

Aflevering 9

Lucas Bagge

I artiklen Effects of user age on smartphone and tablet use, measured with an eye-tracker via fixation duration, scan-path duration, and saccades proportion studeres, hvordan brugen af smartphoner og tablets varierer mellem forskellige aldersgrupper. Under brugen følges en persons øjen-bevægelse, og herudfra dannes et mål SPD (scan-path duration, målt i millisekunder), der afspejler en persons evne til at bruge redskabet. I artiklen siges der: “SPD measures global processing of interfaces, where longer SPD indicates less efficient scanning and browsing”. Personer deles op i tre aldersgrupper: unge, midaldrende og ældre. Desuden fordeles personerne på to eksperimenter (Ex1 og Ex2). Hvert eksperiment består af ni opgaver inden for brugen af forskellige smartphone apps, og opgaven hørende til en app er forskellig mellem de to eksperimenter. Data er i filen Smartphone.csv der har tre søjler med henholdsvis eksperiment, aldersgruppe og SPD-målingen

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.3      v purrr 0.3.4
## v tibble 3.1.0       v dplyr 1.0.5
## v tidyr 1.1.3        v stringr 1.4.0
## v readr 1.4.0        v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()

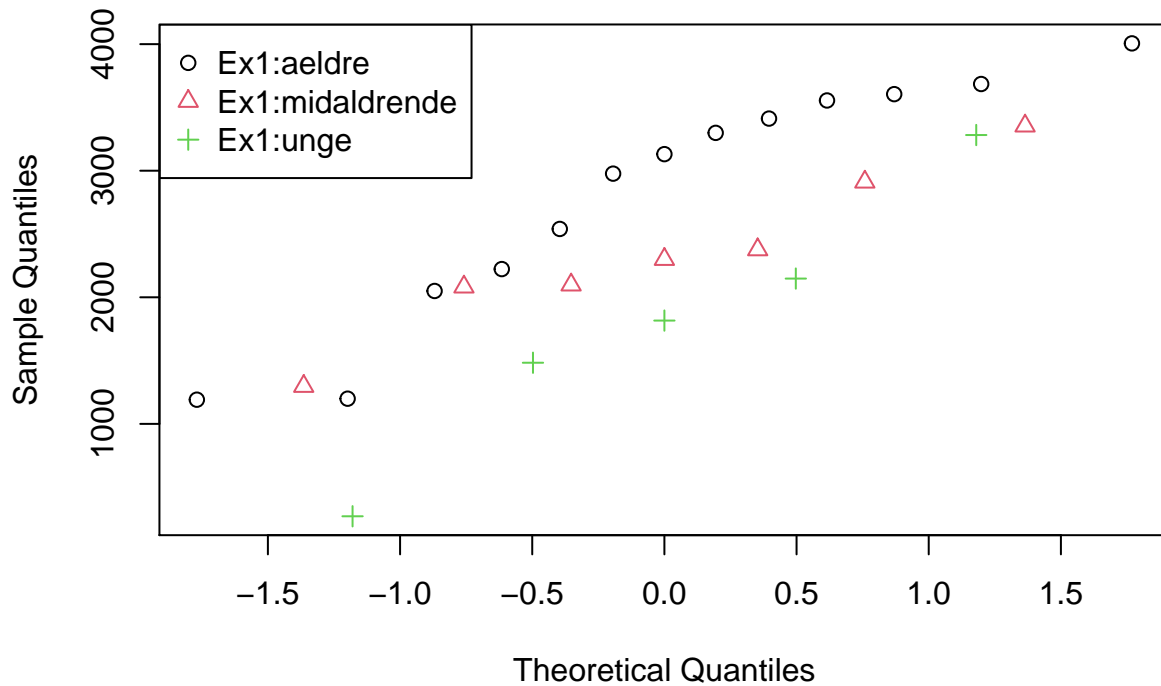
data <- read.csv("MatStat-R/data/JLJfiler/Smartphone.csv")
head(data)

##   Eksperiment   Alder SPD
## 1          Ex1     unge 3282
## 2          Ex1     unge 270
## 3          Ex1     unge 1483
## 4          Ex1     unge 2148
## 5          Ex1     unge 1816
## 6          Ex1 midaldrende 2098

source("MatStat-R/source/Rfunktioner.R")
data_eksperiment_1 <- data %>%
  filter(Eksperiment == "Ex1")
eksperiment <- as.factor(data_eksperiment_1$Eksperiment)
alder <- as.factor(data_eksperiment_1$Alder)
spd <- data_eksperiment_1$SPD

