

Khiin neliö (Kaius Sinnemäki)

Transkriptio

Tässä videossa käsittelen Excel-taulukkoa, jonka avulla khiin neliö -testiin liittyviä arvoja voi laskea melko lailla automaattisesti. Näitä arvoja ovat khiin neliön kertymäarvo, odotusarvot, vapausasteet ja merkitsevyys eli p-arvo. Käytän tässä esimerkkinä omaa kielitypologista tutkimustani, jossa olen analysoinut 50:stä kielestä sijan merkintään ja sanajärjestykseen liittyviä tietoja. Nämä tiedot ovat tässä muistiossa. Valitsen ne täältä, kopioin ja tuon suoraan tähän näkymään, noin.

Sanajärjestys tässä tarkoittaa sitä, että se saa arvoksi "no" silloin, kun subjektin ja objektin keskinäinen sanajärjestys ei erottele niiden syntaktisia funktioita toisistaan. Ja sitten taas rigid_order saa arvoksi "yes", kun sanajärjestyksestä käytetään subjektin ja objektin erottelemiseksi. Sijan merkintä "case" tässä tarkoittaa vastaavia funktioita eli "no" tarkoittaa sitä, että sijaa ei käytetä subjektin ja objektin erottelemiseksi toisistaan, ja "yes" tarkoittaa, niitä käytetään tähän funktioon.

Ristiintaulukoidaan tämä aineisto Insert PivotTable, ja nyt se tarjoaa... Excel tarjoaa meille nämä kolme saraketta täältä automaattisesti. Sanotaan OK, sitten vedetään tuolta rigid_order riveille ja sanotaan, että se on "Rigid order". Sitten sija eli "case" sarakkeisiin, ja sanotaan, että sarakkeen nimi on Case. Ja sitten kielet vedetään tuolta tuonne Values-kohtaan. Ja näin me ollaan ristiintaulukoitu tämä aineisto.

Seuraavaksi voidaan sitten sijoittaa tänne havaintoarvoihin, mitä me ollaan saatu täällä ristiintaulukoimalla tuo aineisto. Eli voidaan tänne laittaa 7, 18, 15 ja 10. Ja nyt huomataan, että nämä sarakesummat ja rivisummat on aivan samat kuin mitä täällä alkuperäisessä ristiintaulukoidussa taulukossa. Tämä taulukko laskee odotusarvot automaattisesti jokaiselle solulle, samoin se laskee khiin neliön kertymäarvot jokaiselle solulle. Täältä nähdään khiin neliön kertymäarvo koko taulukolle, vapausasteet nähdään tuosta ja sitten taulukko laskee myös p-arvon automaattisesti tälle aineistolle.

Tässä seuraavalla rivillä on khiin neliön kriittinen arvo tapaukselle 0,05, eli tuossa on 3,84 ja risat, joka tarkoittaa sitä, että tuossa kun khiin neliön kertymäarvo on tuon suuruinen, niin silloin $p=0,05$.

Seuraavalla rivillä nähdään khiin neliön kertymäarvon kriittinen arvo tapaukselle $p=0,01$. Eli kun khiin neliön kertymäarvo on 6,63, tällöin $p=0,01$ kun vapausasteet tällaiselle taulukolle on 1.

Seuraava rivi tässä näyttää suurimman kertymäarvon näistä khiin neliön kertymäarvoista. Eli suurin khiin neliön kertymäarvo on 1,45. Ja sitten vielä seuraavalla rivillä nähdään aineiston koko, joka tässä tapauksessa on 50.

Täällä lisäksi olen antanut muutamia aputyövälineitä. Eli jos $p=0,05$ tai alle, niin silloin voidaan suurin piirtein tämmöistä tekstiä käyttää ja ottaa täältä näistä soluista arvot, kopioida tänne. Eli 5,19 tuonne, $df...$ solusta N3 tuonne, ja solusta N4 p eli 0,02. Ja tämän tekstin voisi sitten kopioida tekstinkäsittelyohjelmaan, ja vähän vielä muokata sitä, antaa vähän lisätietoa, ja sitä voi käyttää siihen, kun tuloksia raportoidaan.

Jos p sitä vastoin on suurempi kuin 0,05, tällöin vastaavasti voidaan raportoida muuttujat ovat tilastollisesti toisistaan riippumattomia, ja samoin täydentää tänne vastaavat soluarvot.

Tällä välilehdellä nämä kaikki laskennat perustuu siihen, että meillä on tämmöinen ristiintaulukoitu aineisto, jossa on kaksi riviä ja kaksi saraketta, eli molemmilla muuttujilla on kaksi luokkaa.

Seuraavalla välilehdellä nähdään tapaus, jossa on vapausasteet 2 eli meillä on toisella muuttujalla kolme luokkaa, toisella muuttujalla kaksi luokkaa. Täällä vapausasteet on 2, ja kaikki nämä relevantit arvot päivittyy, kun tänne muutetaan havaintoarvoihin jotain. Sama, jos on vapausasteet 3, eli tällöin tapaus voisi olla, että toisella muuttujalla on kaksi luokkaa ja toisella muuttujalla neljä luokkaa. Sitten vapausasteet 4, 5 ja 6. Kun mitä suurempiin vapausasteisiin mennään, niin silloin tulee vastaan myös se, että molemmilla muuttujilla voi olla useampia luokkia kuin kaksi. Eli tässä tapauksessa toisella muuttujalla, A-muuttujalla, on kolme luokkaa, ja B-muuttujalla neljä luokkaa. Jos tarvitset suurempia taulukoita kuin nämä, eli jos vapausasteet ovat vielä suurempia, niin voit itse täydentää näitä havaintoarvo-, odotusarvo, khiin neliön kertymäarvo -sarakkeita.

Mutta tämän taulukon avulla pääsee hyvin vauhtiin.