

Tilastollisten mallien peruskurssi – H2

2a.

```
> leveneTest(oma_tulo~taltyyt)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  3.2282 0.02357 *
      197
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> # Yksisuuntainen varianssianalyysi, hajonnat ovat yhtäsuuret
>
> malli1<-lm(oma_tulo~taltyyt)
> malli_lm<-aov(malli1)
> summary (malli_lm)
      Df    Sum Sq Mean Sq F value    Pr(>F)
taltyyt    3  8254276  2751425   18.18 1.88e-10 ***
Residuals 197 29820026   151371
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Levenen testin p-arvo=0.024, on tilastollisesti merkitseviä eroja

F-testin p-arvo<0.001, on tilastollisesti merkitseviä eroja

```
> #Tukey HSD-testi
>
> tukey.test <- TukeyHSD(malli_lm)
> tukey.test
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = malli1)

$taltyyt
              diff            lwr            upr      p adj
melko tyytyväinen-erittäin tyytyväinen -50.67857 -333.2553  231.898176 0.9666501
melko tyytymaton-erittäin tyytyväinen -298.82051 -594.2775  -3.363572 0.0463104
erittäin tyytymaton-erittäin tyytyväinen -530.44667 -827.2236 -233.669733 0.0000389
melko tyytymaton-melko tyytyväinen -248.14194 -426.0240  -70.259846 0.0021479
erittäin tyytymaton-melko tyytyväinen -479.76810 -659.8342 -299.702040 0.0000000
erittäin tyytymaton-melko tyytymaton -231.62615 -431.2983  -31.953997 0.0157221
```

Erittäin tyytyväiset vs melko tyytyväiset p=0.967, ei ole tilastollisesti merkitsevää eroa

Erittäin tyytyväiset vs melko tyytymättömät p=0.046, on tilastollisesti merkitsevä ero

Erittäin tyytyväiset vs erittäin tyytymättömät p<0.001, on tilastollisesti merkitsevä ero

Melko tyytyväiset vs melko tyytymättömät p=0.002, on tilastollisesti merkitsevä ero

Melko tyytyväiset vs erittäin tyytymättömät p<0.001, on tilastollisesti merkitsevä ero

Melko tyytymättömät vs erittäin tyytymättömät p=0.016, on tilastollisesti merkitsevä ero

2b.

Kuvailevia tunnuslukuja:

```
> # keskiarvot soluittain
>
> with(dat,tapply(oma_tulo,list(tyotilan,taltyyt),mean))
      erittain tyytyvainen melko tyytyvainen melko tyytymaton erittain tyytymaton
tyoton      1201.500      999.5769      946.3667      825.3488
tyossa      1437.077      1514.3103      1325.6818      1181.5714
> # mediaanit soluittain
>
> with(dat,tapply(oma_tulo,list(tyotilan,taltyyt),median))
      erittain tyytyvainen melko tyytyvainen melko tyytymaton erittain tyytymaton
tyoton      1201.5      935.5      945      730
tyossa      1326.0      1468.0      1292      1097
> # keskihajonnat soluittain
>
> with(dat,tapply(oma_tulo,list(tyotilan,taltyyt),sd))
      erittain tyytyvainen melko tyytyvainen melko tyytymaton erittain tyytymaton
tyoton      79.90307      255.1887      275.4617      351.3359
tyossa      288.84842      406.4278      355.6563      217.9532
```

Työssäkäyvien normaalijakaumatesti Shapiro-Wilkin testillä:

```
> raj.ain1 <- select(filter(dat, tyotilan=="tyossa"),c(tyotilan,taltyyt,oma_tulo))
>
> with(raj.ain1,tapply(oma_tulo,list(taltyyt),shapiro.test))
$`erittain tyytyvainen`

      shapiro-wilk normality test

data:  x[[i]]
W = 0.81151, p-value = 0.00937

$`melko tyytyvainen`

      shapiro-wilk normality test

data:  x[[i]]
W = 0.9753, p-value = 0.2825

$`melko tyytymaton`

      shapiro-wilk normality test

data:  x[[i]]
W = 0.92419, p-value = 0.09287

$`erittain tyytymaton`

      shapiro-wilk normality test

data:  x[[i]]
W = 0.90218, p-value = 0.3444
```

Työttömien normaalijakaumatesti Shapiro-Wilkin testillä:

