Tampereen yliopisto  
Informaatioteknologian ja viestinnän tiedekunta  
Matematiikan ja tilastollisen data-analyysin tutkinto-ohjelma

**OPINTOJAKSOKUVAUS**

**Opintojakson nimi: DATA.STAT.210 EMPIIRINEN PROJEKTI**

**Laajuus:** 5 op

**Oletustyyppi:** aineopinnot

**Osaamistavoitteet:**

Opintojakson suoritettuaan opiskelija osaa toteuttaa tilastollisen tutkimuksen kaikki eri vaiheet. Opiskelija oppii hankkimaan hyvän tutkimusaineiston ja ratkaisemaan siitä rajaamiaan tutkimusongelmia tilastollisin tutkimusmenetelmin käyttäen hyväksi tilastotieteen perusopintojaksoilla saamiaan tietoja. Opinnäytteellä opiskelija osoittaa hallitsevansa keskeisimpien tilastollisten menetelmien oikeaoppisen käytön empiirisessä tilastollisessa tutkimuksessa ja osaavansa ymmärrettävästi raportoida tutkimustuloksia.

**Sisältö:**

1. TUTKIMUSONGELMAN MÄÄRITTÄMINEN

2. SUUNNITELMA ONGELMAN RATKAISEMISEKSI

3. AINEISTON HANKKIMINEN, TALLENNUS JA TARKISTUS

4. AINEISTON KUVAILU

5. AINEISTON ANALYSOINTI

6. TULOSTEN RAPORTOINTI

**Vaadittavat opintosuoritukset:**

Harjoitustyö

**Arviointi:** hyväksytty / hylätty

**Suositeltavat edeltävät opinnot:** Perusopintojaksot

**Pakolliset edeltävät opinnot:**

**Oppimateriaalit** (huomatkaa e-kirjat ja suositukset)**:** ks. Perusopintojaksot

**YKSITYISKOHTAISEMPIA OHJEITA**

*Tilastotieteen harjoitustyön ensimmäisenä vaiheena on sopia aiheesta ohjaajan kanssa ja laatia kirjallinen* ***yhden sivun mittainen*** *tutkimussuunnitelma. Tutkimussuunnitelmassa tulee esittää kohtien 1 ja 2 asiat. Suunnitelman voi palauttaa sähköisesti.*

# 1. TUTKIMUSONGELMAN MÄÄRITTÄMINEN

Empiirinen tutkimus aloitetaan tutkimusongelman**/**tavoitteiden määrittämisellä. Tutkimuksen tulisi olla sellainen, että tavoitteet voidaan saavuttaa perusopintojaksoilla opituin tilastollisin keinoin.

# 2. SUUNNITELMA ONGELMAN RATKAISEMISEKSI

Ongelman määrittämisen jälkeen mietitään mitä tietoja ongelman ratkaisemiseen tarvitaan ja miten tiedoista muodostetaan tilastollinen havaintoaineisto. Havaintoaineiston hankinnassa on mietittävä, millä tavalla tarvittavien tietojen keruu järjestetään (haastattelemalla, havainnoimalla, mittaamalla...) vai voidaanko tutkimuksessa hyödyntää valmiiksi kerättyä aineistoa. Älä käytä aikasarjatyyppisiä aineistoja, jos niitä ei voi käsitellä normaalin havaintoaineiston tavoin. (Jos havaintojen järjestyksellä ei ole merkitystä analysointiin ja tuloksiin, on aineisto tällöin sopiva). Älä myöskään valitse aineistoksi mitään "valmisaineistoa", ts. aineistoa, jota on jo tilastollisesti tutkittu (esim. SPSS:ssä olevat valmisaineistot). Aineiston lähteet on mainittava. Jotta aineistosta voisi tehdä järkeviä analyyseja, olisi suotavaa **havaintoja** olevan **noin** **100** ja **muuttujia** (kategorisia ja jatkuvia) yhteensä **noin** **10**.

Suunnitelmavaiheessa mietitään myös millaisia ongelmia pyritään ratkaisemaan tilastollisin menetelmin. On muistettava, että useimmat yleisimmistä ja käyttökelpoisimmista menetelmistä vaativat vähintään intervalliasteikollisen mittauksen. Tässä vaiheessa riittää ongelmien ratkaisujen olevan muotoa "*tutkitaan onko sukupuolella merkitystä opintomenestykseen, vaikuttaako ikä pituuteen*" yms. Myöhemmin voidaan päättää, millaisilla testeillä ongelmia ratkaistaan.

