

# oefeningen hoofdstuk 6 - Analyse op 2 Variabelen

*TijsMartens*

*12 april 2019*

## voorbeeld bij 6.2: Het opstellen van een kruistabel

data invoeren en opmaken van de tabel

```
waarderingen_m <- matrix(c(9,8,5,0,8,10,5,4), ncol = 2)

rownames(waarderingen_m) <- c("Goed", "Voldoende", "Onvoldoende", "Slecht")
colnames(waarderingen_m) <- c("Man", "Vrouw")
waarderingen <- as.table(waarderingen_m)
```

marginale totalen berekenen

```
margin.table(waarderingen, 1) # rijtotalen
```

```
##      Goed  Voldoende Onvoldoende      Slecht
##      17      18      10      4
```

```
margin.table(waarderingen, 2) # kolomtotalen
```

```
##  Man Vrouw
##  22    27
```

```
margin.table(waarderingen) # aantal waarnemingen (n)
```

```
## [1] 49
```

percentages van rijen berekenen

```
waarderingen_pct <- prop.table(waarderingen, 2)
waarderingen_pct
```

```
##           Man      Vrouw
## Goed      0.4090909 0.2962963
## Voldoende  0.3636364 0.3703704
## Onvoldoende 0.2272727 0.1851852
## Slecht     0.0000000 0.1481481
```

berekenen chi-kwadraat

```
#verwacht <- as.array(margin.table(waarderingen, 1)) %*%t(as.array(margin.table(waarderingen, 2))) / ma
```

```
#afwijking <- (waarderingen - verwacht) ^ 2 / verwacht
```

```
#sum(afwijking)
```

```
## ^ is het zelfde als volgende code
```

```
summ <- summary(waarderingen)
chi_sq <- summ$statistic
```

```
chi_sq
```

```
## [1] 3.810518
```

berekenen van Cramér's V

```
k <- min(nrow(waarderingen), ncol(waarderingen))
V <- sqrt(chi_sq/margin.table((waarderingen) * (k-1)))
```

```
k
```

```
## [1] 2
```

```
V
```

```
## [1] 0.278865
```

bekomen resultaten

```
t(waarderingen)
```

```
##      Goed Voldoende Onvoldoende Slecht
## Man      9         8           5      0
## Vrouw    8        10           5      4
```

```
waarderingen
```

```
##      Man Vrouw
## Goed      9    8
## Voldoende  8   10
## Onvoldoende 5    5
## Slecht    0    4
```

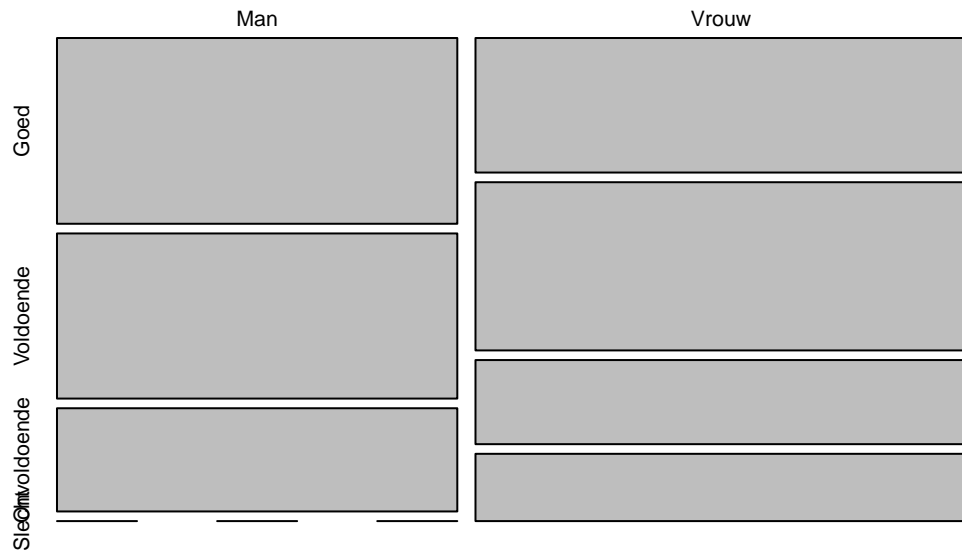
```
waarderingen_pct
```

```
##      Man      Vrouw
## Goed    0.4090909 0.2962963
## Voldoende 0.3636364 0.3703704
## Onvoldoende 0.2272727 0.1851852
## Slecht    0.0000000 0.1481481
```

maken tabel

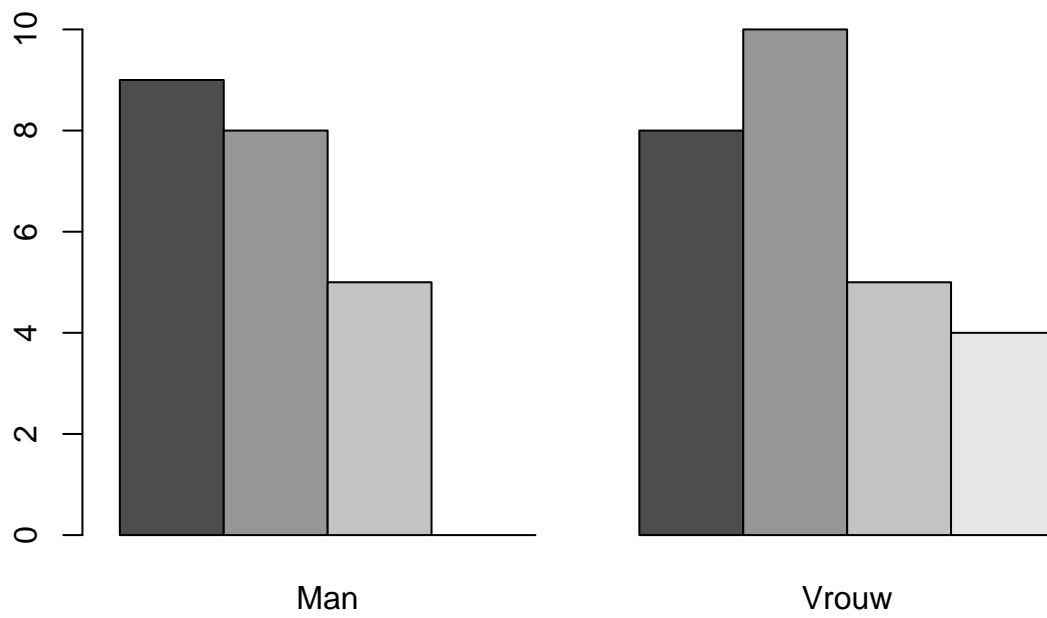
```
plot(t(waarderingen))
```

## t(waarderingen)



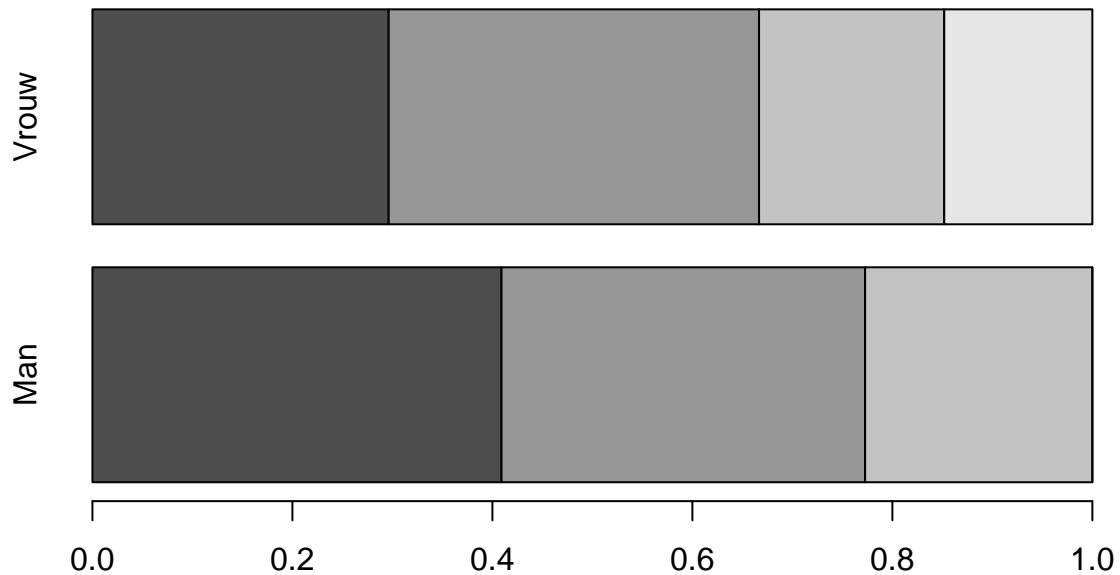
maken geclustere bar chart

```
barplot(waarderingen, beside = TRUE)
```



