

Opinnäytetyö (Turun Ammattikorkeakoulu)

Tuotantotalouden koulutus

2020

Joonas Nikkanen

TILASTOLLISET MENETELMÄT -OPINTOJAKSON KEHITTÄMINEN

TURKU AMK 
TURKU UNIVERSITY OF
APPLIED SCIENCES

OPINNÄYTETYÖ (AMK) | TIIVISTELMÄ

TURUN AMMATTIKORKEAKOULU

Tuotantotalous

2020 | 35 sivua, 7 liitesivua

Joonas Nikkanen

TILASTOLLISET MENETELMÄT -OPINTOJAKSON KEHITTÄMINEN

Tämä työ käsittelee Turun ammattikorkeakoulun insinööriopiskelijoiden Tilastolliset menetelmät -opintojakson kehittämistä. Opinnäytetyön tavoitteena oli luoda ja kehittää kurssilla käytettävää oppimateriaalia insinööriopiskelijoiden tarpeisiin ja vastata lisääntyneeseen etäopiskelun tarpeeseen.

Työn teoriaosuus koostui tilastollisesta ajattelusta ja lukutaidosta, tilastollisesta tutkimuksesta ja peruskäsitteistä sekä Excelin käytöstä tilastotieteessä. Lisäksi käsiteltiin tilastojen ja tilastollisten menetelmien käyttöä jokapäiväisessä elämässä ja eritoten insinöörin työssä. Työn käytännön osa toteutettiin laatimalla kysymyksiä ja tehtäviä tilastollisesta ajattelusta, tilastollisten menetelmien peruskäsitteistä, tilastollisesta tutkimuksesta sekä Excel -ohjelman käytöstä tilastotieteessä.

Opinnäytetyössä esille nousi kehitysidea siitä, että kurssia voitaisiin tulevaisuudessa toteuttaa yhteistyössä yrityssektorin kanssa, jotta tilastollisten menetelmien käytön merkitys insinööreille tulisi paremmin esille.

ASIASANAT:

insinöörikoulutus, kehittäminen, tilasto, tilastolliset menetelmät

BACHELOR'S THESIS | ABSTRACT

TURKU UNIVERSITY OF APPLIED SCIENCES

Industrial engineering and management

2020 | 35 pages, 7 pages in appendices

Joonas Nikkanen

IMPROVING THE STUDY MODULE “STATISTICAL METHODS” IN ENGINEERING EDUCATION

This thesis discusses the improvement of “Statistical methods”- study module for engineering students of Turku University of Applied Sciences. The aim of this thesis is to create learning material and improve already existing course materials. The viewpoint of engineering studies and the viewpoints of independent and online learning are also considered.

Theory part of the thesis consists of statistical thinking, statistical analysis and the basic concepts of statistical methods. Also, the use of Excel software in statistics is briefly discussed as well as the use of statistics in one's everyday life and more importantly in the work of an engineer.

Thesis was carried out by creating questions and assignments about statistical thinking, statistical analysis, basic concepts of statistical methods and the use of Microsoft Excel.

Thesis brought up an idea for improvement for the study module, which was the integration of actual data from companies into the study module. This would probably highlight the importance of statistics in the work of engineers.

KEYWORDS:

development, engineering education, statistical methods, statistics

SISÄLTÖ

1 JOHDANTO	6
2 TILASTOT JA NIIDEN HYÖDYNTÄMINEN	7
2.1 Tilastot jokapäiväisessä elämässä	7
2.2 Tilastolliset menetelmät insinöörin työssä	7
3 TILASTOLLISET MENETELMÄT	9
3.1 Tilastollinen ajattelu ja tilastollinen lukutaito	9
3.2 Tilastollisen ajattelun peruskäsitteet	12
3.3 Tilastollinen eli kvantitatiivinen tutkimus	17
3.4 Excel-ohjelman käyttö tilastotieteessä	25
4 TYÖN KÄYTÄNNÖN TOTEUTUS	28
4.1 Kysymysten laadinta	28
4.2 Excel – tehtävien laadinta	29
5 POHDINTA	32
LÄHTEET	34

LIITTEET

- Liite 1. Laaditut kysymykset
- Liite 2. Excel tehtävänanto

KUVAT

Kuva 1. Katkaistu pystyakseli.	10
Kuva 2. Oikeaoppinen pystyakseli alkaa nollost.	11
Kuva 3. Esimerkki tilastollisen tutkimusprosessin kulusta.	19
Kuva 4. Otanta.	23
Kuva 5. Erilaisia kuvaajia.	25
Kuva 6. Microsoft Excel.	26
Kuva 7. Excel tehtävässä 1 luotu kuvaaja.	30

TAULUKOT

Taulukko 1. Havaintomatriisi.	12
Taulukko 2. Esimerkki tuloksista.	14

KAAVAT

Kaava 1. Keskihajonta.	15
------------------------	----

1 JOHDANTO

Tämän opinnäytetyön aiheena on Turun ammattikorkeakoulun tarjoaman insinööriopiskelijoiden Tilastolliset menetelmät -opintojakson kehittäminen. Opintojakson tarkoituksena on antaa opiskelijalle perusvalmiudet ja osaaminen tilastollisten perusmenetelmien käyttöön sekä data-analyysin suorittamiseen. Työ suoritetaan laatimalla kysymyksiä ja tehtäviä, jotka soveltuvat tukemaan oppimista ja antavat aineistoa kurssin arviointiin. Kysymyksiä ja tehtäviä laaditaan neljästä aihealueesta, jotka ovat tilastollinen ajattelu, tilastolliset peruskäsitteet, tilastollinen tutkimus sekä Excelin käyttö tilastotieteessä.

Työn tavoitteena on vastata Turun ammattikorkeakoulussa esille nousseeseen tarpeeseen kehittää ja päivittää kyseisen opintojakson oppimateriaaleja insinööriopintojen näkökulmasta sekä vastata lisääntyneeseen etäopetuksen ja itsenäisen opiskelun tarpeeseen. Aihe on merkityksellinen myös työelämään valmistavasta näkökulmasta, koska nykyajan teollisuudessa datan määrä ja tarve sen analysoinnille tilastollisia menetelmiä käyttäen on kasvanut. Datan analysointi on perustavanlaatuinen osa insinöörin työtä, ja tilastotieteen menetelmien hallitseminen lisää valmiutta insinöörin tehokkaaseen työskentelyyn. (Bisgaard 1991.)

Työn teoriaosuus rajataan käsittelemään tilastotiedettä ja tilastollisia menetelmiä kyseisen opintojakson oppimistavoitteiden puitteissa. Laadittujen tehtävien ja kysymysten osalta oppimisalusta rajataan tavoitteiden ulkopuolelle ja tilasto-ohjelmia tarvitsevilla tehtävillä rajoitetaan Microsoft Excelin käyttöön.

Työssä käydään ensin läpi tilastojen käyttöä jokapäiväisessä elämässä sekä tehdään katsaus tilastotieteen ja tilastollisten menetelmien käyttöön insinöörin työssä. Opinnäytteen teorialuvuissa tarkastellaan tilastollista ajattelua, tilastollisia peruskäsitteitä, tilastollisen tutkimuksen tekemistä sekä Excelin käyttöä tilastotieteessä. Lopuksi käydään läpi kurssia varten tuotettua oppimateriaalia, tuloksia sekä johtopäätöksiä.

2 TILASTOT JA NIIDEN HYÖDYNTÄMINEN

Tässä luvussa käsitellään sitä miten tilastot ja tilastolliset menetelmät näkyvät jokapäiväisessä elämässä, sekä perehdytään tilastollisten menetelmien ja taitojen merkittävyyteen erityisesti insinööriyössä.

2.1 Tilastot jokapäiväisessä elämässä

Monet opiskelijat saattavat tilastokursseja opiskellessaan miettiä, mihin tilastotiedettä ja tilastollisia menetelmiä oikeastaan tarvitaan jokapäiväisessä elämässä. Onko tilastollisten menetelmien osaaminen ja kyky tilastolliseen ajatteluun tarpeellista normaalissa elämässä?

Tilastoja käytetään kuitenkin useampaan tarkoitukseen kuin äkkiseltään voisi kuvitella. Tilastojen avulla voidaan tehdä ennusteita tulevasta. Lääkärit käyttävät tilastoja apuvälineenä ennakoimaan tulevia tapahtumia, kuten esimerkiksi tautien kehittymistä. Sään ennustamisessa tilastoilla ja tilastollisilla menetelmillä on suuri merkitys. Poliitikassa tilastoja käytetään seuraamaan kannatuslukuja eri alueilla. Vakuutusyhtiöt käyttävät tilastoja hyväkseen miettiessään sopivia vakuutuksia. Urheilussa tilastoilla voidaan helposti seurata joukkueiden ja pelaajien suorituksia. (Statanalytica 2020.)

Voidaan siis todeta, että tilastot ovat läsnä ympärillä ja tilastojen, ja tilastollisten menetelmien ymmärtäminen on hyödyllinen taito jokaiselle meistä nykypäivän yhteiskunnassamme.

2.2 Tilastolliset menetelmät insinöörin työssä

Tilastollisten menetelmien hallinta on tärkeä taito nykypäivän insinööreille. Insinöörit käyttävät tilastoja hyödyksi esimerkiksi arvioidessaan käynnissä olevien projektien onnistumista ja määrittääkseen, kuinka kauan projektin valmistumisessa kestää. Myös esimerkiksi laadunvalvonnassa käytetään hyödyksi tilastoja; kunnollinen laadunvalvonta on harvoin edes mahdollista ilman tilastollista kontrollia. (Statanalytica 2020.) Yksi hyvä esimerkki tilastollisten menetelmien käytöstä teollisuudessa on *Statistical Process Control* (SPC), eli tilastollinen prosessinohjaus. Tilastollisen prosessinohjauksen avulla

uudelleentekemisen tarve sekä viallisten tuotteiden määrä vähenee huomattavasti. (Kankaanranta 2012, 74.) Teollisuuden kehittyessä ja datan määrän kasvaessa myös tilastotieteen merkitys kasvaa, sillä tuotettu data olisi hyvä saada tilastollisten analyysien avulla muutettua hyödylliseksi informaatioksi. Suuria datamääriä on osattava käsitellä, mutta pelkät ohjelmat eivät riitä siihen, vaan taustalla pitää olla tilastollisten menetelmien ymmärrys ja datan analysointitaito. (Kankaanranta 2012, 87.) Kankaanranta selvitti väitöstutkimuksessaan (2012) sitä miten tilastotiedettä käytetään hyväksi teollisuudessa. Tutkimuksen mukaan tilastotiedettä ei hyödynnetä riittävän tehokkaasti teollisuudessa ja monet teollisuustilastotieteilijät joutuvat todistelemaan työnsä lisäarvoa ja merkitystä. (Turun Sanomat 2012.)

Insinöörit suorittavat kokeita ja analysoivat dataa jokatapauksessa, oli heillä tilastollista koulutusta tai ei. Ne insinöörit, joilla tilastollisten menetelmien ja tilastollisen ajattelun osaamista löytyy, pystyvät todennäköisesti tekemään työnsä tehokkaammin ja paremmin. Bisgaardin mukaan tilastollisten menetelmien osaamisella ja hallinnalla insinöörien keskuudessa voi olla suuri merkitys tulevaisuuden modernille teollisuudelle. (Bisgaard 1991.)

