

Havaintoja khiin neliö -testin käytöstä amk-opinnäytteissä

17.6.2025

Tilastollisten testien käyttö opinnäytteessä antaa numeeristen tulosten analyysiin syvyyttä ja tuo esiin tekijän kyvyn soveltaa käytäntöön tilastotieteen kurssilla käsiteltyjä asioita. Tilastomatemaattinen vertailu laajentaa näkökulmaa määrällisiin tuloksiin: opinnäytteen tekijä voi esimerkiksi tutkia, onko kyselyssä esiin tullut ero naisten ja miesten asenteissa yleistettävissä laajemmin kohdeperusjoukkoon. Tilastollisia menetelmiä on käytetty amk-opinnäytteissä jo pidempään (Kilpelä, 2021). Tämän artikkelin kirjoittajan omakohtaisen kokemuksen perusteella niiden käyttö on kuitenkin verraten harvinaista; valtaosassa töitä numeerisia eroja ryhmien välillä kuvataan pelkästään prosenttiluvuin.

Eräs yleisimmin käytettyjä ja kursseilla useimmin käsiteltyjä testejä on Pearsonin khiin neliö -testi (χ^2-testi), jolla tarkastellaan, esiintyykö aineistossa tilastollisesti merkitseviä eroja. Vertailu perustuu odotettujen ja havaittujen frekvenssien väliseen eroon tai yhtäläisyyteen. Mitä suurempi odotettujen ja havaittujen frekvenssien ero on, sitä todennäköisempää on, että syynä on jokin muu kuin pelkkä sattuma. Lähtöoletuksena on nollahypoteesi, eli että tilastollisesti merkitsevää eroa ei ole, jolloin saatujen arvojen jakauma vastaa χ^2-jakaumaa, joka puolestaan perustuu vapausasteiden (*degrees of freedom*, lyh. df) lukumäärään: (rivien määrä –1) x (sarakkeiden määrä –1). χ^2-testillä tutkitaan havaintojen lukumäärien eron merkitsevyyttä siitä näkökulmasta, että kuinka suuri on riski, että ryhmien välillä ei todellisuudessa olisikaan eroa vaan saatu tulos selittyisi sattumalla – eikä siten olisi yleistettävissä laajempaan populaatioon.

Tilastollista merkitsevyyttä mitataan riskitasolla, joka on tavallisimmin 5 %. Se ilmaistaan p-arvolla (*probability value*), joka ilmaisee todennäköisyyttä saada yhtä merkittävä ero tuloksessa nollahypoteesiin perustuen. Tavallisin kynnys tilastolliselle merkitsevyydelle on $p < 0,05$, joka vastaa 5 %:n riskitasoa, mutta tutkittavasta asiasta ja tieteenalasta riippuen merkitsevyysraja voidaan asettaa pienemmäksikin.

Khiin neliö -testin käyttöön liittyviä ongelmia opinnäytteissä

Aihe tämän artikkelin kirjoittamiseen lähti kirjoittajan omista kokemuksista. Päättyneen lukuvuoden aikana on tullut vastaan kaksi tekeillä ollutta opinnäytettä, joissa oli käytetty χ^2-testiä ilman, että testin käyttöehdot ovat täyttyneet. Molemmissa töissä ongelma liittyi odotettuihin (teoreettisiin) frekvensseihin. Testin käyttöehto on, ettei 2 x 2 -taulukossa saa olla yhtäkään alle viiden (5) suuruista odotettua frekvenssiä eikä suuremmissa taulukoissa tällaisia saa olla useammassa kuin viidesosassa soluista (20 %), joissa ei kuitenkaan saa olla lainkaan alle yhden (1) suuruisia odotettuja frekvenssiä. Tämä vaatimus saattaa unohtua testiä vähän käyttäneeltä, etenkin kun testi tuottaa tuloksia, vaikka nämä olisivat virheellisiä.

Edellä esitetyn perusteella kirjoittajaa kiinnosti tarkastella, esiintyisikö havaittua muuallakin amk-opinnäytteissä, eli onko kyseessä laajempikin ilmiö. Siten 11.6.2025 kerättiin tarkasteltavaksi 15 amk-opinnäytettä, joissa testiä oli käytetty. Opinnäytteet haettiin Google-hauilla, ja hakukenttään kirjoitettiin ”khiin neliö -testi” ja ”Theseus”. Tarkasteluun otettiin vastauksissa esiin tulleet 15 ensimmäistä työtä. Useissa töissä esitettyjen tulosten alhaisten solufrekvenssien (kuten 0, 1, 2 jne.) perusteella heräsi epäilys, että odotetut frekvenssien perusteella testin käyttöedellytykset eivät mahdollisesti täyttyisi, ja sama testi tehtiin tuloksille uudelleen. Tarkastelun tulokset ovat seuraavat.

Ensimmäisessä tarkastellussa työssä kiinnitetään huomiota odotettuihin frekvensseihin siten, että taulukosta, jonka tuloksiin testiä oli käytetty, todetaan, että tulokset eivät ole luotettavia, koska 58 % teoreettisista solufrekvensseistä on alle 5.

Toisessa työssä ei mainita rajoituksista, mutta työstä löytyy taulukko, jossa alle viiden (5) suuruisia frekvenssejä on 22,2%, eli edellytys ei olisi täyttynyt.

Kolmannesta työstä löytyy taulukko, jossa alle viiden suuruisia odotettuja frekvenssejä on peräti 77,8%.

Neljännessä ja **viidennessä** työssä ei ole ongelmia.