maken van percentage chart

```
barplot(waarderingen_pct, horiz = TRUE)
```



## voorbeeld 6.4.

voorbeeld van een regressierechte

```
data <- read.csv("C:\\Users\\tijsm\\Google Drive\\HoGent 2018-2019\\2e semester\\Onderzoekstechnieken\\")
attach(data)
data
```

```
## x.y.regression.residue
## 1 0\t177\t189.9643\t12.9643
## 2 10\t231\t229.2143\t1.7857
## 3 20\t249\t268.4643\t19.4643
## 4 30\t348\t307.7143\t40.2857
## 5 40\t361\t346.9643\t14.0357
## 6 50\t384\t386.2143\t2.2143
## 7 60\t404\t425.4643\t21.4643
```

code is juist om regressierechte te tekenen. Maar hij kent kolom “x” en “y” smh niet

```
#x = eiwitgehalte
#y = gewichtstoename

#plot(x, y, main = 'gewichtstoename', xlab = 'eiwitgehalte', ylab = 'gewichtstoename')

##regressie berekenen
```

```
#regr <- lm(y ~ x)
#regr

#abline(regr, col="red")
```

## oefening 6.1.

(oefening zelf gemaakt, geen oplossing)

### opgave

Marktonderzoek toont aan dat achtergrondmuziek in een supermarkt invloed kan hebben op het aankoopgedrag van de klanten. In een onderzoek werden drie methoden met elkaar vergeleken: geen muziek, Franse chansons en Italiaanse hits. Telkens werd het aantal verkochte ??essen Franse, Italiaanse en andere wijnen geteld (Ryan, Northrop-Clewes, Knox & Thurnham, 1998). De onderzoeksdata bevindt zich in het csv-bestand MuziekWijn.

1. Stel de correcte kruistabel op. Gebruik hiervoor het R-commando table om de frequentietabel te bekomen.
2. Bepaal de marginalen.
3. Bepaal de verwachte resultaten.
4. Bereken manueel de ??2 toetsingsgrootheid.
5. Bereken manueel de Cramér's V. Wat kan je hieruit besluiten?

### oplossing

inlezen csv bestand

```
data <- read.csv("C:\\Users\\tjism\\Google Drive\\HoGent 2018-2019\\2e semester\\Onderzoekstechnieken\\")
data
```

```
##      Muziek      Wijn
## 1      Geen      Franse
## 2      Franse     Franse
## 3      Franse  Italiaanse
## 4  Italiaanse     Andere
## 5  Italiaanse     Franse
## 6  Italiaanse     Andere
## 7      Geen      Franse
## 8      Geen      Andere
## 9      Geen      Franse
## 10     Geen  Italiaanse
## 11     Franse     Franse
## 12     Franse     Andere
## 13     Franse     Franse
## 14     Geen      Franse
## 15     Geen      Andere
## 16  Italiaanse     Franse
## 17     Geen  Italiaanse
## 18     Franse     Franse
## 19     Geen      Franse
## 20     Franse     Andere
## 21     Geen  Italiaanse
```

## 22	Geen	Franse
## 23	Geen	Andere
## 24	Franse	Franse
## 25	Geen	Italiaanse
## 26	Franse	Andere
## 27	Franse	Franse
## 28	Geen	Andere
## 29	Italiaanse	Franse
## 30	Italiaanse	Italiaanse
## 31	Italiaanse	Andere
## 32	Franse	Andere
## 33	Geen	Franse
## 34	Geen	Andere
## 35	Geen	Italiaanse
## 36	Geen	Franse
## 37	Italiaanse	Franse
## 38	Italiaanse	Italiaanse
## 39	Geen	Andere
## 40	Franse	Franse
## 41	Franse	Andere
## 42	Geen	Franse
## 43	Geen	Italiaanse
## 44	Geen	Italiaanse
## 45	Geen	Andere
## 46	Italiaanse	Franse
## 47	Franse	Andere
## 48	Geen	Franse
## 49	Italiaanse	Franse
## 50	Geen	Andere
## 51	Italiaanse	Italiaanse
## 52	Geen	Italiaanse
## 53	Italiaanse	Andere
## 54	Geen	Franse
## 55	Italiaanse	Andere
## 56	Franse	Andere
## 57	Italiaanse	Franse
## 58	Italiaanse	Italiaanse
## 59	Geen	Andere
## 60	Geen	Franse
## 61	Italiaanse	Italiaanse
## 62	Italiaanse	Andere
## 63	Geen	Italiaanse
## 64	Franse	Andere
## 65	Italiaanse	Franse
## 66	Geen	Franse
## 67	Franse	Franse
## 68	Italiaanse	Andere
## 69	Italiaanse	Italiaanse
## 70	Italiaanse	Franse
## 71	Geen	Andere
## 72	Geen	Franse
## 73	Franse	Andere
## 74	Italiaanse	Franse
## 75	Italiaanse	Andere

## 76	Italiaanse	Italiaanse
## 77	Geen	Italiaanse
## 78	Geen	Franse
## 79	Franse	Andere
## 80	Italiaanse	Franse
## 81	Italiaanse	Andere
## 82	Italiaanse	Italiaanse
## 83	Geen	Andere
## 84	Geen	Franse
## 85	Italiaanse	Franse
## 86	Italiaanse	Andere
## 87	Italiaanse	Italiaanse
## 88	Franse	Franse
## 89	Franse	Andere
## 90	Geen	Andere
## 91	Italiaanse	Franse
## 92	Geen	Franse
## 93	Italiaanse	Andere
## 94	Italiaanse	Italiaanse
## 95	Geen	Andere
## 96	Italiaanse	Franse
## 97	Geen	Italiaanse
## 98	Franse	Andere
## 99	Geen	Franse
## 100	Italiaanse	Franse
## 101	Geen	Andere
## 102	Italiaanse	Italiaanse
## 103	Italiaanse	Andere
## 104	Franse	Andere
## 105	Geen	Franse
## 106	Franse	Franse
## 107	Italiaanse	Franse
## 108	Geen	Franse
## 109	Italiaanse	Italiaanse
## 110	Geen	Andere
## 111	Italiaanse	Franse
## 112	Italiaanse	Andere
## 113	Franse	Andere
## 114	Geen	Franse
## 115	Franse	Franse
## 116	Italiaanse	Franse
## 117	Geen	Andere
## 118	Italiaanse	Italiaanse
## 119	Franse	Andere
## 120	Italiaanse	Andere
## 121	Italiaanse	Italiaanse
## 122	Italiaanse	Franse
## 123	Geen	Franse
## 124	Franse	Franse
## 125	Italiaanse	Andere
## 126	Geen	Andere
## 127	Italiaanse	Italiaanse
## 128	Franse	Andere
## 129	Geen	Franse



## 130	Italiaanse	Franse
## 131	Geen	Andere
## 132	Italiaanse	Andere
## 133	Franse	Franse
## 134	Franse	Andere
## 135	Italiaanse	Andere
## 136	Geen	Andere
## 137	Italiaanse	Franse
## 138	Geen	Franse
## 139	Italiaanse	Italiaanse
## 140	Franse	Andere
## 141	Italiaanse	Franse
## 142	Geen	Andere
## 143	Italiaanse	Andere
## 144	Geen	Franse
## 145	Franse	Franse
## 146	Italiaanse	Franse
## 147	Italiaanse	Italiaanse
## 148	Geen	Andere
## 149	Italiaanse	Andere
## 150	Italiaanse	Franse
## 151	Franse	Andere
## 152	Italiaanse	Andere
## 153	Geen	Franse
## 154	Franse	Franse
## 155	Italiaanse	Italiaanse
## 156	Italiaanse	Franse
## 157	Franse	Andere
## 158	Geen	Franse
## 159	Italiaanse	Andere
## 160	Geen	Andere
## 161	Italiaanse	Andere
## 162	Italiaanse	Franse
## 163	Franse	Franse
## 164	Franse	Andere
## 165	Italiaanse	Italiaanse
## 166	Franse	Franse
## 167	Geen	Franse
## 168	Geen	Andere
## 169	Italiaanse	Andere
## 170	Franse	Andere
## 171	Italiaanse	Andere
## 172	Franse	Franse
## 173	Geen	Franse
## 174	Geen	Andere
## 175	Italiaanse	Andere
## 176	Franse	Andere
## 177	Italiaanse	Andere
## 178	Franse	Franse
## 179	Franse	Andere
## 180	Italiaanse	Andere
## 181	Franse	Franse
## 182	Geen	Franse
## 183	Italiaanse	Andere