3 TILASTOLLISET MENETELMÄT

Tässä luvussa tutustutaan tilastollisten menetelmien teoriaan. Luvussa käydään läpi tilastollisen ajattelun konseptia, tilastollisten menetelmien peruskäsitteitä, tilastollisen tutkimuksen tekemistä, sekä Excel ohjelman käyttöä tilastollisten toimenpiteiden tekemiseen.

3.1 Tilastollinen ajattelu ja tilastollinen lukutaito

”Tilastollinen ajattelukyky on ihmiselle jonakin päivänä yhtä tärkeä kuin kyky lukea ja kirjoittaa” – H.G. Wells

Tilastollisten menetelmien ymmärtäminen on tarpeellista ja hyödyllistä lähes jokaiselle meistä. Mediaa seurattaessa törmää erilaisiin graafisiin kuvioihin ja tutkimustuloksiin. Näiden kuvioiden, tutkimusten ja tilastojen ymmärtäminen sekä tulosten arvioiminen edellyttävät tilastotieteen perusteiden hallintaa ja kykyä tilastolliseen ajatteluun. Välttämättömäksi tilastollisten menetelmien hallitseminen muuttuu, kun tavoitteena on itse tutkia ja perehtyä tutkimustuloksiin perinpohjaisesti. (Nummenmaa, Holopainen & Pulkkinen. 2016, 9.) Kykyä ymmärtää tilastoja, tilastollisia tuloksia ja niistä tehtyjä johtopäätöksiä voidaan kutsua *tilastolliseksi lukutaidoksi*. Tilastollisen lukutaidon omaava osaa olla kriittinen tarkastellessaan tilastollisia aineistoja ja tuloksia. (Tilastokeskus 2020a.)

Uskottavien ja kelvollisten tilastojen erottaminen arvelluttavista ja kyseenalaisista on tärkeä taito oppia ja tilastojen sekä tutkimustulosten kyseenalaistaminen ja epäileminen on suotavaa. Tutkimusraportteja ja tuloksia tarkastellessa on syytä ottaa huomioon muun muassa

- tutkimuksen tekijä ja tilaaja
- tutkimuksen tarkoitus
- otantamenetelmät ja otoksen koko
- kuvat ja niiden mahdollinen harhaanjohtavuus
- kysymysten selkeys (Heikkilä 2014, 73–77.)

Tutkimuksen tilaajan ja tutkimuksen tekijän omat agendat saattavat ohjata tutkimuksen tuloksia tukemaan omaa näkökantaansa. Tärkeää tutkimustuloksia ja tilastoja

tulkittaessa onkin, että tiedetään, mikä oli tutkimuksen tarkoitus, sillä se saattaa vaikuttaa tutkimuksesta saatuihin tuloksiin. Tutkimuksen perusjoukolla ja käytetyillä otantamenetelmillä on myös merkittävä vaikutus tutkimustuloksiin. Esimerkiksi jos suomalaisten joulunviettotottumuksia tutkittaessa kyselytutkimukseen on vastannut huomattavasti enemmän toisen sukupuolen edustajia, ei tutkimustuloksia voida pitää koko Suomen kansaa edustavina. Tutkimuksessa käytettävän otoksen koko vaikuttaa myös merkittävästi saatuihin tuloksiin ja niiden tarkkuuteen. Kolikon heitossa molempien mahdollisten lopputulosten mahdollisuus on 50 %. Lähemmäksi tätä 50 %:n tulosta päästään kuitenkin heittämällä testissä kolikkoa 10 000 kertaa kuin heittämällä 100 kertaa. Otokoko siis vaikuttaa tulosten tarkkuuteen riippuen tutkittavasta asiasta. Pylväskaavioiden y-akselin katkaisulla taas saadaan kuva kertomaan täysin erilaista ja liioiteltua tarinaa totuuteen verrattuna. Onkin tärkeää että asteikko alkaa nolasta, jotta vääristynyttä käsitystä ei pääse syntymään. (Heikkilä 2014, 73–77.)

Kuvia 1 ja 2 vertaamalla näkee, miten erilaisen kuvan y-akselin katkaisu antaa pylväskaavion esittämille tuloksille.



Kuva 1. Katkaistu pysty akseli. (Heikkinen 2019)



Kuva 2. Oikeaoppinen pysty akseli alkaa nollost. (Heikkinen 2019)

Syy-seuraussuhde eli kausaliiteetti sekä korrelaatio

Syy-seuraussuhteella eli kausaliiteetillä tarkoitetaan kahden erillisen tapahtuman suhdetta keskenään. On jokin tapahtuma (syy), joka johtaa tiettyyn tulokseen eli seuraukseen. Korrelaatiolla sen sijaan tarkoitetaan kahden asian riippuvuutta toisistaan. Kun olemassa on voimakas korrelaatio kahden eri asian välillä, voidaan virheellisesti luulla, että syy-seuraussuhde on syntynyt. Kun asioita yhdistellään riittävästi, saadaan aikaan satunnaisia tapahtumia, jotka vaikuttavat liittyvän toisiinsa vaikka todellisuudessa taustalla on yleensä jokin muu vaikuttava tekijä. (Tilastokeskus 2020b.)

Ehkäpä yleisimpänä esimerkkinä tästä käytetään kesäajan jäätelönsyönnin vaikutusta hukkumistapauksiin. Niiden välillä voi olla paljonkin korrelaatiota, mutta jäätelönsyönti ei silti lisää mahdollisuutta hukkua. Sekä jäätelönsyöntiä että uimista harrastetaan paljon lämpimällä ja aurinkoisella säällä. Lämmin sää toimii siis yhdistävänä tekijänä jäätelönsyönnin ja hukkumistapausten korrelaatiolle. (Tilastokeskus 2020b.) Syy-seuraussuhteen tekemisessä pelkän korrelaation perusteella on oltava siis varovaisia, sillä merkittäväkään korrelaatio ei takaa syy-seuraussuhteen syntymistä (Heikkilä 2014, 91).

3.2 Tilastollisen ajattelun peruskäsitteet

Tässä luvussa käsitellään muutamia keskeisiä tilastollisia käsitteitä. Tilastotieteessä on todella paljon erilaisia käsitteitä ja tähän lukuun on yritetty koota opintojakson oppimistavoitteiden näkökulmasta tärkeitä käsitteitä. Kattavamman listan tilastotieteen käsitteistä voi löytää esimerkiksi Tilastokeskuksen nettisivuilta.

Muuttuja, havainto ja havaintomatriisi

Muuttujalla tarkoitetaan ominaisuutta, jota halutaan tarkastella ja mitata. Se normaalisti vaihtelee eri henkilöillä. Muuttuja voi olla vaikka henkilön paino tai pituus. Havainto taas tarkoittaa mittaustulosta; jos muuttujana toimii esimerkiksi paino ja mitattava henkilö painaa 80 kg. on tämä tulos silloin havainto. (Nummenmaa ym. 2016, 17.)

Havaintomatriisiin avulla voidaan esitellä havaintoaineistoa. Havaintomatriisi on taulukko, joka tehdään aina samalla tavalla. Jokaisella rivillä on yhden tilastoyksikön muuttujien arvot ja joka sarakkeelta on nähtävissä yhden muuttujan kaikki arvot. (Nummenmaa ym. 2016, 38.) Taulukossa 1 on esimerkki havaintomatriisista. Tässä matriisissa on kolme muuttujaa: paino, pituus ja ammatti. Havaintoyksiköitä matriisissa on kolme.

Taulukko 1. Havaintomatriisi. (Tilastokeskus 2020d, mukaillen).

Paino	Pituus	Ammatti
80 kg	180 cm	Leipuri
76 kg	165 cm	Varastomies
87 kg	190 cm	Insinööri

Mitta-asteikot

Tilastollista tutkimusta voidaan tehdä niin monista eri asioista, että erilaisia muuttujia on paljon. Kaikkia muuttujia ei voi kuitenkaan mitata täysin samalla tavoin, ja siksi on kehitetty erilaisia mitta-asteikoita:

1. laatueroasteikko eli luokitteluasteikko
2. järjestysasteikko
3. välimatka-asteikko
4. suhdeasteikko. (Nummenmaa ym. 2016, 18–19.)

Luokitteluasteikossa tarkasteltavat yksiköt voivat kuulua ainoastaan yhteen tiettyyn luokkaan. Yksiköt voidaan jakaa esimerkiksi veriryhmän tai sukupuolen mukaan luokkiin.

Järjestysasteikossa luokilla on tietty järjestys. Esimerkiksi iän mukaan jaoteltuna alle 20-vuotiaat, 20–50 vuotiaat ja yli 50-vuotiaat.

Välimatka-asteikolla luokkien väliset matkat ovat yhtä suuria, esimerkiksi celsius asteikko lämpötilan mittaamiseen. Välimatka-asteikko soveltuukin lähinnä lämpötilan mittaamiseen. Välimatka-asteikkoa käytetään siis todella harvoin muihin mitta-asteikkoihin verrattuna.

Suhdeasteikko on samankaltainen kuin välimatka-asteikko, mutta asteikossa on olemassa myös nollapiste. Tämä mahdollistaa myös tulosten välisien suhteiden laskemisen. Esimerkkinä liikevaihto ja euromääräiset tulot. (Nummenmaa ym. 2016, 18–19.)

Perusjoukko, tilastoyksikkö ja otos

Perusjoukolla tarkoitetaan tutkimuksen tai mittauksen kohteena olevaa joukkoa. Perusjoukko voi olla esimerkiksi Suomen kansa. Tämä joukko koostuu tutkittavista yksiköistä eli tilastoyksiköistä. Kaikkien perusjoukossa olevien tilastoyksiköiden tutkiminen on usein mahdotonta tai vaatii liikaa resursseja. Tällöin perusjoukosta voidaan valita pienempi ryhmä tilastoyksiköitä. Tätä ryhmää kutsutaan *otokseksi*. Otoksen avulla saadut tulokset voidaan sitten yleistää koko perusjoukkoon. Otoksen on täytettävä tietyt vaatimukset. Otosta tehtäessä on tärkeää, että kuka tahansa perusjoukosta voi tulla valituksi otokseen. Jos tämä perusehto ei täyty, on kyseessä näyte. Näyte on kyseessä silloin, kun tilastoyksiköitä valitaan harkinnanvaraisesti esimerkiksi työtovereista eikä satunnaisesti koko perusjoukosta. (Nummenmaa ym. 2016, 10, 17, 26, 33.)

Tunnusluvut

Erilaisia tunnuslukuja käytetään kuvaamaan muuttujan jakaumaa. Keskikohtaa yleisesti kuvataan keskiarvolla tai mediaanilla. Arvojen sijoittumista keskiarvon ympärille ja niiden vaihtelua taas kuvataan hajontaluvuilla. (Tilastokeskus 2020c.)

Keskiluvut

Keskiarvo on tulosten esimerkiksi pituuksien yhteenlaskettu summa jaettuna niiden lukumäärällä. Keskiarvoa käyttämällä saadaan tiivistettyä suuria aineistoja helpommin esimerkiksi kuvaajaksi.

Moodia käytetään, kun tarkastellaan tuloksia, jotka ovat laatueroasteikolla, eli tulokset eivät ole järjestyksessä. Moodi on siis arvo, joka esiintyy aineistossa useimmiten. Moodia käytetään yleisesti, kun keskikohtaa ei voida laskea. Esimerkiksi jos tutkimusaineistossa vastaukset on annettu kuten taulukossa 2, on moodi eli yleisin vastaus ”jokseenkin samaa mieltä”.

Taulukko 2. Esimerkki tuloksista.

