

## Aineistokoko ja voima-analyysi

### Johdanto

Aineisto- eli otoskoon arviointi ja tutkimuksen voima-analyysi ovat tilastollisen tutkimuksen suunnittelussa keskeisimpiä asioita. **Otoskoon arvioinnilla** voidaan ennen varsinaisen tutkimuksen aloittamista selvittää sopiva aineistomäärä tavoitteiden toteuttamiseksi. Vaikka tutkimuksella oltaisiinkin osoittamassa merkittävää ja lähes itsestään selvää eroa, liian pienellä aineistolla tämän eron osoittaminen ei onnistu. Isoilla aineistoilla lähes kaikki erot ovat tilastollisesti merkitseviä, mutta turhan suuren aineiston kerääminen ei ole ajallisesti ja taloudellisesti kannattavaa. Lisäksi on huomioitava eettinen näkökulma: osoitettaessa uuden hoitomuodon toimivuutta ylimääräisten potilaiden hoitaminen on turhaa.

Useat lehdet vaativat **voimalaskelmien** raportointia tutkimuksen tuloksia julkaistaessa. Haettaessa tutkimukselle rahoitusta voimalaskelmilla voidaan osoittaa, että tutkijalla on käsitys, mitä on osoittamassa todeksi ja kuinka suurella aineistolla tämä tulee onnistumaan.

Tässä oppaassa kuvataan, kuinka otoskoon määrittäminen tehdään tyypillisimmässä lääketieteellisessä tutkimusasetelmassa. Tyypillisin tilanne on kahden ryhmän tutkimus, jossa on kyseessä joko keskiarvojen tai osuuksien vertailu.

### Toteutus

Tutkimuksessa on yleensä aina nolla- ja vastahypoteesi. Nollahypoteesi on esimerkiksi muotoa: ”ryhmät eivät eroa toisistaan” tai ”käsittelyllä ei ole vaikutusta”. Vastahypoteesi on nollahypoteesin vastakohta: ”ryhmät eroavat toisistaan” tai ”käsittelyllä on vaikutusta”.

Määritettäessä aineiston kokoa toteutus on yleensä samankaltainen. Valitaan tai oletetaan tutkimukseen p-arvo, voima (englanniksi *power*) ja efektin koko. Näiden tietojen avulla tietokone ratkaisee matemaattisen yhtälön, ja tuloksena saadaan otoskoko N. On tärkeää huomioida, että määritetty aineiston koko on arvio, joka pätee silloin kun oletukset toteutuvat.

Käytännössä p-arvo tarkoittaa todennäköisyyttä, että nollahypoteesi on totta. Jos on havaittu pieni p-arvo, tapahtuman ”ryhmät eivät eroa toisistaan” tai ”käsittelyllä ei ole vaikutusta” todennäköisyys on pieni. Käsite voima tarkoittaa: mikäli on havaittu tilastollisesti merkitsevä p-arvo, niin todennäköisyydellä *voima* havaittu ero on todellinen.

Käytännössä tutkimuksen suunnitteluvaiheessa voiman voidaan ajatella olevan todennäköisyys sille, että saadaan tilastollisesti merkitsevä p-arvo.

Lääketieteellisessä tutkimuksessa tutkitaan pääsääntöisesti keskiarvojen tai osuuksien eroja. Nämä on voitava olettaa. Osuuksien ollessa kyseessä oletetaan tutkimusryhmiin esimerkiksi parantuneiden osuudet, aikaisempi menetelmä parantaa tietyn prosenttiosuuden ja testattava menetelmä jonkin verran enemmän. Osoitettaessa keskiarvojen erojen olemassaoloa oletetaan ryhmiin keskiarvot ja yleensä yhteinen keskihajonta. Lisäksi kannattaa pohtia, onko kehitteillä oleva menetelmä kliinisesti merkitsevästi parempi kuin vanhempi menetelmä. Esimerkiksi kivun hoidossa voi olla, että koetun kivun on laskettava vähintään 20 %, jotta kivun alenemista voidaan pitää kliinisesti merkitseväenä. Näiden keskiarvojen tai osuuksien avulla lasketaan efektin koko. Efektin koon laskentaa ei tarvitse tehdä käytännössä, sillä laskentaohjelmat tekevät sen annettujen lukujen perusteella.