## 184	Franse	Franse
## 185	Geen	Andere
## 186	Italiaanse	Franse
## 187	Italiaanse	Andere
## 188	Franse	Andere
## 189	Geen	Andere
## 190	Franse	Franse
## 191	Italiaanse	Franse
## 192	Geen	Andere
## 193	Franse	Franse
## 194	Franse	Andere
## 195	Italiaanse	Andere
## 196	Franse	Franse
## 197	Geen	Andere
## 198	Italiaanse	Franse
## 199	Franse	Franse
## 200	Geen	Andere
## 201	Italiaanse	Andere
## 202	Franse	Franse
## 203	Franse	Andere
## 204	Geen	Andere
## 205	Italiaanse	Andere
## 206	Geen	Andere
## 207	Italiaanse	Franse
## 208	Franse	Franse
## 209	Franse	Andere
## 210	Geen	Andere
## 211	Italiaanse	Andere
## 212	Geen	Andere
## 213	Italiaanse	Andere
## 214	Franse	Franse
## 215	Franse	Andere
## 216	Geen	Andere
## 217	Franse	Franse
## 218	Italiaanse	Andere
## 219	Geen	Andere
## 220	Franse	Franse
## 221	Franse	Andere
## 222	Geen	Andere
## 223	Franse	Franse
## 224	Franse	Andere
## 225	Geen	Andere
## 226	Franse	Franse
## 227	Geen	Andere
## 228	Geen	Andere
## 229	Franse	Franse
## 230	Franse	Andere
## 231	Geen	Andere
## 232	Franse	Franse
## 233	Franse	Andere
## 234	Geen	Andere
## 235	Franse	Franse
## 236	Geen	Andere
## 237	Franse	Andere

```
## 238      Franse      Franse
## 239      Geen      Andere
## 240      Franse      Franse
## 241      Franse      Franse
## 242      Franse      Franse
## 243      Franse      Andere
```

## 1.

opstellen van de frequentietabel

```
kruistabel <-table(data)
kruistabel
```

```
##           Wijn
## Muziek      Andere  Franse  Italiaanse
##   Franse      35     39      1
##   Geen       43     30     11
##   Italiaanse  35     30     19
```

## 2.

bepalen van de marginalen

```
margin.table(kruistabel, 1)
```

```
## Muziek
##   Franse      Geen Italiaanse
##      75      84      84
```

```
margin.table(kruistabel, 2)
```

```
## Wijn
##   Andere      Franse  Italiaanse
##      113      99      31
```

```
prop.table(kruistabel)
```

```
##           Wijn
## Muziek      Andere      Franse  Italiaanse
##   Franse  0.144032922 0.160493827 0.004115226
##   Geen   0.176954733 0.123456790 0.045267490
##   Italiaanse 0.144032922 0.123456790 0.078189300
```

## 3.

bepalen van de verwachte waarden

```
library(gmodels)
```

```
## Warning: package 'gmodels' was built under R version 3.5.3
```

```
library(lsr)
```

```
CrossTable(kruistabel)
```

```
##
##
##   Cell Contents
```

```
## |-----|
## |              N |
## | Chi-square contribution |
## |      N / Row Total |
## |      N / Col Total |
## |      N / Table Total |
## |-----|
##
##
## Total Observations in Table:  243
##
##
##           | Wijn
##      Muziek |      Andere |      Franse |      Italiaanse |      Row Total |
## -----|-----|-----|-----|-----|
##      Franse |           35 |           39 |           1 |           75 |
##           |           0.000 |           2.334 |           7.672 |           |
##           |           0.467 |           0.520 |           0.013 |           0.309 |
##           |           0.310 |           0.394 |           0.032 |           |
##           |           0.144 |           0.160 |           0.004 |           |
## -----|-----|-----|-----|-----|
##      Geen |           43 |           30 |           11 |           84 |
##           |           0.397 |           0.521 |           0.008 |           |
##           |           0.512 |           0.357 |           0.131 |           0.346 |
##           |           0.381 |           0.303 |           0.355 |           |
##           |           0.177 |           0.123 |           0.045 |           |
## -----|-----|-----|-----|-----|
##      Italiaanse |           35 |           30 |           19 |           84 |
##           |           0.422 |           0.521 |           6.404 |           |
##           |           0.417 |           0.357 |           0.226 |           0.346 |
##           |           0.310 |           0.303 |           0.613 |           |
##           |           0.144 |           0.123 |           0.078 |           |
## -----|-----|-----|-----|-----|
## Column Total |           113 |           99 |           31 |           243 |
##           |           0.465 |           0.407 |           0.128 |           |
## -----|-----|-----|-----|-----|
##
##
```

#### 4.

bereken manueel de chi kwadraat

met r:

```
cramersV(kruistabel)
```

```
## [1] 0.193937
```

manueel:

$$\sum \frac{(a-e)^2}{e}$$

a = geobserveerde waarde e = verwachte waarde

- verwachte waarde berekenen

$$e = \left( \frac{\text{rijtotaal}}{n} \times \frac{\text{kolomtotaal}}{n} \times n \right) e = \frac{\text{rijtotaal} \times \text{kolomtotaal}}{n}$$

```
kruistabel
```

```
##           Wijn
## Muziek      Andere Franse Italiaanse
## Franse      35      39      1
## Geen        43      30     11
## Italiaanse  35      30     19
```

```
aantalKol <- ncol(kruistabel)
aantalKol
```

```
## [1] 3
```

```
aantalRij <- nrow(kruistabel)
aantalRij
```

```
## [1] 3
```

```
n <- nrow(data)
n
```

```
## [1] 243
```

```
e = (aantalKol * aantalRij) / n
e
```

```
## [1] 0.03703704
```

- chi kwadraat berekenen

$$\sum \frac{(a-e)^2}{e}$$

5.

```
chisq.test(kruistabel)
```

```
##
## Pearson's Chi-squared test
##
## data:  kruistabel
## X-squared = 18.279, df = 4, p-value = 0.001088
```

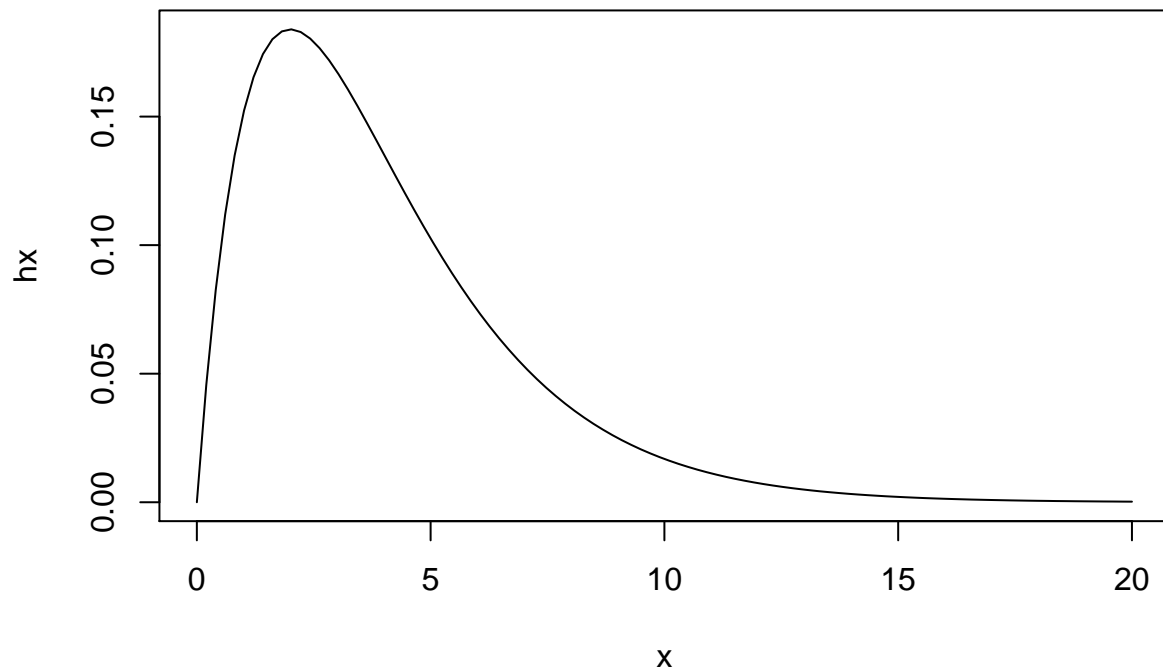
```
qchisq(0.95, df = 4)
```

```
## [1] 9.487729
```

```
qchisq(1-0.001088, df = 4)
```

```
## [1] 18.27978
```

```
x <- seq(0, 20, length=100)
hx <- dchisq(x, df = 4)
plot(x, hx, type="l")
```



## oefening 6.2.

*(oefening zelf gemaakt, geen oplossing)*

### opgave

Gebruik dezelfde data.

