

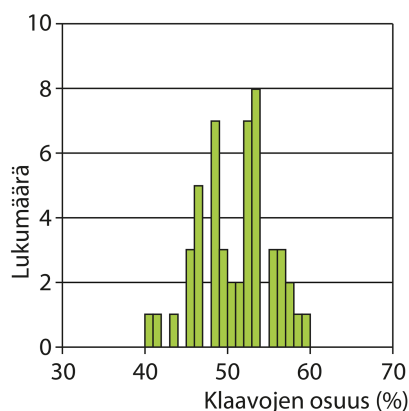
# Suhteellisen osuuden jakauma

## Esimerkki 1

Mari testasi heittämällä virtuaalista kolikkoa, kuinka monta prosenttia heitoista antaa klaavan. Hän selvitti asiaa viidellä testillä.

*Testi 1:* Mari heitti sadan heiton sarjan, jossa klaavoja tuli 57 kappaletta ja kruunia 43 kappaletta. Klaavojen osuus oli 57 %.

*Testi 2:* Mari heitti 50 sadan heiton sarjaa, jolloin klaavojen osuudet jakoutuivat alla olevan kuvan mukaisesti.



Esimerkki 2 Laskentaohjelmasta saadaan odotusarvo 0,4985 ja keskihajonta 0,0693 prosenttiyksikköä.

% ja osuuksien keskihajonta 4,985 prosenttiyksikköä. Koulun opiskelijoista 40 % on poikia. Koulun opiskelijoista satunnaisilmion toistokokeissa tietyllä tapahtumalle saadut osuudet noudattavat normaali-jakaumaa, jossa todennäköisyydellä vähintään 0,075 on olemassa poikia. Osuuksien keskiarvo on sama kuin satunnaisilmion odotusarvo ja osuuksien keskihajonta on

Ratkaisu voidaan havaita, että suuremmalla nettosarjojen määrällä osuudet sijoittuvat yhä

50 prosenttiin poikien osuudet.

Keskijonon odotusarvo on 50 %, jos otos on suuri, jätetään josta vain 5 prosenttiyksikköä. Osuuksien

Esimerkissä 1 osuuksien keskiarvoksi testissä 5 saatiin jakauma muistuttaa siis normaali-jakaumaa  $N(0,5; 0,05)$ .

49,98 % kun tapahtuman "saadaan klaava" odotusarvo on

50 %. Osuuksien keskihajonnaksi saatiin 4,985

prosenttiyksikköä.

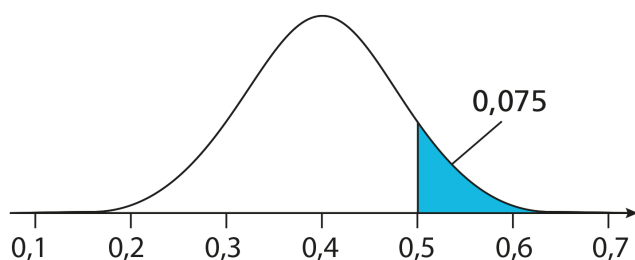
prosenttiyksikköä, kun teorian mukaan keskihajonta olisi

Syötetään odotusarvo 0,4 ja keskihajonta 0,0693

laskentaohjelmaan.  $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0,5(1-0,5)}{100}} = 0,05 = 5$  prosenttiyksikköä.

Laskentaohjelmasta saadaan

$P(X \geq 0,5) = 0,07451... \approx 0,075$ .



Vastaus

Vähintään puolet otoksen opiskelijoista on poikia todennäköisyydellä 0,075.

## Suhteellisen osuuden vaihteluvälit

Normaalijakauman mukaisesti jakautuneelle otoksesta määritetylle osuudelle voidaan kirjoittaa seuraavat

vaihteluvälit:

95 % osuuksista on välillä

$$\left[ p - 1,96 \cdot \sqrt{\frac{p(1-p)}{n}} ; p + 1,96 \cdot \sqrt{\frac{p(1-p)}{n}} \right]$$

99 % osuuksista on välillä

$$\left[ p - 2,58 \cdot \sqrt{\frac{p(1-p)}{n}} ; p + 2,58 \cdot \sqrt{\frac{p(1-p)}{n}} \right]$$

99,9 % osuuksista on välillä

$$\left[ p - 3,29 \cdot \sqrt{\frac{p(1-p)}{n}} ; p + 3,29 \cdot \sqrt{\frac{p(1-p)}{n}} \right]$$

Kaavoissa

$p$  on osuuden odotusarvo

$n$  on otoskoko.