```
> raj.ain2 <- select(filter(dat, tyotilan=="tyoton" & taltyyt!="erittain tyytyväinen"),c(tyotilan,taltyyt,oma_tulo))
>
> with(raj.ain2,tapply(oma_tulo,list(taltyyt),shapiro.test))
$`erittain tyytyväinen`
NULL

$`melko tyytyväinen`
      shapiro-wilk normality test

data:  x[[i]]
W = 0.83256, p-value = 0.0006651

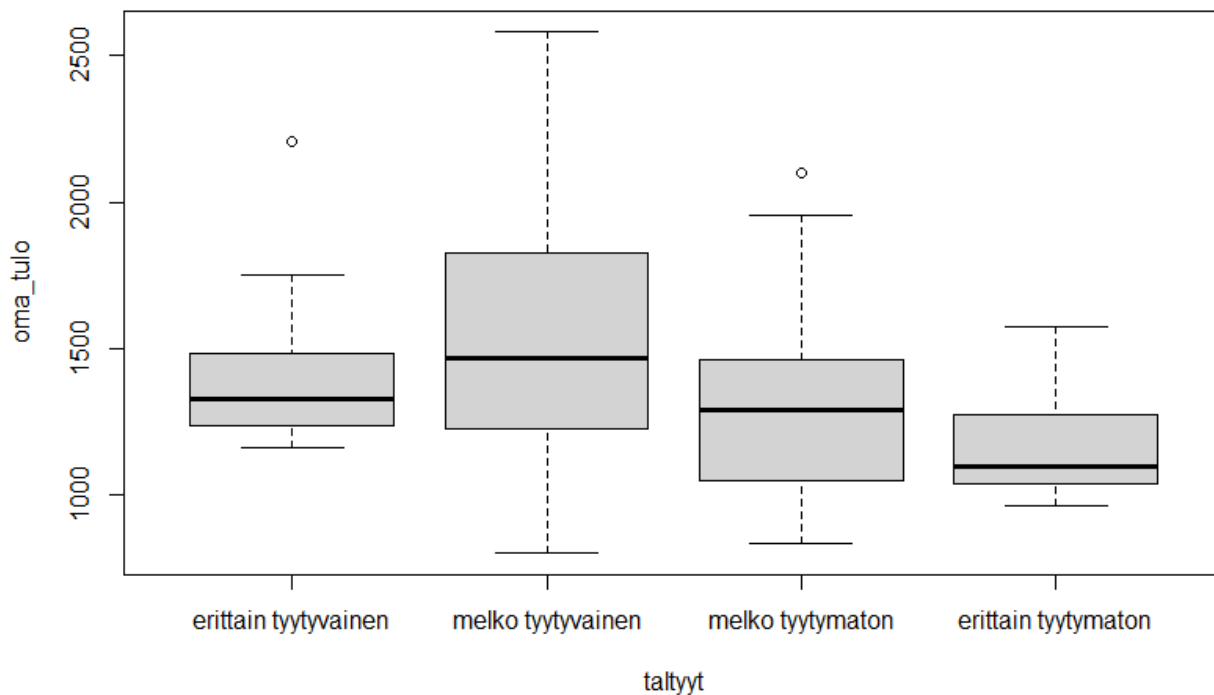
$`melko tyytymaton`
      shapiro-wilk normality test