Vastaus	Lukumäärä
Samaa mieltä	10
Jokseenkin samaa mieltä	59
Jokseenkin eri mieltä	30
Eri mieltä	18

Mediaanin avulla aineiston tulokset voidaan jakaa kahteen yhtäsuureen osaan. Puolet havainnoista on pienempiä ja puolet suurempia kuin mediaani. Mediaani on siis keskimäinen arvo suuruusjärjestyksen asetetuista arvoista. Jos arvoja on parillinen määrä käytetään mediaanina jompaakumpaa keskimmäisistä arvoista. (Nummenmaa ym. 2016, 69 – 72.) Yläpuolella olevan taulukon 2. aineiston mediaani on ”jokseenkin samaa mieltä.” $(10+59+30+18)/2 = 58.5$, ja 58 sekä 59 havainto sijoittuvat ”jokseenkin samaa mieltä”-kategoriaan.

Hajontaluvut

Hajontaluvuilla saadaan selville kuinka suurelle tai pienelle välille saadut arvot sijoittuvat. Niillä mitataan yleensä myös sitä kuinka tiheästi saadut tulokset sijoittuvat keskiluvun ympärille. Hajontaluvuista kaikkein yksinkertaisin on vaihteluväli. Esimerkiksi pituuksia tarkastellessa pisimmän ja lyhyimmän havainnon välinen ero. (Nummenmaa ym. 2016, 79 – 81.)

Mediaanin jakaessa tulokset kahteen yhtä suureen osaan, kvartiilit taas jakavat aineiston siten että 25% havainnoista on suurempia kuin yläkvartiili ja 25% havainnoista on pienempiä kuin alakvartiili. Kvartiileja ja mediaania hyväksikäyttämällä aineisto saadaan jaettua neljään osaan, jotka ovat yhtä suuria. Havaintoaineiston aluetta alakvartiilista yläkvartiiliin kutsutaan kvartiiliväliksi. Kvartiilivälin avulla voidaan tarkastella aineiston keskimmäisen 50% sijoittumista. (Nummenmaa ym. 2016, 82.)

Keskihajonnalla kuvataan saatujen havaintoarvojen hajontaa. Keskihajonta kuvaa sitä miten kaukana havaintoarvot ovat keskiarvosta. Mittayksikkö pysyy samana kuin aiemminkin mitatussa muuttujassa. Jos pituuksia mitattiin metreinä niin keskihajontakin ilmoitetaan metreinä. (Nummenmaa ym. 2016, 82 – 83.) Alla keskihajonnan kaava:

Kaava 1. Keskihajonta.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}, \text{ missä}$$

\bar{x} = havaintoarvojen keskiarvo

n = havaintoarvojen lukumäärä

Otantamenetelmät

Otantamenetelmiä on paljon erilaisia ja eri tilanteisiin sopivat erilaiset otantamenetelmät. Otannassa pyritään yleisesti satunnaistamiseen, koska siten saadaan varmistettua että otos kuvastaa mahdollisimman hyvin koko perusjoukkoa. (Nummenmaa ym. 2016, 27.) Otantamenetelmän valintaan vaikuttaa muun muassa se mitä tutkimuksella halutaan

saavuttaa, perusjoukon sijainti, käytettävissä olevat luettelot ja rekisterit sekä rahallinen budjetti (Heikkilä 2014, 33). Yleisimpiä otantamenetelmiä ovat:

1. yksinkertainen satunnaisotanta
2. systemaattinen otanta
3. ositettu otanta
4. ryväotanta

Yksinkertaisessa satunnaisotannassa perusideana on että jokaisella populaation eli perusjoukon tilastoyksiköllä on sama mahdollisuus tulla valituksi osaksi otosta. Menetelmää käytetään esimerkiksi kyselytutkimuksissa. (Tilastokeskus 2007, 57.) Satunnaisotanta voidaan suorittaa vaikka niin että perusjoukon yksiköiden nimet kirjoitetaan lapulle ja laput laitetaan pussiin. Pussista sitten vedetään halutun otoksen verran lappuja ja heidät valitaan otokseen. Suuremmissa joukoissa voidaan numeroida yksiköt ja valita satunnaisia numeroita halutun otokseen verran. (Nummenmaa ym. 2016, 28.)

Systemaattisessa otannassa yksiköt valitaan tasaisin välein koko perusjoukosta. Otantaväli voi olla vaikka 20. Tällöin joka 20. yksikkö valitaan otokseen. Tämä otantamenetelmä on erityisen käyttökelpoinen silloin kun perusjoukko on valmiiksi jossain järjestyksessä, esimerkiksi jonottamassa kassalle tai lipunmyyntiin. (Nummenmaa ym. 2016, 29 – 30.) Tutkittaessa vaikkapa asuinalueesta riippuvaa ilmiötä väestössä, systemaattisella otannalla saadaan oikein jakautunut otos kun perusjoukko järjestetään kotipaikkatunnusten mukaan (Tilastokeskus 2007, 57).

Ositettu otanta toimii siten että perusjoukko jaetaan ositteisiin. Jokaisesta ositteesta valitaan yksiköitä otokseen. Jokaisesta ositteesta voidaan sitten valita joko yhtä monta tilastoyksikköä (tasainen kiintiöinti), prosentuaalisesti yhtä paljon yksiköitä (suhteellinen kiintiöinti), tai käyttää erilaista otantasuhdetta eri ositteisiin ottamalla huomioon muuttujia kuten yksikkökustannukset tai ositteen koko. (optimaalinen kiintiöinti). Ositetulla otannalla saadaan haluttaessa varmistettua esimerkiksi vähemmistöjen edustajien osuus otoksessa. (Heikkilä 2014, 35 – 37.)

Kun perusjoukko jaetaan osajoukkoihin on kyseessä ryväotanta. Ryväotannassa perusjoukko voidaan jakaa vaikka vaalipiireittäin tai yrityksittäin. Ryväotannan etuihin kuuluu sen helppous ja menetelmän edullisuus. Satunnaisotannalla poimitut yksiköt saattavat sijoittua monille eri paikkakunnille ja tätä varten tutkijat joutuvat matkustamaan useampaan paikkaan saadakseen tuloksia. Ryväisiin jaetussa otannassa monta

tilastoyksikköä on samassa sijainnissa mikä helpottaa tutkijan työtä. (Nummenmaa ym. 2016, 32.) Ryväotannan heikkouksiin kuuluu se että usein samasta rypäystä olevien yksiköiden vastaukset korreloivat keskenään ja täten tuloksille ei saada niin paljoa vaihtelua (Tilastokeskus 2007, 58).

Induktiivisuus ja deduktiivisuus

Tieteellisiä johtopäätöksiä voidaan tehdä deduktiivisesti tai induktiivisesti. Deduktiivinen päättely tarkoittaa sitä että aloitetaan tarkastelemalla asioita jotka tiedetään varmasti todeksi, näitä kutsutaan premiseiksi. Näistä premiseista sitten johdetaan uusia johtopäätöksiä tarkasteltavaksi. Premisseinä voi olla esimerkiksi että planeetat kiertävät tähtiä ja maa on planeetta. Tästä voidaan tehdä johtopäätös että maa kiertää tähteä. Induktiivisessa päättelyssä liikkeelle lähdetään jostain havainnosta ja edetään yleistyksen. Esimerkiksi kun näemme valkoisen joutsenen kerta toisensa jälkeen, saa se aikaan käsityksen siitä että kaikki joutsenet ovat valkoisia. Vaikka se ei suoraan todista kaikkien joutsenten olevan valkoisia, voidaan kuitenkin sanoa että suurella todennäköisyydellä luonnossa nähty joutsen on valkoinen. (Nummenmaa ym. 2016, 12 – 13 .)

Validiteetti, reliabiliteetti ja objektiivisuus

Validiteetilla tarkoitetaan mittauksen pätevyyttä. Onko mittauksella mitattu sitä mitä oli tarkoitus mitata. Mitattavat käsitteet ja muuttujat tulee määritellä tarkasti jotta saadaan aikaan valideja tuloksia. Mittaustulosten luotettavuutta ja tarkkuutta kuvaa reliabiliteetti. Liian suuri sattumanvaraisuus tuloksissa tarkoittaa että tulokset eivät ole luotettavia. Tulosten luotettavuutta laskee esimerkiksi jos tutkimuksen otoskoko on liian pieni. Objektiivisuudella tarkoitetaan puolueettomuutta. Yksinkertaisesti selitettynä tutkijan omat mielipiteet eivät saisi vaikuttaa tutkimuksen tuloksiin. Tulokset ovat objektiivisiä jos samoihin tuloksiin päästään vaikka tutkijaa vaihdettaisiin. (Heikkilä 2014, 27 – 29.)

3.3 Tilastollinen eli kvantitatiivinen tutkimus

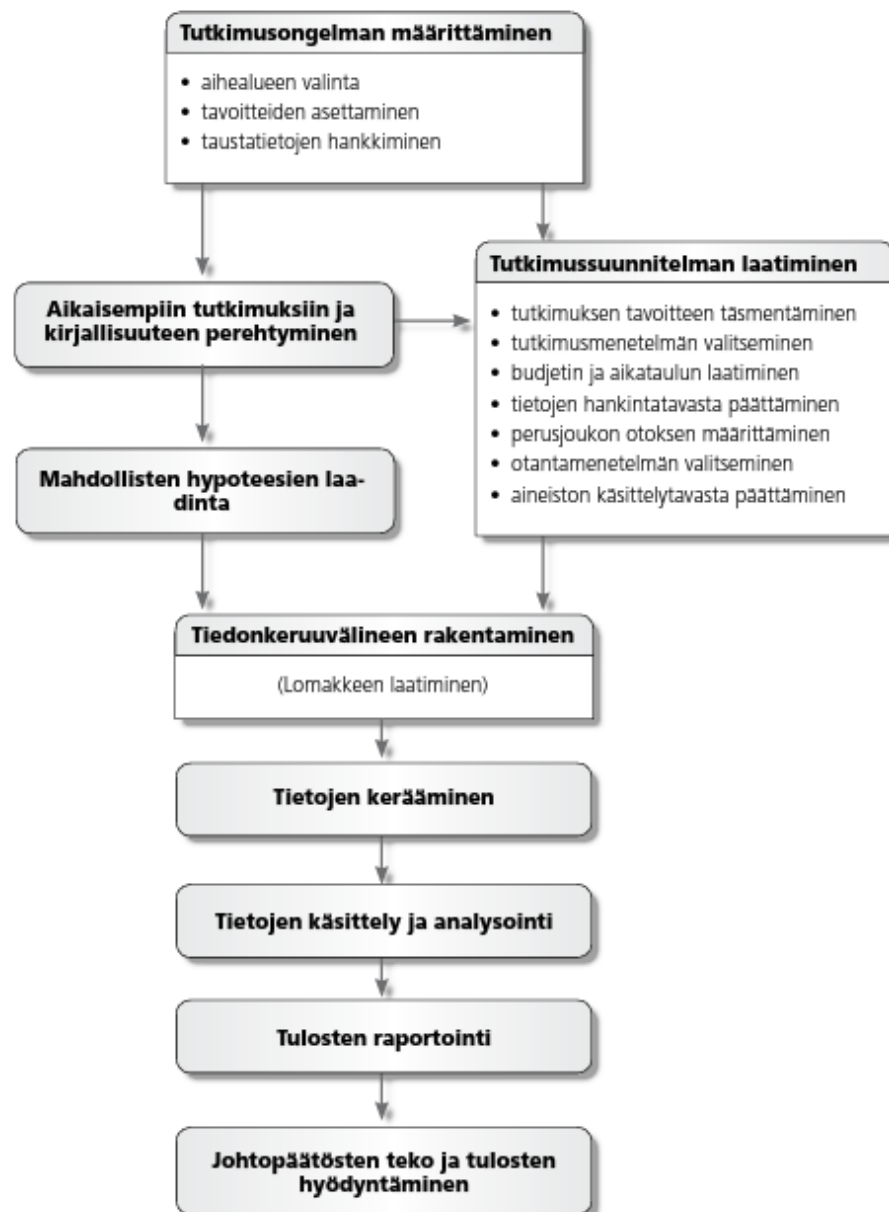
Kvantitatiivisella eli tilastollisella tutkimuksella pyritään selvittämään prosenttiosuuksiin ja lukumääriin liittyviä kysymyksiä. Tilastollisen tutkimuksen tekeminen edellyttää otosta,

