# Regression I R

Sammenhæng mellem alder og forbrug

Som vanligt: importere vi filen

library(readxl)

bar <- read\_excel("HVOR FILEN NU ER PLACERET/bar.xls")

View(bar)

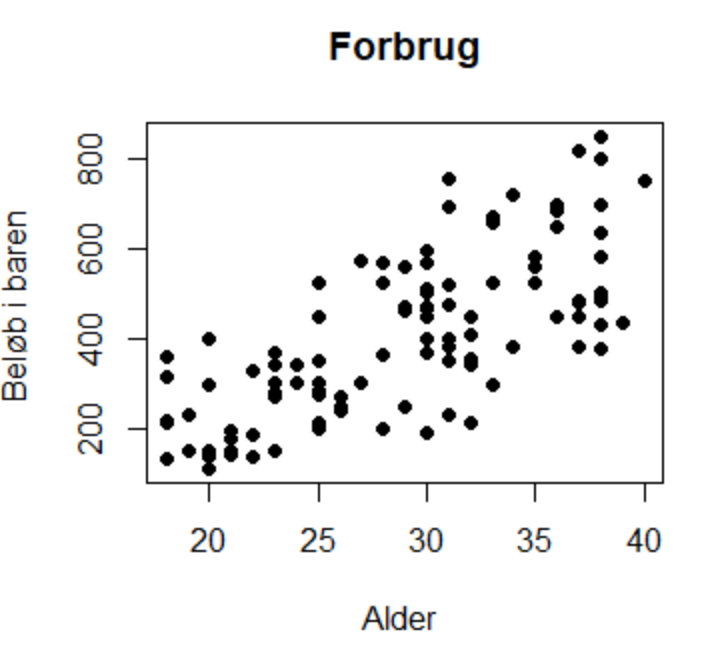
Altid den god ide at lave et scatter plot for at vurdere data.

attach(bar)

plot(Alder, Beløb, main="Forbrug",

xlab="Alder ", ylab="Beløb i baren ", pch=19)

#mere viden om pch her: <https://www.datanovia.com/en/blog/pch-in-r-best-tips/>



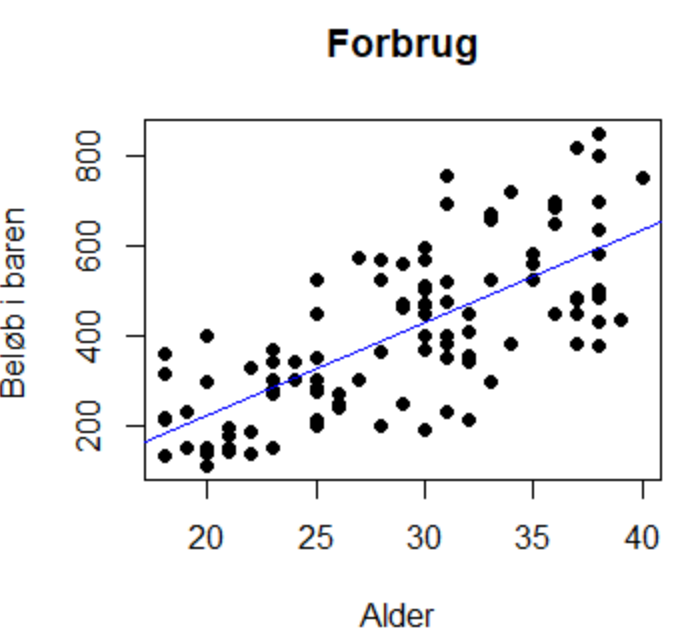
Ligner at der er en sammenhæng mellem alder og forbruget at penge på et værtshus, dog med en stor spredning.

VIGTIGT: Læg mærke til at vi med plot har x,y hvor man NATURLIGVIS gør det modsat i regression med y~x.

Så altid fokus på hvem der er den afhængige og uafhængige variable.

