

Luottamusvälit

Normaalijakauma johonkin kohtaan

Perusjoukko ja otanta

Jos halutaan tutkia esimerkiksi Suomessa elävien naarashirvien painoa, se voidaan (periaatteessa) tehdä kahdella tavalla:

1. tutkimalla jokainen Suomessa elävä naarashirvi, eli tutkimalla koko **populaatio** eli **perusjoukko**
2. valitsemalla perusjoukosta otanta, esimerkiksi 1000 hirveä, jotka tutkimalla luodaan arvio koko populaation painojakaumalle

Koska koko perusjoukon tutkiminen on usein mahdotonta, tutkitaan monia asioita otannan avulla.

Otoskeskiarvo ja odotusarvo, mitä eroa?

Otoskeskiarvo \bar{x} tarkoittaa otoksesta laskettua keskiarvoa.

Perusjoukon odotusarvo μ tarkoittaa koko populaation todellista keskiarvoa.

Perusjoukon odotusarvo on usein mahdottoman työlästä selvittää ja täten tuntematon, siksi käytetään arviota, eli otoskeskiarvoa.

Kun otos on tarpeeksi suuri, otoskeskiarvo on hyvä estimaatti perusjoukon odotusarvolle.

Perusjoukon tunnusluvun arvioiminen otannan avulla

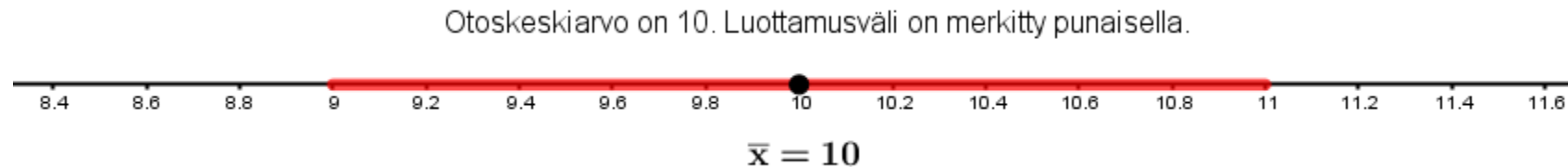
Otannasta voidaan laskea joku tilastollinen tunnusluku, esimerkiksi painon keskiarvo. Otannasta laskettu keskiarvo on estimaatti koko populaation keskiarvolle, eli **perusjoukon odotusarvolle**.

Otannasta laskettu tunnusluku ei kuitenkaan varmasti ole pilkulleen oikea, eli tasan tarkkaan sama kuin vastaava perusjoukon tunnusluku.

Tuhannen hirvinaaraan otannasta laskettu painon keskiarvo ei varmasti ole tasan yhtä suuri kuin kaikkien Suomessa elävien hirvinaaraiden keskipaino.

Luottamusväli

Otannasta lasketun otoskeskiarvon \bar{x} lisäksi voidaan ilmoittaa joku otoskeskiarvon lähiympäristö, jossa todellinen arvo (perusjoukon odotusarvo) todennäköisesti on.



Mitä suurempi väli, sitä suuremmalla todennäköisyydellä odotusarvo osuu kyseiselle välille. Ja toisaalta mitä suuremmalla todennäköisyydellä haluamme odotusarvon osuvan välille, sitä suuremman välin tarvitsemme.

Perusjoukon odotusarvon μ luottamusväli

- 95 %:n luottamustasolla on väli $\left[\bar{x} - 1,96 \frac{s}{\sqrt{n}} ; \bar{x} + 1,96 \frac{s}{\sqrt{n}} \right]$
- 99 %:n luottamustasolla on väli $\left[\bar{x} - 2,58 \frac{s}{\sqrt{n}} ; \bar{x} + 2,58 \frac{s}{\sqrt{n}} \right]$
- 99,9 %:n luottamustasolla on väli $\left[\bar{x} - 3,29 \frac{s}{\sqrt{n}} ; \bar{x} + 3,29 \frac{s}{\sqrt{n}} \right]$

Kaavoissa s on satunnaismuuttujan keskihajonta, \bar{x} satunnaismuuttujan otoksesta laskettu otoskeskiarvo ja n otoksen koko.

(Satunnaismuuttuja, esimerkiksi hirvinaaraan massa.)

Tärkeä tarkennus

Tarkalleen ottaen Maolissa lukee:

*Kaavoissa s on satunnaismuuttujan keskihajonta, \bar{x} satunnaismuuttujan otoksesta laskettu otoskeskiarvo ja n otoksen koko. Kun hajonta estimoidaan otoksesta, korvataan **kertoimet** 1,96; 2,58 ja 3,29 t -jakauman arvoilla t_p , kun $p = 0,05$; $p = 0,01$ ja $p = 0,001$ sekä vapausasteluku $n - 1$.*

- Keskihajonta tarkoittaa koko populaatiosta laskettua hajontaa. (Populaatio ja **hajonta tarkasti tunnettu.**)
- Kun hajonta estimoidaan otoksesta, tarkan koko populaation keskihajonnan tilalla on epätarkempi otoskeskihajonta. Tällöin myös luottamusvälien pitäisi olla leveämpiä ja **kertoimien** suurempia.