data:  x[[i]]
W = 0.96417, p-value = 0.3939

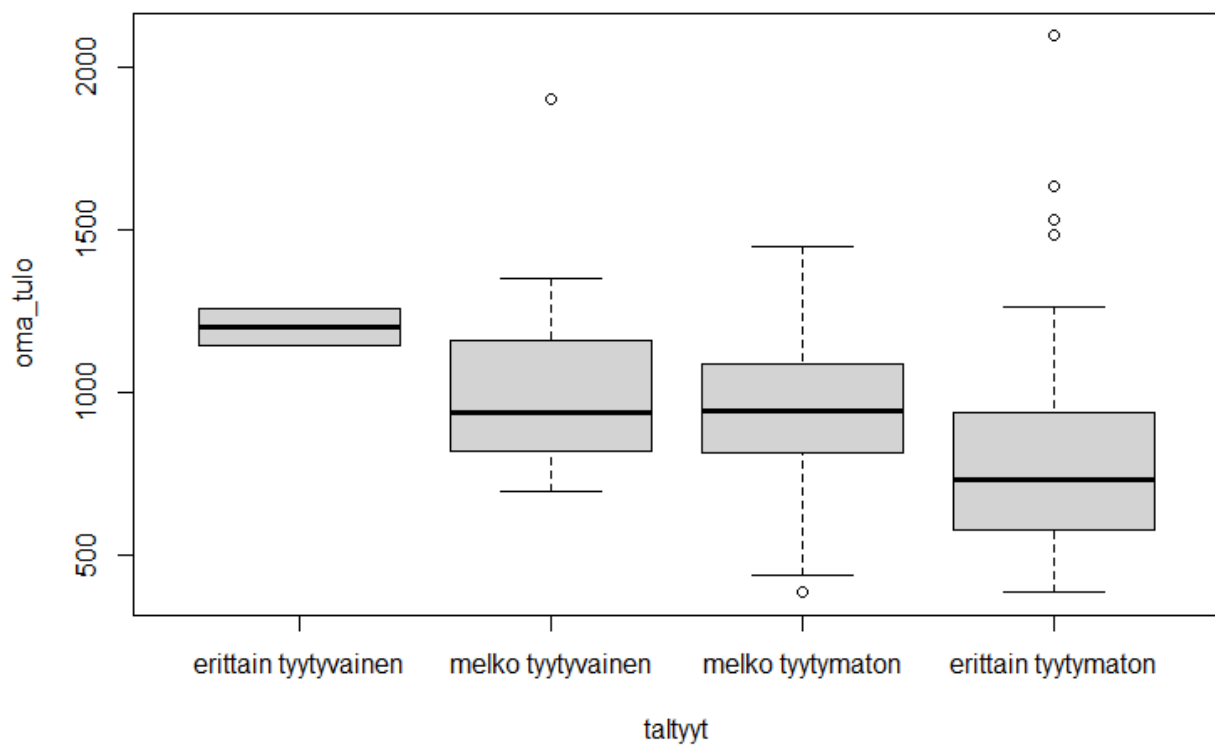
$`erittain tyytymaton`
      shapiro-wilk normality test

data:  x[[i]]
W = 0.83702, p-value = 2.501e-05
```

Työssäkäyvien laatikko-janakuvio:



Työttömien laatikko-janakuvio:



3.

Hajontojen yhtäsuuruustestaus:

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  7  2.2532 0.03171 *
      193
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Levenen testin p-arvo=0.032, eli yhtäsuuruusoletus ei ole voimassa

```
> #kaksisuuntainen varianssianalyysi
>
> anova(lm(dat$soma_tulo~taltyyt*tyotilan))
Analysis of Variance Table

Response: dat$soma_tulo
      Df  Sum Sq Mean Sq F value    Pr(>F)
taltyyt    3  8254276 2751425 23.7306 4.061e-13 ***
tyotilan    1  7183216 7183216 61.9540 2.465e-13 ***
taltyyt:tyotilan  3   259544   86515  0.7462  0.5258
Residuals 193 22377265 115944
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Onko taltyyt tilastollisesti merkitsevä selittäjä? (F-testin $p < 0.001$) On

Onko tyotilan tilastollisesti merkitsevä selittäjä? (F-testin $p < 0.001$) On

Onko yhdysvaikutus tilastollisesti merkitsevä? (F-testin $p = 0.526$) Ei ole

4.

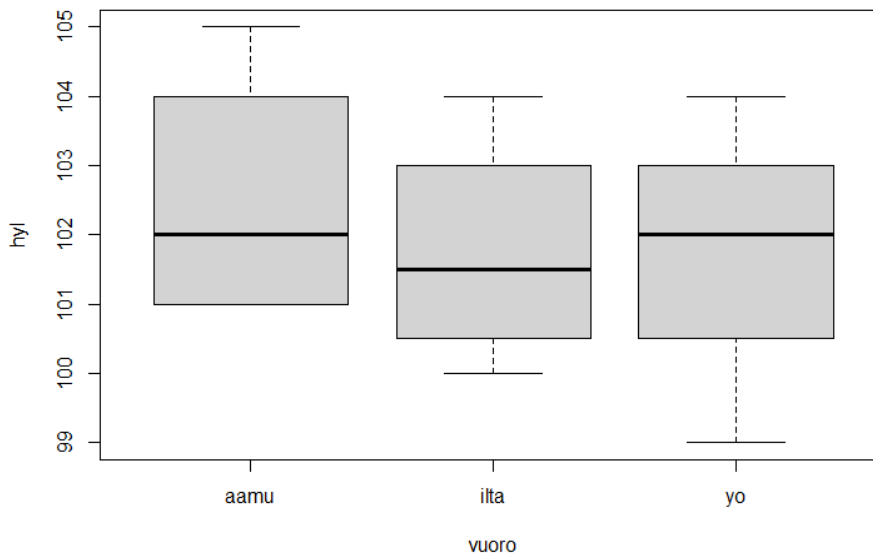
```
> # keskiarvot soluittain
>
> with(dat2, tapply(hyl, list(tehdas, vuoro), mean))
      aamu   ilta   yo
tehdas1 102.5 101.75 101.75
tehdas2 100.5 109.50 117.25
```

