

Aflevering 10

Lucas Bagge

I denne opgave skal I studere sammenhængen mellem amerikanske firmaers markedsværdi og omsætning. I alt har vi data for 35 firmaer fordelt på tre sektorer. Mere præcist er 10 af firmaerne beskæftiget med detailhandel, 15 firmaer er indenfor energisektoren og de resterende 10 firmaer er i industrisektoren.

I skal beskrive data med en model, hvor der for hver sektor er en lineær sammenhæng mellem markedsværdien og omsætningen. Data findes i filen Firmaer.csv der har fire søjler: observationsnummer, sektor angiver sektoren, værdi angiver firmaets markedsværdi, og omsaet firmaets omsætning. Husk i spørgsmålene nedenfor at opskrive den statistiske model.

```
data <- read.csv("MatStat-R/data/AmFirma.csv")
head(data)
```

```
##   Obs  omsaet  vaerdi sektor
## 1    1 9.08171 8.39344 detail
## 2    2 7.37838 6.75926 detail
## 3    3 7.71735 6.43455 detail
## 4    4 8.79710 7.38275 detail
## 5    5 9.74824 8.33759 detail
## 6    6 8.53307 8.14381 detail
```

```
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()
```

```
distinct(data, sektor)
```

```
##      sektor
## 1    detail
## 2    energi
## 3  industri
```

Lav et plot af omsætning mod værdi, hvor hver sektor har sin egen farve, og indtegn en regressionslinje til hver sektor.

```

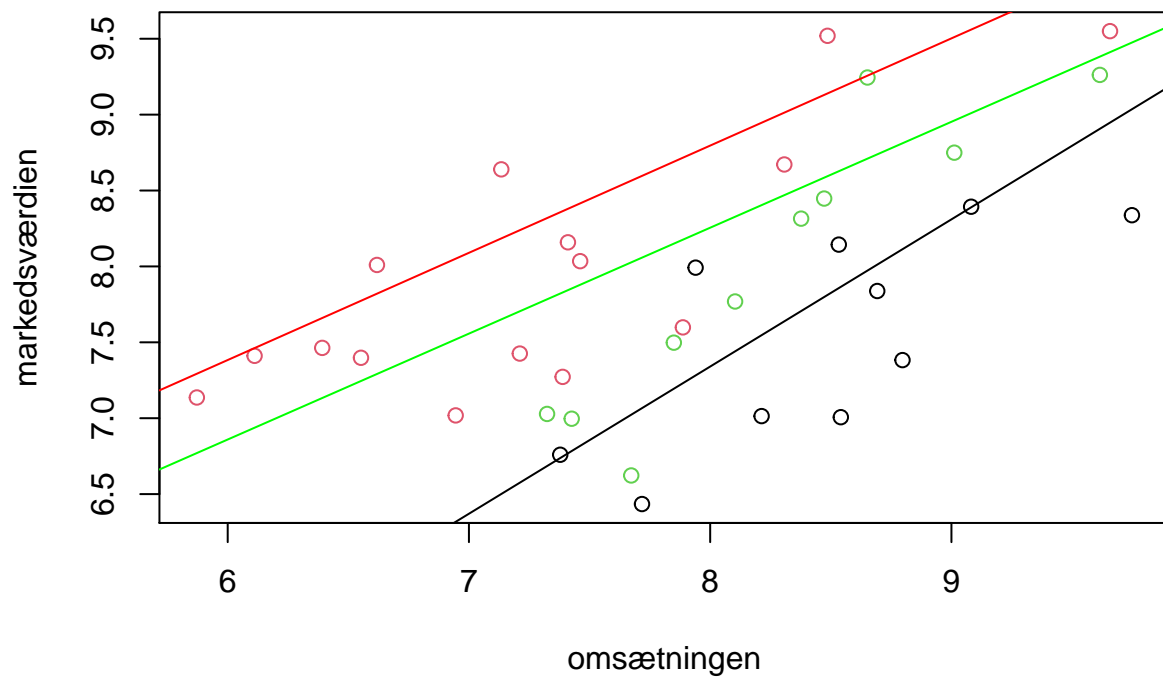
sektor <- factor(data$sektor)
markedsværdien <- data$vaerdi
omsætningen <- data$omsaet

detailhandel <- subset(data, sektor == "detail")
energisektoren <- subset(data, sektor == "energi")
industriektoren <- subset(data, sektor == "industri")

lmUD_detailhandel <- lm(detailhandel$omsaet ~ detailhandel$vaerdi)
lmUD_energisektoren <- lm(energisektoren$omsaet ~ energisektoren$vaerdi)
lmUD_industriektoren <- lm(industriektoren$omsaet ~ industriektoren$vaerdi)

plot(omsætningen, markedsværdien, col = sektor)
abline(lmUD_detailhandel, col = "red")
abline(lmUD_energisektoren)
abline(lmUD_industriektoren, col = "green")

```



Ovenstående viser et plt af omsætning mod markedsværdien for de tre sektor og vi har også indtegnet en regressionslinje.