1. Stel de percentages verkochte wijnen voor in een staafdiagram met de muziekconditie= Geen.
2. Stel de percentages verkochte wijnen voor in een geclusterd staafdiagram (clustered bar chart).
3. Stel de percentages verkochte wijnen voor in rependiagram (stacked bar chart) ## oplossing

```
library(vcd) assocstats(data)
```

```
data <- read.csv("C:\\Users\\tijsm\\Google Drive\\HoGent 2018-2019\\2e semester\\Onderzoekstechnieken\\")
```

```
data
```

```
##      Muziek      Wijn
## 1      Geen      Franse
## 2      Franse      Franse
## 3      Franse      Italiaanse
## 4      Italiaanse      Andere
## 5      Italiaanse      Franse
## 6      Italiaanse      Andere
## 7      Geen      Franse
## 8      Geen      Andere
```

## 9	Geen	Franse
## 10	Geen	Italiaanse
## 11	Franse	Franse
## 12	Franse	Andere
## 13	Franse	Franse
## 14	Geen	Franse
## 15	Geen	Andere
## 16	Italiaanse	Franse
## 17	Geen	Italiaanse
## 18	Franse	Franse
## 19	Geen	Franse
## 20	Franse	Andere
## 21	Geen	Italiaanse
## 22	Geen	Franse
## 23	Geen	Andere
## 24	Franse	Franse
## 25	Geen	Italiaanse
## 26	Franse	Andere
## 27	Franse	Franse
## 28	Geen	Andere
## 29	Italiaanse	Franse
## 30	Italiaanse	Italiaanse
## 31	Italiaanse	Andere
## 32	Franse	Andere
## 33	Geen	Franse
## 34	Geen	Andere
## 35	Geen	Italiaanse
## 36	Geen	Franse
## 37	Italiaanse	Franse
## 38	Italiaanse	Italiaanse
## 39	Geen	Andere
## 40	Franse	Franse
## 41	Franse	Andere
## 42	Geen	Franse
## 43	Geen	Italiaanse
## 44	Geen	Italiaanse
## 45	Geen	Andere
## 46	Italiaanse	Franse
## 47	Franse	Andere
## 48	Geen	Franse
## 49	Italiaanse	Franse
## 50	Geen	Andere
## 51	Italiaanse	Italiaanse
## 52	Geen	Italiaanse
## 53	Italiaanse	Andere
## 54	Geen	Franse
## 55	Italiaanse	Andere
## 56	Franse	Andere
## 57	Italiaanse	Franse
## 58	Italiaanse	Italiaanse
## 59	Geen	Andere
## 60	Geen	Franse
## 61	Italiaanse	Italiaanse
## 62	Italiaanse	Andere

## 63	Geen	Italiaanse
## 64	Franse	Andere
## 65	Italiaanse	Franse
## 66	Geen	Franse
## 67	Franse	Franse
## 68	Italiaanse	Andere
## 69	Italiaanse	Italiaanse
## 70	Italiaanse	Franse
## 71	Geen	Andere
## 72	Geen	Franse
## 73	Franse	Andere
## 74	Italiaanse	Franse
## 75	Italiaanse	Andere
## 76	Italiaanse	Italiaanse
## 77	Geen	Italiaanse
## 78	Geen	Franse
## 79	Franse	Andere
## 80	Italiaanse	Franse
## 81	Italiaanse	Andere
## 82	Italiaanse	Italiaanse
## 83	Geen	Andere
## 84	Geen	Franse
## 85	Italiaanse	Franse
## 86	Italiaanse	Andere
## 87	Italiaanse	Italiaanse
## 88	Franse	Franse
## 89	Franse	Andere
## 90	Geen	Andere
## 91	Italiaanse	Franse
## 92	Geen	Franse
## 93	Italiaanse	Andere
## 94	Italiaanse	Italiaanse
## 95	Geen	Andere
## 96	Italiaanse	Franse
## 97	Geen	Italiaanse
## 98	Franse	Andere
## 99	Geen	Franse
## 100	Italiaanse	Franse
## 101	Geen	Andere
## 102	Italiaanse	Italiaanse
## 103	Italiaanse	Andere
## 104	Franse	Andere
## 105	Geen	Franse
## 106	Franse	Franse
## 107	Italiaanse	Franse
## 108	Geen	Franse
## 109	Italiaanse	Italiaanse
## 110	Geen	Andere
## 111	Italiaanse	Franse
## 112	Italiaanse	Andere
## 113	Franse	Andere
## 114	Geen	Franse
## 115	Franse	Franse
## 116	Italiaanse	Franse



## 117	Geen	Andere
## 118	Italiaanse	Italiaanse
## 119	Franse	Andere
## 120	Italiaanse	Andere
## 121	Italiaanse	Italiaanse
## 122	Italiaanse	Franse
## 123	Geen	Franse
## 124	Franse	Franse
## 125	Italiaanse	Andere
## 126	Geen	Andere
## 127	Italiaanse	Italiaanse
## 128	Franse	Andere
## 129	Geen	Franse
## 130	Italiaanse	Franse
## 131	Geen	Andere
## 132	Italiaanse	Andere
## 133	Franse	Franse
## 134	Franse	Andere
## 135	Italiaanse	Andere
## 136	Geen	Andere
## 137	Italiaanse	Franse
## 138	Geen	Franse
## 139	Italiaanse	Italiaanse
## 140	Franse	Andere
## 141	Italiaanse	Franse
## 142	Geen	Andere
## 143	Italiaanse	Andere
## 144	Geen	Franse
## 145	Franse	Franse
## 146	Italiaanse	Franse
## 147	Italiaanse	Italiaanse
## 148	Geen	Andere
## 149	Italiaanse	Andere
## 150	Italiaanse	Franse
## 151	Franse	Andere
## 152	Italiaanse	Andere
## 153	Geen	Franse
## 154	Franse	Franse
## 155	Italiaanse	Italiaanse
## 156	Italiaanse	Franse
## 157	Franse	Andere
## 158	Geen	Franse
## 159	Italiaanse	Andere
## 160	Geen	Andere
## 161	Italiaanse	Andere
## 162	Italiaanse	Franse
## 163	Franse	Franse
## 164	Franse	Andere
## 165	Italiaanse	Italiaanse
## 166	Franse	Franse
## 167	Geen	Franse
## 168	Geen	Andere
## 169	Italiaanse	Andere
## 170	Franse	Andere

## 171	Italiaanse	Andere
## 172	Franse	Franse
## 173	Geen	Franse
## 174	Geen	Andere
## 175	Italiaanse	Andere
## 176	Franse	Andere
## 177	Italiaanse	Andere
## 178	Franse	Franse
## 179	Franse	Andere
## 180	Italiaanse	Andere
## 181	Franse	Franse
## 182	Geen	Franse
## 183	Italiaanse	Andere
## 184	Franse	Franse
## 185	Geen	Andere
## 186	Italiaanse	Franse
## 187	Italiaanse	Andere
## 188	Franse	Andere
## 189	Geen	Andere
## 190	Franse	Franse
## 191	Italiaanse	Franse
## 192	Geen	Andere
## 193	Franse	Franse
## 194	Franse	Andere
## 195	Italiaanse	Andere
## 196	Franse	Franse
## 197	Geen	Andere
## 198	Italiaanse	Franse
## 199	Franse	Franse
## 200	Geen	Andere
## 201	Italiaanse	Andere
## 202	Franse	Franse
## 203	Franse	Andere
## 204	Geen	Andere
## 205	Italiaanse	Andere
## 206	Geen	Andere
## 207	Italiaanse	Franse
## 208	Franse	Franse
## 209	Franse	Andere
## 210	Geen	Andere
## 211	Italiaanse	Andere
## 212	Geen	Andere
## 213	Italiaanse	Andere
## 214	Franse	Franse
## 215	Franse	Andere
## 216	Geen	Andere
## 217	Franse	Franse
## 218	Italiaanse	Andere
## 219	Geen	Andere
## 220	Franse	Franse
## 221	Franse	Andere
## 222	Geen	Andere
## 223	Franse	Franse
## 224	Franse	Andere

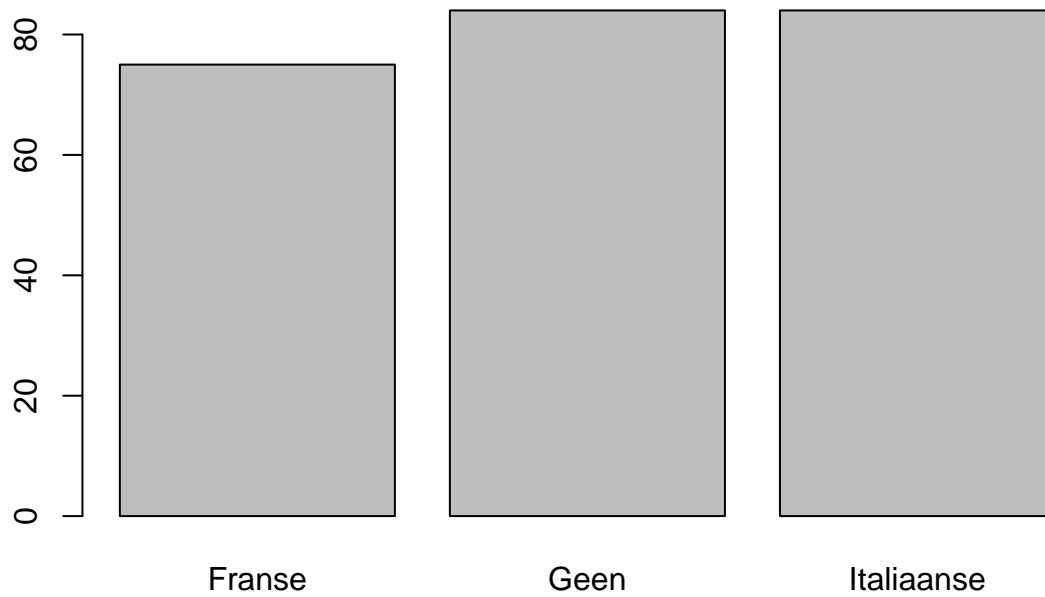
```
## 225      Geen      Andere
## 226     Franse     Franse
## 227      Geen      Andere
## 228      Geen      Andere
## 229     Franse     Franse
## 230     Franse     Andere
## 231      Geen      Andere
## 232     Franse     Franse
## 233     Franse     Andere
## 234      Geen      Andere
## 235     Franse     Franse
## 236      Geen      Andere
## 237     Franse     Andere
## 238     Franse     Franse
## 239      Geen      Andere
## 240     Franse     Franse
## 241     Franse     Franse
## 242     Franse     Franse
## 243     Franse     Andere
```

### 1. staafdiagram

```
tabel <- table(data$Muziek)
tabel
```

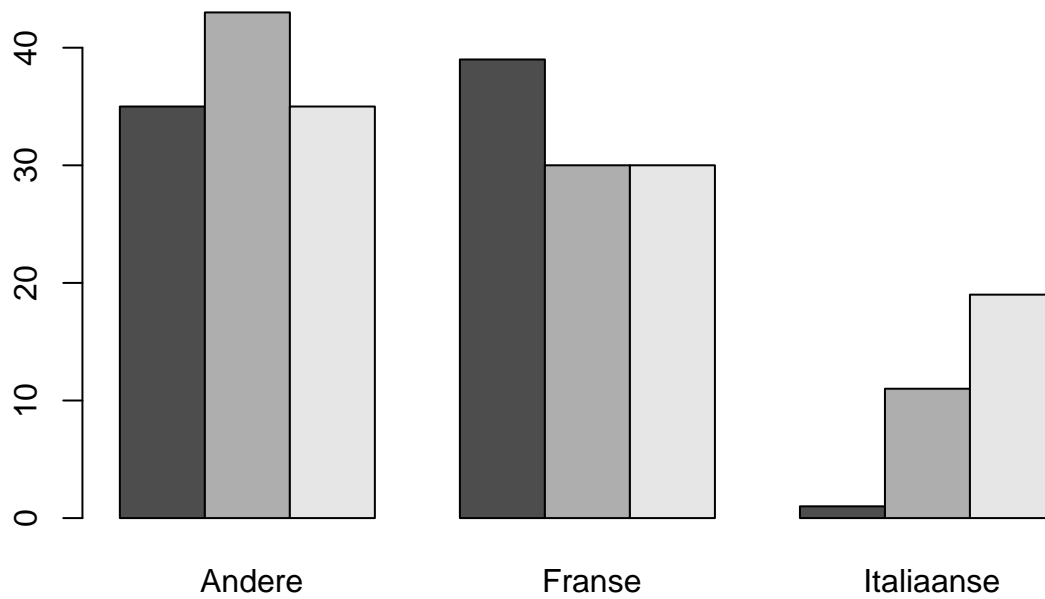
```
##
##      Franse      Geen Italiaanse
##          75          84          84
```

```
barplot(tabel)
```



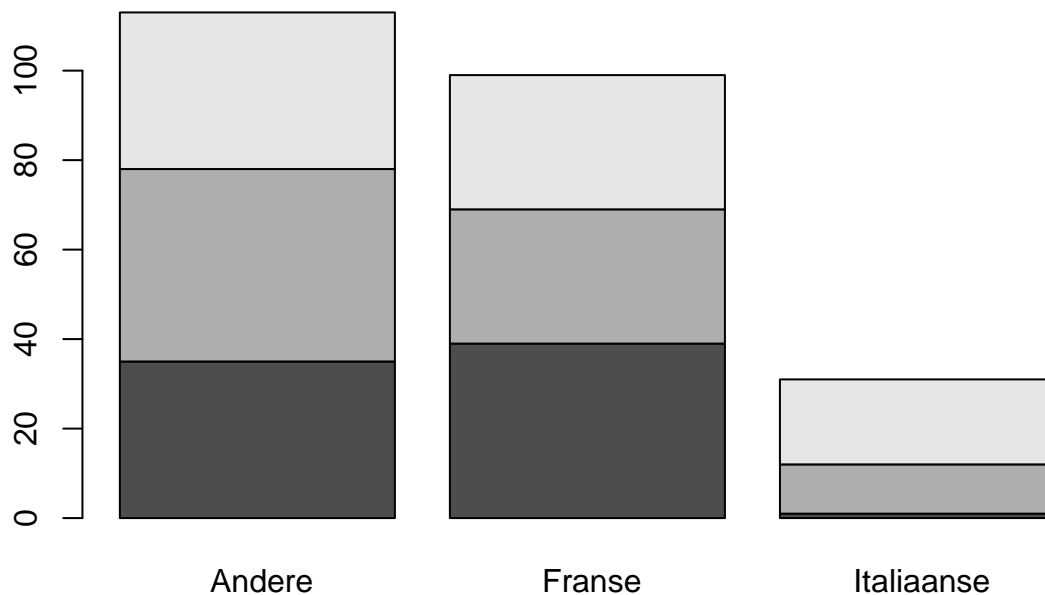
## 2. geclusterde staafdiagram

```
tabel <- table(data)  
barplot(tabel, beside = TRUE)
```



### 3. rependiagram - stacked bar chart

```
tabel <- table(data)
barplot(tabel)
```



## oefening 6.3.

*(oefening zelf gemaakt, geen oplossing)*

### opgave

Lees het databestand “Aardbevingen.csv” in.

1. Maak een histogram en een boxplot van de variabele “Magnitudes”.
2. Maak een lijngraaf met het aantal aardbevingen per maand.
3. Onderzoek of er een verband bestaat tussen de variabelen “Type” en “Source”. Bereken ook de Cramér’s V-waarde. Wat is de conclusie?