joka on sekä riittävän suuri että edustava. Tutkimuksissa pyritään yleensä selvittämään eri asioiden välillä vallitsevia riippuvuuksia tai muutoksia tutkittavassa ilmiössä. Tuloksia kuvataan numeerisesti ja havainnollistetaan kuvioiden ja taulukoiden avulla. (Heikkilä 2014, 15.) Melkein mitä tahansa pystytään nykyään mittaamaan numeerisesti, ja tästä syystä hyvin suuri osa nykyisin tehtävistä tutkimuksista ovat tutkimustavaltaan juuri kvantitatiivisia (Nummenmaa ym. 2016, 16). Kvantitatiiviseen tutkimukseen tarvittava aineisto voidaan hankkia joko keräämällä itse, tai hyödyntämällä erilaisia valmiita muiden kokoamia tilastoja ja tietokantoja. Suomessa tärkein tilastojen tuottaja ja ylläpitäjä on tilastokeskus. (Heikkilä 2014, 16.)

Hyvän tilastollinen tutkimuksen tulee Heikkilän (2014, 27 – 30), mukaan olla:

- Validi
 - o On tutkittu sitä, mitä pitikin tutkia
 - o Ei systemaattisia virheitä
- Reliaabeli
 - o Tarkat tulokset, joissa ei ole sattumanvaraisuutta
- Objektiivinen
 - o Tulokset eivät riipu tutkijan omista mielipiteistä
 - o Puolueettomuus
- Tehokas ja taloudellinen
 - o Hyöty ja kustannukset oikeassa suhteessa
 - o Laadukkaat tulokset
- Avoin
 - o Tutkittaville pitää selvittää tutkimuksen tarkoitus ja tietojen käyttötapa
- Tietosuojaan huomioon ottava
 - o Kenenkään yksityisyyttä ei loukata
 - o Yksittäistä vastaajaa ei voi tunnistaa tuloksista
- Hyödyllinen ja käyttökelpoinen
 - o Jotakin uutta esiin tuovat tulokset
 - o Esimerkiksi tutkimuslomakkeiden kysymysten tarpeellisuus tulee miettiä tarkkaan
- Sopivalla aikataululla tehty
 - o Tiedot käytössä silloin kun niitä tarvitaan
 - o Tiedot ovat tuoreita ja täsmällisiä
 - o Aikataulussa pysytään

Tilastollinen tutkimus koostuu kolmesta vaiheesta jotka ovat suunnitteluvaihe, operationaalinen vaihe ja arviointivaihe. Suunnitteluvaiheessa määritetään tutkimuksen sisältö, tutkimusongelma, kohdejoukko ja tiedonkeruumenetelmät. Operationaalisessa vaiheessa suoritetaan otoksen poiminta ja tiedonkeruu, käsitellään ja valmistellaan aineistoa sekä analysoidaan tuloksia. Arviointivaiheessa tarkastellaan miten asetetut kriteerit ja tavoitteet on saavutettu. (Tilastokeskus 2007, 29 – 31.)



Kuva 3. Esimerkki tilastollisen tutkimusprosessin kulusta. (Heikkilä 2014, 23)

Suunnitteluvaihe

Tutkimuksen tekeminen alkaa aina suunnitelulla. Tutkimusprosessin ensimmäinen vaihe on määrittää tutkimusongelmat ja ottaa selvää mahdollisista aikaisemmista tutkimuksista, jotka liittyvät aiheeseen. Tutkimusongelmilla tarkoitetaan niitä asioita joihin pyritään saamaan vastaus tutkimuksella. Tutkimusongelmat voidaan esittää erilaisina kysymyksinä ja hypoteeseina, esimerkiksi ”Vaikuttaako jäätelön syöminen hukkumiskuolemiin?”. (Heikkilä 2014, 20 – 21.) Tutkimusongelmien huolellinen ja tarkka määrittäminen on tärkeää tutkimuksen onnistumisen kannalta (Nummenmaa ym. 2016, 14).

Tutkimusongelmien määrittämisen jälkeen seuraa tutkimussuunnitelman laatiminen. Tutkimuksen suunnittelun keskeinen tavoite on mahdollistaa mahdollisimman tarkan, ajantasaisen ja täsmällisen tutkimustuloksen aikaansaaminen. Suunnitteluvaiheessa tulee tähdätä myös tunnistamaan mahdollisia ongelmia, joita tutkimuksen aikana saattaa esiintyä. (Tilastokeskus 2007, 31.) Tutkimussuunnitelmasta tulee käydä yksityiskohtaisesti ilmi Heikkilän (2014, 23) mukaan:

- Mitä tutkitaan?
- Miksi sitä tutkitaan?
- Tutkimusmenetelmät
- Tiedonkeruumenetelmät
- Otantamenetelmät
- Budjetti ja aikataulu

Tutkimuksen tekeminen

Hyvän suunnitelun jälkeen on aika siirtyä operationaaliseen vaiheeseen, jossa suoritetaan tiedonkeruu, otanta, aineiston käsittely ja valmistus sekä analyysit ja raportointi (Tilastokeskus 2007, 29).

Tiedonkeruu

Tilastollisen tutkimuksen toteuttamiseen tarvitaan aineistoa. Aineisto voi olla jo valmiiksi olemassa olevaa ja monet tutkimusongelmat voidaan ratkaista jo olemassa olevilla

aineistoilla. Yleisimpiä tutkimuksissa käytettäviä valmiita aineistoja ovat arkistoidut tutkimusaineistot, tilastot, rekisterit ja tietokannat. Esimerkiksi Suomen yhteiskuntatieteelliseen tietoarkistoon on kerätty tuhansia tallennettuja aineistoja. Tilastoja löytää esimerkiksi tilastokeskuksesta, johon on kerätty satoja tilastoja erilaisilta aihealueilta kuten asuminen, koulutus tai työmarkkinat. Myös terveyden ja hyvinvoinnin laitoksen (THL) tuottamista tilastoista suurin osa on vapaasti käytettävissä tutkimuksia varten. Mikäli valmiit tilastot tai aineistot eivät sovellu palvelemaan tutkimusta on aineisto silloin kerättävä itse. Silloin aineisto voidaan kerätä esimerkiksi kyselytutkimuksilla, haastattelututkimuksilla tai havainnointitutkimuksilla. (KvantiMOTV 2010.)

Kyselytutkimukset

Kyselytutkimukset eli surveytutkimukset tehdään kyselylomakkeen avulla. Valmiiksi mietityt kysymykset ja vastausvaihtoehdot sisältävä kyselylomake toimitetaan havaintoyksiköille yleensä postin tai internetin välityksellä. (KvantiMOTV 2010.)

Postikysely toteutaan lähettämällä kohdehenkilöille postin välityksellä kyselylomake, saatekirje sekä palautuskuori postimerkillä. Huonoja puolia postikyselyssä on esimerkiksi se että ei voida olla varmoja onko kyselyyn vastannut juuri se henkilö jolle kysely oli tarkoitettu. Myöskin vastausprosentit saattavat olla matalia postikyselyissä. Vastausprosenttia voidaan pyrkiä parantamaan esimerkiksi käyttämällä palkintoja ja lähettämällä muistutuskirjeitä vastaamatta jättäneille. (Heikkilä 2014, 63.) Internetkyselyllä suoritettava kyselytutkimus eroaa postikyselystä muun muassa siten että aineisto on käsiteltävissä ohjelmilla heti keräämisen jälkeen. Internetkysely on huomattavasti nopeampi tapa kerätä vastaustuloksia. Internetkyselyiden toteuttamista varten on erityisiä tiedonkeruuhjelmia kuten Webropol. (Heikkilä 2014, 66.)

Haastattelututkimukset

Haastatteluilla toteutettavissa tutkimuksissa vastaaja on suorassa vuorovaikutuksessa tutkijan kanssa. Haastattelut suoritetaan yleensä puhelimitse tai käyntihaastatteluina, haastateltavan kotona. (KvantiMOTV 2010.) Haastattelututkimusten etuna on yleensä hyvinkin korkea jopa >90% vastausprosentti. Haastattelijalla on myös mahdollisuus tarkentaa epäselviä kohtia vastaajalle, sekä kysyä kysymykset haluamassaan järjestyksessä. Haittapuolina henkilökohtaisissa haastatteluissa ovat haastattelijan

vaikutus vastaajan vastauksiin ja suuri haastattelijamäärän tarve joka voi koitua kustannuksellisesti haastavaksi. (Heikkilä 2014, 64.)

Havainnointitutkimukset

Havainnointitutkimuksissa tarvittavat tiedot kerätään havainnoimalla tilanteita, ihmisiä tai tapahtumia. Havainnointitutkimuksissakin käytetään yleensä hyödyksi valmista havainnointilomaketta. (KvantiMOTV 2010.) Hyvä esimerkki havainnointitutkimuksesta on kuluttajien käyttäytymistutkimukset, kuten kuinka usein asiakkaat katselevat vaikkapa näyteikkunaa kaupassa (Heikkilä 2014, 17).

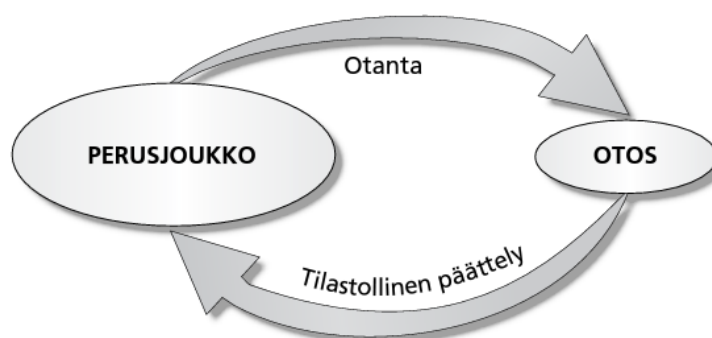
Kyselylomake

Kyselylomakkeen huolellinen suunnittelu ja laatiminen on tärkeä tekijä tutkimuksen onnistumisen kannalta. Hyväkin tutkimus on mahdollista pilata huonosti toteutetulla kyselylomakkeella. (Heikkilä 2014, 45.)

Hyvän tutkimuslomakkeen tunnuspiirteitä Heikkilän (2014, 47) mukaan ovat muun muassa että lomake on visuaalisesti houkuttelevan näköinen, kysymykset on aseteltu hyvin ja ne etenevät loogisesti ja ohjeet vastaamiseen ovat selkeät. Lomake ei myöskään saa olla liian pitkä. Kysymysten tarpeellisuus on harkittava tarkkaan ja lomakkeen olisi hyvä saada vastaaja tuntemaan omien vastaustensa olevan tarpeellisia ja hyödyllisiä.