qqnormFlere(spd, eksperiment:alder)
```

a) Lav en figur med 3 delplots med qqplots af SPD for de tre aldersgrupper for eksperiment 1. Kommenter på figuren. Opskriv den statistiske model, hvor hver gruppe bestemt af aldersgruppe og eksperiment har sin egen middelværdi og sin egen varians af SPD, og data er normalfordelt. Lav et test for hypotesen, at der er samme varians i de 6 grupper.



Ud fra det ovenstående så er den eneste som ikke ser normalt fordelt ud Ex1:unge hvor vi har en observation der falder meget ud fra lingen.

Opskriv nu modellen for at hver faktor har sin egen middelværdi og varians.

$$M_1 : SPD \sim N(\mu_{alder}, \sigma_{alder}^2), (\mu_1, \dots, \mu_6, \sigma_1^2, \dots, \sigma_6^2)$$

Nu skal vi have det andet eksperiment med i selve data så vi bruger bare det fulde data, hvor vi laver en Bartlett's test

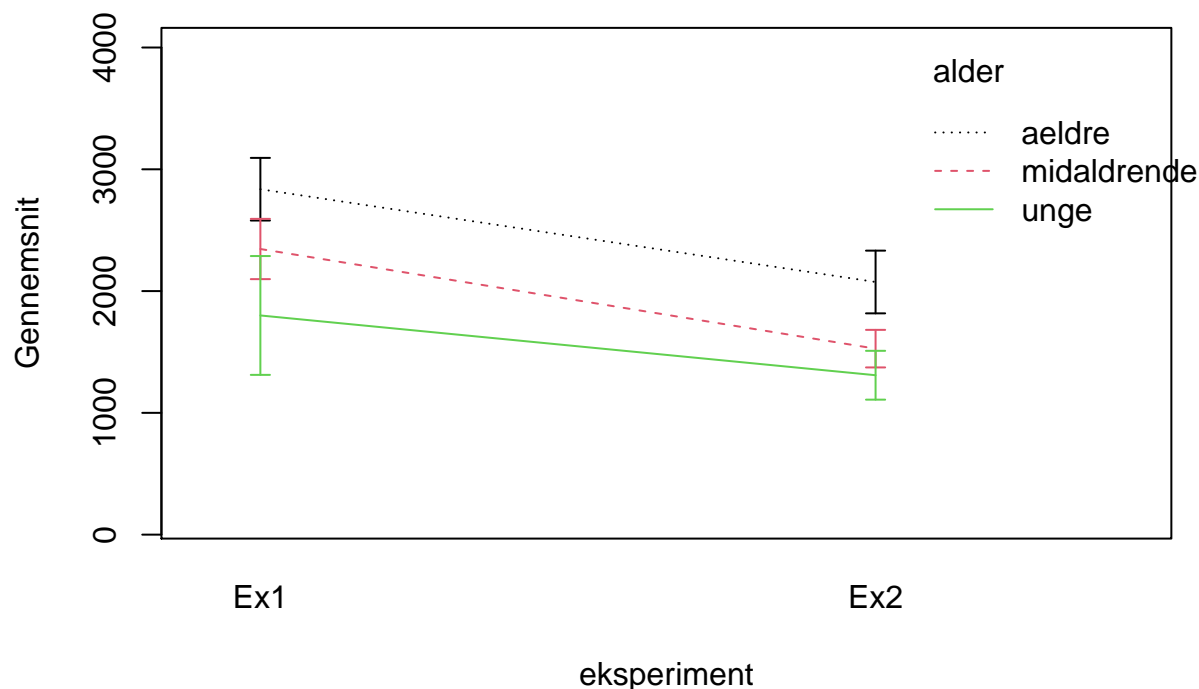
```
eksperiment <- as.factor(data$Eksperiment)
alder <- as.factor(data$Alder)
spd <- data$SPD
bartlett.test(spd, eksperiment:alder)
```

```
##
## Bartlett test of homogeneity of variances
##
## data: spd and eksperiment:alder
## Bartlett's K-squared = 7.1414, df = 5, p-value = 0.2103
```

Vi ser p-værdien er over 0.05 og derfor ikke statistisk signifikant dermed accepterer vi H_0 og vores varianser er ens.

b) Lav et interaktionsplot, og kommenter på hvad du ser i figuren. Opskriv modellen, hvor hver gruppe bestemt af aldersgruppe og eksperiment har sin egen middelværdi af SPD, og alle grupperne har den samme varians. Opskriv hypotesen om en additiv struktur af middelværdien med et bidrag fra aldersgruppe og fra eksperiment. Lav F-testet for hypotesen om additivitet. Lav til en start en interaktionsplot:

```
additivitetsPlot(eksperiment, alder, spd)
```



spd måler hvor god en person er til at browse gennem en hjemmeside. Desto lavere værdi desto bedre.

Her ser vi at de unge er bedre til at browse gennem en hjemmeside end de midaldrende og ældre. Desuden ser vi at der er en forskel mellem de to eksperimenter.

Opskriv modellen for forskellig middelværdi men ens varians:

$$spd \sim N(\mu_{E_i, A_i}, \sigma^2)$$

Herefter skal vi opskrive den additive hypotese:

$$M_2 : spd \sim \eta_{E_i} + \gamma_{A_i}, (\eta_1, \dots, \eta_k, \gamma_1, \dots, \gamma_m, \sigma^2)$$

En additiv hypotese er egentlig bare hvor vi ser på at hvis vi har flere predictives features som kan være med til at forklare vores regressor. Det er egentlig umiddelbar multiple regression.

Hypotesen:

$$H_0 : alder + eksperiment = alder : eksperiment$$

Jeg forstår denne hypotese sådan at hvis vi accepter H_0 så har den additive model samme middelværdi for alder og eksperiment.

Nu skal vi således la

