

# Odotusarvojen laskeminen (Kaius Sinnemäki)

## Transkriptio

Tässä videossa käsittelen khiin neliö -testiä. Käsitellään odotusarvojen laskeminen ja khiin neliön kertymäarvojen laskeminen. Tehdään nämä käsin tässä videossa. Lisämateriaaliin olen ladannut videon ja Excel-taulukon, jonka avulla khiin neliö -testin eri testisuureet ja arvot saadaan laskettua automaattisesti.

Käytän tässä tätä samaa kuvitteellista aineistoa, mitä ristiintaulukointia koskevassa videossa. Käyn tässä vähän nopeammin nyt tämän ristiintaulukoinnin tekemisen. Eli PivotTable-komennon avulla luodaan tämä ristiintaulukoitu aineisto tänne. Noin. Ja sit täältä nuo blank-rivi ja -sarake pois, ja sitten muutetaan tänne näin muutamien solujen sisältö.

Tehdään tässä niin, että kopioidaan tämä taulukko alemmas ja liitetään se niin, että pelkät arvot liitetään ja laitetaan tänne "Havaintoarvot". Sitten laitetaan tänne alapuolelle "Odotusarvot" ja sit aletaan laskea näitä odotusarvoja.

Eli odotusarvothan lasketaan kullekin solulle sillä tavalla, että kerrotaan sitä solua vastaava sarakesumma sitä solua vastaavalla rivisummalla ja jaetaan otoskoolla. Eli jos otetaan tämä, että miehet eivät tupakoi, niin kerrotaan sitä saraketta vastaava sarakesumma sitä riviä vastaavalla rivisummalla ja jaetaan otoskoolla. Niin saadaan sinne, että odotusarvoisesti 26 miestä ei tupakoi. Sitten sama tehdään naisten kohdalle. Tuolta sarakesumma kerrotaan rivisummalla, jaetaan sadalla. Samaten odotusarvoisesti 26 naista ei tupakoi tämän aineiston perusteella. Sitten otetaan niille, jotka tupakoi, niin tuolta kerrotaan sarakesumma rivisummalla ja jaetaan jälleen otoskoolla. Ja sama sitten tälle viimeiselle solulle myös. Ja näin saatiin tuonne odotusarvo. Lasketaan näille myös summat, niin tämä pikkusen selventää sitä, mitä tässä on tapahtunut.

Eli jos tarkastellaan havaintoarvoja, niin meillä on siellä sarakesummissa 52 ja 48. Ja odotusarvoissa on myös täysin samat sarakesummat. Rivisummina on 50 ja 50 sekä havainto- että odotusarvoissa.

Khiin neliön kertymäarvot lasketaan nyt sitten vertaamalla odotusarvoja havaintoarvoihin ja tarkastelemalla, miten paljon havaintoarvot poikkeavat näistä odotusarvoista, jotka on laskettu pelkästään näitten sarake- ja rivisummien eli taulukon marginaalien perusteella.

Luodaan tänne siis tälleen nämä sarakkeiden rivet nimet kopioidaan tänne ja nyt kirjoitetaan tänne "Khiin kertymäarvot".

Khiin neliön kertymäarvot saadaan vähentämällä havaintoarvoista odotusarvo, korottamalla se toiseen potenssiin ja jakamalla odotusarvolla. Tämä voidaan kopioida kaikkiin näihin soluihin. Jos halutaan koko taulukkoa koskeva khiin neliön kertymäarvo, niin se saadaan laskemalla yhteen nämä kaikki solukohtaiset kertymäarvot. Eli tälle havaintoaineistolle khiin neliön kertymäarvo on 5,76 tai 5,77 tarkemmin ottaen. Näin siis voidaan laskea khiin neliön odotusarvot ja kertymäarvot manuaalisesti Excelissä.

Lasketaan tästä nyt vielä, mikä on tuo p-arvo ja sitten vielä vapausasteet. Eli vapausasteet lasketaan sillä tavalla [epäselvää]... sitä nyt ei tässä tarvi välttämättä edes laskea, koska tämä on niin yksinkertainen aineisto. Eli meillä on kaksi saraketta ja kaksi riviä, ja tällaisen taulukon vapausasteet on 1. Ja saadaan laskemalla siten, että otetaan sarakkeiden määrä, vähennetään siitä 1, ja sitten otetaan rivien määrä, vähennetään siitä 1 ja otetaan näiden tulo. Eli saadaan tuloksi  $1*1=1$ .

Sitten tuo p-arvo, tämä voidaan tehdä kahdella eri tavalla Excelissä. Yksi tapa on käyttää komentoa CHISQ.DIST.RT, jolle annetaan ensimmäiseksi arvoksi tuo koko taulukon khiin neliön kertymäarvo ja sitten puolipiste ja sitten vapausasteet. Niin saadaan tuo p-arvo tällä tavalla. Toinen tapa laskea p-arvo on käyttää komentoa CHISQ.TEST. Tässä ensimmäiseksi muuttujaksi maalataan nämä havaintoarvot – ei rivisummia eikä sarakesummia, ainoastaan havaintoarvot! – erotetaan puolipisteellä ja sitten lasketaan tai maalataan täältä, valitaan täältä nämä odotusarvot ja jälleen ei valita sarake- tai rivisummia. Sulkeet kiinni, enteriä, niin saadaan tätä havaintoaineistoa vastaava khiin neliön kertymäarvo ja p-arvo.

Tästä huomataan se, että riippumatta siitä, että kummalla tavalla tämä p-arvo lasketaan, niin meidän täytyy ristiintaulukoida aineisto ja laskea nämä odotusarvot. Khiin neliön kertymäarvot voidaan myös laskea ja oikeastaan tietenkin tulosten raportointiin täytyy tietää khiin neliön kertymäarvo, vapausasteet ja sitten p-arvo.

Tässä videossa en käy nyt läpi enempää. Tosiaan lisämateriaalissa näytän, miten käytetään tämmöistä valmista khiin neliö -taulukkoa pohjana, kun lasketaan khiin neliön kertymäarvoa, vapausasteita ja p-arvoa tietynlaisille arvoille.