Esimerkki 3

Esimerkissä 2 poikien osuudeksi perusjoukossa tiedettiin 40 %. Kun otoskoko oli 50 opiskelijaa, saatiin otoksesta määritettävien osuuksien keskihajonnaksi 6,93 prosenttiyksikköä. Mille osuuden odotusarvon suhteen symmetriselle välille sijoittuu 95 % otoksista saaduista poikien osuuksista, kun otoskoko on 50?

Ratkaisu

95 % osuuksista on välillä

$$\left[ p - 1,96 \cdot \sqrt{\frac{p(1-p)}{n}} ; p + 1,96 \cdot \sqrt{\frac{p(1-p)}{n}} \right].$$

Sijoitetaan osuuden vaihteluvälin ala- ja ylärajan lausekkeisiin  $p = 0,4$  ja  $n = 50$ .

Vaihteluvälin alarajaksi saadaan

$$0,4 - 1,96 \cdot \sqrt{\frac{0,4(1-0,4)}{50}} = 0,26420... \approx 0,264 = 26,4 \, \%.$$

Vaihteluvälin ylärajaksi saadaan

$$0,4 + 1,96 \cdot \sqrt{\frac{0,4(1 - 0,4)}{50}} = 0,53579 \dots \approx 0,536 = 53,6 \, \%.$$

Kun otoskoko on 50, poikien osuudet otoksissa sijoittuvat 95 prosentin todennäköisyydellä välille 26,4 %...53,6 %.

Tulos tarkoittaa myös 5 prosentin todennäköisyyttä sille, että otoksesta saatu poikien osuus olisi alarajaa 26,4 % pienempi tai ylärajaa 53,6 % suurempi.

Vastaus

Väli on 26,4 %...53,6 %.

## Otoksesta määritettyyn osuuteen liittyvä virhemarginaali

Esimerkkien 2 ja 3 tarkastelut perustuivat siihen, että poikien osuus koulun opiskelijoista tunnettiin. Tällöin voitiin arvioida mahdollisia otoksista saatavia osuuksia ja niiden todennäköisyyksiä. Käytännössä tilanne on yleensä päinvastainen, jolloin todellista osuutta ei tunneta. Tällöin otoksesta määritetyn osuuden perusteella pyritään arvioimaan eli estimoimaan todellista osuutta perusjoukossa.

Otoksesta määritettyyn osuuteen  $p$  liittyvä virhemarginaali lasketaan seuraavasta kaavasta.

$$\text{virhemarginaali} = \text{kriittinen arvo} \cdot \sqrt{\frac{p(1 - p)}{n}},$$

missä

$p$  on otoksesta määritetty osuus

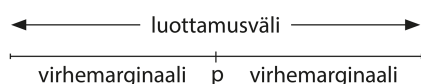
$n$  on otoskoko.

Kriittinen arvo valitaan luottamustason mukaan seuraavasta taulukosta.

Luottamusvälin osuudelle  
 otoksen koon määrityksessä, eikä se myöskään vaikuta  
 virhemarginaalin suuruuteen.

Virhemarginaalin avulla voidaan muodostaa luottamusväli  
 perusjoukossa esiintyvälle osuudelle. Osuus perusjoukossa

sijoittuu luottamusvälin sisään luottamustasoa vastaavalla  
 todennäköisyydellä. Yleisimmin luottamusvälit  
 määritetään 95 prosentin luottamustasolla. Mitä  
 korkeampi luottamustaso luottamusväliin liittyy, sitä  
 suurempi on virhemarginaali ja luottamusvälin pituus.



Kun otoskoko on riittävän suuri eli yli 30, osuudelle  
 perusjoukossa voidaan kirjoittaa seuraavat luottamusvälit:

#### Osuus perusjoukossa on

95 prosentin todennäköisyydellä välillä

$$\left[ p - 1,96 \cdot \sqrt{\frac{p(1-p)}{n}} ; p + 1,96 \cdot \sqrt{\frac{p(1-p)}{n}} \right]$$

99 prosentin todennäköisyydellä välillä

$$\left[ p - 2,58 \cdot \sqrt{\frac{p(1-p)}{n}} ; p + 2,58 \cdot \sqrt{\frac{p(1-p)}{n}} \right]$$

99,9 prosentin todennäköisyydellä välillä

$$\left[ p - 3,29 \cdot \sqrt{\frac{p(1-p)}{n}} ; p + 3,29 \cdot \sqrt{\frac{p(1-p)}{n}} \right]$$

missä

$p$  on otoksesta määritetty osuus

$n$  on otoskoko.

