

# Tilastotieteen harjoitustyö 2018

Samuli Virtanen

spevir@utu.fi

op. nro. 511178

## 1. Numeeristen vastemuuttujien mallinnus

Elinolo2018 (Tilastokeskuksen elinolotutkimuksen aineisto, N=2199)

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

```
Transform-Random Number Generators...
Set Starting Point-Fixed Value-annetaan oma opiskelijanumero-0k
DATA>SELECT CASES>RANDOM SAMPLE OF CASES>
EXACTLY 600 CASES FROM THE FIRST 2199 CASES
CONTINUE
UNSELECTED CASES ARE FILTERED>
```

### 1.1. Varianssianalyysi

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

#### 1.1.1. Suunnitelma

Kyseessä on kaksi kategorista selittävää muuttujaa, sekä yksi numeerinen selitettävä muuttuja, joten käytetään kaksisuuntaista varianssianalyysiä.

#### 1.1.2. Normaalijakaumaoletus

Ensimmäiseksi tutkitaan jakaumia sukupuolittain asumisahtauden perusteella.

Jaetaan data kahteen osaan sukupuolen perusteella.

```
SORT CASES BY supu.
```

```
SPLIT FILE SEPARATE BY supu.
```

Luodaan tarvittavat kuviot ja tiedot SPSS:n Explore-työkalulla.

```
EXAMINE VARIABLES=pala BY ahtas
/PLOT BOXPLOT NPLOT
/COMPARE GROUPS
/STATISTICS DESCRIPTIVES
```

```
/CINTERVAL 95  
/MISSING LISTWISE  
/NOTOTAL.
```

Tulostedokumentista löytyvän Shapiro-Wilk-testin tulosten perusteella neljä ryhmää ei ole normaalisti jakautunut:

- mies \* normaali
- mies \* tilava
- nainen \* normaali
- nainen \* tilava

Näiden ryhmien havaintojen suuresta määrästä johtuen voidaan kuitenkin käyttää parametristä testiä. Myöskin visuaalinen tarkastelu osoittaa ryhmien olevan jokseenkin normaalijakautuneita.

### 1.1.3. Kaksisuuntainen varianssianalyysi

Tehdään kaksisuuntainen varianssianalyysi SPSS:n Univariate-työkalulla.

```
SPLIT FILE OFF.
```

```
UNIANOVA pala BY supu ahtas  
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/PLOT=PROFILE(ahtas*supu)  
/PRINT=DESCRIPTIVE HOMOGENEITY  
/CRITERIA=ALPHA(.05)  
/DESIGN=supu ahtas supu*ahtas.
```

Tuloksista nähdään että selittävien muuttujien päävaikutukset ovat merkitseviä merkitsevyystasolla  $p < 0,05$ , mutta yhdysvaikutus ei ole merkitsevä. ( $p \sim 0.638$ )

Myöskään hajontojen yhtäsuuruusoletus ei näytä pitävän paikkaansa (Levenen testin  $p$ -arvo  $\sim 0.014$ )

Voidaan tulkita että miehillä on keskimäärin suuremmat asunnot kuin naisilla, ja asunnon pinta-ala kasvaa koetun asumisahtauden mukaan järjestyksessä ahdas -> normaali -> tilava (pienestä pinta-alasta suurempaan). Koska yhdysvaikutuksella ei ole merkitystä, voidaan tulkita että sukupuolella ei ole väliä asumisahtauden eroissa.

### 1.1.4. Monivertailut

Tutkitaan vielä monivertailut estimoiduilla keskiarvoilla.

```
UNIANOVA pala BY supu ahtas  
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE
```

```

/PLOT=PROFILE(ahtas*supu)
/EMMEANS=TABLES(supu) COMPARE ADJ(SIDAK)
/EMMEANS=TABLES(ahtas) COMPARE ADJ(SIDAK)
/EMMEANS=TABLES(supu*ahtas)
/PRINT=ETASQ DESCRIPTIVE HOMOGENEITY
/CRITERIA=ALPHA(.05)
/DESIGN=supu ahtas supu*ahtas.

```

Tulostedokumentista löytyvän *Pairwise comparisons* -taulukon perusteella voidaan todeta kaikkien asumisahtaustasojen välillä olevan tilastollisesti merkittäviä eroja. ( $p < 0.001$ )

## 1.2. Regressiomalli

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

*Ratkaisu: Sovitetaan regressiomallia, jossa selittäjinä ovat on kuluttajayksiköiden lukumäärä, asumismenot yhteensä sekä alueella asumisaika. Selitettävänä muuttujana on asunnon pinta-ala.*

Tutkitaan sirontakuvioita, joka löytyy tulostedokumentista.

GRAPH

```

/SCATTERPLOT(MATRIX)=rkyks asmenot alaika pala
/MISSING=LISTWISE.

```

Alueella asumisaika näyttäisi kasvaessaan vaikuttavan pienentäväsi asumismenoihin. Muiden selittävien muuttujien välistä korrelaatiota ei visuaalisesti tarkasteltuna esiinny, jokaisen selittävän muuttujan välinen sirontakuviokuva on siis suunnilleen ellipsin muotoinen.

Kovin selkeitä korrelaatioita ei esiinny. Näyttäisi kuitenkin hieman siltä että kodin kuluttajayksiköiden määrän kasvaessa myös asunnon pinta-ala kasvaa ja toisaalta samoin käy myös asumismenojen kasvaessa. Alueella asumisaika ei sen sijaan vaikuta vaikuttavan asunnon pinta-alaan.

Lasketaan Pearsonin ja Spearmanin korrelaatiokertoimet.

CORRELATIONS

```

/VARIABLES=alaika asmenot rkyks pala
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE.

```

NONPAR CORR

```

/VARIABLES=alaika asmenot rkyks pala
/PRINT=SPEARMAN TWOTAIL NOSIG
/MISSING=PAIRWISE.

```

Tulostedokumentin tuloksista huomataan, että alueella asumisajan ja asumismenojen yhteensä välillä on merkittävä (Pearsonin korrelaatiokerroin  $\sim 0.29$ ) yhteys. Huomataan myös että alueella asumisajan ja asunnon pinta-alan välillä sekä kotitalouden kuluttajayksiköiden määrän ja pinta-alan välillä on vielä merkittävämpi yhteys.

Vastaavat Spearmanin korrelaatiokertoimet ovat hieman korkeampia kuin Pearsonin, mutta eivät paljon. Suhde ei siis ole täydellisen lineaarinen, mutta ei epälineaarinenkaan.

Luodaan useamman selittäjän lineaarinen malli ja lasketaan sille kertoimet.

```
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT pala
/METHOD=ENTER alaika asmenot rkyks
/SAVE ZRESID.
```

Mallin kaavaksi saadaan  $23.58 + 24.83 * rkyks + 0.364 * alaika$ . Selitysaste on vaatimattomat 27,3%. Asumismenojen kerroin ja p-arvo on  $< 0.001$  joten se jätetään mallista pois.

### 1.3. Toistomittausmalli

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.