




Tarja Heikkilä

Tilastomenetelmien lopputyö

Lopputyössä on esimerkkejä erilaisista tilastomenetelmistä. Datatiedosto *Harjoitusdata.sav* on muokattu tätä harjoitusta varten, joten se ei vastaa kaikkien muuttujien osalta todellista aineistoa. Harjoitustyö on suunniteltu suoritettavaksi SPSS-ohjelmalla, mutta tehtävät voi tehdä muillakin tilasto-ohjelmilla. Myös Excel-ohjelmaa voi käyttää, erityisesti kuvioiden piirtäminen onnistuu sillä hyvin. Harjoitustyössä harjoitellaan yleisimpiä yksinkertaisia tilastomenetelmiä.

Tehtävät ovat seuraavista aiheista:

1. Muuttujien määrittely lomakkeen kysymyksistä
2. Aineistoon tutustuminen
3. Datatiedoston muuttujien arvojen tarkistus
4. SPSS-tiedostojen tallennus muiden ohjelmien käyttöön
5. Taustatietojen kuvaaminen
 - Frekvenssitaulukot
 - Tunnusluvut
 - Luokittelu
 - Pylväskuviot
6. Samaan kysymykseen kuuluvien vastausten yhdistäminen ja tulosten esittäminen
7. Kertausta edellisten tehtävien tilastomenetelmistä
8. Kysymysryhmien tulosten esittäminen
9. Ristiintaulukointi ja khiin neliö -testi
10. Tunnusluvut kokonaistuloksista
11. Korrelaatiokertoimet
12. Kahden ryhmän keskiarvojen vertailu ja t -testi
13. Usean ryhmän keskiarvojen vertailu ja siihen liittyvät testit
14. Avointen kysymysten vastausten esittäminen



Suorita harjoitustyön tehtävät. Kerää tehtävien 5–14 tulokset yhteen ja kirjoita niistä tutkimusraportti. Käytä raportissasi sanallisia otsikoita tehtävänumeroiden sijaan.

1. Määrittele muuttujat lomakkeesta. Katso erillinen tehtävä *Muuttujien määrittely*.
2. Avaa valmiiksi syötetty datatiedosto *Harjoitusdata.sav*. Vertaa omia muuttujien määrittelyjäsi tiedoston muuttujiin. Tutustu aineistoon.
3. Tarkista *Analyze/Descriptive Statistics/Descriptives* -komennolla, ovatko muuttujien arvot oikealla välillä.
4. Tallenna datatiedosto Excel-tiedostoksi *File/Save as* -komennolla valitsemalla tallennustyyppi *Save as type* -valikosta.

Tallenna *Output*-tiedosto Word- tai Excel-tiedostoksi komennolla *File/Export* valitsemalla haluamasi tiedostotyyppi kohdasta *Type* ja tallennuskansio *Browse*-painikkeella. Ennen tallennusta kannattaa ikkunan ylälaidasta valita kohta *All visible*, jolloin vain *Output*-tiedostossa olevat tiedot tallentuvat.

5. Raportoi vastaajien taustatiedot.

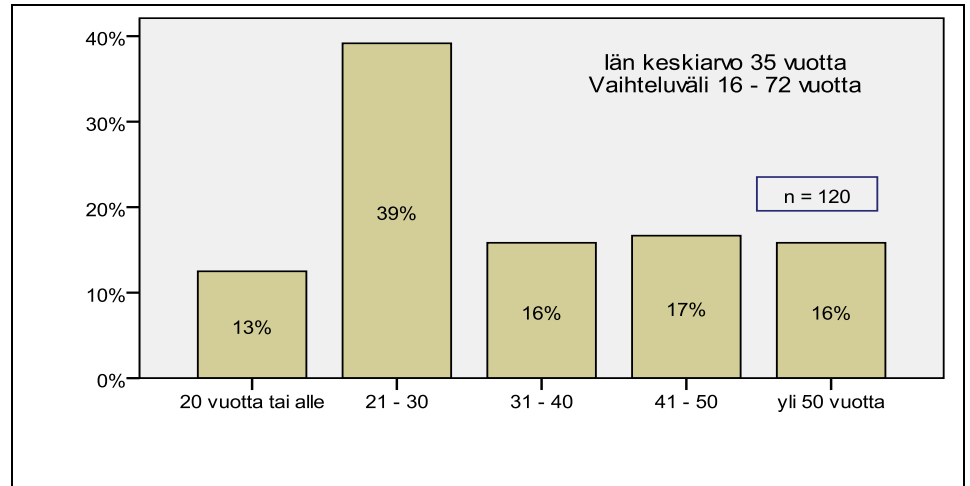
Vinkkejä:

