

## Harjoitustyö TILM3558

*Työn voi tehdä haluamallaan tilastollisella ohjelmistolla.*

Kirjoita harjoitustyön raportti erilliseen dokumenttiinsa, jonka voit tallentaa esim. word- tai pdf-muodossa. Tulosten tulkintaan kannattaa kiinnittää huomiota, esim. pelkkä huomio siitä, että jokin yhteys on merkitsevä, ei riitä. Tulosteet liitetään erikseen joko yhtenä tai useampana liitteenä. **Työn palautus Moodlen palautuslaatikkoon.**

### **Numeeristen vastemuuttujien mallitus:**

**Elinolo2020** (Tilastokeskuksen elinolotutkimuksen aineisto, N=2199)

*Kukin opiskelija puolestaan poimii 700 kokoisen satunnaisotoksen kyseisestä datasta seuraavalla tavalla:*

**SPSS:llä**

Transform-Random Number Generators...

Set Starting Point-Fixed Value-annetaan oma opiskelijanumero-Ok

DATA>SELECT CASES>RANDOM SAMPLE OF CASES>

EXACTLY **700** CASES FROM THE FIRST **2199** CASES

CONTINUE>

UNSELECTED CASES ARE FILTERED>

## **R:llä**

```
library(foreign)
ht1.dat<-read.spss("elinolo2020.sav", to.data.frame=TRUE)
attach(ht1.dat)

# Suluissa olevan 1:n tilalle oma opiskelijanumero
set.seed(1)

# 700 kokoinen otos
oma.otos1<-ht1.dat[sample(nrow(ht1.dat), 700), ]
attach(oma.otos1)
```

### **1. Varianssianalyysi**

Tutki, onko sukupuolella ja asumisahtaudella yhteyttä asunnon pinta-alaan.

### **2. Regressiomalli**

Tutki, onko kotitalouden kuluttajayksiköiden lukumäärällä, asumismenoilla yhteensä ja alueella asumisajalla yhteyttä asunnon pinta-alaan.

### 3. Toistomittausmalli

**Toistomittausaineisto2020** (7 maasta kerätty aineisto potilaan ohjauksesta, N=1299)

*Kukin opiskelija puolestaan poimii 600 kokoisen satunnaisotoksen kyseisestä datasta seuraavalla tavalla:*

**SPSS:llä**

Transform-Random Number Generators...

Set Starting Point-Fixed Value-annetaan oma opiskelijanumero-Ok

DATA>SELECT CASES>RANDOM SAMPLE OF CASES>

EXACTLY **600** CASES FROM THE FIRST **1299** CASES

CONTINUE>

UNSELECTED CASES ARE FILTERED>

**R:llä**

```
library(foreign)
```

```
ht2.dat<-read.spss("Toistomittausaineisto2020.sav", to.data.frame=TRUE)
```

```
attach(ht2.dat)
```

```
# Suluissa olevan 1:n tilalle oma opiskelijanumero
```

```
set.seed(1)
```

```
# 600 kokoinen otos
```

```
oma.otos2<-ht2.dat[sample(nrow(ht2.dat), 600), ]
```

```
attach(oma.otos2)
```

*Tutkijalla on hypoteesi, että potilaan mielestä saatu ohjaus leikkauksen jälkeen toiminnallista seikoista (Functional\_M2) on ollut vähäisempää kuin odotettu ennen leikkausta (Functional\_M1). Eli keskiarvo toisessa mittauksessa on matalampi. Lisäksi kiinnostaa se, onko tuo ero mittausten välillä erilainen sukupuolittain.*

Tutki saavatko nämä tutkimushypoteesit tukea mallittamalla aineisto toistettujen mittausten varianssianalyysillä.

*Huom. Numeeristen vastemuuttujien mallituksessa on varianssianalyysien osalta tehtävä tarvittava kuvaileva tarkastelu ja regressiomallissa yhteyksien suoraviivaisuuksien tarkastelu ja jäännöstarkastelu.*

## Kategoristen vastemuuttujien mallitus:

**EK2011** (eduskuntavaaliaineisto vuodelta 2011, N=1318)

*Kukin opiskelija puolestaan poimii 900 kokoisen satunnaisotoksen kyseisestä datasta seuraavalla tavalla:*

### SPSS:llä

Transform-Random Number Generators...