Yleensä tutkimuksissa käytetään p-arvona lukemaa 0,05 ja voiman määränä lukemaa 0,80. Tutkimusta suunniteltaessa kannattaa varmistaa, ovatko nämä arvot nimenomaisen tieteenalan käytössä. Tehtäessä arvioita otoskoosta kannattaa kokeilla eri arvoja näille kahdelle parametrille, etenkin eri voimatasoja, ja katsoa, kuinka arvio aineistonkoosta muuttuu. Jos tutkimuksessa on useita tavoitteita, niin aineistokokolaskelmat on tehtävä kaikille lopputulosmuuttujille erikseen. Tutkimuksissa on yleensä aina myös katoa. Tämä on otettava huomioon päätettäessä aineiston koosta. Jos tipahtamisosuus on  $p$ , aineiston koon on oltava  $N/(1-p)$ . Jos tipahtamisosuus olisi 30 %, tällöin aineiston koon olisi oltava noin 1,4-kertainen, jotta aineistokoko olisi riittävä aloitettaessa aineiston analysointia.

Voimalaskelmia tehdään myös toiseen suuntaan tutkimuksessa saatujen tulosten perusteella. Tällä tarkoitetaan sitä, että tutkimuksessa saaduille tuloksille lasketaan voima. Tässä vaiheessa ovat tiedossa keskiarvot hajontoineen tai osuudet ja aineistonkoko. Näiden avulla on mahdollista laskea havaitun efektin voima. Voimaa ratkaistaessa kiinnitetään p-arvoksi yleensä lukema 0,05 ja syötetään otoskoko ja havaitut käsittelyvaikutukset. On tärkeää tehdä voiman laskenta sitä tilastomenetelmää vastaavalla menetelmällä, jolla tutkimuksen vertailutkin on tehty. Jos tutkimuksen keskiarvovertailut on tehty t-testillä, käytetään t-testin voimalaskuria. Jos tästä ratkaistu tutkimuksen tulosten voima on korkea, esimerkiksi yli 0,80, voidaan olla suhteellisen varmoja, että havaittu käsittelyvaikutus on todella olemassa. Tämä laskettu voima on hyvä raportoida julkaisun tulososiossa.

Aineistokoko- ja voimalaskelmia varten löytyy netistä laskureita. Hakusanoina keskiarvojen tapauksissa toimii ”*sample size calculator mean*” ja osuuksille ”*sample size calculator proportion*”. Laskureista kannattaa valita sellainen, johon voi laittaa myös halutun voimatason. Paras tapa tehdä voimalaskelmia on käyttää yleisessä käytössä olevia ohjelmistoja, esimerkiksi R-ohjelmaa. Joissakin laskureissa on mahdollisuus säätää vastahypoteesi yksisuuntaiseksi, eli uudempi menetelmä on parempi kuin vanhempi. Tä-

mä muutos pienentää arviota otoskoosta. Kaksisuuntainen hypoteesi, eli ”käsittelyt eroavat toisistaan”, on kuitenkin kaikista yleisin ja suositeltavin vaihtoehto.

## Esimerkkejä

**Esimerkki 1.** Tutkitaan 80-vuotiaiden kuolleisuutta. Halutaan osoittaa, että sukupuolten välillä on eroa kuolleisuudessa. Oletetaan, että miehistä kuolee 33 % ja naisista 22 % seuraavan viiden vuoden kuluessa. Valitaan tutkimukseen p-arvoksi 0,05 ja käytetään voimatasona lukemia 0,80 ja 0,90. Näin olettaen vaaditut otoskoot per sukupuoli ovat 258 ja 345 kahdella eri voimatasolla. Kokemuksesta tiedetään, että seuranta-ajan päätyttyä elossaolotietoa ei ole saatavilla viidesosalta henkilöistä. Tämä tieto huomioiden saadaan otoskokoarvioiksi 323 ja 432. Voimatasolla 0,80 ja käyttäen p-arvoa 0,05 tulisi 80-vuotiaita miehiä olla mukana 323 ja naisia 323.