Tutkimussuunnitelmasta tulee käydä ilmi:

– mistä/miten aineisto on hankittu

– millainen aineisto on (paljonko havaintoja, millaisia muuttujia ym.)

– millaisiin ongelmiin tutkimus pyrkii löytämään vastauksen

Kyselyä tai haastattelua tehtäessä kysymykset on laadittava siten, että ne mittaavat tarkasti niitä asioita, joita on tarkoitus selvittää. Kyselylomakkeen laadinnassa tärkein periaate on, että vastaaja saadaan vastaamaan kaikkiin kysymyksiin. Yksi mahdollisuus kyselytutkimukselle on tehdä tutkimus jollekin yritykselle (esim. asiakaskysely).

*Hyväksytyn tutkimussuunnitelman jälkeen voidaan aloittaa varsinainen tutkimusvaihe.*

# 3. AINEISTON HANKKIMINEN, TALLENNUS JA TARKISTUS

# 4. AINEISTON KUVAILU

Kuvailu käsittää havaintoaineiston esittelyn, tarvittaessa muuttujien luokittelun ja taulukoinnin, graafiset esitykset ja tunnuslukujen laskemisen. Kuvaileva osa selkeyttää tutkittavaa ongelmaa ja käytettyjä muuttujia, sekä muokkaa havaintomatriisin informaatiota havainnollisempaan muotoon. Pelkkä muuttujien luettelu/ esittely ei riitä, vaan muuttujista tehtyjä kuvia ja/tai taulukoita pitää myös selittää. Tarkoitus on kuvailla pääasiassa muuttujia, joita käytetään jatkoanalyyseissa.

# 5. AINEISTON ANALYSOINTI

Tämän kohdan tavoitteena on varsinaisen tutkimusongelman ratkaiseminen. Tilastollisia menetelmiä on käytettävä oikein. Tutkimusraportista on ilmettävä, että käyttäjä hallitsee käyttämiensä menetelmien perusteet. Harjoitustyössä on käytettävä **kolmea** **eri** **menetelmää** (menetelmät ovat **χ**:n neliötesti, t-testi, varianssi- ja lineaarinen regressioanalyysi), **kutakin vain kerran ja jokaista eri tutkimusongelman ratkaisuun**. **Yhtenä menetelmänä pitää kuitenkin olla lineaarinen regressioanalyysi vähintään kahdella selittäjällä tutkien**. Tärkeää on selittää analyysin tulos ei-tilastollisin termein, siten että lukija ilman tilastotieteen koulutustakin ymmärtää tuloksen.

*Vaiheen 5 jälkeen kirjoitetaan varsinainen tutkimusraportti ja jätetään se tarkistukseen.*

# 6. TULOSTEN RAPORTOINTI

Lopullisessa tutkimusraportissa tulee olla selvitys tutkimuksen keskeisistä vaiheista, yksityiskohtaiset tutkimustulokset ja arvio tulosten luotettavuudesta. Raportin tulee olla selkeä ja havainnollinen. Tuloksien tulee olla luettavissa ja ymmärrettävissä raportista myös sellaiselle henkilölle, jolla ei ole varsinaista tilastotieteen tuntemusta. **Tulokset pitää siis esittää kansantajuisessa muodossa.**

Toisaalta kyse on tilastollisesta raportista, joten päätelmiin johtaneet esim. hypoteesit ja testitulosteet on esitettävä (sopivasti tiivistettynä). Ohjaajan tulee raportin perusteella pystyä tarkistamaan raportissa esitettyjen asioiden oikeellisuus. Mahdollinen kyselylomake sekä varsinaisista **analyyseista saadut ohjelmatulostukset** laitetaan tutkimusraportin **liitteeksi tai muokkaamattomana raportin textin sekaan**. **Raportti palautetaan syksystä 2024 sähköisesti pdf-versiona.** Aineistoa ei tarvitse antaa**.**

Hyvä harjoitustyö koostuu seuraavista osista:

* Johdanto
* Kuvaileva osa
* Analyysiosa
* Yhteenveto tuloksista