Otanta

Tutkimuksella pyritään selvittämään jonkin joukon ominaisuuksia. Kun tutkimusta aletaan tekemään voidaan se tehdä joko selvittämällä koko tämän perusjoukon jokaisen yksikön ominaisuudet, jolloin on kyse kokonaistutkimuksesta. Tai tehdä otantatutkimus, jossa koko perusjoukosta valitaan sopivan kokoinen otos kuvaamaan tätä joukkoa. (Nummenmaa 2016, 25 – 26.) Kaikkien yksiköiden tutkiminen ei aina ole mahdollista, eikä myöskään tarpeellista. Silloin otantatutkimuksen käyttäminen kokonaistutkimuksen sijaan on järkevää. (Tilastokeskus 2007, 56.) Joukosta otetun otoksen tehtävä on kuvata pienoiskoossa koko perusjoukkoa (Heikkilä 2014, 31).



Kuva 4. Otanta. (Heikkilä 2014, 33)

Heikkilän (2014, 33) mukaan otanta aloitetaan määrittämällä perusjoukko. Perusjoukon määrittämisen jälkeen tulee etsiä rekisteri joka kuvaa kyseistä perusjoukkoa. Tätä seuraa otosyksikön määrittäminen, esimerkiksi onko kyseessä yritys vai henkilö. Seuraavaksi valitaan sopiva otantamenetelmä (ks. luku 2) ja ratkaistaan otoksen koko. Sitten voidaan suunnitella ja suorittaa otanta.

Sopivan otoskoon selvittämiseen on olemassa laskukaavoja. Yleensä otoskoon ratkaiseminen onnistuu kuitenkin ilman kaavaa käytännön kokemuksta hyödyksi käyttäen. Esimerkiksi valtakunnallisessa kuluttajatutkimuksessa sopiva otoskoko on 500-1000 yksikköä. (Heikkilä 2014, 43.)

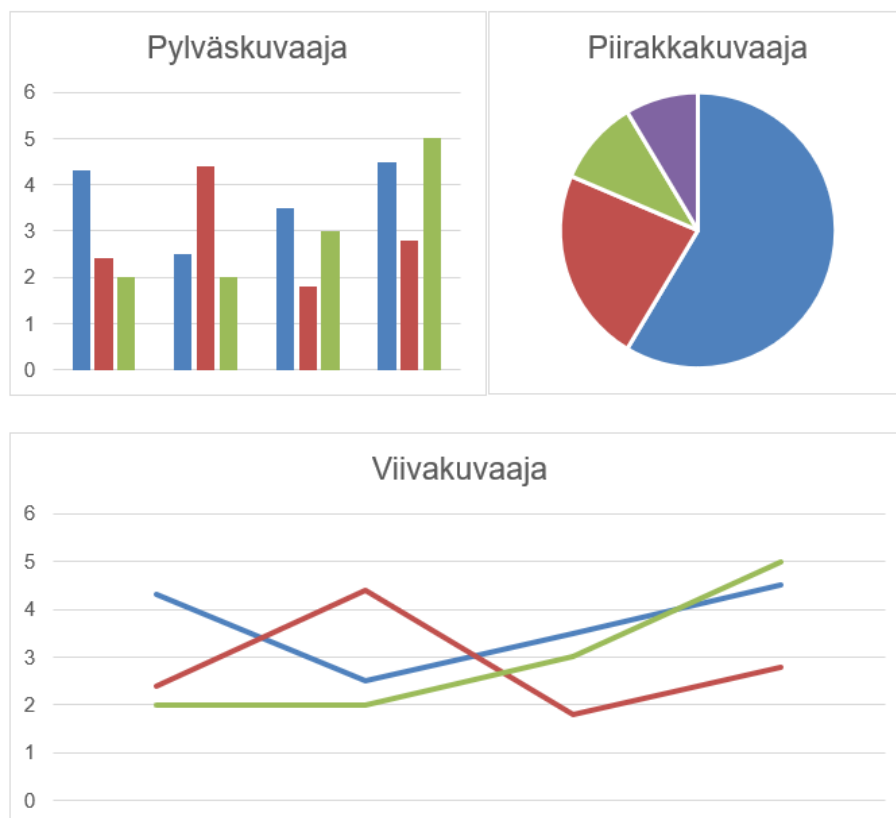
Analysointi, esittäminen ja raportointi

Kun aineisto on kerätty havaintoyksiköiltä tiedonkeruumenetelmillä täytyy saatua dataa analysoida. Sopivan analysointimenetelmän valitseminen ei ole helppoa ja monesti voi joutua testaamaan montaa eri sopivaa menetelmää ennen haluttuihin tuloksiin pääsemistä (Heikkilä 2014, 174). Otannalla tehtyjen tutkimusten tuloksissa ei yleensä olla kovinkaan kiinnostuneita otoksesta, vaan siitä miten se kuvaa koko perusjoukkoa. Tähän tarvitaan tilastollista päättelyä. Tilastollisen päättelyn avulla arvioidaan esimerkiksi sitä miten hyvin tulokset on yleistettävissä koko perusjoukkoon. (KvantiMOTV 2004b.) Kun tarkastellaan yksittäistä muuttujaa ja sen arvoja, käytetään yleensä keskilukuja, hajontalukuja ja graafista tarkastelua tilastollisen jakauman kuvailuun. (KvantiMOTV 2003.) Keskiarvotestejä tekemällä voidaan testata esimerkiksi sitä poikkeavatko eri ryhmien keskiarvot toisistaan vai onko kyse vain sattumasta (Heikkilä 2014, 174). Kahden muuttujan analysointia voidaan tehdä esimerkiksi

ristiintaulukoinnilla. Ristiintaulukoinnin tarkoituksena on selvittää muuttujien välistä yhteyttä ja sitä millainen vaikutus niillä on toisiinsa. (Heikkilä 2014, 198.)

Graafinen esittäminen

Ilmiöiden havainnollistaminen on yleensä helpompaa kun käytetään erilaisia kuvioita. Kuvioiden tarkoitus on tuoda informaatio helpommin hahmotettavaksi (KvantiMOTV 2004a). Kuvio ja taulukko välittää tietoa täysin eri tavalla, ja molemmilla on oma tehtävänsä. Kuvio välittää tiedon nopeasti katselijalle, kun taas taulukosta on nähtävillä tarvittava tarkkuus. Hyvän tilastokuvion ominaisuuksia ovat kiinnostuksen herättäminen, tiedon visuaalinen välittäminen, suuren datamäärän esittäminen pienemmässä tilassa ja kokonaisuuden muodostaminen. (Heikkilä 2014, 148.) Nummenmaan ym. (2016, 44) mukaan yleisimmät kuvaajatyytit ovat pylväskuvaajat, viivakuvaajat ja sektorikuvaajat. Erilaisia kuvaajavaihtoehtoja on olemassa niin paljon, että sopivan valitseminen saattaa joskus osoittautua vaikeaksi. Turhan hienojen ja näyttävien kuvioiden tekeminen ei välttämättä ole järkevää ja on yleensä parempi pysyä yksinkertaisemmissa tiedon paremmin välittävissä vaihtoehtoissa. (Heikkilä 2014, 149.)



Kuva 5. Erilaisia kuvaajia.

Yllä näkyvässä kuvassa 5. on kuvattu erilaisia havainnolistamiseen käytettäviä kuvaajia.

Raportointi

Tutkimusraportin tarkoitus on olla yhteenveto ja kuvaus siitä mitä tutkimuksessa on tehty ja millaisiin tuloksiin on päästy. Raportista tulisi käydä ilmi tutkimuksen tarkoitus, millaisia menetelmiä on käytetty, tulokset ja johtopäätökset. Tutkimusraportissa tulisi myös miettiä sitä onko tutkimuksen tulokset luotettavia, kuinka paljon virheitä tuloksissa on ja vaikuttavatko nämä virheet merkittävästi tuloksiin. (Heikkilä 2014, 71 – 73.)

3.4 Excel-ohjelman käyttö tilastotieteessä

Tilastollisia ohjelmia on olemassa todella paljon SPSS, SAS, STATA, STATISTICA, TIXEL jne. Kaikilla ohjelmilla on omat tunnuspiirteensä ja niitä käytetään tilastollisten aineistojen käsittelyyn ja analysointiin. (Heikkilä 2014, 118.) Keskitymme tässä työssä

kuitenkin pelkästään Excel taulukkolaskentaohjelmaan, ja miten sitä voi hyödyntää tilastollisissa toimenpiteissä.

Microsoft Excel on Microsoftin luoma taulukkolaskentaohjelma, jonka avulla voidaan organisoida dataa, suorittaa laskelmia, käyttää graafisia työkaluja ja luoda kuvioita. Excel koostuu soluista jotka on järjestetty riveihin ja sarakkeisiin. (Techopedia 2019.)



Kuva 6. Microsoft Excel.

Excel soveltuu mainiosti esimerkiksi tiedon visualisoimiseen. Tilastollisen tutkimuksen tuloksia ja dataa on yleensä hyvä visualisoida erilaisten kuvaajien avulla. Excelin avulla voidaan tehdä erilaisia kuvaajia kuten pylväskuvaajia, piirakkakuvaajia, viivakuvaajia, ja pivot-kaavioita. Pivot-kaaviot soveltuvat erinomaisesti havainnollistamaan yhteenvetoja suurista aineistoista. (Taanila 2020a.) Excelin avulla voidaan suorittaa myös esimerkiksi tunnuslukujen kuten keskiarvon, keskihajonnan tai mediaanin laskemista datasta funktioiden avulla. (Taanila 2019).

Hyödyllisiä Excel funktioita

- =AVERAGE, Keskiarvo
- =STDEV, Keskihajonta
- =MEDIAN, Mediaani
- =MAX, maksimi (suurin arvo)
- =MIN, minimi (pienin arvo)
- =PERCENTILE, (prosenttipiste esimerkiksi ylä- ja alakvartiilien laskemiseen)

Aki Taanilan Olennaiset Excel-taidot (2020b.) blogikirjoituksen mukaan taulukkolaskennan välttämättömiä perustaitoja ovat muun muassa:

- Taulukoissa liikkuminen, sekä taulukon osien järjestely siirtämällä ja kopioimalla
- Sarakkeiden ja rivien lisääminen
- Lukujen muotoilu

- Summien laskeminen
- Kaavojen hallitseminen
- Funktioiden hallitseminen ja pääperiaatteet

Excelin tilastokäsittelyyn käyttämisen etuja ovat muun muassa se että ohjelma on yleisesti käytössä lähes joka paikassa. Excelin kaavio- ja kuviovalikoima on erittäin monipuolinen ja helposti käytettävä. Ohjelma sisältää myös analyysityökaluja ja tilastollisia apuohjelmia. Excel toimii myös mainiosti tukemaan muiden tilasto-ohjelmien käyttöä ja tiedonsiirto muista tilasto-ohjelmista Exceliin on yleensä helppoa. (Heikkilä 2014b.)

