

oving2.R

Kasper

2021-09-21

Del 2:

```
data <- read.table("fatherSon.txt", header = TRUE)
```

```
#Vektor
```

```
father <- data$Father
```

```
son <- data$Son
```

```
summary(father)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      60.51  67.29   69.27   69.19   71.10   76.93
```

```
summary(son)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      60.01  68.43   70.12   70.18   71.97   79.86
```

```
#Gjør om til cm og rund av
```

```
father.cm <- father * 2.54
```

```
son.cm <- son * 2.54
```

```
father.round = round(father.cm)
```

```
son.round = round(son.cm)
```

```
summary(father.round)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      154.0   171.0   176.0   175.7   181.0   195.0
```

```
summary(son.round)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      152.0   174.0   178.0   178.3   183.0   203.0
```

```
#Spredningsplott
```

```
plot(father, son)
```

```
#Gjennomsnitt og standardavvik
```

```
father.mean = mean(father)
```

```
son.mean = mean(son)
```

```
father.sd = sd(father)
```

```
son.sd = sd(son)
```

```
#Nye observasjoner
```

```
neighbour = rnorm(1078, father.mean, father.sd)
```

```

data$neighbour = neighbour

#Korrelasjonsmatrise
kor <- cor(data)
summary(kor)

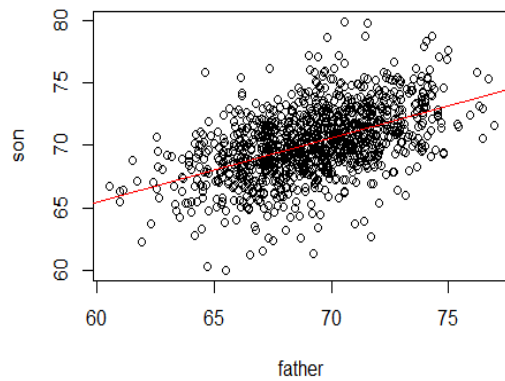
##           Father           Son           neighbour
## Min.      :0.00879   Min.      :-0.007729   Min.      :-0.0077290
## 1st Qu.:0.25506   1st Qu.: 0.246805   1st Qu.: 0.0005304
## Median :0.50134   Median : 0.501338   Median : 0.0087897
## Mean     :0.50338   Mean     : 0.497870   Mean     : 0.3336869
## 3rd Qu.:0.75067   3rd Qu.: 0.750669   3rd Qu.: 0.5043948
## Max.     :1.00000   Max.     : 1.000000   Max.     : 1.0000000

#Linær regresjon
regresjon <- lm(son~father, data=data)
summary(regresjon)

##
## Call:
## lm(formula = son ~ father, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.8772 -1.5144 -0.0079  1.6285  8.9685
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.61544    1.87289   18.48  <2e-16 ***
## father       0.51409    0.02705   19.01  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.437 on 1076 degrees of freedom
## Multiple R-squared:  0.2513, Adjusted R-squared:  0.2506
## F-statistic: 361.2 on 1 and 1076 DF, p-value: < 2.2e-16

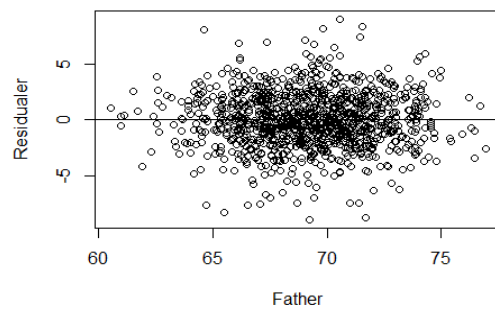
#Plot with regresjonslinje
plot(father,son)
abline(regresjon, col = "red")

```



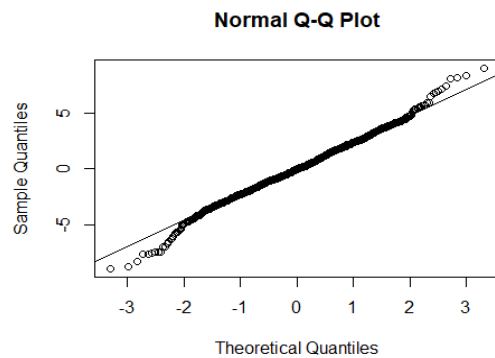
#Spredningsplott

```
regresjon.residualer = regresjon$residuals
plot(father, regresjon.residualer, xlab="Father", ylab="Residualer")
abline(0,0)
```



#Normalfordelingsplott

```
qqnorm(regresjon.residualer)
qqline(regresjon.residualer)
```



Del 3:

1.

Vi ser at dataene har en lineær sammenheng. Når x stiger, stiger også y. Positiv sammenheng. Vi har dataene samlet i et stort kluster, men det er en del «outliers» utenfor dette klusteret, men en klar samling inn mot sentrum av klusteret.

2.

	Father	Son	Neighbour
Father	1.00000000	0.50133831	-0.03548335
Son	0.50133831	1.00000000	-0.01935408
Neighbour	-0.03548335	-0.01935408	1.00000000

Korrelasjonen mellom father and son er sterk, da den er på 0.501. Dette er en positiv sammenheng. Sammenhengen er helt lik fra tommer til cm, man bare bruker andre verdier på x og y aksene.

3.

Negativ sammenheng på både neighbour-son og neighbour-father med -0.036 og -0.019. Ikke veldig stor forskjell men det er en liten en. Ville forventet forskjeller her.

4.

Fordi her sjekker du korrelasjonen til x med x, altså de samme verdiene med hverandre. Da vil verdiene alltid være 1.

5.

Vi har en lineær sammenheng, i positiv retning. Linjen går gjennom der dataen samler seg, men vi har en del data som er langt i fra linja, «outliers».

6.

7.

Regresjonsmodellen passer. Mesteparten av dataene samler seg nær linjen. Vi har noen outliers som legger seg litt ifra, men ikke i nærheten av nok til å ha mye å si.

8.

I normalfordelingsplottet vårt ser vi minimale avvik fra linjen, de aller fleste verdier ligger seg på den rette linjen og er normalfordelt.

9.