Tutkimusta lähdettiin toteuttamaan suunnitelman mukaisesti eli 323 henkeä per sukupuoli. Viiden vuoden jälkeen tieto oli saatavissa 251 miehestä, joista oli kuollut 82 (32,7 %). Naisista tieto oli saatavilla 260:stä ja heistä oli kuollut 60 (23,1 %). Tilastollinen merkitsevyys tälle havaitulle osuuksien erolle on  $p=0,02$  ja voima 0,68. Tutkimuksessa saatiin merkitsevä ero sukupuolten väliselle kuolleisuudelle, mutta voima jäi hieman tavoitetasoa alhaisemmaksi. Kuvassa 1 on esitetty, kuinka voima voidaan laskea nettilaskurilla. Linkki laskuriin:

<http://www.stat.ubc.ca/~rollin/stats/ssize/b2.html>

### Inference for Proportions: Comparing Two Independent Samples

(To use this page, your browser must recognize JavaScript.)

Choose which calculation you desire, enter the relevant population values (as decimal fractions) for p1 (proportion i  
After making your entries, hit the **calculate** button at the bottom.

- Calculate Sample Size (for specified Power)
- Calculate Power (for specified Sample Size)

Enter a value for p1:

Enter a value for p2:

- 1 Sided Test
- 2 Sided Test

Enter a value for  $\alpha$  (default is .05):

Enter a value for desired power (default is .80):

The sample size (for each sample separately) is:

Kuva 1. Voiman laskeminen. Laskuriin on lisätty kuolleiden osuudet havaitusta aineistosta, p-arvoksi on valittu 0,05 ja keskiarvo on otettu ryhmäkooista. Näiden avulla on laskettu voima sukupuolten väliselle kuolleisuuden erolle.

**Esimerkki 2.** Synnytyssalissa on havaittu, että uudelleen synnyttäjien lapset ovat painavampia kuin ensisynnyttäjien lapset. Halutaan osoittaa tämä huomio todeksi. Oletetaan, että ensisynnyttäjien lapset painavat keskimäärin 3400 grammaa ja uudelleen synnyttäjien 3600 grammaa. Lisäksi oletetaan keskihajonnan olevan 500 grammaa per synnyttäjärühmä. Valitaan tutkimuksen p-arvoksi 0,05 ja voimatasoksi 0,80. Lisäksi tiedetään, että 10 %:lta vastasyntyneistä ei saada mitattua painoa tutkimusprotokollan mukaisesti. Näiden oletusten vallitessa synnytyksiä tulee olla 99 per synnyttäjärühmä. Huomioiden painon mittaustuloksen puute ensisynnytyksiä tulee olla 110 ja uudelleen synnyttäjiä 110. Kuvassa 2 on laskettu, kuinka otoskoon voi laskea tällaisessa tilanteessa. Linkki laskuriin: <http://www.stat.ubc.ca/~rollin/stats/ssize/n2.html>

Kun ennalta määritetty otoskoko oli saavutettu, ensisynnyttäjien lapsia oli punnittu 98 ja uudelleen synnyttäjien lapsia 100. Ensisynnyttäjien lapset painoivat keskimäärin 3329 grammaa ja uudelleen synnyttäjien 3589 grammaa. Hajonta oli likimäärin 520 grammaa molemmissa ryhmissä. P-arvo tälle erolle oli  $p < 0,001$  ja voima 0,94. Tässä tutkimuksessa saatiin voimatasoksi korkea lukema, joten voidaan olla suhteellisen varmoja, että uudelleen synnyttäjien lapset ovat painavampia kuin ensisynnyttäjien.

### Inference for Means: Comparing Two Independent Samples

(To use this page, your browser must recognize JavaScript.)

Choose which calculation you desire, enter the relevant population values for  $\mu_1$  (mean of population 1),  $\mu_2$  relevant. After making your entries, hit the calculate button at the bottom.

- Calculate Sample Size (for specified Power)
- Calculate Power (for specified Sample Size)

Enter a value for  $\mu_1$ :

Enter a value for  $\mu_2$ :

Enter a value for sigma:

- 1 Sided Test
- 2 Sided Test

Enter a value for  $\alpha$  (default is .05):

Enter a value for desired power (default is .80):

The sample size (for each sample separately) is:

Kuva 2. Otoksoon laskeminen. Laskuriin on laitettu keskiarvot ja -hajonta, p-arvo ja voimataso. Kun nämä on syötetty, voidaan laskea otoskoko per ryhmä.

## Lähteet

Matthews D.E., Farewell V.T. (2007) Using and Understanding Medical Statistics.  
Sarna S. (2011) Kliinisen biostatistiikan peruskurssi.

Tämän tutkimusoppaan on laatinut Tuomas Selander, biostatistikko, KYS Tiedepalvelukeskus, tuomas.selander (at) kuh.fi