Tehdas 2:n ilta- ja yövuoroissa jotakin häikkää.

```
> #Shapiro-wilk -testit, tehdas1
>
> tehdas1.dat <- select(filter(dat2, tehdas=="tehdas1"), c(tehdas, vuoro, hyl))
>
> with(tehdas1.dat, tapply(hyl, list(vuoro), shapiro.test))
$aamu
      Shapiro-Wilk normality test
data:  X[[i]]
W = 0.86337, p-value = 0.2725

$ilta
      Shapiro-Wilk normality test
data:  X[[i]]
W = 0.97137, p-value = 0.85

$yo
      Shapiro-Wilk normality test
data:  X[[i]]
W = 0.92614, p-value = 0.5719
```



Normaalijakaumaoletus on voimassa

```

> #shapiro-wilk -testit,tehdas2
>
> tehdas2.dat <- select(filter(dat2, tehdas=="tehdas2"),c(tehdas,vuoro,hyl))
>
> with(tehdas2.dat,tapply(hyl,list(vuoro),shapiro.test))
$aamu

      Shapiro-Wilk normality test

data:  x[[i]]
W = 0.99291, p-value = 0.9719

$ilta

      Shapiro-Wilk normality test

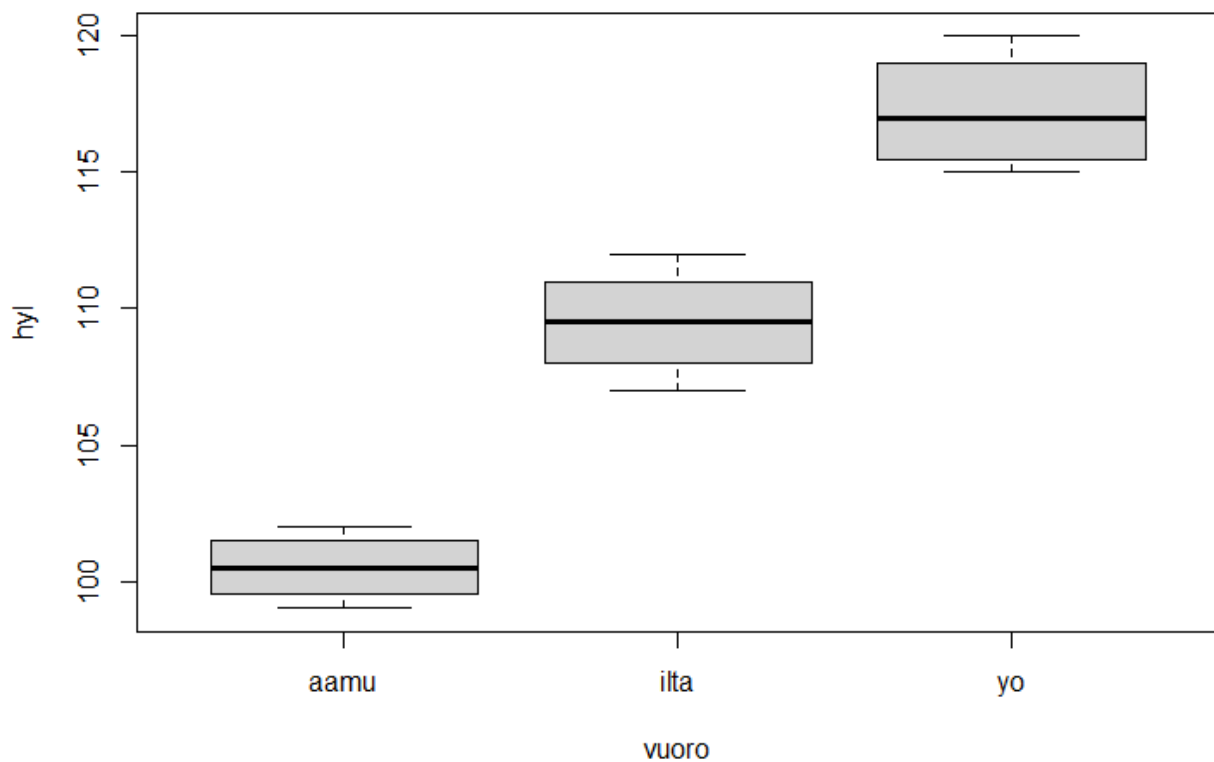
data:  x[[i]]
W = 0.9984, p-value = 0.9951

$yo

      Shapiro-Wilk normality test

data:  x[[i]]
W = 0.96307, p-value = 0.7982