Tulosta frekvenssijakaumat vastaajien taustatiedoista (kysymykset 1–3) *Analyze/Descriptive Statistics/Frequencies* -komennolla. Kuvaile vastaajien taustatietoja ja kirjoita tärkeimmät tulokset tekstiksi. Seuraavassa on ohjeita taustatietojen esittämiseen.

- a) Selvitä, kuinka monta prosenttia vastaajista oli naisia ja kuinka monta miestä. Tähän riittää sanallinen selitys; ei tarvita kuviota tai taulukkoa. Ilmoita sekä lukumäärät että prosenttiluvut. Jos haluat havainnollistaa tuloksia kuviolla, valitse piirakkakuvio.
- b) Tutki vastaajien ikäjakaumaa. Laske sopivia tunnuslukuja alkuperäisestä ikämuuttujasta. Käytä *Analyze/Descriptive Statistics/Frequencies* -komentoa ja tee tunnuslukuvalinnat *Statistics*-ikkunassa. Luokittele ikä *Transform/Visual Binning* -komennolla ikäluokkiin:

– 20, 21–30, 31–40, 41–50, 51 +.

Esitä ikäjakauma pylväskuviona kuten seuraavassa kuvassa. Lisää prosenttiluvut pylväisiin ja pyöristä kokonaisen prosenttiyksikön tarkkuudelle. Lisää kuvioon prosenttilukujen perusarvo ja halutessasi iän keskiarvo ja vaihteluväli. Siirrä kuvio raporttiin ja analysoi vastanneiden ikäjakaumaa kuvion ja tunnuslukujen perusteella.



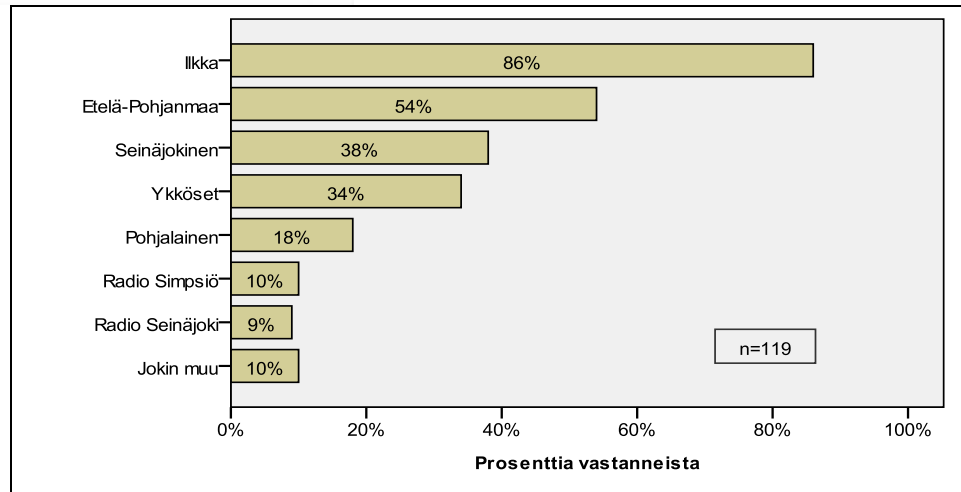
Kuvio 1. Vastaajien ikäjakauma

- c) Muokkaa asemaa työmarkkinoilla esittävä taulukko raporttia varten. Kuvaa vastaajien asemaa myös palkkikuvion avulla. Muokkaa kuvio selkeäksi. Valitse raporttiin joko taulukko tai kuvio ja kerro tulos myös sanallisesti tekstinä. Tutki myös, mitä vastauksia tuli avoimeen kohtaan ”Jokin muu” ja kerro tulos lukijallekin.