### oplossing

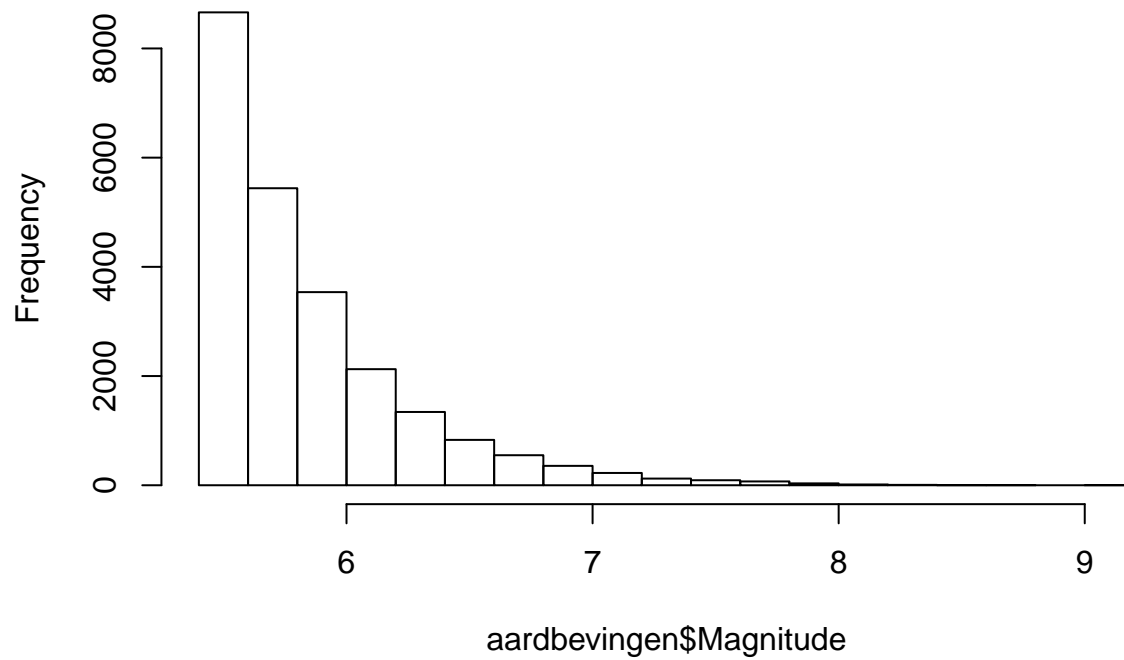
inlezen data

```
aardbevingen <- read.csv("C:\\Users\\tijsm\\Google Drive\\HoGent 2018-2019\\2e semester\\Onderzoekstechniek\\Aardbevingen.csv")
View(aardbevingen)
```

histogram van “magnitudes”

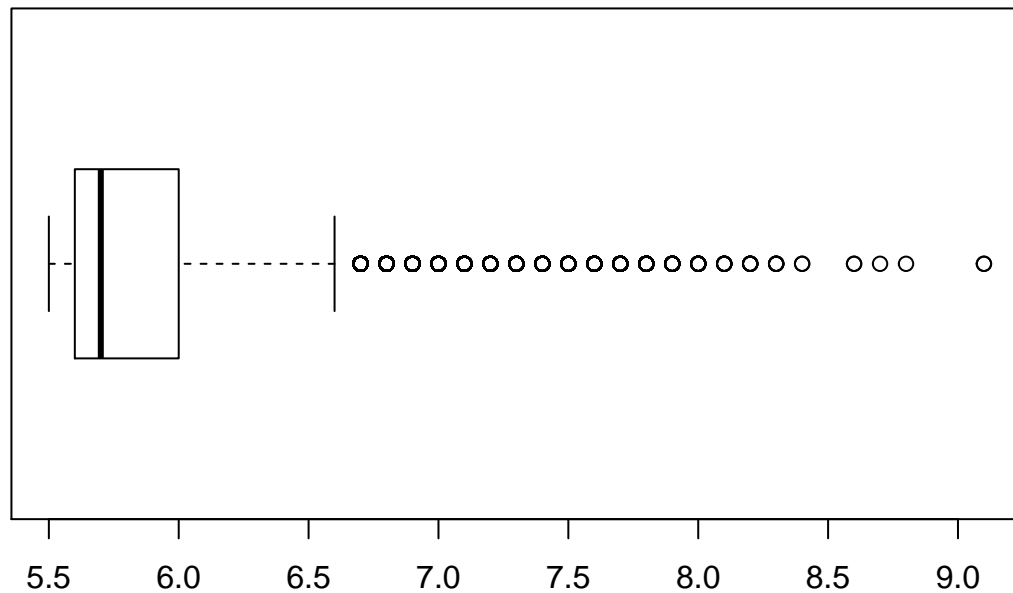
```
hist(aardbevingen$Magnitude)
```

## Histogram of aardbevingen\$Magnitude



boxplot van magnitudes

```
boxplot(aardbevingen$Magnitude, horizontal = TRUE)
```



lijngrafiek aardbevingen per maand

```
typeof(aardbevingen$Date)
```

```
## [1] "integer"
```

```
# omzetten integer naar data
```

```
aardbevingen$Date <- as.Date(aardbevingen$Date,"%Y/%m/%d")
```

```
typeof(aardbevingen$Date)
```

```
## [1] "double"
```

```
str(aardbevingen)
```

```
## 'data.frame':   23412 obs. of  21 variables:
## $ Date          : Date, format: "0001-02-19" "0001-04-19" ...
## $ Time          : Factor w/ 20470 levels "00:00:03","00:00:04",...: 11786 9867 15449 160...
## $ Latitude      : num  19.25 1.86 -20.58 -59.08 11.94 ...
## $ Longitude     : num  145.6 127.4 -174 -23.6 126.4 ...
## $ Type          : Factor w/ 4 levels "Earthquake","Explosion",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Depth         : num  132 80 20 15 15 ...
## $ Depth.Error   : num  NA NA NA NA NA NA NA NA NA NA ...
## $ Depth.Seismic.Stations : int  NA NA NA NA NA NA NA NA NA NA ...
## $ Magnitude     : num  6 5.8 6.2 5.8 5.8 6.7 5.9 6 6 5.8 ...
## $ Magnitude.Type : Factor w/ 11 levels "", "MB", "MD", "MH",...: 7 7 7 7 7 7 7 7 7 7 ...
```



```
## $ Magnitude.Error      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ Magnitude.Seismic.Stations: int  NA NA NA NA NA NA NA NA NA NA NA ...
## $ Azimuthal.Gap        : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ Horizontal.Distance   : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ Horizontal.Error      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ Root.Mean.Square      : num  NA NA NA NA NA NA NA NA NA NA NA ...
## $ ID                   : Factor w/ 23412 levels "AK11232962","AK11248623",...: 2580 2581 2582 2583 ...
## $ Source                : Factor w/ 13 levels "AK","ATLAS","CI",...: 5 5 5 5 5 5 5 6 5 ...
## $ Location.Source       : Factor w/ 48 levels "AEI","AEIC","AG",...: 21 21 21 21 21 21 21 21 21 ...
## $ Magnitude.Source      : Factor w/ 24 levels "1000","1009",...: 12 12 12 12 12 12 12 12 12 ...
## $ Status                : Factor w/ 2 levels "Automatic","Reviewed": 1 1 1 1 1 1 1 1 1 1 ...
```

```
aardbevingen$month <- as.Date(cut(aardbevingen$Date, breaks = "month"))
```

```
#####
```

onderzoek als er een verband is tussen “type” en “source”

*hypothese*  $h_0$  - er is geen verband  $h_1$  - er is wel een verband

*cramèrs v*

```
cramersV(aardbevingen$Type, aardbevingen$Source)
```

```
## Warning in chisq.test(...): Chi-squared approximation may be incorrect
```

```
## [1] 0.06709128
```

de *cramèrs V* is 0.067

een *cramèrs v* van 0 wil zeggen dat er geen verband is tussen de variabelen een *cramèrs v* van 1 wil zeggen dat er een perfect verband is tussen de variabelen

In dit geval kunnen we dus stellen dat er geen verband is tussen het type en de oorzaak. We kunnen de nulhypothese dan ook NIET verwerpen

## oefening 6.4.

(oefening zelf gemaakt, geen oplossing)

### opgave

- Bepaal handmatig de regressierechte  $y = a + bx$ .
- Bepaal handmatig de correlatie- en determinatiecoëfficiënt ( $R, R^2$ )
- Geef uitleg bij de gevonden statistieken *##* oplossing

### oplossing

#### 1. handmatig

```
x <- c(10, 12, 8, 13, 9, 10, 7, 14, 11, 6)
y <- c(11, 14, 9, 13, 9, 9, 8, 14, 10, 6)
```

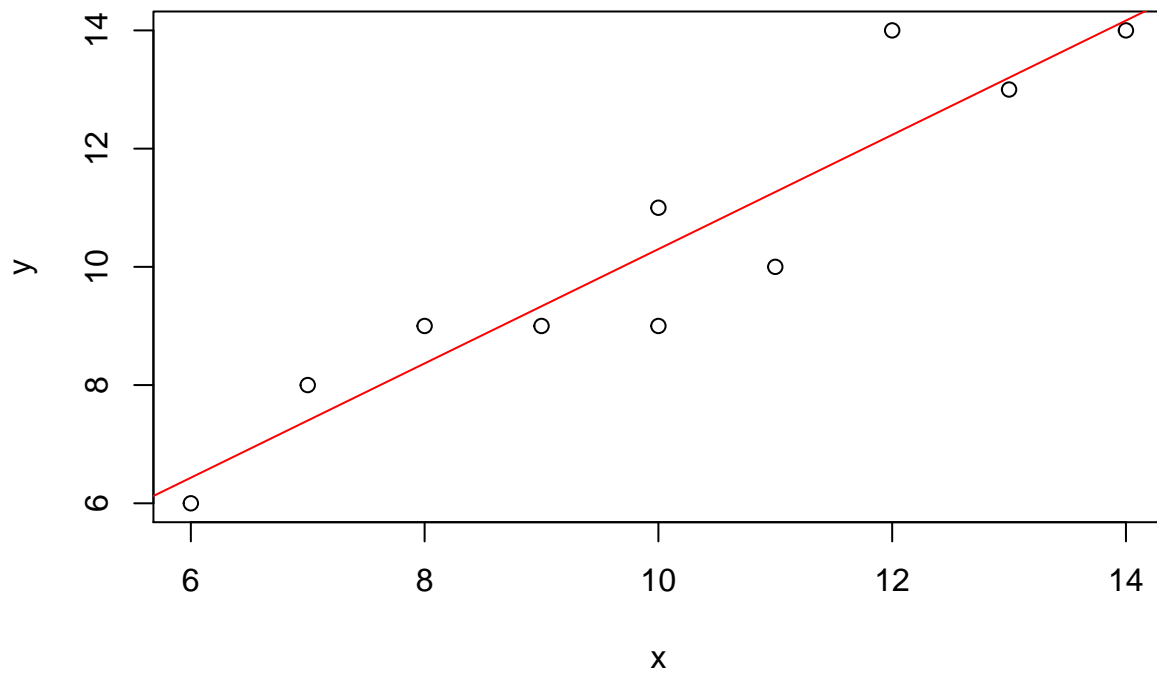
```
mx <- mean(x)
my <- mean(y)
```

```
xx <- x - mx
yy <- y - my
beta1 <- sum(xx*yy)/sum(xx^2)
beta0 <- my - beta1 * mx
```

2.

```
plot(x, y)

regr <- lm(y ~ x)
abline(regr, col="red")
```



### 3. uitleg

hoe dicht de punten bij de lijn ligt hoe beter het verband

### 4. bepaal handmatig r en r kwadraat

bepalen r

```
cov(x = x, y = y)
```

```
## [1] 6.444444
```

pearson's correlatie co

```
rr <- cor (x, y)
rr
```

```
## [1] 0.9352406
```

in het kwadraat

= +- 81% 81% van het gemiddelde wordt bepaald door ?????????????? de rest door toeval

## oefening 6.5.

/

*(oefening zelf gemaakt, geen oplossing)*

### opgave

Gegeven 6 scatterplots in volgende ???guur en onderstaande correlatiecoëf???ciënten. Match de coëf???ciënten met de scatterplots. Er is dus één scatterplot waarvan geen correlatie gegeven staat hieronder.

### oplossing

als r

- = 0 ==> geen verband te zien
- = 1 ==> stijgend verband
- = -1 ==> dalend verband

r1 = E

- 0.6
- de algemene trend van de waarden is stijgend, r moet dus groter zijn dan 0
- liggen niet heel dicht gegroepeerd, r dus ook nie “dicht” bij 1

r2 = D

- 0
- 0 betekend dat er geen relatie is, de waarden van grafiek D liggen random door elkaar

r3 = C

- -0.9
- negatieve waarde = dalend verband
- dicht bij -1 dus de waarden moeten dicht bij elkaar liggen

r4 = A

- 0.9
- positieve waarde = stijgend verband
- dicht bij 1 dus de waarden moeten dicht bij elkaar liggen

r5 = F

- 0.3
- waarde dicht bij 0, er is dus geen duidelijk verband
- in grafiek F liggen de waarden ver uit elkaar en zij ze op het eerste zicht “onvoorspelbaar”. Het valt wel op dat er een algemene stijgende trend is. Dit komt overeen met de 0.3

## oefening 6.6.

*(oefening zelf gemaakt, geen oplossing)*

### opgave

Lees het databestand “Cats.csv” in. 1. Voer een lineaire regressieanalyse uit op de variabelen Lichaamsgewicht (Bwt, afhankelijke variabele) en Gewicht hart (Hwt, onafhankelijke variabele). 2. Maak een spreidingsdiagram van beide variabelen. 3. Bereken en teken de regressielijn. 4. Bereken de correlatie- en de determinatiecoëfficiënt. 5. Geef een interpretatie van deze resultaten.

### oplossing

bestand inladen

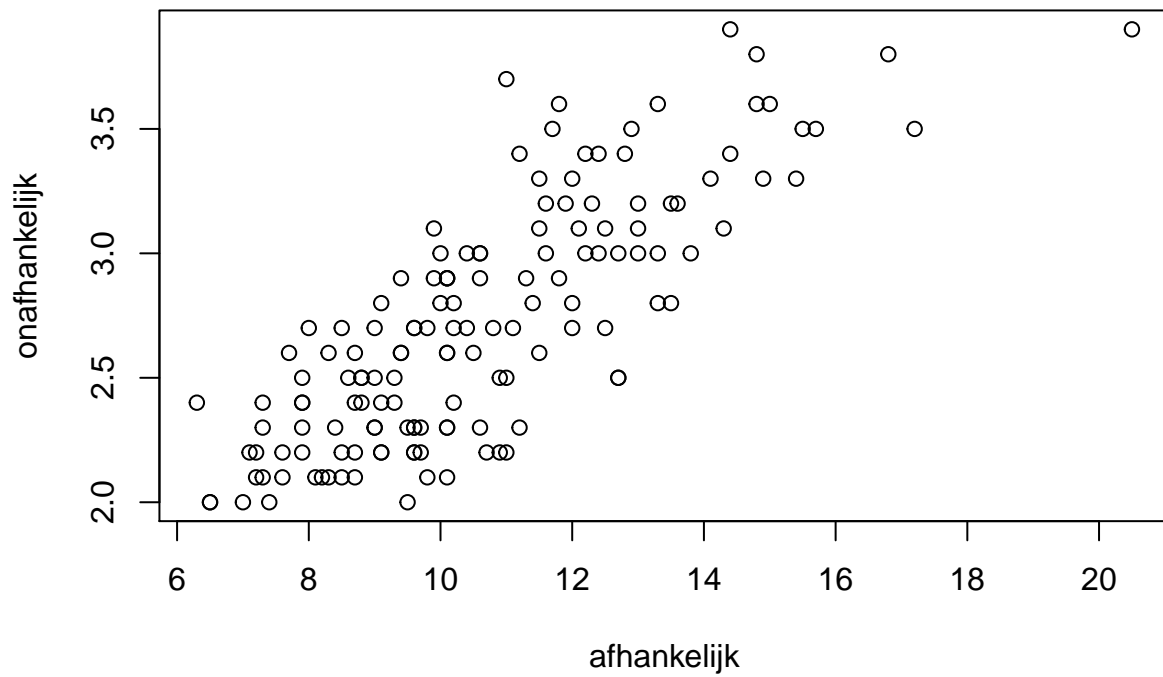
```
cats <- read.csv("C:\\Users\\tijsm\\Google Drive\\HoGent 2018-2019\\2e semester\\Onderzoekstechnieken\\  
View(cats)
```