```



Normaalijakaumaoletus on voimassa

5.

Hajontojen yhtäsuuruustestaus Levenen testillä, $p=0.937$ eli yhtäsuuruusoletus on voimassa

```
> levenetest(hyl~tehdas*vuoro)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  5  0.2444 0.9372
      18
```

```
> #Kaksisuuntainen varianssianalyysi
>
> anova(lm(hyl~tehdas*vuoro))
Analysis of Variance Table

Response: hyl
      Df Sum Sq Mean Sq F value    Pr(>F)
tehdas   1  301.04   301.042    83.046 3.656e-08 ***
vuoro     2  256.08   128.042    35.322 5.870e-07 ***
tehdas:vuoro 2  307.58   153.792    42.425 1.540e-07 ***
Residuals 18   65.25     3.625
```

Onko tehdas tilastollisesti merkitsevä selittäjä? (F-testin $p<0.001$) [On](#)

Onko vuoro tilastollisesti merkitsevä selittäjä? (F-testin $p<0.001$) [On](#)

Onko yhdysvaikutus tilastollisesti merkitsevä? (F-testin $p<0.001$) [On](#)

6.

Tukeyn testin vuorojen väliset vertailut:

\$`tehdas:vuoro`	diff	lwr	upr	p adj
tehdas2:aamu-tehdas1:aamu	-2.00	-6.27856	2.27856	0.6771101
tehdas1:ilta-tehdas1:aamu	-0.75	-5.02856	3.52856	0.9926410
tehdas2:ilta-tehdas1:aamu	7.00	2.72144	11.27856	0.0007328
tehdas1:yo-tehdas1:aamu	-0.75	-5.02856	3.52856	0.9926410
tehdas2:yo-tehdas1:aamu	14.75	10.47144	19.02856	0.0000000
tehdas1:ilta-tehdas2:aamu	1.25	-3.02856	5.52856	0.9338558
tehdas2:ilta-tehdas2:aamu	9.00	4.72144	13.27856	0.0000367
tehdas1:yo-tehdas2:aamu	1.25	-3.02856	5.52856	0.9338558
tehdas2:yo-tehdas2:aamu	16.75	12.47144	21.02856	0.0000000
tehdas2:ilta-tehdas1:ilta	7.75	3.47144	12.02856	0.0002320
tehdas1:yo-tehdas1:ilta	0.00	-4.27856	4.27856	1.0000000
tehdas2:yo-tehdas1:ilta	15.50	11.22144	19.77856	0.0000000
tehdas1:yo-tehdas2:ilta	-7.75	-12.02856	-3.47144	0.0002320
tehdas2:yo-tehdas2:ilta	7.75	3.47144	12.02856	0.0002320
tehdas2:yo-tehdas1:yo	15.50	11.22144	19.77856	0.0000000

Tehdas 1 aamuvuoro vs iltavuoro $p=0.993$, ei ole merkitsevä

Tehdas 1 aamuvuoro vs yövuoro $p=0.992$, ei ole merkitsevä

Tehdas 1 iltavuoro vs yövuoro $p=1.000$, ei ole merkitsevä

Tehdas 2 aamuvuoro vs iltavuoro $p<0.001$, on merkitsevä

Tehdas 2 aamuvuoro vs yövuoro $p<0.001$, on merkitsevä

Tehdas 2 iltavuoro vs yövuoro $p<0.001$, on merkitsevä