Siis keskihajonta vai otoskeskihajonta?

Käytännössä otoskeskihajonta vastaa hyvin perusjoukon keskihajontaa, kun otos on yli 30. Oletamme siis, että otoskeskihajonta on tarpeeksi lähellä keskihajontaa ja käytämme kertoimia 1,96; 2,58 ja 3,29.

Yo-kokeessa voisi olla hyvä kirjata jotain tämän tyyliä ”*Käytetään otoskeskihajontaa keskihajonnan estimaattina, vaikka ovatkin eri asia.*”

Esimerkki 1

Hirvinaaraiden otoksesta ($n = 100$) laskettu keskipaino oli 265 kg.
Määritä keskipainon 95 %:n luottamusväli, jos keskihajonta $s = 42$ kg.

95 %:n luottamustasolla on väli $\left[\bar{x} - 1,96 \frac{s}{\sqrt{n}} ; \bar{x} + 1,96 \frac{s}{\sqrt{n}} \right]$

$$\text{alaraja} = 265 - 1,96 \cdot \frac{42}{\sqrt{100}} = 256,768 \approx 257 \text{ (kg)}$$

$$\text{yläraja} = 265 + 1,96 \cdot \frac{42}{\sqrt{100}} = 273,232 \approx 273 \text{ (kg)}$$

Keskipainon 95 %:n luottamusväli on siis $[257, 273]$

Esimerkki 2

Kauppiaas väitti asiakkaalle, että kaupattavien kananmunien keskipaino on 50 g. Asiakas poimi 36 munan otoksen ja sai keskiarvon 45 g hajonnan ollessa 6 g. Tuntuuko kauppiaan väite uskottavalta?

Ratkaisu:

Väite on uskottava, jos 50 g osuu 95 %:n luottamusvälille.

$$\text{alaraja} = 45 - 1,96 \cdot \frac{6}{\sqrt{36}} = 43,04 \approx 43 \text{ (kg)}$$

$$\text{yläraja} = 45 + 1,96 \cdot \frac{6}{\sqrt{36}} = 46,96 \approx 47 \text{ (kg)}$$

50 g ei kuulu välille [43, 47], joten kauppiaan väite ei ole uskottava.

Suhteellinen osuus p

Voimme myös tutkia populaation tietyn tyyppisten alkioiden prosenttiosuutta p .

Esimerkiksi kuinka monta prosenttia 50 kappaleen otoksessa on viallisia kahvinkeitinimiä.

Kuinka monta prosenttia 3000 ihmisen otoksessa kannattaa demareita.

Kuinka monta prosenttia tuhannen hirven otoksessa on ruskeasilmäisiä.

Suhteellisen osuuden p luottamusväli

- 95 %:n luottamustasolla on väli $[\hat{p} - 1,96s ; \hat{p} + 1,96s]$
- 99 %:n luottamustasolla on väli $[\hat{p} - 2,58s ; \hat{p} + 2,58s]$
- 99,9 %:n luottamustasolla on väli $[\hat{p} - 3,29s ; \hat{p} + 3,29s]$

Kaavoissa \hat{p} ja s ovat satunnaismuuttujan laskettavia tunnuslukuja:

\hat{p} = otoksesta laskettu suhteellinen osuus (siis suhteellinen frekvenssi!),
 n = otoksen koko,

$s = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ = suhteellisen osuuden otosjakauman keskihajonta.

HUOM! \hat{p} kaavaan lukuna välillä $[0, 1]$, ei prosenttilukuna.

Esimerkki 3

Taloustutkimuksen viimeisimmän puolueiden kannatusmittauksen mukaan SDP on suosituin puolue 21,0 % kannatuksella. Määritä SDP:n kannatuksen 95 %:n luottamusväli, kun otos $n = 3000$.

Ratkaisu: $[\hat{p} - 1,96s ; \hat{p} + 1,96s]$ ja $s = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$$\text{Alaraja} = 0,21 - 1,96 \sqrt{\frac{0,21(1-0,21)}{3000}} = 0,1954 \dots \approx 19,5 \%$$

$$\text{Yläraja} = 0,21 + 1,96 \sqrt{\frac{0,21(1-0,21)}{3000}} = 0,2245 \dots \approx 22,5 \%$$

SDP:n kannatuksen 95 %:n luottamusväli on $[19,5; 22,5]$.