6. Esitä tulokset paikallismedioita koskevasta kysymyksestä (lomakkeen kysymys 4) palkkikuviona (prosentteina kysymykseen vastanneista). Huomaa, että kysymys on monivalintakysymys, josta on määritelty useita muuttujia.

Vinkkejä:

Yhdistä kysymyksen 4 muuttujat *Analyze/Multiple Response/Define Variable Sets* -komennolla ja tulosta sen jälkeen frekvenssijakauma *Analyze/Multiple Response/Frequencies* -komennolla. Muokkaa taulukko ja luo kuvio taulukon viimeisen sarakkeen prosenttiluvuista pikavalikon *Create Graph/Bar* -komennolla. Prosenttiluvut tulee olla valittuna ennen komentoa. Muokkaa kuvio. Lisää prosenttilukujen perusarvo eli kysymykseen vastanneiden määrä.



Kuvio 2. Vastaajien säännöllisesti seuraamat paikallismediat.

Kuvalle tärkeimpiä tuloksia tekstissä. Selvitä myös, minkälaisia vastauksia tuli kohtaan ”Jokin muu”.

7. Esitä kysymysten 5–7 tulokset selkeästi ja havainnollisesti. Kirjoita tulokset tekstiksi.

Vinkkejä:

Selvitä ensin, kuinka monta prosenttia vastaajista oli järjestänyt juhlia kodin ulkopuolella. Saat avoimeen kohtaan tulleet vastaukset juhlapaikoista *Analyze/Frequencies*-komennolla muuttujasta *Missä kodin ulkopuolella*. Esitä myös avointen vastausten tulos raportissa sanallisesti.

Esitä kysymyksen 6 (”Mistä saa tietoa juhlapaikoista?”) tulokset taulukkona tai kuviona. Huomaa, että kysymyksestä on määritelty kolme muuttujaa. Yhdistä muuttujat samalla tavalla kuin edellisessä tehtävässä *Analyze/Multiple Response/Define Variable Sets* -komennolla ennen frekvenssijakauman tulostamista. Muokkaa frekvenssitaulukko raporttia varten tai piirrä kuvio. Käytä tuloksissa *Percent of cases* -prosenttilukuja (ei siis vastausten yhteismäärästä laskettuja). Järjestä taulukko tai kuvio prosenttilukujen mukaiseen järjestykseen. Selitä tärkeimmät tulokset tekstinä.

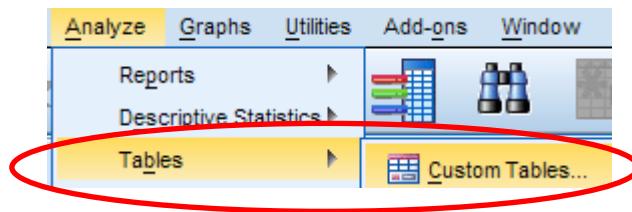
Esitä kysymyksen 7 (”sopiva hinta juhlaruokailulle”) tulokset vaikkapa pylväs- tai palkkikuviona. Kohtaan *muu hinta* tulleet vastaukset saat *Analyze/Frequencies* -komennolla.

8. Esitä kysymyksen 8 väittämien tulokset.

Vinkkejä:

Esitä tulokset kahtena ryhmänä: henkilökuntaa koskevat ja ruokaa koskevat. Harjoittele tulosten esittämistä sekä taulukkona että kuviona. Esitä Rinneravintolan henkilökuntaa koskevien väittämien tulokset yhtenä taulukkona ja ruokaan liittyvien väittämien tulokset kuviona (katso malleja seuraavassa). Kommentoi tuloksia myös sanallisesti.

Käytä *Analyze/Tables/Custom Tables* -komentoa taulukoiden tekemiseen, jos käyttämässäsi ohjelmassa on lisämoduuli. Taulukon tekemisestä on erillinen ohje (SPSS\Kysymysryhmät).