#### Esimerkki 4

Laaduntarkkailussa 2 000 kappaleen tuote-erästä löydettiin  
 52 viallista tuotetta. Määritä luottamusväli viallisten  
 tuotteiden osuudelle 95 prosentin luottamustasolla.

## Ratkaisu

2 000 kappaleen otoksessa viallisten tuotteiden osuus oli

$$\frac{52}{2\,000} = 0,026 = 2,6 \, \%.$$

Viallisten tuotteiden osuus on 95 prosentin todennäköisyydellä välillä

$$\left[ p - 1,96 \cdot \sqrt{\frac{p(1-p)}{n}} ; p + 1,96 \cdot \sqrt{\frac{p(1-p)}{n}} \right].$$

Sijoitetaan osuuden vaihteluvälin ala- ja ylärajan lausekkeisiin  $p = 0,026$  ja  $n = 2\,000$ .

Viallisten tuotteiden osuuden luottamusvälin alarajaksi saadaan

$$0,026 - 1,96 \cdot \sqrt{\frac{0,026(1-0,026)}{2\,000}} = 0,01902... \approx 0,019 = 1,9 \, \%.$$

Luottamusvälin ylärajaksi saadaan

$$0,026 + 1,96 \cdot \sqrt{\frac{0,026(1-0,026)}{2\,000}} = 0,03297... \approx 0,033 = 3,3 \, \%.$$

95 prosentin luottamusväli viallisten tuotteiden osuudelle tuotannossa on 1,9 %...3,3 %. Toisin sanoen 95 prosentin todennäköisyydellä viallisten tuotteiden osuus on tällä välillä.

Useimmat laskentaohjelmat laskevat luottamusvälin rajat, kun lähtöarvoina annetaan luottamustaso, frekvenssi ja otoskoko. Toiminto voi olla esimerkiksi suhteen Z-estimaatti tai 1 väittämän z-väli.

## Vastaus

Luottamusväli on 1,9 %...3,3 %.

### Esimerkki 5

Puolueiden kannatustutkimuksessa haastateltiin 3 419 yli 18-vuotiasta. Puolueen kannatukseksi saatiin 14,8 %.

a) Kuinka suuri on kannatusprosentin virhemarginaali 95 prosentin luottamustasolla?

b) Muodosta luottamusväli puolueen kannatukselle.

c) Puolueen kannatus oli edellisissä vaaleissa 16,4 %.

Voidaanko tutkimuksen osoittamaa kannatuksen laskua pitää luotettavana?

### Ratkaisu

a) Lasketaan otoksesta määritettyyn osuuteen liittyvä virhemarginaali 95 prosentin luottamustasolla eli 5 prosentin riskitasolla.

Virhemarginaaliksi saadaan

$$1,96 \cdot \sqrt{\frac{0,148(1 - 0,148)}{3\,419}} = 0,01190\dots \approx 0,012 = 1,2\%$$

prosenttiyksikköä.

b) Kannatusprosentin luottamusvälin alarajaksi saadaan

$$14,8 \% - 1,2 \% = 13,6 \%$$

Luottamusvälin ylärajaksi saadaan

$$14,8 \% + 1,2 \% = 16,0 \%$$

Puolueen todellinen kannatus on 95 prosentin todennäköisyydellä välillä 13,6 %...16,0 %.

c) Koska aiempi 16,4 prosentin kannatus ei sisälly luottamusväliin 13,6 %...16,0 %, puolueen kannatuksen muutosta voidaan pitää luotettavana.

Valitusta luottamustasosta seuraa, että päätelmään sisältyy alle 5 prosentin todennäköisyys, että kannatuksessa ei olisikaan tapahtunut muutosta.

Vastaus

a) Virhemarginaali on 1,2 prosenttiyksikköä.

b) Luottamusväli on 13,6 %...16,0 %.

c) Voidaan pitää luotettavana.

Avainkäsitteet

- osuuden virhemarginaali
- osuuden luottamusväli