Set Starting Point-Fixed Value-annetaan oma opiskelijanumero-Ok

```
DATA>SELECT CASES>RANDOM SAMPLE OF CASES>  
EXACTLY 900 CASES FROM THE FIRST 1318 CASES  
CONTINUE>  
UNSELECTED CASES ARE FILTERED>
```

### R:llä

```
library(foreign)
```

```
ht3.dat<-read.spss("EK2011.sav", to.data.frame=TRUE)
```

```
attach(ht3.dat)
```

```
# Suluissa olevan 1:n tilalle oma opiskelijanumero
```

```
set.seed(1)
```

```
# 900 kokoinen otos
```

```
oma.otos3<-ht3.dat[sample(nrow(ht3.dat), 900), ]
```

```
attach(oma.otos3)
```

## Muuttujien riippuvuus rakenne

1. Tarkastellaan muuttujia sukupuoli (d2), työttömyys viimeisen 12 kuukauden aikana (k32) ja oman sukupuolen 2011 eduskuntavaaleissa äänestäminen (k23). Tee ensin yksiulotteiset frekvenssijakaumat ja kolmen muuttujan ristiintaulu.

Onko 3-ulotteisessa ristiintaulussa nollasoluja ?

2. Tarkastele kolmen muuttujan välisiä riippuvuuksia loglineaaristen mallien avulla. Ota mukaan muuttujista vain ne luokat, joissa havaintoja on yli 10.

Millaiset riippuvuudet muuttujien välillä askeltavan menetelmän avulla valittuun malliin jäivät ?

Mikä on mallin generoiva luokka?

Mikä on mallin yhteensopivuustestin p-arvo ?

Mikä on standardoitujen jäännösten vaihteluväli ?

3. Tee mallin mukainen yhteyksien jatkotarkastelu ristiintauluin ja tulkitse malli riviprosenttien avulla.

## Kaksiluokkainen selitettävä muuttuja

1. Tutki muuttujien sukupuoli (d2) ja ikä yhteyttä työttömyyteen viimeisen 12 kuukauden aikana (k32) käyttämällä logistista regressiomallia.

2. Mitkä muuttujat selittävät työttömyyttä ?

Tulkitse yhteydet OR:ien avulla. Raportoi myös luottamusvälit OR:ille.

Mikä on mallin Nagelkerke selitysaste ?

## Monimuuttujamenetelmät:

**pankkiotos2020 (todellinen asiakasaineisto, N=2453)**

*Kukin opiskelija puolestaan poimii 1500 kokoisen satunnaisotoksen kyseisestä datasta seuraavalla tavalla:*

### SPSS:llä

Transform-Random Number Generators...

Set Starting Point-Fixed Value-annetaan oma opiskelijanumero-Ok

```
DATA>SELECT CASES>RANDOM SAMPLE OF CASES>  
EXACTLY 1500 CASES FROM THE FIRST 2453 CASES  
CONTINUE>  
UNSELECTED CASES ARE FILTERED>
```

### R:llä

```
library(foreign)
```

```
ht4.dat<-read.spss("pankkiotos2020.sav", to.data.frame=TRUE)
```

```
attach(ht4.dat)
```

```
# Suluissa olevan 1:n tilalle oma opiskelijanumero
```

```
set.seed(1)
```

```
# 1500 kokoinen otos
```

```
oma.otos4<-ht4.dat[sample(nrow(ht4.dat), 1500), ]
```

```
attach(oma.otos4)
```

## **Muuttujien ryhmittely**

1. Muodosta pääkomponenttianalyysillä luokitelluista muuttujista (41 kpl :  
autom\_lainan\_perinta\_luok - kulutusluotot1\_luok) pääkomponentteja ominaisarvokriteerin  
mukaan. ( Promax-rotaatio)
2. Talleta havaintomatriisiin uusiksi muuttujiksi pääkomponenttipistemäärät.
3. Nimeä uudet muuttujat (pääkomponentteihin latautuneiden muuttujien mukaisesti).

## **Havaintojen ryhmittely**

4. Käytä näitä uusia muuttujia klusterianalyysissä, jossa muodostat asiakasryhmiä K-means  
menetelmällä lähtien kahdesta klusterista viiteen tai kuuteen klusteriin saakka. Kuvaile  
muodostamiasi ryhmiä.