \right)^2$$

- Huomaa, että tarvittava otoskoko *saavuttaa maksiminsa*

$$n = \left(\frac{z_{\alpha/2}}{2A} \right)^2$$

kun

$$p = 1/2$$

Bernoulli-jakauman odotusarvon luottamusväli

Otoskoon määrääminen 5/5

- Huomautus:

Tarvittavan otoskoon kaavasta nähdään, että mitä lyhyempää luottamusväliä toivotaan, sitä suurempi otos on poimittava:

Jos esimerkiksi luottamusvälin pituus halutaan puolittaa, pitää havaintoja kerätä 4 kertaa enemmän.