

# 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.

```

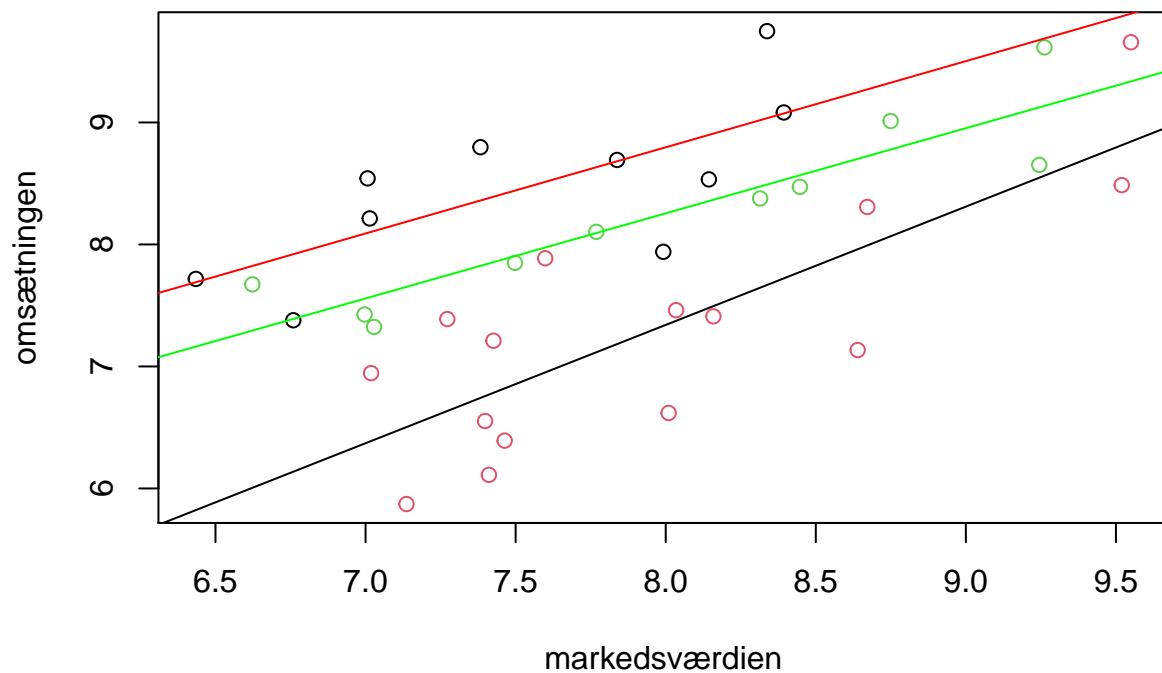
sektor <- factor(data$sektor)
markedsværdien <- data$vaerdi
omsætningen <- data$omsaet

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:

$$omsaetning \sim N(\alpha_{sektor_i} + \beta_{sektor_i} markedsv\_rdi, \sigma_{sektor_i}^2), 1, \dots, 35$$

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 H0 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:

$$oms\_tning_{sektor_i} \sim N(\alpha_{sektor_i} + \beta_{sektor_i} lmarkedsv\_rdi_{sektor_i}, \sigma^2)$$

$$\beta_{detail} = \beta_{energi} = \beta_{industri}$$

```
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.

Derved kan vi konkluderer de tre sektorer godt kan have fælles hældning.

Nu vil jeg undersøge samme for skæring

```
anova(lm(omsætningen ~ sektor + markedsværdien),
       lm(omsætningen ~ markedsværdien))
```

```
## Analysis of Variance Table
##
## Model 1: omsætningen ~ sektor + markedsværdien
## Model 2: omsætningen ~ markedsværdien
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      31  8.2962
## 2      33 22.2764 -2    -13.98 26.119 2.244e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Her får vi en lav p værdi og vi kan konkludere at de ikke har samme skæring. Det stemmer overens med hvad vi så i vores plot.

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

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

```
##
## Call:
## lm(formula = omsætningen ~ markedsværdien + sektor)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.90102 -0.36075 -0.01226  0.37861  1.06092
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.3275     0.8486   2.743  0.0100 *
## markedsværdien  0.8150     0.1106   7.370 2.68e-08 ***
## sektorenergi   -1.5139     0.2163  -6.998 7.44e-08 ***
## sektorindustri -0.5912     0.2370  -2.495  0.0181 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5173 on 31 degrees of freedom
## Multiple R-squared:  0.7465, Adjusted R-squared:  0.7219
## F-statistic: 30.42 on 3 and 31 DF,  p-value: 2.299e-09
```

Ud fra ovenstående koefficienter kan det ses at den fælles hældning ikke kan være nul.

**Angiv et 95%-konfidensinterval for spredningen på omsætningen.**

Hertil kan vi bruge formlen i afsnit 4.8 vedrørende skønnet over spredningen.

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

```
## [1] 0.5173204
```

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

```
## [1] 31
```

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

```
## [1] 0.4147375 0.6877669
```

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
```

Her er det interval vi skal bruge den for markedsværdien, hvor vi ser at værdien er mellem 0.58 og 1.04: