

2.3.2 Merkitsevyytestit

Merkitsevyytesti on keskeisessä roolissa, kun tehdään tilastollisia päätelmiä aineistosta.

Merkitsevyytesti palauttaa todennäköisyyden (p , engl. *probability*), jolla nollahypoteesi on voimassa. Toisin sanoen, p antaa todennäköisyyden eli kertoo riskin sille, että nollahypoteesin hylkääminen on väärä päätelmä. P saa arvoja suljetulla välillä $[0, 1]$. Mitä suurempi p on, sitä suuremmalla todennäköisyydellä havaittu jakauma on syntynyt sattumalta; tällöin nollahypoteesin hyväksyminen on oikea päätelmä. Vastavuoroisesti mitä pienempi p on, sitä pienempi on todennäköisyys sille, että havaittu jakauma olisi syntynyt pelkästään sattumalta ja että nollahypoteesin hyväksyminen olisikin oikea päätelmä. Eli mitä pienempi merkitsevyytestin palauttama todennäköisyysarvo on, sitä pienempi riski on sille, että nollahypoteesin hylkääminen olisikin väärä päätelmä.

On tärkeä pitää mielessä, että päätelmä tehdään aina otoksen perusteella. Tämän vuoksi emme voi olla täysin varmoja siitä, että jos samasta perusjoukosta otettaisiin uusi otos useita kertoja, päätyisimme aina samaan johtopäätökseen. Päätelmän tekeminen nollahypoteesin hylkäämisestä sisältää siis aina riskin ja jotta riski olisi tarpeeksi pieni, p -arvon tulisi vastaavasti olla riittävän pieni. Päätelmien tekemisessä käytetään usein sovittuja raja-arvoja tälle riskille:

- $p \leq 0,05$. Tällöin katsotaan, että aineisto ei tue nollahypoteesia tarpeeksi vaan se hylätään ja hyväksytään vastahypoteesi. Tilastotieteessä sanotaan, että tulos on tällöin tilastollisesti melkein merkitsevä ja sitä merkitään joskus yhdellä tähdellä (*).
- $p \leq 0,01$. Tällöin, katsotaan, että aineisto ei tue nollahypoteesia vaan se hylätään ja hyväksytään vastahypoteesi. Tilastotieteessä sanotaan, että tulos on tällöin tilastollisesti merkitsevä ja sitä merkitään toisinaan kahdella tähdellä (**)
- $p \leq 0.001$. Tällöin katsotaan, että aineisto ei lainkaan tue nollahypoteesia vaan se hylätään ja hyväksytään vastahypoteesi. Tilastotieteessä sanotaan, että tulos on tällöin tilastollisesti erittäin merkitsevä ja sitä merkitään joskus kolmella tähdellä (***)

Ihmistieteissä käytetään yleisesti raja-arvona $p = 0,05$. Tällaiset raja-arvot ovat kuitenkin sopimuksenvaraisia eikä niihin kannata hirttäytyä. Mitä enemmän olen itse tarkastellut erilaisia aineistoja tilastollisesti, sitä vakuuttuneemmaksi olen tullut siitä, että tilastollisen merkitsevyyden lisäksi on tärkeä selvittää tuloksen herkkyyttä virheille analyysissa tai muutoksille otannassa. Jos aineistoni koostuu esimerkiksi 30 havainnosta ja saan merkitsevyydestin tulokseksi $p = 0,045$, voi olla, että virhe jo yhdessäkin havainnossa muuttaa todennäköisyyden ei-merkitseväksi (esim. $p = 0,051$). Vaikka siis raja-arvot ovatkin tärkeitä, ei niiden varaan kannata laskea liiaksi.