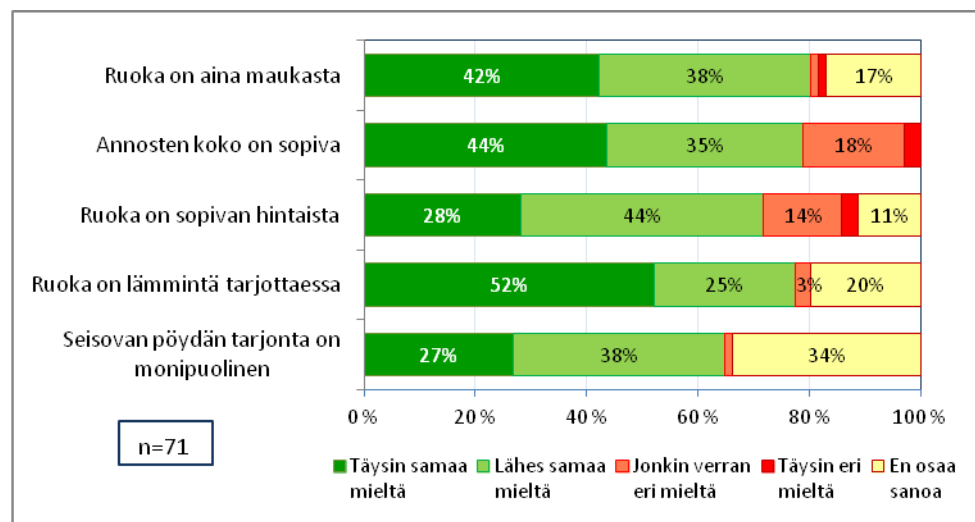


Jos *Tables*-komentoa ei ole käytettävissä, saa vastaavan taulukon keräämällä yksittäisten väittämien frekvenssitaulukoista *Valid Percent* -arvot samaan taulukkoon.

Taulukko 1. Vastaajien mielipiteet Rinneravintolan henkilökuntaan liittyvistä väittämistä.

n=120	Täysin samaa mieltä	Lähes samaa mieltä	Jonkin verran eri mieltä	Täysin eri mieltä	En osaa sanoa
Rinneravintolan palvelu on mielestäni ystävällistä	67 %	28 %	3 %	1 %	2 %
Henkilökunta on ammattitaitoista	60 %	28 %	1 %	0 %	11 %
Henkilökunta on palvelualtista	63 %	29 %	3 %	1 %	5 %

Kuvion piirtämistä varten taulukko vietään Exceliin ja piirretään kuvio Excelin pinottuna palkkikuviona.



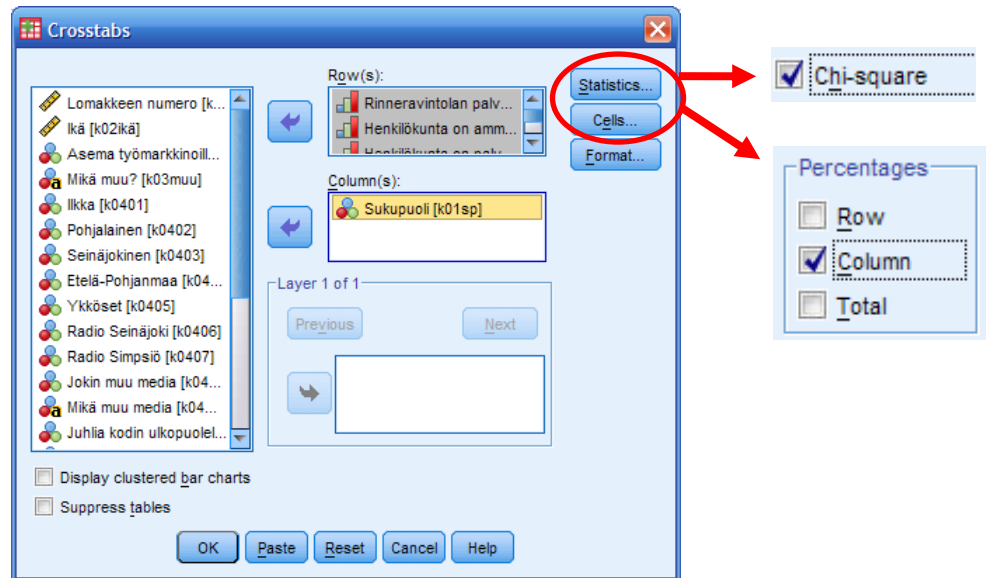
Kuvio 3. Vastaajien mielipiteet ruokaa koskevista väittämistä.