Vis, at det kan antages, at variansen ikke afhænger af sektoren.

Opstil modellen:

$$oms_tning \sim N(\alpha_{sektor_i} + \beta_{sektor_i} markedsv_rdisektor_i, \sigma_{sektor_i}^2), 1, \dots, 35$$

check om samme varians:

```
bartlett.test(list(lmUD_detailhandel, lmUD_energisektoren, lmUD_industrisektoren))
```

```
##
## Bartlett test of homogeneity of variances
##
## data: list(lmUD_detailhandel, lmUD_energisektoren, lmUD_industrisektoren)
## Bartlett's K-squared = 4.0517, df = 2, p-value = 0.1319
```

```
#
```

Vi ser at p værdien er 0.1319 og dermed ikke statistik signifikant dermed accepter vi H_0 og vores varianser er ens.

Derfor afhænger variansen ikke af sektoren.

Vis, at det kan antages, at der er samme hældning på de tre regressionslinjer.

Opstil model:

$$oms_tning_{sektor_i} \sim N(\alpha_{sektor_i} + \beta_{sektor_i} lmarkedsv_rdi_{sektor_i}, \sigma^2)$$

$$anova(lm(x \sim gr + t), lm(x \sim gr * t))$$

```
M1_til_M2_alpha <- anova(lm(omsætningen ~ sektor + markedsværdien),
                          lm(omsætningen ~ sektor * markedsværdien))
M1_til_M2_alpha
```

```
## Analysis of Variance Table
##
## Model 1: omsætningen ~ sektor + markedsværdien
## Model 2: omsætningen ~ sektor * markedsværdien
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      31 8.2962
## 2      29 7.9097  2   0.38657 0.7087 0.5006
```

Data strider altså ikke mod hypotesen:

$$\beta_U = \beta_M$$

Vi kan godt beskrive ved den additive model.

Vi laver altså en reduktion fra M1 til M2 som er den model hvor skæring er den samme, og undersøger om skæringen kan komme fra den samme foreling p væriden siger det kan godt lade sig gøre.

Nu vil jeg undrsøge sammde for skræing

```
M1_til_M2_beta <- anova(lm(omsætningen ~ sektor : markedsværdien),
                        lm(omsætningen ~ sektor * markedsværdien))
M1_til_M2_beta
```

```
## Analysis of Variance Table
##
## Model 1: omsætningen ~ sektor:markedsværdien
## Model 2: omsætningen ~ sektor * markedsværdien
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      31 8.8275
## 2      29 7.9097  2   0.91782 1.6826 0.2035
```

samme forklaring. Modellen kan beskrives additivt.

Vis, at det ikke kan antages, at den fælles hældning er 0.

```
confint(lm(omsætningen ~ markedsværdien))
```

```
##                2.5 %   97.5 %
## (Intercept)  -0.08516603 5.379872
## markedsværdien 0.32343764 1.016414
```

Umiddelbart ud fra konfidens interval så vil jeg ikke sige at den fælles hældning vil være nul, men derimod mellem 0.32 og 1.01.

Angiv et 95%-konfidensinterval for spredningen på omsætningen.

```
lmSUM <- summary(lm(omsætningen ~ markedsværdien))
lmSUM$sigma
```

```
## [1] 0.8216089
```

```
df <- lmSUM$df
df[2]
```

```
## [1] 33
```

```
sqrt(df[2] * lmSUM$sigma^2 /
      qchisq(c(0.975, 0.025), df[2]))
```

```
## [1] 0.6626905 1.0814656
```

Finder konfidensinterval for spredning på omsætning.

Angiv et 95%-konfidensinterval for den fælles hældning på de tre regressionslinjer.

```
confint(lm(omsætningen~sektor+markedsværdien))
```

```
##              2.5 %      97.5 %  
## (Intercept)  0.5967555  4.0582560  
## sektorenergi -1.9551102 -1.0726982  
## sektorindustri -1.0745116 -0.1079832  
## markedsværdien  0.5894214  1.0404821
```

Ovenstående viser konfidensinterval for den fælles hældning for de tre regressionslinjer: