

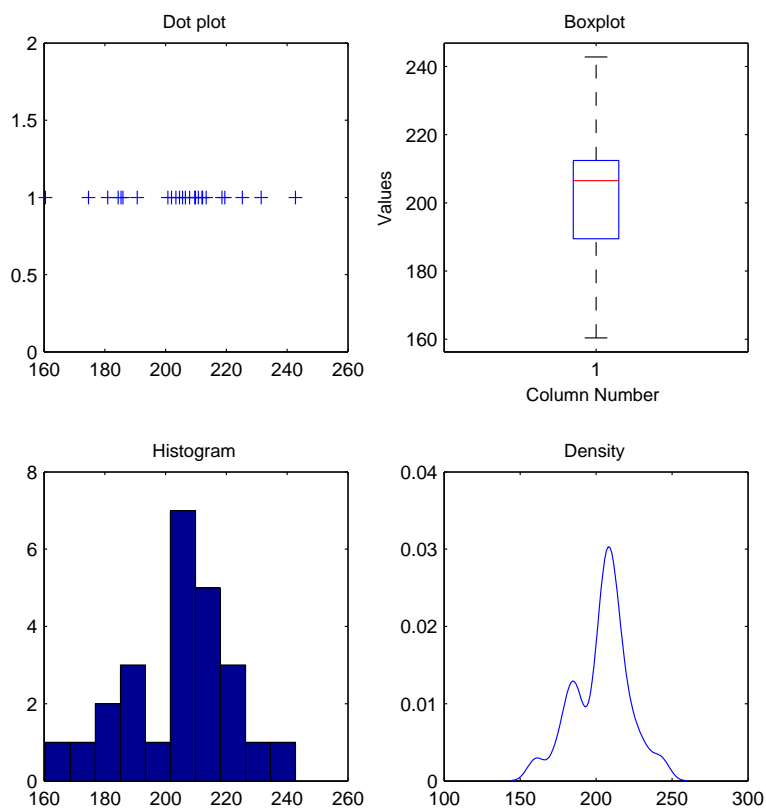
6. Beskrivning av datamängder

Beskrivning av data genom att använda figurer

Tre datamängder

D1 Styrkan av spänningen (tensile strength) (enhet "1000 pounds per square inch") av materialet som används för att bygga flygplan.

Data (styrkan av spänningen): 203.41 209.58 213.35 218.56 242.76
185.97 190.67 207.88 210.80 231.46 184.41 200.73 206.51 209.84 212.15
160.44 180.95 201.95 204.60 219.51 174.63 185.34 205.59 212.00 225.25



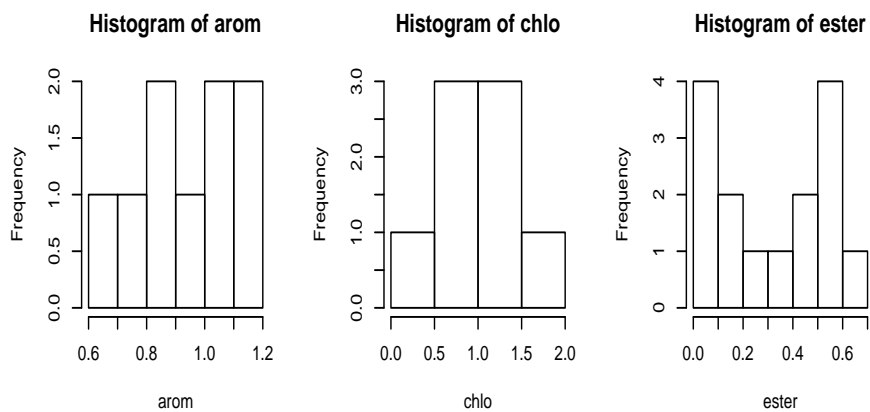
D2 Studie av kemikaliska egenskaper av tre olika typer av farliga organiska lösningsmedel, som används för att putsa metalledar: aromater (aromatics), kloroalkaner (chloroalkanes) och ester (esters). En av variablerna som undersöktes var "sorption rate" mätt i molandel (mole percentage).

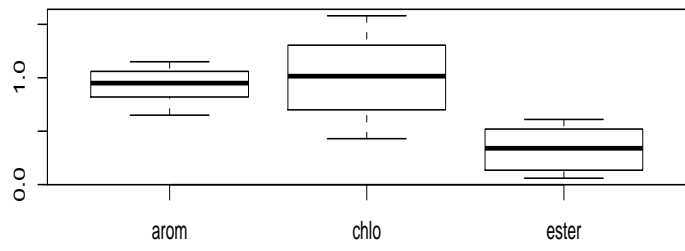
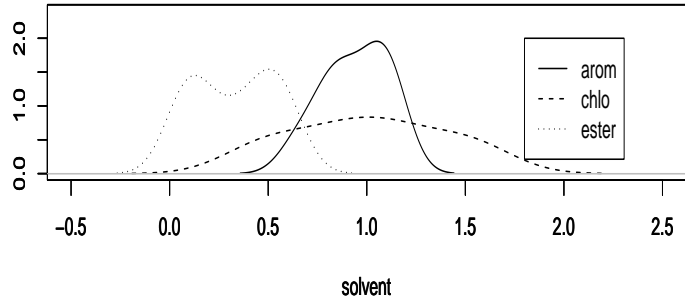
Data (sorption rate):

aromater: 1.06 0.79 0.82 0.89 1.05 0.95 0.65 1.15 1.12

kloroalkaner: 1.58 1.45 0.57 1.16 1.12 0.91 0.83 0.43

ester: 0.29 0.06 0.44 0.61 0.55 0.43 0.51 0.10 0.34 0.53 0.06 0.09 0.17 0.60 0.17