Vi tilfører en regressionslinje

abline(lm(Beløb~Alder), col="blue")

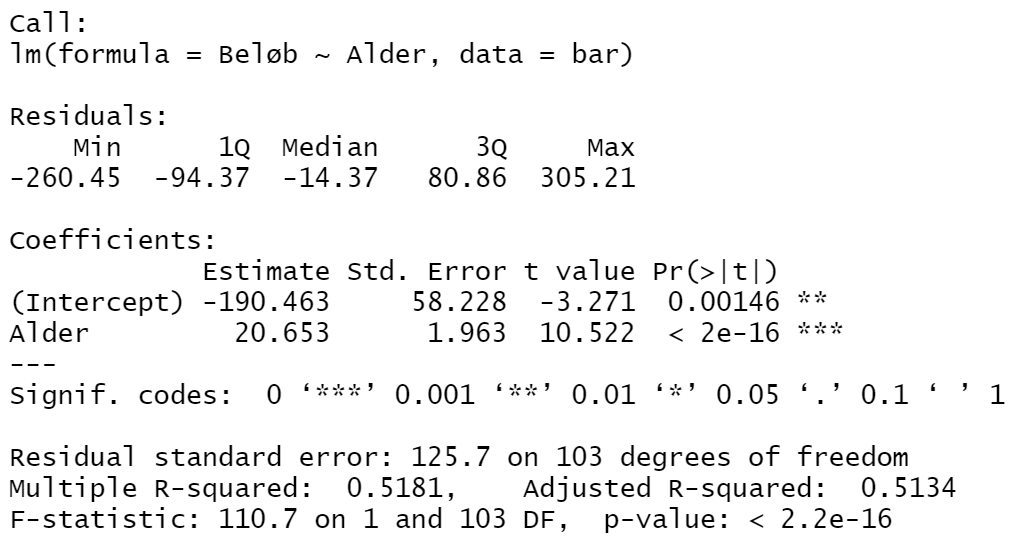


Der er en fin sammenhæng, nu tester vi

lmBar <- lm(Beløb~Alder, data=bar)

summary(lmBar)

NÆSTE SIDE





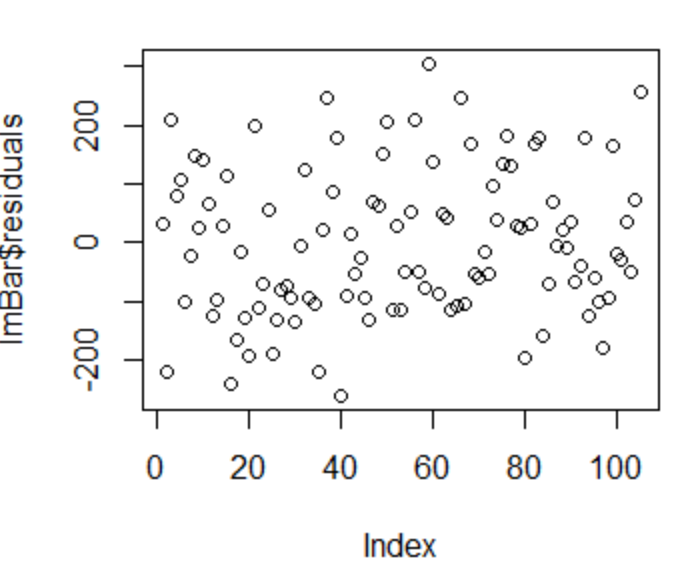
Vi kan se at vi bruger en 20 kroner mere i baren for hvert år vi bliver ældre, og at det er ret lave p-værdier, så det er signifikante koefficienter.

Vi kan også se at sammenhængskraften ikke er imponerende 0,51 men virkeligheden egner sig sjældent til terori :D så det er faktisk en fin model.

Residualerne, bliver naturligvis lnagt fra 0, da data er ret spredt, så ofte er det mere givende at se på et residualplot, og vurdere om der er tale om ’white noice’

Så det gør vi

plot(lmBar$residuals)



Med et rimeligt boldøje ser det nu ud til at der er en jævn fordeling i forhold til 0, så den tager vi