9. Vertaile miesten ja naisten vastausten prosenttijakaumia kysymysryhmän 8 väittämiin.

Vinkkejä:

Ristiintaulukoi kysymyksen 8 väittämät sukupuolen kanssa (*Analyze/Descriptive Statistics/Crosstabs*). Testaa khiin neliö -testillä, riippuvatko mielipiteet sukupuolesta tilastollisesti merkitsevästi.

Valitse raporttiin ainakin sellaiset tulokset, joissa naisten ja miesten välillä on tilastollisesti merkitseviä eroja. Kommentoi myös muita vertailuja. Kirjoita tulokset raporttiin omaksi luvuksi tai lisää ne edellisen tehtävien tulosten perään sopivin alaotsikoin.




Jos *Sig.*-arvo viittaa tilastollisesti merkitsevään riippuvuuteen, tarkista testin edellytysten voimassa olo. Jos pieniä odotettuja frekvenssejä on liikaa, kokeile auttaisiko uudelleen luokittelu.

10. Esitä kysymyksen 9 kokonaisarvosanoja koskevat tulokset. Käytä frekvenssi- tai prosenttijakaumia ja laske tunnuslukuja kokonaisarvosanoista. Valitse tunnusluvuiksi minimi, maksimi, moodi, mediaani, keskiarvo ja keskihajonta *Analyze/Frequencies*-komennon *Statistics*-ikkunassa. Halutessasi voit valita muitakin tunnuslukuja. Esitä tulokset mahdollisimman havainnollisesti.

11. Tutki korrelaatiokertoimien avulla (*Analyze/Correlate*), vaikuttaako ikä rinneravintolasta ja rinnepalveluista annettuihin kokonaisarvosanoihin. Käytä alkuperäistä ikä-muuttujaa. Korreloivatko rinneravintola- ja rinnepalveluarvosanat keskenään? Kommentoi tuloksia.

Jos haluat tutkia korrelaatiokertoimen avulla, onko vastaajien iällä vaikutusta kysymyksen 8 väittämien vastauksiin, on kysymyksen 8 muuttujille määriteltävä "0 = en osaa sanoa" puuttuvaksi tiedoksi. Tämä tehdään *Variable View* -näytymän



kohdassa *Missing*. Näin on tehtävä myös, jos haluaa laskea keskiarvoja kysymyksen 8 mielipiteistä. Vasta tämän jälkeen valitaan korrelaatiokertoimien tai tunnuslukujen laskemista koskeva komento.

12. Vertaa miesten ja naisten rinneravintolalle ja rinnepalveluille antamien kokonaisarvosanojen keskiarvoja. Testaa *t*-testillä, ovatko erot tilastollisesti merkitseviä (*Analyze/Compare Means/Independent-Samples T Test*).

Vertaa myös miesten ja naisten keskimääräisiä mielipiteitä kysymyksen 8 väittämiin. Huomaa, että ennen keskiarvojen vertaamista, on 0 määriteltävä puuttuvaksi tiedoksi, kuten edellisessä tehtävässä neuvottiin. Tarkista vielä *t*-testin antamat tilastollisesti merkitsevät tulokset ei-parametrisellä Mann-Whitneyn *U*-testillä (*Analyze/Nonparametric Tests/Legacy Dialogs/2 Independent Samples*).

13. Vertaa eri ikäryhmien kokonaisarvosanojen keskiarvoja *Analyze/Compare Means/Means* -komennolla. Testaa erojen tilastollinen merkitsevyys varianssianalyysilla ja tarkista ei-parametrisella Kruskal-Wallis testillä.

14. Esitä raportissa viimeisen avoimen kysymyksen kehittämisehdotuksia koskevat tulokset. Saat tulostettua kaikki vastaukset frekvenssikomennon avulla. Kirjoita vastauksista yhteenveto raporttiin. Suorat lainaukset voit laittaa raportin liitteeksi.