

Tilastollisten mallien peruskurssi – H1

1.

	id	supu	ika	koulutus	siviilis	työtilan	omistusa	oma_tulo	kotittul
1	1	mies	22	kansakoulu/peruskoulu	naimaton	työton	on	575	658
2	2	mies	32	korkeakoulututkinto	avo/avioliitossa	työssä	ei	2579	4242
3	3	mies	52	koulututkinto	naimaton	työton	on	1012	1240
4	4	nainen	44	koulututkinto	naimaton	työssä	on	987	1076
5	5	mies	20	lukio	avo/avioliitossa	työton	on	799	2545
6	6	nainen	32	koulututkinto	avo/avioliitossa	työton	on	793	2428

tvsaasto	evsaasto	asuntol	muulaina	yltyyt	taltyyt	ihmtyyt
1250	3000	46757	632	melko tyytymaton	melko tyytymaton	melko tyytymaton
5188	4341	0	93554	melko tyytyväinen	melko tyytyväinen	erittäin tyytyväinen
14168	19008	20313	6667	melko tyytymaton	erittäin tyytymaton	melko tyytymaton
9580	11959	44717	5191	melko tyytyväinen	melko tyytyväinen	melko tyytyväinen
2463	5519	85232	2045	erittäin tyytyväinen	melko tyytyväinen	erittäin tyytyväinen
2655	5000	8926	17356	melko tyytyväinen	melko tyytymaton	melko tyytyväinen

2.

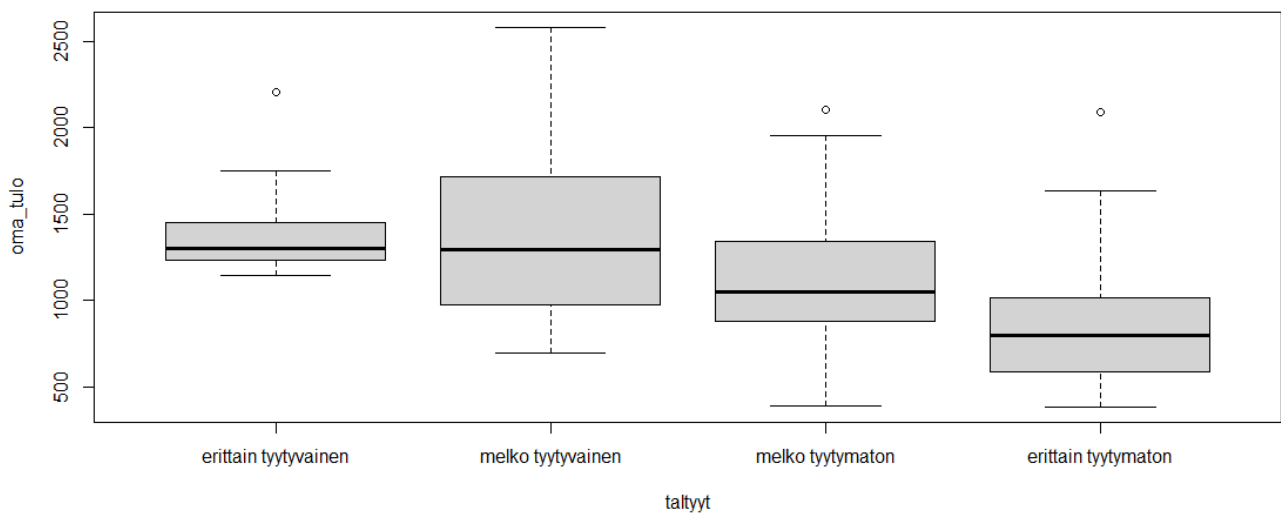
```
> # kuvailevat tunnusluvut ryhmittäin
> with(dat, tapply(oma_tulo, taltyyt, summary))
$`erittäin tyytyväinen`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1145   1235   1302   1406   1454   2209

$`melko tyytyväinen`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 697.0   978.2  1295.0  1355.0  1718.2  2579.0

$`melko tyytymaton`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 388.0   880.2  1051.0  1106.8  1333.8  2102.0

$`erittäin tyytymaton`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 385.0   587.2   796.5   875.2  1006.8  2093.0
```

tyytyväisyys taloudelliseen tilanteeseen



```

> # Normaalijakaumatestaus
> with(dat, tapply(oma_tulo, taltyyt, shapiro.test))
$`erittain tyytyväinen`

      Shapiro-Wilk normality test

data:  x[[i]]
W = 0.79511, p-value = 0.003184

$`melko tyytyväinen`

      Shapiro-Wilk normality test

data:  x[[i]]
W = 0.94832, p-value = 0.002051

$`melko tyytymätön`

      Shapiro-Wilk normality test

data:  x[[i]]
W = 0.95938, p-value = 0.07354

$`erittain tyytymätön`

      Shapiro-Wilk normality test

data:  x[[i]]
W = 0.89492, p-value = 0.0003278

```

Ainoastaan melko tyytymättömien p-arvo ylittää 0.05, joten kaikki muut poikkeavat normaalista.