Kuten aiemmin luvussa kaksi jo mainittiin, insinöörit suorittavat testejä, analysoivat, sekä käsittelevät dataa työssään. Erilaisten tilastollisten testien ja analyysien tulokset on pystyttävä myös esittämään selkeästi ja loogisesti. Tähän voidaan käyttää hyödyksi juuri Excelin kaltaisia ohjelmia. Excelin avulla voidaan tehdä yhteenvetoja erilaisista aineistoista ja esittää tuloksia graafisesti ja helposti ymmärrettävässä muodossa. Exceliä voidaan hyödyntää moneen tarkoitukseen insinööriyössä. Esimerkiksi asiakastytyväisyyttä tutkittaessa data voidaan kerätä Exceliin ja saadut tulokset esittää graafisesti käyttämällä hyödyksi Excelin laajaa kuvaajavalikoimaa. Omien insinööriopintojen aikana olen nähnyt Exceliä käytettävän yrityksissä muun muassa varastosaldojen kirjaamiseen. Excel on mainio työkalu insinöörille ja juuri siksi koen, että sen osaamista on hyvä harjaannuttaa tällä opintojaksolla.

4 TYÖN KÄYTÄNNÖN TOTEUTUS

Opinnäytetyön käytännön toteutus liittyi kurssin materiaalin kehittämiseen. Tämä tarkoitti käytännössä olemassa olevan materiaalin eteenpäin kehittämistä ja tarkistamista. Olemassa oleva toteutus tavarataloa koskevasta case-aineistosta ja esimerkistä tarkistettiin. Osa käytännön toteutuksesta toteutui myös kurssilla jatkossa mahdollisesti käytettävien uusien tehtävien tekemisen kautta. Tehtäviä tehtiin tilastollisen tutkimuksen tekemisestä, Ylen Tiedeykkösen tilastoja käsittelevästä radio-ohjelmasta sekä tilastollisten menetelmien peruskäsitteistä. Apuna käytettiin muun muassa Tilastollisten menetelmien perusteet (Nummenmaa ym. 2016) kirjaa, joka on hyvä materiaali kurssia suorittaville opiskelijoille. Apuna käytettiin myös olemassa olevaa ja luonnosteltua kurssimateriaalia.

4.1 Kysymysten laadinta

Tilastollista ajattelua käsittelevä Tiedeykkösen radio-ohjelma kuunneltiin ja siitä laadittiin kysymyksiä, joita voidaan käyttää ohjelman sisällön kertaamiseen mikäli kyseistä radio-ohjelmaa käytetään kurssin oppimateriaalina. Tiedeykkösen radio-ohjelmassa ”Tilastotieteessä pitää osata tulkita – Jäätelönsyönti ei lisää haiden hyökkäyksiä” keskustellaan tilastollisten tutkimusten tulkinnasta ja tulosten luotettavuuden arvioinnista. Radio-ohjelmasta tehdyt kysymykset ovat avoimia kysymyksiä ja niihin vastaaminen edellyttää ainoastaan radio-ohjelman kuuntelemista.

Tilastollisesta tutkimuksesta ja peruskäsitteistä kysymyksiä tehtiin lähinnä monivalintamuotoisina. Monivalintakysymykset ovat hyviä, koska opiskelijoiden osaamista voidaan mitata nopeasti, innostavalla tavalla sekä kiinnostusta herättäen (esim. Hemminki ym. 2013, 42). Kysymyksiä pyrittiin tekemään siitä näkökulmasta, että ne eivät olisi liian haastavia vaan enemmänkin innostavia. Lisäksi otettiin huomioon se, että kysymyksiä voi käyttää esimerkiksi verkkoalustalla, ja niiden tarkistaminen on helppoa. Kuitenkin, vaikka materiaalin tuottamisessa pyrittiin pääosin monivalintakysymysten tekemiseen, monipuolisuuden nimissä tehtiin myös hieman haastavampia avoimia kysymyksiä.

Kysymysten tekemisessä pyrittiin ottamaan huomioon opiskelijoiden oppimisprosessi. Siksi oikeisiin vastauksiin laitettiin mukaan myös selityksiä. Näin ollen vastauksia

voidaan mahdollisesti käyttää esimerkiksi verkkoalustoilla niin, että opiskelija saa välittömästi palautteen vastauksestaan sekä ymmärtää myös, millä perusteella hänen vastauksensa on mennyt oikein tai väärin. Opintojakson materiaaliksi laaditut kysymykset löytyvät tämän työn liitteistä. Kaikkiin luotuihin kysymyksiin, tehtäviin ja toteutuksiin tehtiin myös oikeat ratkaisut ja ne toimitettiin opintojaksosta vastaavalle henkilölle.

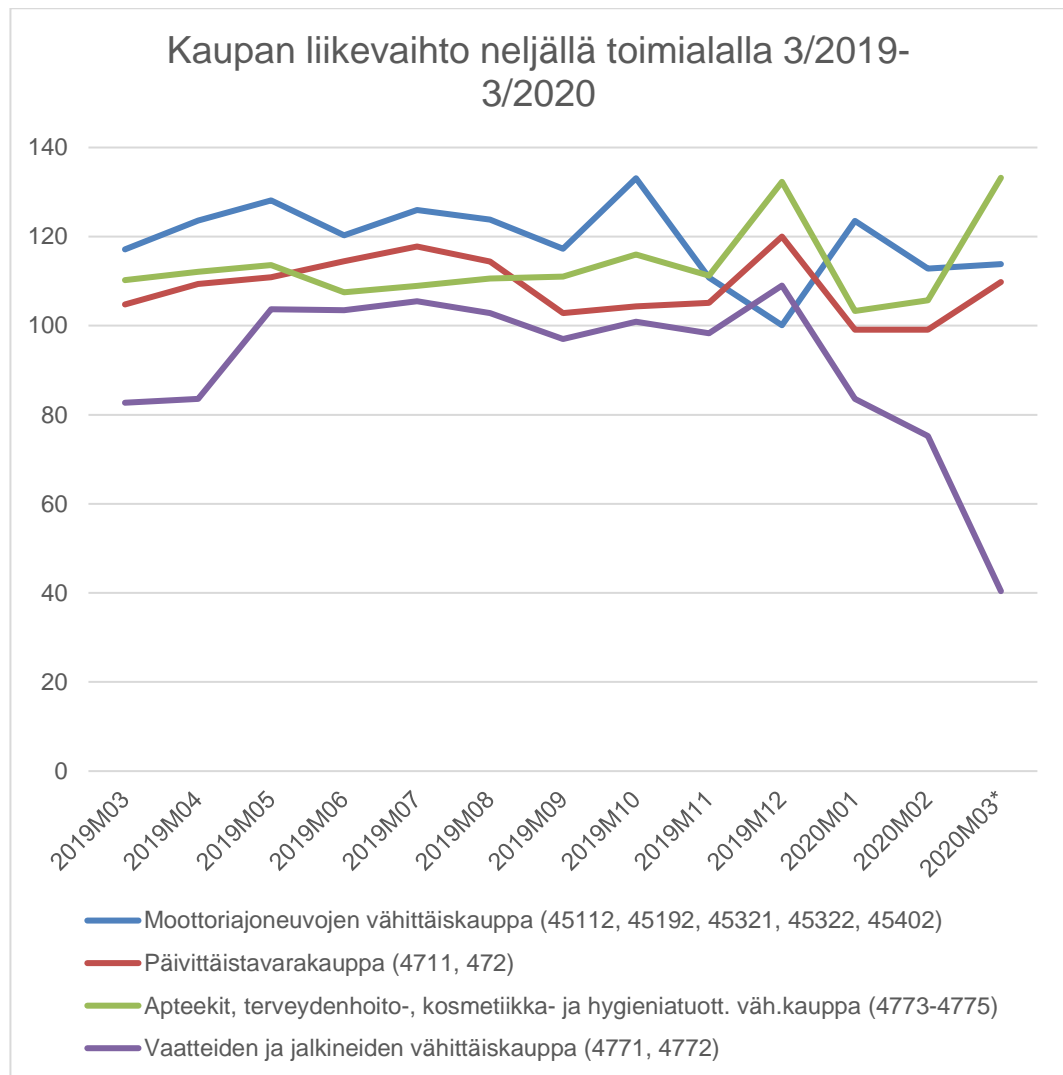
4.2 Excel – tehtävien laadinta

Tilastollisten menetelmien perusteet -kirjassa (Nummenmaa ym. 2016) on liitteenä todelliseen tutkimukseen liittyvä tapausesimerkki tavaratalon palvelutasotutkimuksesta. Toteutuksessa tarvittava aineisto (Kysely) on saatavissa verkosta Sanoma Pro:n verkkosivuilta. Kurssin valmiina materiaalina oli Excel-toteutus kyseisestä palvelutasotutkimuksesta, joka tarkistettiin. Lisäksi materiaalia täydennettiin lisäämällä siihen hintojen esillä olemista ja liikkeen aukioloaikoja koskevat toteutukset. Liikkeen aukioloaikoja koskevat toteutukset vaativat muuttujien uudelleen luokittelua. Muuttujien uudelleen luokittelu toteutettiin luokittelemalla ikä-muuttuja ja liikkeen aukioloaikojen sopivuutta mittaavat muuttujat kolmiluokkaisiksi.

Lisäksi kurssille laadittiin Excelissä toteutettavia tehtäviä, joihin haettiin aineisto Tilastokeskuksen avoimesta PxWeb-tilastotietokannasta sekä Sanoma Pro:n sivuilta. Aineistona käytettiin Tilastokeskuksen avoimen datan kaupan liikevaihtokuvaajaa sekä Sanoma Pro:n sivuilta saatavaa Tilastollisten menetelmien perusteet -kirjan materiaalina käytettävää Asiakas-dataa. Kaupan liikevaihtokuvaaja kuvaa kaupan yritysten liikevaihdon kehitystä ja siitä tehdyissä tehtävissä opiskelijoilta pyydettiin muun muassa graafista mallinnusta sekä keskiarvojen laskemista. Asiakas-datasta tehdyissä tehtävissä opiskelijoilta pyydettiin muun muassa mediaanin ja frekvenssien määrittämistä sekä ristiintaulukoinnin tekemistä. Excel tehtävät löytyvät työn liitteestä 2.

Laadittuihin Excel tehtäviin pyrittiin tuomaan käytännön näkökulmaa käyttämällä mahdollisimman ajankohtaista dataa, josta voi tehdä päätelmiä ajankohtaisista ilmiöistä. Aineistona käytettiin Kaupan liikevaihtokuvaajaa, joka kuvaa kaupan yritysten liikevaihdon kehitystä. Tehtävässä käytössä oleva aineisto kuvaa kaupan liikevaihdon kehitystä vuoden 2019 maaliskuusta vuoden 2020 maaliskuuhun. Mielestäni ajankohtaisten ja tuoreiden tilastoaineistojen käyttö lisää tehtävien mielenkiintoisuutta, sillä opiskelija voi aineistojen perusteella huomata nykypäivän ilmiöitä.

Graafinen esittäminen, selkeiden kuvaajien ja yhteenvetojen teko on yksi taito joka insinöörillä täytyy olla hallussa. Siksi tehtävissä tehdään aineistoista erilaisia kuvaajia, jotka kertovat tilastoista paljon enemmän kuin pelkät numerot. Alla olevassa kuvassa 7. voidaan nähdä esimerkki yhdestä tehtävissä luotavista graafisista kuvaajista. Laskut ja nousut on helpommin havaittavissa graafisesta esityksestä kuin pelkästä taulukosta täynnä numeroita.



Kuva 7. Excel tehtävässä 1 luotu kuvaaja.