```
anova(lm(spd ~ eksperiment:alder), lm(spd ~ eksperiment + alder))
```

```
## Analysis of Variance Table
##
## Model 1: spd ~ eksperiment:alder
## Model 2: spd ~ eksperiment + alder
##   Res.Df      RSS Df Sum of Sq    F Pr(>F)
## 1      45 28938030
## 2      47 29128146 -2    -190116 0.1478  0.863
```

Vi ser vores p værdi er over 0.05 dermed kan vi ikke forkaste H_0 . Dermed tyder det på den middelværdien for den additive model er ens for hypotesen fra forrigere.

c) Undersøg, om det kan antages, at aldersgruppe ikke har nogen effekt påSPD. Undersøgogså, om eksperiment har nogen effekt påSPD. Husk at skrive modellerne op. Opskriv modellen for aldersgruppen

$$spd \sim N(\mu_E, \sigma^2)$$

Hypotesen:

$$H_0 : \mu_E = 0$$

Test om en effekt, hvor jeg bruger en t test:

```
t.test((spd ~ eksperiment))
```

```
##
## Welch Two Sample t-test
##
## data:  spd by eksperiment
## t = 3.1662, df = 45.646, p-value = 0.002751
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   277.4799 1246.6432
## sample estimates:
## mean in group Ex1 mean in group Ex2
##      2491.600      1729.538
```

Vi får en p værdi under 0.05 dermed kan vi forkaste H_0 . Det betyder altså at eksperiment har en betydning for værdien for spd.

Opskriv modellem for eksperiment

$$spd \sim N(\mu_A, \sigma^2)$$

Hypotesen:

$$H_0 : \mu_A = 0$$

Test om en effekt:

d) Angiv skøn over middelværdien blandt de ældre for eksperiment Ex1. Angiv et 95%-konfidensinterval for forskellen i middelværdi af SPD mellem gruppen af unge og gruppen af ældre inden for den additive model. Angiv skøn over spredningen på SPD i den additive model Angiv et skøn over middelværdien blandt ældre for eksperiment 1:

```
summary(lm(SPD ~ Alder, data = data %>% filter(Eksperiment == "Ex1")))
```

```
##
## Call:
## lm(formula = SPD ~ Alder, data = data %>% filter(Eksperiment ==
##      "Ex1"))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1645.38  -316.80   30.57   575.62  1482.20
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2836.4      248.4   11.419 1.03e-10 ***
## Aldermidaldrende  -491.0      419.9   -1.169   0.2548
## Alderunge        -1036.6      471.3   -2.199   0.0387 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 895.6 on 22 degrees of freedom
## Multiple R-squared:  0.1881, Adjusted R-squared:  0.1143
## F-statistic: 2.548 on 2 and 22 DF,  p-value: 0.1011
```

Jeg er lidt i tvivl om hvorfor jeg i mit summary ikke ser Ældre, men jeg antager det er mit intercept. Det gør jeg da det vil give mening i forhold til min intuition at hvis du er ældre så vil din SPD stige, og vi lægge mærke til værdien er 2836. Modsat kan vi se at for de unge så er middelværdien -1036, som giver mening at jo yngre du er desto lavere vil din spd værdi være.

Angiv 95% konfidensinterval for forskellen i middelværdi af SPD mellem unge og ældre indenfor den additive model.

```
confint(lm(SPD ~ Alder + Eksperiment, data = data))
```

```
##              2.5 %      97.5 %
## (Intercept)  2434.553 3198.300709
## Aldermidaldrende -1041.071  -5.501166
## Alderunge      -1465.283 -317.782086
## EksperimentEx2  -1164.532 -275.745743
```

Her ser vi at konfidens båndet for de ældre er (2434, 3198) og for de unge der det (-1465, -317)

Angiv et skøn over spredning på SPD i den additive model

```
summary(lm(SPD ~ Alder + Eksperiment, data = data))$sigma
```

```
## [1] 787.2406
```

Vi ser således at spredning for den additive model er 787.

Generelt kan man at vi nok approximativ sige at aldergruppen er normalt fordelt på trods af jeg nævnte tidligere at de yngre falder lidt ud.

Desuden ser vi at eksperiment og aldersgruppen har betydning for SPD. Angående eksperiment så tænker jeg vi egentlig gerne så at det ikke havde en betydning, da det tyder på at vores sample af stikprøven ikke er random.