

Merkitsevyystestit (Kaius Sinnemäki)

Transkriptio

Tässä videossa käsitellään merkitsevyystestejä. Tilastollista päättelyä koskevat testit voidaan karkeasti jakaa kolmeen ryhmään. Ensinnäkin muuttujien välisiä suhteita mittaavat testit ja myös muuttujien jakaumia mittaavat testit muodostavat ensimmäisen ryhmän, toiseksi ovat merkitsevyystestit ja kolmanneksi luotettavuustestit.

Muuttujien välisiä suhteita mittaavat testit tyypillisesti palauttavat semmoisen testisuureen, joka kertoo muuttujien välisen korrelaation voimakkuudesta tai muuttujien välisen suhteen voimakkuudesta tai muuttujien välisestä riippuvuudesta. Merkitsevyystestit palauttavat niin sanotun todennäköisyysarvon eli p-arvon, ja tämä p-arvo liittyy sellaisiin päätelmiin, mitä me tehdään muuttujien välisestä suhteesta ja tähän päätelmään liittyvästä riskistä. Luotettavuustestit puolestaan liittyvät tulosten virhemarginaaliin. Käsitellään siis merkitsevyystestejä tässä.

Tutustutaan tutkimuskirjallisuuteen, jossa on käsitelty... käytetty jotakin menetelmää kuten tässä Spearmanin järjestyskorrelaatiota, ja sitten on saatu jokin tulos, ja sitten todetaan, että se tulos on tilastollisesti merkitsevä kuten tässä. Eli on saatu jonkinlainen korrelaatio ja sitten annetaan p-arvo ja sitten kerrotaan, että tämä tulos on tilastollisesti merkitsevä.

Toisinaan tämä p-arvo ja sen tulkinta voi vähän vaikuttaa mysteeriltä, että mitä se oikein on. Ja joskus voi erehtyä luulemaan, että se tarkoittaa muuttujien välisen suhteen voimakkuutta, mitä se kuitenkin ei tarkoita. Selvitän hetken päästä, mitä se tarkoittaa.

Toisinaan, kun tarkastellaan tutkimuskirjallisuutta, niin siellä sitten taas sanotaan, että p-arvo on jotain vaikkapa kuten tässä 0,17, ja sen perusteella todetaan, että muuttujien välillä ei ole merkitsevää tilastollista yhteyttä. Eli tässä tapauksessa p-arvo on sen verran suuri, että tehdään päätelmä muuttujien välisestä yhteydestä, että ei siellä semmoista todellista yhteyttä kuitenkaan ole.

Mikä tämä merkitsevyystesti sitten on, ja mitä sillä tehdään? Merkitsevyystestin avulla ensinnäkin... sillä tehdään päätelmiä aineistosta, päätelmiä niistä tuloksista mitä on saatu vaikkapa korrelaatiotestin avulla tai vaikkapa khiin neliö -testin avulla. Ja testi palauttaa todennäköisyyden eli p-arvon sille, että nollahypoteesi on voimassa. Toisin sanoen p-arvo

kertoo riskin siitä, että nollahypoteesin hylkääminen on väärä päätelmä. Koska merkitsevyydestä liittyy nollahypoteesiin ja siitä tehtävään päätelmään, niin on tärkeitä muodostaa nollahypoteesi mielekkäällä tavalla siinä vaiheessa jo, kun suunnitellaan tutkimusta, niin suunnitellaan ja mietitään sitä, mikä olisi tai mitkä olisivat sellaisia järkeviä nollahypoteeseja, joita sitten voisi testata ja joiden paikkansapitävyyttä voi näillä tilastollisilla testeillä sitten arvioida.

No, eli kun p kertoo riskin sille, että nollahypoteesin hylkääminen on väärä päätelmä, niin katotaan sitä vähän tarkemmin. Ensinnäkin p saa arvoja 0:n ja 1:n välillä, ja tyypillisesti sellaisena raja-arvona pidetään, että $p = 0,05$. Tätä käytetään yleisesti humanistisissa tieteissä, yleisesti ihmistieteissä. Eli raja-arvona sille, että pidetäänkö tulosta jotenkin tilastollisesti merkitsevänä vai ei. Jos p on aika pieni eli se on pienempi kuin tuo 0,05, niin sitä pienempi on riski sille, että nollahypoteesin hylkääminen on väärä päätelmä. Eli vaikkapa täällä, missä käytettiin Spearmanin järjestyskorrelaatiota ja korrelaation voimakkuudeksi tuli -0,18 ja p on 0,028, niin tällöin voidaan todeta, että kun p on alle 0,05, niin on aika pieni riski sille, että kun me tämän tuloksen perusteella hylätään nollahypoteesi, niin me tehtäis väärä päätelmä.

Nollahypoteesi tässä tapauksessa olisi se, että muuttujien välillä ei ole mitään korrelaatiota. Vastahypoteesi olisi se, että muuttujien välillä esiintyy korrelaatio. Koska p -arvo on niin pieni, niin me tehdään tilastollinen päätelmä, että hylätään nollahypoteesi siitä, että muuttujien välillä ei esiintyisi mitään korrelaatiota, ja hyväksytään vastahypoteesi, että niiden välillä esiintyy tilastollisesti merkitsevä yhteys. Tämä tilastollisesti merkitsevä yhteys tässä tarkoittaa sitä, että se on jotenkin huomionarvoinen merkityksellinen, että siellä muuttujien välillä oikeasti on jonkinlainen havaittavissa oleva suhde ja että se ei ole rajoittunut pelkästään siihen otokseen, vaan että se esiintyy myös siellä perusjoukossa.

Jos taas p on suuri, jos se on suurempi kuin 0,05 tai vielä sitäkin suurempi, niin sen perusteella todetaan, että sitä todennäköisemmin havaittu jakauma on syntynyt sattumalta. Ja tällöin, jos me hyväksyttäisiin nollahypoteesi, niin se olisi aivan oikea päätelmä. Mutta vastahypoteesin hyväksyminen sitä vastoin olisi väärä päätelmä. Eli palataan tähän khiin neliö -esimerkkiin. Täällä khiin neliön kertymäärä saa arvoksi 6,39 ja p on 0,17 eli tässä tapauksessa p on yli tuon raja-arvon 0,05, joten me tehdään päätelmä, että nollahypoteesi tässä tapauksessa jää voimaan. Tässä khiin neliön testissä, kun testataan muuttujien välistä riippuvuutta, niin nollahypoteesi on se, että muuttujien välillä ei ole mitään riippuvuutta eli muuttujat eivät riipu toisistaan. Vastahypoteesi taas olisi sellainen, että muuttujien arvot

riippuvat toisistaan. Mutta koska p on noinkin suuri, niin me tehdään päätelmä, että nollahypoteesi jää voimaan, ja koska se on tuon suuruinen tuo p -arvo, niin riski sille, että me tehtäisiin tällöin väärä päätelmä, on aika pieni.

Eli toivottavasti tämä selvitys vähän auttaa ymmärtämään sitä, että mitä p -arvo on, mitä se tarkoittaa ja kun sitten käydään tarkemmin käsittelemään khiin neliö -testiä, niin silloin ehkä paremmin myös aukeaa se, että millaisia päätelmiä me voidaan tehdä sen meidän aineiston perusteella.