Koska opinnäytetyön toimeksiannossa harjoitusten toteutusmuoto oli rajattu Exceliin, ei kovin monimutkaisiin tilastollisiin malleihin ollut tarpeellista mennä. Tämä ei ollut tarpeellista myöskään siitä näkökulmasta, että kurssi keskittyy tilastollisen ajattelun ja menetelmien perusteiden hallitsemiseen. Jatkokehitysideana esille voisi nostaa sen, että ammattikorkeakoulussa tarjottaisiin kurssia, jossa harjoitellaan haastavampien

tilastollisten analyysien suorittamista. Tilastollisen analyysin tekeminen olisi Exceliä helpompaa tilasto-ohjelmistojen, kuten SPSS:n, STATA:n tai TIXEL:n, avulla, joten näitä ohjelmia ja niiden käytön opettelua voisi harkita osaksi insinöörien opintoja.

5 POHDINTA

Tässä opinnäytetyössä kehitettiin Turun ammattikorkeakoulun insinööriopiskelijoiden kolmen opintopisteen Tilastolliset menetelmät -opintojaksoa. Työn tehtävänantona oli laatia kysymyksiä ja tehtäviä, jotka soveltuvat tukemaan oppimista ja antavat aineistoa kurssin arviointiin, sekä päivittää olemassa olevaa kurssimateriaalia. Työn tavoitteena oli myös etäopiskelun ja itsenäisen opiskelun näkökulman sekä insinööriopiskelijoiden näkökulman huomioiminen.

Työn teoriaosuus koostui tilastollisesta ajattelusta ja lukutaidosta, tilastollisesta tutkimuksesta ja peruskäsitteistä sekä Excelin käytöstä tilastotieteessä. Työn teoriaosuudessa käsiteltiin niitä aiheita, joista myös tehtiin kurssille materiaalia. Tilastollisesta ajattelusta, tutkimuksen tekemisestä ja tilastotieteen peruskäsitteistä tehtiin kysymyspatteristoja, joita voidaan käyttää oppimisen testaamisessa. Kurssille laadittiin myös Excelissä toteutettavia tehtäviä sekä kehitettiin jo olemassa olevaa Excel-toteutusta tavaratalon palvelutasotutkimuksesta.

Mielestäni opinnäytetyön tavoitteisiin päästiin ja kurssin materiaalia saatiin kehitettyä eteenpäin. Tuotetut tehtävät vastasivat esille tulleeseen etäopiskelun tarpeeseen, koska ne ovat käytettävissä alustasta riippumatta esimerkiksi verkkoympäristöissä kertaavina kysymyksinä etäopiskeltuihin aiheisiin. Tilastollisten menetelmien hallitseminen on tärkeä osa insinöörien jokapäiväistä työskentelyä, ja tuotetut kurssimateriaalit edistävät tilastollisen ajattelun perusteiden hallintaa. Kuitenkin, koska tuotetun materiaalin vaikeusaste on kohtuullisen helppo, voivat tehtävät olla joillekin opiskelijoille liian helppoja.

Insinööriopiskelijoiden käytännön näkökulman huomioiminen kurssin materiaalin tuottamisessa osoittautui yllättävän hankalaksi. Tilastomatematiikan ja tilastollisten menetelmien hyötyjä voi olla haastavaa yhdistää insinööriyöhön tai normaaliin elämään, joten käytännön näkökulmien tuominen esille voisi olla tärkeää. Näin kurssia ei suoritettaisi vain läpi pääseminen mielessä vaan opiskelijat voisivat nähdä siitä konkreettisesti työelämässäkin saatavan hyödyn.

Kurssin tulevaisuutta ajatellen kehitysehdotuksena voisi olla mahdollisimman käytännönläheiset opetusmuodot, jotka tukisivat tilastojen ja niiden analysoimisen tarpeellisuutta insinööriyössä. Ammattikorkeakouluissa toimitaan läheisessä

yhteistyössä erilaisten yritysten kanssa, jonka vuoksi tulevaisuudessa voisi olla mielekäästä, jos opiskelijat pääsisivät analysoimaan tilastollisten menetelmien kursseilla aitoa yrityksiltä saatavaa dataa. Lisäksi haastavampien tilastollisten menetelmien jatkokurssin tarjoaminen vapaaehtoisena opintojaksona voisi edistää insinöörien tilastotieteellisiä taitoja. Tähän kurssiin voisi sisällyttää myös tilasto-ohjelmistojen käytön harjoittelua.

LÄHTEET

Bisgaard, S. 1991. Teaching Statistics to Engineers. *The American Statistician*. Vol. 45, No. 4, 274–283.

Heikkilä, T. 2014. Tilastollinen tutkimus. 9. painos. Helsinki: Edita Publishing Oy.

Heikkilä, T. 2014b. Kvantitatiivinen tutkimus. Verkkomateriaali. Viitattu 5.6.2020 <http://www.tilastollinentutkimus.fi/1.TUTKIMUSTUKI/KvantitatiivinenTutkimus.pdf>.

Heikkinen, S. 2019. Tilastojen seitsemän kuolemansyntiä: Näin sinua harhautetaan. Yle 9.4.2019. Viitattu 12.5.2020 <https://yle.fi/aihe/artikkeli/2019/04/09/tilastojen-seitseman-kuolemansyntia-nain-sinua-harhautetaan>.

Hemminki, M.; Leppänen, M. & Valovirta, T. 2013. Innostu ja onnistu opetuksessa. Aalto-yliopiston julkaisusarja CROSSOVER 15/2013.

Kankaanranta, J. 2012. Applications of industrial statistics. Turun yliopiston julkaisuja – Annales Universitatis Turkuensis. Sarja – ser. A 1 osa – tom. 450. Väitöskirja. Luonnontieteiden ja matematiikan tiedekunta. Informaatioteknologian laitos. Turku: Turun yliopisto. Viitattu 27.5.2020 <https://www.utupub.fi/bitstream/handle/10024/86540/AnnalesA1450Kankaanranta.pdf?sequence=1&isAllowed=y>.

KvantiMOTV. 2003. Menetelmäopetuksen tietovaranto. Tampere: Yhteiskuntatieteellinen tietoarkisto. Viitattu 28.5.2020 <https://www.fsd.tuni.fi/menetelmaopetus/menetelma/menetelmatyypit.html>.

KvantiMOTV. 2004a. Menetelmäopetuksen tietovaranto. Tampere: Yhteiskuntatieteellinen tietoarkisto. Viitattu 29.5.2020 <https://www.fsd.tuni.fi/menetelmaopetus/kuviot/kuviot.html>.

KvantiMOTV. 2004b. Menetelmäopetuksen tietovaranto. Tampere: Yhteiskuntatieteellinen tietoarkisto. Viitattu 29.5.2020 <https://www.fsd.tuni.fi/menetelmaopetus/paattely/paattely.html>.

KvantiMOTV. 2010. Menetelmäopetuksen tietovaranto. Tampere: Yhteiskuntatieteellinen tietoarkisto. Viitattu 28.5.2020 <https://www.fsd.tuni.fi/menetelmaopetus/tutkimus/aineistotyypit.html>.

Nummenmaa, L.; Holopainen, M. & Pulkkinen, P. 2016. Tilastollisten menetelmien perusteet. 1.–2. painos. Helsinki: Sanoma Pro Oy.

Statanalytica 2020. Top 10 Uses of Statistics In Our Day to Day Life. Viitattu 21.5.2020 <https://statanalytica.com/blog/uses-of-statistics/>.

Taanila, A. 2020a. Olennaiset Excel-taidot: Tiedon visualisointi. Aki Taanilan menetelmäblogi. Viitattu 29.5.2020 <https://excelapu.wordpress.com/tiedon-visualisointi/>.

Taanila, A. 2020b. Olennaiset Excel-taidot: Perustaidot. Aki Taanilan menetelmäblogi. Viitattu 1.6.2020. <https://excelapu.wordpress.com/perustaidot/>.

Taanila, A. 2019. Tunnuslukuja. Aki Taanilan menetelmäblogi. Viitattu 29.5.2020 <https://tilastoapu.wordpress.com/2011/10/19/tunnuslukuja/>.

Techopedia. 2019. Microsoft Excel. Viitattu 30.5.2020 <https://www.techopedia.com/definition/5430/microsoft-excel>.

Tilastokeskus. 2007. Laatusa tilastoissa. Käsikirjoja 43. 2. uudistettu painos. Helsinki: Yliopistopaino. http://www.stat.fi/meta/qg_2ed.pdf.

Tilastokeskus. 2020a. Tilastokoulu: Johdatus tilastotieteeseen. Tilastotieteen peruskäsitteitä: Tilastollinen lukutaito. Viitattu 14.5.2020
https://tilastokoulu.stat.fi/verkkokoulu_v2.xql?course_id=tkoulu_tilaj&lesson_id=2&subject_id=17&page_type=sisalto.

Tilastokeskus. 2020b. Tilastokoulu: Johdatus tilastotieteeseen. Vaihtelu ja riippuvuus. Viitattu 13.5.2020
https://tilastokoulu.stat.fi/verkkokoulu_v2.xql?course_id=tkoulu_tilaj&lesson_id=4&subject_id=6&page_type=sisalto.

Tilastokeskus. 2020c. Tilastokoulu: Johdatus tilastotieteeseen. Tunnusluvut. Viitattu 14.5.2020
https://tilastokoulu.stat.fi/verkkokoulu_v2.xql?course_id=tkoulu_tilaj&lesson_id=3&subject_id=0&page_type=sisalto.

Tilastokeskus. 2020d. Tilastokoulu: Johdatus tilastotieteeseen. Tilastotieteen peruskäsitteitä: Havaintomatriisi. Viitattu 13.5.2020
https://tilastokoulu.stat.fi/verkkokoulu_v2.xql?course_id=tkoulu_tilaj&lesson_id=2&subject_id=4&page_type=sisalto.

Turun Sanomat. 2012. Väitös: Teollisuus hyötyisi tilastotieteen tehokkaammasta käytöstä. Viitattu 5.6.2020
<https://www.ts.fi/uutiset/kotimaa/426883/Vaitos+Teollisuus+hyotyisi+tilastotieteen+tehokkaammasta+kaytosta>.

Laaditut kysymykset

Kysymykset Radio-ohjelmasta

1. Mitä tarkoitetaan kausaliteetilla?
2. Voimakas korrelaatio tarkoittaa että syy-seuraussuhde on syntynyt?
 - a. Oikein
 - b. Väärin
3. Ohjelmassa puhutaan siitä että korrelaation taustalta voi joskus löytyä yhteinen selittävä tekijä, mikä on yhteinen selittävä tekijä haiden hyökkäyksissä ja jäätelön myynnin kasvussa ohjelman esimerkissä?
4. Mihin tulisi kiinnittää huomiota kun tarkastellaan tutkimustuloksia? Luettele X määrä.
5. Mitä tarkoitetaan 95% luottamusvälillä?
6. Kun p-arvo on pienempi kuin 0.05 voidaan eroa pitää tilastollisesti merkitseväenä?
7. Onko 5% raja tuloksellisessa merkitsevyydessä aivan ehdoton.
8. Ohjelmassa puhutaan lukioiden vertailusta, kertooko pelkästään parhaat lopputulokset siitä että opetus kyseisessä lukiossa on parempaa? Miksi?
9. Yksi mahdollisuus on mitata sitä mikä lukio on parantanut tuloksiaan eniten? Mikä tässä voi tulla ongelmaksi?
10. Mikä on ohjelmassa esille tullut kaksoissokkokoe?

