

Ristiintaulukointi (Kaius Sinnemäki)

Transkriptio

Tässä videossa käsittelen ristiintaulukointia Excelissä. Olen luonut tänne Exceliin kuvitteellisen aineiston, jossa on kaksi muuttujaa: henkilön sukupuoli ja tupakoiko henkilö vai ei. Nämä molemmat muuttujat ovat luokitteluasteikollisia, joten tällaiselle aineistolle sopisi khiin neliö -testi.

Tällä aineistolla voisi tutkia esimerkiksi kysymystä, esiintyykö tupakoinnin ja sukupuolen välillä jonkinlainen riippuvuussuhde, esimerkiksi tupakoivatko miehet todennäköisemmin vai naiset. Nollahypoteesi olisi se, että sukupuoli ja tupakointi ovat toisistaan riippumattomia.

No ennen kuin khiin neliö -testiä tällaiselle aineistolle voisi alkaa soveltaa, niin tämä havaintoaineisto täytyisi ristiintaulukoida. Kerrataan vielä, mitä ristiintaulukointi tarkoittaa. Eli se tarkoittaa sitä, että me lasketaan kaikille näitten muuttujien arvojen yhdistelmille frekvenssit. Eli meillä on tässä "sukupuoli", se saa arvoksi tässä rivissä "nainen" ja "tupakoi" saa tällä rivillä arvoksi "ei", eli laskettaisiin "nainen" ja "ei tupakoi" -yhdistelmän frekvenssi eli montako naista ei tupakoi. Sitten laskettaisiin yhdistelmä "nainen" ja "kyllä tupakoi" -frekvenssi, jolloin me saataisiin laskettua frekvenssi sille, miten monta naista tämän aineiston perusteella tupakoi tai miten monta tässä otoksessa tupakoi. Sama sitten "mies" ja "ei", "mies" ja "kyllä tupakoi", saadaan näille kaikille frekvenssit laskettua.

Miten tämä sitten tapahtuu? No, ensin valitaan täältä jokin solu, mihin se ristiintaulukoitu aineisto halutaan sitten luoda eli vaikka tuosta D3. Sitten mennään täältä ylhäältä kohdasta Insert tai suomenkielisessä Excelissä Lisää ja PivotTable. Sitten avautuu tällainen valikko, jossa ensin pyydetään valitsemaan se alue, josta lasketaan tämä ristiintaulukoitu aineisto. Eli tässä on selvillä se aineisto, joten mennään vaan soluun A, pidetään Shift pohjassa, sitten nuoli oikealle, sitten pistetään Ctrl pohjaan ja sitten nuoli alas, niin se vie suoraan tuonne havaintoaineiston loppuun. Ja painetaan OK:ta. Ja sitten vieritetään ylöspäin. Eli tänne se on nyt luonut tällaisen taulukkopohjan.

Tänne oikealle avautuu kentät, joissa on nyt nämä meidän muuttujat "sukupuoli" ja "tupakoi" ja sitten rivit, sarakkeet ja arvot. Eli valitaan tuosta "sukupuoli", vedetään se tuonne rivi-kohtaan, ja "tupakointi" tuonne sarake-kohtaan ja vaikka "tupakointi" tuonne arvot-kohtaan. Eli näin tämä PivotTable-funktio automaattisesti laski meille näitten muuttujien arvojen

yhdistelmien lukumäärät. Eli montako miestä ei tupakoi, montako tupakoi, montako naista ei tupakoi, ja montako tupakoi.

Näitä solujen sisältöjä voi täällä pikkusen muokata. Eli esimerkiksi tässä on englanniksi annettu nämä... tässä on Grand Total niin mun pitäisi laittaa sinne "summa". Ja tässä on Row Labels, niin voitais siihen laittaa "sukupuoli". Ja tuonne "tupakoi" ja tätä kohtaa Count of tupakoi mä kokeilen deletellä poistaa, niin se ei anna poistaa sitä, niin voidaan siihen vaikka laittaa "taulukko A" tai jotain.

Tämmöisen ristiintaulukoinnin voi suorittaa myös ihan asteen verran yksinkertaisemmin. Eli näytän sen seuraavaksi. Eli valitaan tuosta vaikka solu D11, mennään kohtaan Insert ja sieltä PivotTable ja sitten tässä, missä pitää se alue valita, niin valitaan vain tuosta A ja B -sarakkeet ja painetaan OK:ta. Ja sitten samalla lailla tehdään: vedetään "sukupuoli" tuonne kohtaan Rows, ja "tupakointi" kohtaan Columns ja sitten tuonne Values-kohtaan "tupakoi". No, koska me valittiin sarakkeet kokonaisuudessaan, niin se tänne nyt sitten herjaa, että täällä on tämmöinen blank-kohta täällä sarakkeessa ja riveissä tässä ristiintaulukoidussa aineistossa myös. Me voidaan mennä tänne vaikkapa tuosta Column Labels, avataan valikko siitä ja sit klikataan tuosta blank-kohta pois ja painetaan OK. Niin silloin se poisti sieltä ne tyhjät arvot nyt tästä. Tällä tavalla voi myös luoda tän ristiintaulukoidun aineiston. Poistetaan nämä nyt tästä.

Näytän vielä, miten voi tästä aineistosta kätevästi muodostaa pylväskaavion. Eli mennään tuohon Pivot-taulukkoon, sit mennään tänne ylävalikkoon PivotTable Analyze ja täältä oikealta kohtaan PivotChart. Eli tästä voidaan sitten luoda jonkinlainen kaavio vaikkapa pylväskaavio, jossa on frekvenssit, tai sitten pinottu pylväskaavio, tai sitten tämmöinen suhteellinen pinottu pylväskaavio, jossa on y-akselilla prosentit. Voidaan myös muodostaa vaakaan nämä pylväät, jos nyt vaikka halutaan tuommoinen suhteellinen pinottu pylväskaavio luoda, niin tuosta sen voisi luoda ja tuohon se sitten tulisi. Noin, ja sit voidaan sen kokoa vähän muuttaa. Ja edelleen sen voisi tästä siirtää vaikkapa Wordiin tai jonnekin muuhun tekstinkäsittelyohjelmaan.

Tämä tällä erää tästä asiasta.