Kuudennessa työssä on taulukko, jossa 25 %:ssa soluista odotetut frekvenssit jäävät alle viiden, mutta samassa työssä raportoidaan tuloksia myös oikein suorituista testeistä.

Seitsemännessä työssä asiaa ei voi tutkia, koska taulukoissa on esitetty pelkät prosentuaaliset osuudet.

Kahdeksannessa työssä ei ole ongelmia.

Yhdeksännessä työssä on tiedostettu testin käyttörajoitukset, ja testiä ei ole käytetty silloin, kun se ei ole ollut sallittua. Samoin on menetelty myös **kymmenennessä** työssä.

Yhdennestätoista työstä löytyy taulukko, jossa odotettu frekvenssi jää alle viiden 57,1 %:ssa soluista.

Kahdestoista työ on mielenkiintoinen: asiasta on puhuttu, ja testiä ei vallitsevasti ole käytetty, jos käyttö ei ole sallittu, mutta yhdessä taulukossa 70 %:ssa soluista frekvenssi on alle viisi. Tekijä pitää kuitenkin tulosta validina ja toteaa, että ”tilastollista riippuvuutta ei löytynyt”.

Kolmastoista työ ei tarjoa mahdollisuutta arvioida asiaa, koska absoluuttisia arvoja ei ole esitetty, pelkästään prosenttiarvoja. Toisaalta työssä on taulukko, jossa on sarakeotsikko ”Fraction < 5”, jonka alapuolella kaikki tarkastellut riippumattomat muuttujat saavat arvon 0. Tekijä ei kuitenkaan avaa asiaa lainkaan.

Neljännestätoista työstä löytyy taulukko, jossa 40 %:ssa soluista odotettu frekvenssi jää alle viiden. Lisähavaintona mainittakoon, että tekijä käyttää termiä ”khiin tarkka neliötesti”. On huomattava, että χ^2-testi ei oikeasti ole tarkka testi, kuten Fisherin eksakti (tarkka) testi.

Viidennessätoista työssä ei ole havaittavissa ongelmia.

Johtopäätökset ja pohdinta

Tarkastelu osoittaa, että ongelmia löytyy viidestätoista opinnäytteestä peräti kuudessa (40 %). Vaikka otos on hyvin pieni, jo tämä tulos riittää osoittamaan, että kyseessä on verraten tavallinen ilmiö opinnäytteissä.

Keskeisimmäksi syyksi χ^2-testin käyttöön tilanteissa, joissa sen käyttö ei ole sallittu, voi ensiksikin olettaa testin käyttäjien kokemattomuuden. Jokaisen asian oppiminen vaatii toistoa ja harjoittelua. Asian luonteen takia on epätodennäköistä, että opinnäytettään tekevä opiskelija olisi muissa yhteyksissä aiemmin joutunut tekemisiin testin kanssa, joten testin käytön harjoittelu on mitä todennäköisimmin tapahtunut samalla, kun opinnäytettä on kirjoitettu.

On todennäköistä, että useimmat testiä käyttäneet opiskelijat ovat aiemmin osallistuneet tilastotieteen kurssille. Aihealueena tilastolliset menetelmät ovat kelle tahansa opiskelijalle siinä määrin uusi ja erilainen kokonaisuus, että on ymmärrettävää, että yksittäisen testin soveltaminen käytäntöön ja harjoittelu on saattanut jäädä vähemmälle. Toisaalta juuri χ^2-testin voisi yleisyytensä takia olettaa saavan kursseilla huomiota. Teoriapohjan puutteisiin viittaa myös, ettei opiskelija ole ymmärtänyt SPSS-ohjelman tuottamaa, taulukon alta löytyvää englanninkielistä mainintaa odotettujen frekvenssien määrästä suoraan testin käyttökelpoisuuteen liittyväksi.

Testin ei-sallittuun käyttöön myötävaikuttavat myös pienet otoskoot. Kyselytutkimusten vakio-ongelma on riittävän vastausmäärän saaminen, kun samaan aikaan potentiaalisille vastaajille satelee pyyntöjä osallistua muihinkin kyselyihin. Tällöin opinnäytteen tekijä joutuu helposti toteamaan, että ”tässä vastaukset nyt ovat, ja näillä mennään”.

Ratkaisuksi tarkasteltuun ongelmaan voi esittää opetuksen ja ohjauksen puolelle vaihtoehtoisten tilastollisten menetelmien tuomista esiin. Kirjallisuudessa suositeltavana vaihtoehtona on χ^2-testin korvaaminen eksakteilla testeillä, kuten Fisherin testillä, joita voi käyttää myös pienellä otoskoolla.

Ilpo Kempas
FT, dosentti, lehtori
SEAMK

Ilpo Kempas on filosofian tohtori ja dosentti, joka työskentelee romaanisten kielten ja englannin lehtorina Seinäjoen ammattikorkeakoulussa, opettaen myös tutkimuskirjoittamista. Kempas on kielentutkija, jonka

tutkimusaloja ovat romaaniset kielet ja tieteellinen kirjoittaminen.

Lähteet

Kilpelä, V. (2021). *Tilastollisten menetelmien toteutunut käyttö Liiketalouden koulutusohjelmien AMK-opinnäytetöissä 2017–2019 ja ei-parametristen testien mahdollisuudet pienten aineistojen analyysissä* [amk-opinnäyte, Haaga-Helia ammattikorkeakoulu]. https://www.theseus.fi/bitstream/handle/10024/501979/Kilpela_Vesa.pdf?sequence=2