

Aineistotyyppi tilastollisessa kielentutkimuksessa (Kaius Sinnemäki)

Transkriptio

Tässä videossa käsitellään aineistotyyppiä. Aineistotyyppit jaetaan tyypillisesti neljään pääasteikkoon. On erittäin tärkeä tuntea oman aineistonsa muuttujat ja niiden kunkin muuttujan aineistotyyppi. Tämä johtuu siitä, että kukin tilastollinen testi sopii vain tietyntyyppisille aineistotyypeille. Eli on siis tärkeää tuntea toisaalta oma aineistonsa, mutta toisaalta myös erilaisia tilastollisia menetelmiä, että millaiselle aineistolle tietty menetelmä soveltuu. Jos valitsee sellaisen menetelmän, joka ei sovellu omaan aineistoon tai sen aineistotyyppiin, niin tämä johtaa tietysti harhaanjohtaviin tuloksiin.

Neljä eri aineistotyyppiä ovat luokitteluasteikollinen, järjestyslukuasteikollinen, välimatkaasteikollinen ja suhdeasteikollinen aineistotyyppi. Käsitellään nämä neljä tyyppiä seuraavaksi kukin erikseen.

Luokitteluasteikollinen muuttuja saa arvoksi sellaisia arvoja, jotka voidaan analysoida ainoastaan erilaisiin luokkiin. Nämä arvot ovat yleensä luokkien nimiä merkkijonoina, eikä niitä voi laittaa mihinkään järjestykseen. Luokitteluasteikollinen muuttuja on kielentutkimuksessa hyvin tyypillinen muuttuja. Esimerkkejä luokitteluasteikollisista muuttujista ovat esimerkiksi puhujan äidinkieli, sukupuoli, kotipaikkakunta tai vaikkapa murrealue. Kieliopillisista kategorioista vaikkapa substantiivien luku saa arvoksi luokitteluasteikolliselle muuttujalle tyypillisiä arvoja eli luokkien nimiä, yksikkö tai monikko. Verbin aspekti on myös tyypillinen luokitteluasteikollinen muuttuja, koska se saa arvoksi esimerkiksi perfekti tai imperfekti.

Luokitteluasteikolliselle muuttujalle sopivia tilastollisia menetelmiä ovat mm. khiin neliö - testi, Fisherin tarkka testi tai logistinen regressio. Tässä metodipankin osiossa käsitellään ainoastaan khiin neliö -testi.

Järjestyslukuasteikollinen muuttuja muistuttaa luokitteluasteikollista muuttujaa, mutta se eroaa siitä sillä tavalla, että muuttujan saamat arvot voidaan laittaa johonkin järjestykseen. Eli muuttuja saa arvoksi tyypillisesti luokkia, mutta ne voidaan laittaa johonkin järjestykseen. Tyypillinen esimerkki järjestysasteikollisesta muuttujasta on vaikkapa kilpailussa maaliintulojärjestys. Kielentutkimuksessa erilaiset hierarkiat voidaan tulkita järjestyslukuasteikolliseksi muuttujaksi. Muun muassa vokaalin suppeutta voidaan ajatella

järjestyslukuasteikollisena muuttujana: sen saamat arvot suppea, puolisuppea, puoliväljä ja väljä voidaan laittaa esimerkiksi suppeuden mukaan järjestykseen.

Mielipidemittauksissa käytetään usein niin sanottua Likertin asteikkoa: haastateltavalle esitetään jokin väite, ja sen jälkeen hänelle annetaan vaihtoehtoja, tyypillisesti viisi tai seitsemän, ja nämä eri vaihtoehdot voidaan asettaa tiettyyn järjestykseen, esimerkiksi ääripäinä täysin eri mieltä ja täysin samaa mieltä, ja sitten niiden välissä vaihtoehdot hieman lievempiä arvioita. Sopivia testejä järjestyslukuasteikollisille muuttujille ovat järjestyskorrelaatiotestit, kuten Kendallin tau.