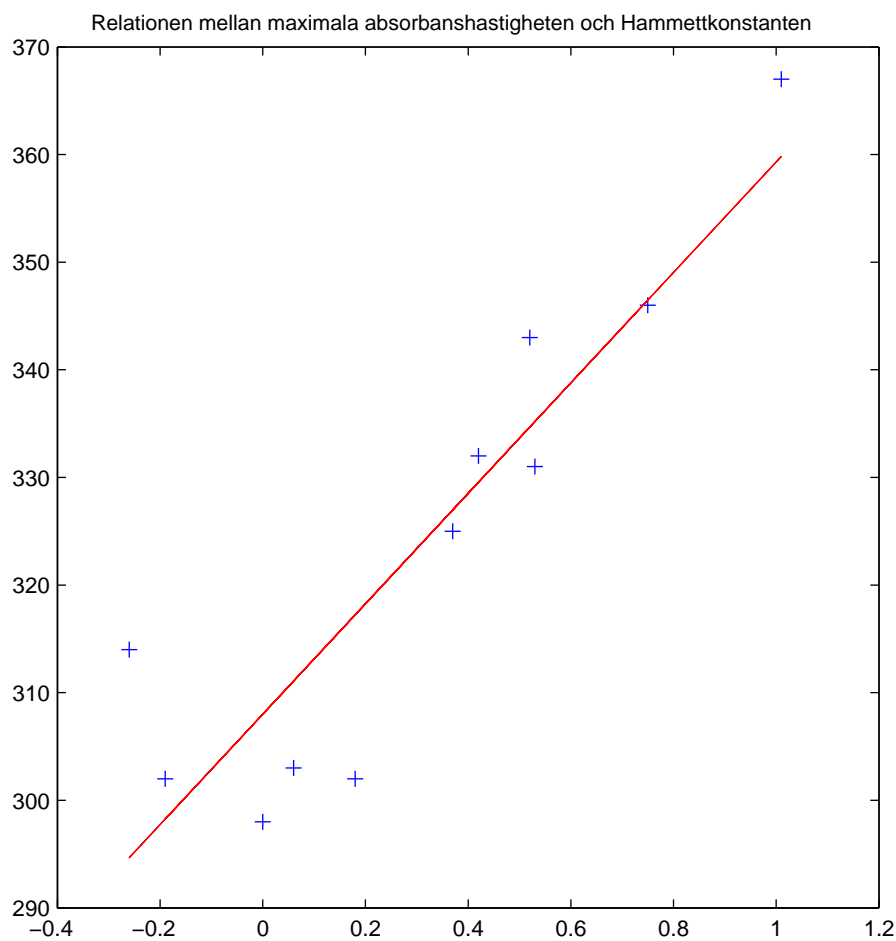




D3 Linjär relation mellan maximala absorptionshastigheten (absorption rate) (i nanomoler) och Hammettkonstanten (the Hammett substituent constant) för metacyclophane förening.

Data:

max absorptionshastighet: 298 346 303 314 302 332 302 343 367 325 331
Hammettkonstanten: 0.00 0.75 0.06 -0.26 0.18 0.42 -0.19 0.52 1.01 0.37 0.53



Numerisk beskrivning av data

Data: x_1, x_2, \dots, x_n

- *Medelvärdet* (aritmetiskt) är

$$\frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

- *Medianen* M är det “mellersta värdet”. Om datamängden är i storleksordning, är medianen
 - a) det mellersta talet, om n är ojämn
 - b) medelvärdet av de två mellersta talen, om n är jämn
- *Variationsbredden* är “största talet - minsta talet”
- *Stickprovsvariansen* är

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

och *stickprovsstandardavvikelsen* är

$$s = \sqrt{s^2}$$

Fel och outliers

Ett *fel* är en observation som inte är korrekt (okorrekt mätt eller kopierat). En *outlier* är en extrem observation vilken inte är konsistent med resten av data.

Obs! Ett fel kan vara en outlier men inte nödvändigtvist, och en outlier kan vara ett fel men inte nödvändigtvist.

Vad gör man med extrema observationer?

- kolla med experter om värdet är omöjligt eller möjligt men osannolikt
- samla mer data om möjligt
- ibland kan man ta bort värdet från datamängden men vanligen rekommenderas detta inte
- en outlier kan vara den mest intressanta observationen
- man kan använda metoder som inte är känsliga mot outliers

Om man inte tror att en outlier är ett fel, kan man göra analysen med och utan det extrema värdet (outlier) och presentera två resultat.