Kysymykset käsitteistä

1. Kiikaroit kaukoputkella taivaalle ja näet planeetan. Tiedät entuudestaan, että kaikki planeetat kiertävät tähtiä. Näin ollen teet johtopäätöksen, että kyseinenkin planeetta kiertää tähteä. Onko kyseessä induktiivinen vai deduktiivinen päättely?

2. Tutkimuksessa mitataan ihmisten kengänkoko. Kengänkoko on muuttuja, ja yhden ihmisen kengänkoko on ...
 - a. Perusjoukko
 - b. Tilastoyksikkö
 - c. Havainto
 - d. Poikkileikkaustutkimus

3. Mitkä ovat neljä mittaamisen tasoa kuvaavaa mitta-asteikkoa?

4. Millä mitta-asteikolla mitataan yrityksen liikevaihtoa?
 - a. Suhdeasteikko
 - b. Välimatka-asteikko
 - c. Järjestysasteikko
 - d. Luokitteluasteikko
5. Validiteetti ilmaisee, missä määrin on kyetty mittaamaan juuri sitä, mitä pitikin mitata? Oikein vai väärin?
 - a. Oikein
 - b. Väärin

6. Mikä seuraavista väitteistä EI pidä paikkaansa?
 - a. Kokonaistutkimuksessa perusjoukon jokainen otantayksikkö on mukana?
 - b. Otoksesta lasketut tunnusluvut poikkeavat tavallisesti hieman perusjoukon tunnusluvuista
 - c. Otantatutkimuksessa koko perusjoukko on tarkasteltavana

7. Teet tutkimusta suomalaisien ammattioppilaitosten opiskelijoiden halukkuudesta hakeutua insinööriopintoihin. Mitä otantamenetelmää käytät?

8. Millä nimellä kutsutaan vastaamattomuutta tai tavoittamattomuutta (esimerkiksi kyselytutkimukseen satunnaisotannalla valikoitunut henkilö jättää vastaamatta kyselyyn)?
 - a. Peitto
 - b. Harha
 - c. Näyte
 - d. Kato

9. Mistä otantamenetelmistä tuloksena on näyte?
 - a. Yksinkertainen satunnaisotanta
 - b. Systemaattinen otanta
 - c. Ositettu otanta
 - d. Ryväsotanta
 - e. Harkinnanvarainen otanta

f. Kiintiötanta

10. Havaintomatriisissa vaakariville laitetaan...?

- a. Jokaiselle vaakariville aina yhteen tilastoyksikköön liittyvien muuttujien arvot
- b. Kaikki tilastoyksiköt sijoitetaan samalle vaakariville
- c. Yhden muuttujan kaikki arvot
- d. Ei laiteta mitään

11. Milloin graafisena kuvaajana käytetään histogrammia?

- a. Kun kuvataan luokiteltuja muuttujia
- b. Kun kuvataan jatkuvia muuttujia
- c. Kun tutkittavia muuttujia on paljon
- d. Kun tutkittavia muuttujia on vähän

12. Minkälaisen graafisen esityksen tekisit osakeyhtiön osakkeen arvon kehityksen kuvaamiseksi?

- a. Viivadiagrammi
- b. Sektorikuvaaja
- c. Histogrammi
- d. Sirontakuvi

13. Haluat tutkia, onko valtion BKT:lla ja kotitalouksien käytettävissä olevilla tuloilla jotain yhteyttä toisiinsa. Minkälaisella graafisella esityksellä voit tarkastella sitä, onko muuttujien välillä mahdollisesti riippuvuutta?

- a. Histogrammi
- b. Viivadiagrammi
- c. Sirontakuvi
- d. Sektoridiagrammi

14. Milloin graafisena kuvaajana käytetään sektoridiagrammia?

15. Moodi, mediaani ja keskiarvo ovat...

- a. Hajontalukuja
- b. Keskilukuja
- c. Muita sijaintilukuja

16. Havaintoaineiston yleisin muuttujan arvo on?

- a. Keskihajonta
- b. Mediaani
- c. Keskiarvo
- d. Moodi

17. Kuinka moneen osaan alakvartiili, mediaani ja yläkvartiili jakavat tilastoaineiston?

- a. Kahteen
- b. Kolmeen
- c. Neljään
- d. Viiteen

18. Olet suorittanut opintokokonaisuuden, jossa on viisi kurssia. Arvosanasi kurseista ovat 3, 2, 3, 4 ja 2. Mikä on opintokokonaisuuden keskiarvo?

- a. 2
- b. 3
- c. 2,8
- d. 2,5

19. Kotitalouden bruttotulot ja elintarvikemenojen osuus kotitalouden menoista korreloivat negatiivisesti keskenään. Mikä seuraavista väittämistä pitää paikkaansa tässä ilmiössä?

- a. Suurituloiset käyttävät suuremman osan tuloistaan ruokaan
- b. Pienituloiset syövät enemmän kuin suurituloiset
- c. Bruttotulojen kasvaessa elintarvikkeisiin käytetyn rahamäärän suhteellinen osuus pienenee

20. Objektivisuudella tarkoitetaan.

- a. Sitä että saatu mittaustulos on luotettava
- b. Sitä että tutkija on pätevä tekemään tutkimuksen
- c. Sitä että tutkimuksen otoskoko on riittävän suuri
- d. Sitä että tutkijan omat henkilökohtaiset mielipiteet eivät vaikuta tutkimustulokseen

Kysymykset tilastollisesta tutkimuksesta

1. Tilastollinen tutkimus on?

- a. Kvantitatiivinen
- b. Kvalitatiivinen
- c. Laadullinen
- d. Määrällinen

2. Haluat tietää millainen menekki keskioluella on lähikaupassasi. Menet perjantai-illalla havainnoimaan virvokehyllyn viereen ja teet johtopäätökset menekistä. Onko tutkimus reliaabeli?

- a. Kyllä
- b. Ei

3. Missä seuraavista tiedonkeruumenetelmistä on yleensä paras vastausprosentti?

- a. Postikysely
- b. Internetkysely
- c. Haastattelututkimus

4. Otanta suoritetaan ennen perusjoukon valintaa?
 - a. Kyllä
 - b. Ei

5. Teet kuluttajatutkimusta suomalaisten ostoskäyttäytymisestä, mikä on sopiva otoskoko tutkimukselle?
 - a. 10
 - b. 10 000
 - c. 100
 - d. 1000

6. Sinulla on lista Turun Ammattikorkeakoulun opiskelijoista, valitset listalta joka viidennen opiskelijan mukaan tutkimukseksi otokseen. Millainen otantamenetelmä on kyseessä?
 - a. Satunnaisotanta
 - b. Systemaattinen otanta
 - c. Ryväotanta
 - d. Ositettu otanta

7. Laita seuraavat tutkimusprosessin vaiheet oikeaan järjestykseen?
 - a. Tutkimusstrategian laatiminen
 - b. Tutkimusaineiston analysoiminen
 - c. Tulosten arviointi
 - d. Aineiston kerääminen

8. Suoritat tilastollisen analyysin jonka p-arvo on 0.02. Mikä seuraavista väittämistä on oikein?
 - a. Tulos on tilastollisesti merkitsevä 95% luottamustasolla.
 - b. Tulosta ei missään nimessä voi yleistää perusjoukkoon
 - c. Tulos on tilastollisesti merkitsevä 99% luottamustasolla
 - d. Tulos on sattumanvarainen

9. Mitkä seuraavista ovat tilastollisia analysointi menetelmiä?
 - a. Khiin neliötesti
 - b. Ristiintaulukointi
 - c. Ryväotanta
 - d. Keskiarvotesti
 - e. Tilastoyksikkö
 - f. Välimatka-asteikko

10. Erästä harvinaista sairautta sairastavia on maailmassa vain 100 kappaletta. He kaikki käyvät säännöllisesti lääkärissä ja heidän sairaudesta kerätään tutkimusdataa. Kumpi tutkimusmenetelmä on kyseessä?
 - a. Kokonaistutkimus
 - b. Otantatutkimus

11. Laita seuraavat tutkimusraportin osat järjestykseen.

- a. Tulokset aihealueittain
- b. Katsaus aiempiin tutkimuksiin ja teoreettinen viitekehys
- c. Johdanto
- d. Johtopäätökset

12. Merkitse ruudukkoon rastilla onko väite OIKEIN vai VÄÄRIN.

	a	b	c	d	e	f	g	h	i
Oikein									
Väärin									

- a. Kysymysten loogisella asettelulla on merkitystä tutkimuslomaketta laadittaessa
- b. Itse keräämisen sijasta tutkimuksen aineisto voidaan hankkia myös valmiista tilastoista ja tietokannoista
- c. Kvantitatiivisessa tutkimuksessa käytetään yleensä suppeaa harkinnanvaraisesti koottua näytettä
- d. Kun kahden tapahtuman välillä on olemassa korrelaatiota on silloin kyse kausaliteetista eli syy-seuraussuhteesta
- e. Korkea vastausprosentti lisää tutkimuksen validiteettia
- f. Satunnaisotannassa perusjoukosta valitaan joka k:s yksikkö. Esim. joka neljäs.
- g. Luotettavan tutkimustuloksen aikaansaamiseksi jokainen perusjoukon jäsen on tutkittava
- h. Postikyselyiden vastausprosentti on yleensä erittäin korkea
- i. Ilmiöiden havainnollistaminen on helpompaa kuvioita käyttämällä

Excel tehtävänanto

Tässä tiedostossa on tehtäviä Tilastokeskuksen avoimesta datasta sekä Tilastollisten menetelmien perusteet kirjan Asiakas-datasta.

ASIAKAS-data: <https://tuotteet.sanomapro.fi/bu561383-tilastollisten-menetelmien-perusteet.html> Kaupan liikevaihtokuvaaja: <https://www.stat.fi/meta/til/klv.html>

Tehtävät 1, 2 ja 3: Aineistona Tilastokeskuksen PxWeb-tilastotietokannasta saatu avoin data. (http://pxnet2.stat.fi/PXWeb/pxweb/fi/StatFin/StatFin__kau__klv/statfin_klv_pxt_111t.px/)

Tehtävä 1. Tee graafinen esitys kaupan liikevaihdosta näiden neljän toimialan osalta.

Tehtävä 2. Mikä voisi selittää liikevaihdon kasvua vuoden 2020 ensimmäisellä kvartaalilla apteekki-, terveydenhuolto- ja kosmetiikka- ja hygienituotteiden osalta ja toisaalta vaatteiden ja jalkineiden vähittäiskaupan liikevaihdon laskua?

Tehtävä 3. Laske moottoriajoneuvojen vähittäiskaupan liikevaihdon vuosikeskiarvo, viimeisen kuuden kuukauden keskiarvo ja viimeisen 3 kuukauden keskiarvo.

Tehtävät 4, 5 ja 6: Aineistona Tilastollisten menetelmien perusteet kirjan Asiakas-data.

Tehtävä 4. Laske koulutuksen mediaani

Tehtävä 5. Laske frekvenssit asumismuodoille ja tee niistä palkkikuvio. Laske myös prosentuaaliset osuudet ja tee niistä sektoridiagrammi. Kuinka monella prosentilla vastaajista on omistusasunto?

Tehtävä 6. a) Tutki sukupuolen yhteyttä tyytyväisyyteen liikkeen tuotevalikoimaan ristiintaulukoinnilla. (Pivot-taulukkotoiminto, vinkki: tulkintaa helpottaa, jos otat riviprosentit.)

Tehtävä 6. b) Ovatko miehet tai naiset olleet useammin "erittäin tyytyväisiä" (arvo 5) tuotevalikoimaan?