Välimatka-asteikollinen muuttuja, sen saamat arvot ovat numeerisia, mutta asteikon nollakohta on sovittu, eikä se ole luonnollinen. Tällaisia muuttujia käytetään aika harvoin kielentutkimuksessa, esimerkkejä ovat vaikkapa lämpötila Celsius-asteikolla, jossa nollakohta on veden jäätymispiste, tai ajanlasku gregoriaanisessa kalenterissa. Sopivia testejä tällaiselle muuttujalle ovat esimerkiksi keskiarvo ja t-testi.

Viimeiseksi on suhdeasteikollinen muuttuja. Sen saamat arvot ovat numeerisia, ja asteikon nollakohta on luonnollisella tavalla motivoitu. Ei siis jotenkin sovittu, vaan se on luonnollisesti motivoitu. Esimerkiksi reaktioaika, sen nollakohta on 0 ja sitten se saa arvoja 0:sta ylöspäin mitattuna sekunteina, minuutteina, tunteina, niin edelleen. Muita esimerkkejä ovat ajanlasku radiohiilivuosina, vaikkapa puhujan ikä, puhujien määrä, kaikki erilaiset frekvenssit, lauseen pituus sanoina, sanan pituus äänteinä ja niin edelleen. Eli käytännössä erilaiset mittaamalla tai laskemalla saadut muuttujat ovat suhdeasteikollisia muuttujia. Tyypillisiä testejä, joita voidaan tällaisille muuttujille käyttää, ovat korrelaatiotesti – nimenomaan lineaarinen korrelaatio – ja t-testi.

Käydään sitten läpi vielä yksi esimerkki koskien Likert-asteikkoa, jota edellä jo sivuttiin. Kuten mainitsin, Likert-asteikkoa käytetään tyypillisesti kyselylomakkeissa, vaikkapa sosiolingvistikissa kyselyissä tai esimerkiksi kielipillisuuden arviointia koskevissa kyselyissä. Pohdi seuraavaksi, mitä aineistotyyppiä seuraava Likert-muuttuja on. Oleta, että haastateltava saa eteensä väittämiä, ja hänen saamat vastausvaihtoehdot on jo annettu tässä. Eli ääripäinä ”täysin eri mieltä” ja ”täysin samaa mieltä” ja sitten jotain siltä väliltä.

No, edellä käsiteltiin jo, että tämäntyyppinen muuttuja voidaan tulkita järjestysasteikolliseksi muuttujaksi. Mutta toisaalta tässä vastausvaihtoehdon viereen on myös annettu numero. Ja nämä numerot voidaan tässä tapauksessa tulkita niin, että ne ohjaavat haastateltavaa ajattelemaan, että nämä eri vastausvaihtoehdot ovat samalla etäisyydellä toisistaan. Tätä

voidaan vahvistaa asettamalla jonkinlainen skaala vastausvaihtoehtojen viereen, joka lisää sitä asteikon mieltämistä jonkinlaisena jatkuvakestoisena skaalana. Tämän tyyppistä Likert-muuttujaa, jossa on annettu numerot vieressä, niin se poikkeaa tämmöisestä järjestysasteikollisesta muuttujasta siinä, että sitä voidaan käsitellä suhdeasteikollisena muuttujana. Ja nyt tämän tyyppiselle Likert-muuttujalle voitaisiinkin käyttää esimerkiksi korrelaatiotestiä.

Tämä on hyödyllistä siinä mielessä, että jos ajatellaan tällaisia järjestyslukuasteikollisia muuttujia, niin niille ei ole olemassa kovin monia erilaisia tilastollisia testejä, joilla niitä voidaan testata. Mutta sitten taas suhdeasteikollisille muuttujille on olemassa useita erilaisia testejä, joilla voidaan niiden jakaumia testata. Joten vinkkinä se, että jos olet muodostamassa jonkinlaista kyselylomaketta ja käytät Likert-asteikkoa, niin käytä ihmeessä numeroita vastausvaihtoehtojen vieressä, jotta voit käsitellä sitten sitä aineistotyyppiä suhdeasteikollisena muuttujana.