3.

```

> # hajontojen yhtäsuuruustestaus
> bartlett.test(oma_tulo~taltyyt)

      Bartlett test of homogeneity of variances

data:  oma_tulo by taltyyt
Bartlett's K-squared = 5.9084, df = 3, p-value = 0.1162

```

Bartlettin testin p-arvo = 0.1162, joka on merkitsevästi > 0.05. Tulkitaan, että testin perusteella populaatiohajontojen välillä ei ole tilastollisesti merkitseviä eroja.

Käytetään yhtä suurien hajontojen varianssianalyysiä.

```
> # Yksisuuntainen varianssianalyysi, hajonnat ovat yhtäsuuret
> anova(lm(oma_tulo~taltyyt))
Analysis of Variance Table

Response: oma_tulo
      Df Sum Sq Mean Sq F value    Pr(>F)
taltyyt   3 8254276 2751425  18.177 1.876e-10 ***
Residuals 197 29820026  151371
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testin perusteella voidaan ajatella, että taloudellisen tyytyväisyyden luokilla on tilastollisesti merkitseviä eroja ja p-arvo < 0.001.

```
> # monivertailu bonferroni-korjauksella, hajonnat yhtäsuuret
> pairwise.t.test(oma_tulo,taltyyt, p.adj="bonferroni")

Pairwise comparisons using t tests with pooled SD

data: oma_tulo and taltyyt

      erittäin tyytyväinen melko tyytyväinen melko tyytymätön
melko tyytyväinen    1.0000                -                -
melko tyytymätön    0.0568                0.0023                -
erittäin tyytymätön 4.0e-05                4.1e-10               0.0180

P value adjustment method: bonferroni
```

Erittäin tyytyväiset vs melko tyytyväiset $p = 1.0000$ ei ole merkitsevä

Erittäin tyytyväiset vs melko tyytymättömät $p = 0.0569$ ei ole merkitsevä

Erittäin tyytyväiset vs erittäin tyytymättömät $p < 0.001$ on merkitsevä

Melko tyytyväiset vs melko tyytymättömät $p = 0.0023$ on merkitsevä

Melko tyytyväiset vs erittäin tyytymättömät $p < 0.001$ on merkitsevä

Melko tyytymättömät vs erittäin tyytymättömät $p = 0.0180$ on merkitsevä

4.

```
> # Epäparametrinen yksisuuntainen malli sijaintien erolle
> kruskal.test(oma_tulo~taltyyt)

      kruskal-wallis rank sum test

data:  oma_tulo by taltyyt
kruskal-wallis chi-squared = 47.432, df = 3, p-value = 2.813e-10
```

$p < 0.001$, joten tyytyväisyysluokkien välillä on tilastollisesti merkitseviä eroja keskimääräisissä sijainneissa

```
> # U-testit holm-korjauksella
> pairwise.wilcox.test(oma_tulo,taltyyt)

      Pairwise comparisons using wilcoxon rank sum test with continuity correction

data:  oma_tulo and taltyyt

melko tyytyväinen      erittäin tyytyväinen melko tyytyväinen melko tyytymätön
melko tyytyväinen      0.4702                -                -
melko tyytymätön      0.0046                0.0062                -
erittäin tyytymätön 2.9e-05                6.4e-09                0.0014

P value adjustment method: holm
```

Erittäin tyytyväiset vs melko tyytyväiset korjattu $p=0.4702$ ei merkitsevää eroa

Erittäin tyytyväiset vs melko tyytymättömät korjattu $p=0.0046$ on merkitsevä ero

Erittäin tyytyväiset vs erittäin tyytymättömät korjattu $p < 0.001$ on merkitsevä ero

Melko tyytyväiset vs melko tyytymättömät korjattu $p=0.0062$ on merkitsevä ero

Melko tyytyväiset vs erittäin tyytymättömät korjattu $p < 0.001$ on merkitsevä ero

Melko tyytymättömät vs erittäin tyytymättömät korjattu $p=0.0014$ on merkitsevä ero

Tuottivatko normaalijakautumaan perustuva ja epäparametrinen saman tuloksen? Kummanko tulokset raportoisit?

Ei tuottanut. Erittäin tyytyväisten ja melko tyytymättömien merkitsevyys muuttui. Parametrisen menetelmän käyttö on suositeltavaa.

5.

```
> #yksinkertaiset kontrastit, LSD
> pairwise.t.test(oma_tulo,taltyyt, p.adj="none")

      Pairwise comparisons using t tests with pooled SD

data:  oma_tulo and taltyyt

      erittäin tyytyväinen melko tyytyväinen melko tyytymaton
melko tyytyväinen    0.64266                -                -
melko tyytymaton    0.00946                0.00038            -
erittäin tyytymaton 6.6e-06                6.8e-11            0.00299

P value adjustment method: none
```

Erittäin tyytymättömät vs erittäin tyytyväiset $p < 0.001$ on merkitsevä

Erittäin tyytymättömät vs melko tyytyväiset $p < 0.001$ on merkitsevä

Erittäin tyytymättömät vs melko tyytymättömät $p = 0.003$ on merkitsevä