#### 1. lineaire regressie

```
afhankelijk <- cats$Bwt  
onafhankelijk <- cats$Hwt  
  
regr <- lm(onafhankelijk ~ afhankelijk)  
regr  
  
##  
## Call:  
## lm(formula = onafhankelijk ~ afhankelijk)  
##  
## Coefficients:  
## (Intercept)  afhankelijk  
##      1.0187      0.1604
```

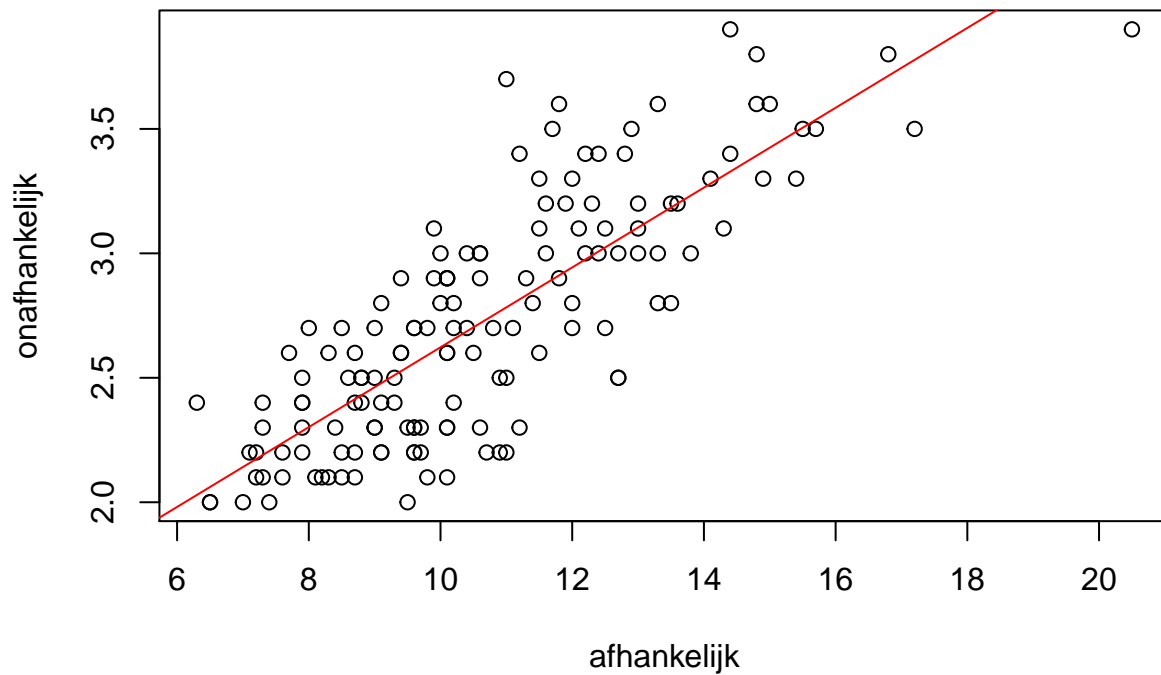
#### 2. spreidingsdiagram

```
plot(afhankelijk, onafhankelijk)
```



### 3. regressierechte

```
plot(afhankelijk, onafhankelijk)
abline(regr, col="red")
```



#### 4. correlatie en determinantcoëfficiënt

*manueel:*

```
covariantie<-cov(onafhankelijk, afhankelijk)
covariantie

## [1] 0.9495697
r = covariantie / (sd(onafhankelijk) * sd(afhankelijk))
correlatie <- r
correlatie

## [1] 0.8041348
determinantcoefficient <- r*r
determinantcoefficient

## [1] 0.6466328
```

*met r:*

```
co <- cor(onafhankelijk,afhankelijk, method = 'pearson')
co

## [1] 0.8041348
determinantcoefficient <-co * co
determinantcoefficient

## [1] 0.6466328
```

## 5. interpretatie

er is een correlatie van 0.80. Dit duidt op een sterk stijgend verband.  
de spreiding rond de regressierechte is vrij groot

## oefening 6.7.

### opgave

1. VoereenlineaireregressieanalyseuitopdevariabelenLichaamsgewicht(Bwt)enGewicht hart (Hwt) per geslacht.
2. Maak een spreidingsdiagram van beide variabelen voor elk van de geslachten.
3. Bereken en teken telkens de regressielijn.
4. Bereken de correlatie- en de determinatiecoëfficiënt.
5. Geef een interpretatie aan deze resultaten.

### oplossing

bestand inladen

```
cats <- read.csv("C:\\Users\\tijsm\\Google Drive\\HoGent 2018-2019\\2e semester\\Onderzoekstechnieken\\  
View(cats)
```

#### 1. regressie

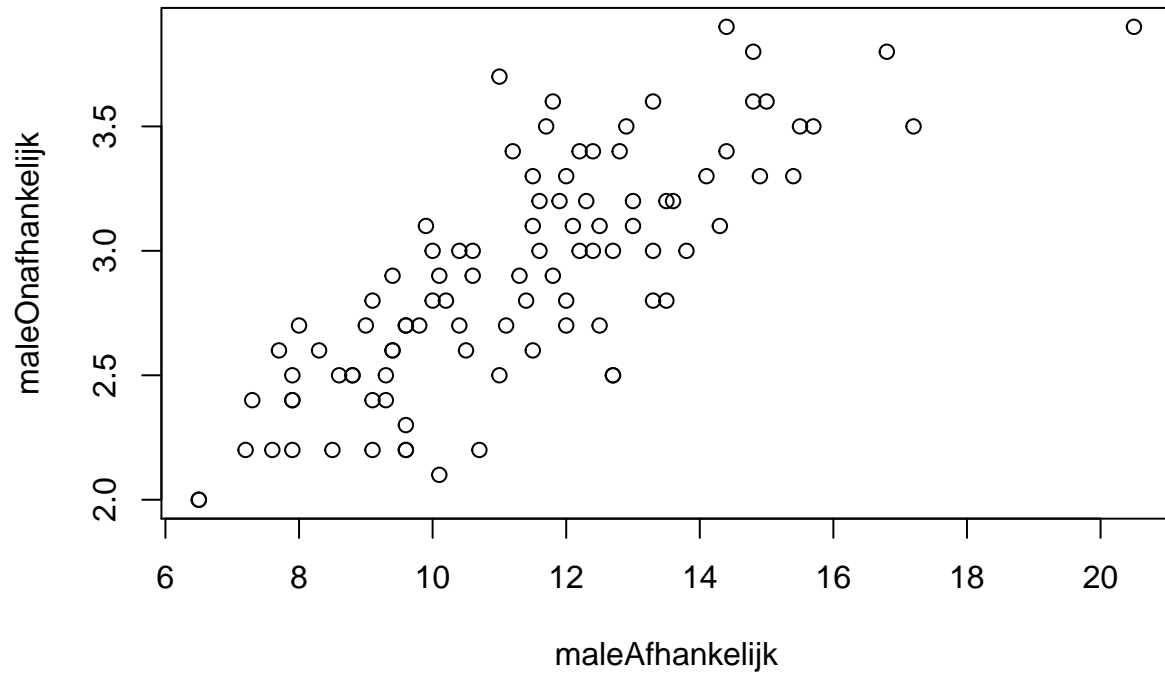
```
maleCats <- subset(cats, cats$Sex=="M")  
femaleCats <- subset(cats, cats$Sex=="F")  
  
#hangt gewicht van een hart of van het lichaamsgewicht?  
maleAfhankelijk <- maleCats$Bwt  
maleOnafhankelijk <- maleCats$Hwt  
  
maleRegressie <- lm(maleOnafhankelijk ~ maleAfhankelijk)  
maleRegressie  
  
##  
## Call:  
## lm(formula = maleOnafhankelijk ~ maleAfhankelijk)  
##  
## Coefficients:  
##      (Intercept)  maleAfhankelijk  
##          1.2479          0.1459  
  
femaleAfhankelijk <- femaleCats$Bwt  
femaleOnafhankelijk <- femaleCats$Hwt  
  
femaleRegressie <- lm(femaleOnafhankelijk ~ femaleAfhankelijk)  
femaleRegressie  
  
##  
## Call:  
## lm(formula = femaleOnafhankelijk ~ femaleAfhankelijk)  
##  
## Coefficients:  
##      (Intercept)  femaleAfhankelijk
```

```
##          1.3715          0.1074
```

## 2. spreiding

male:

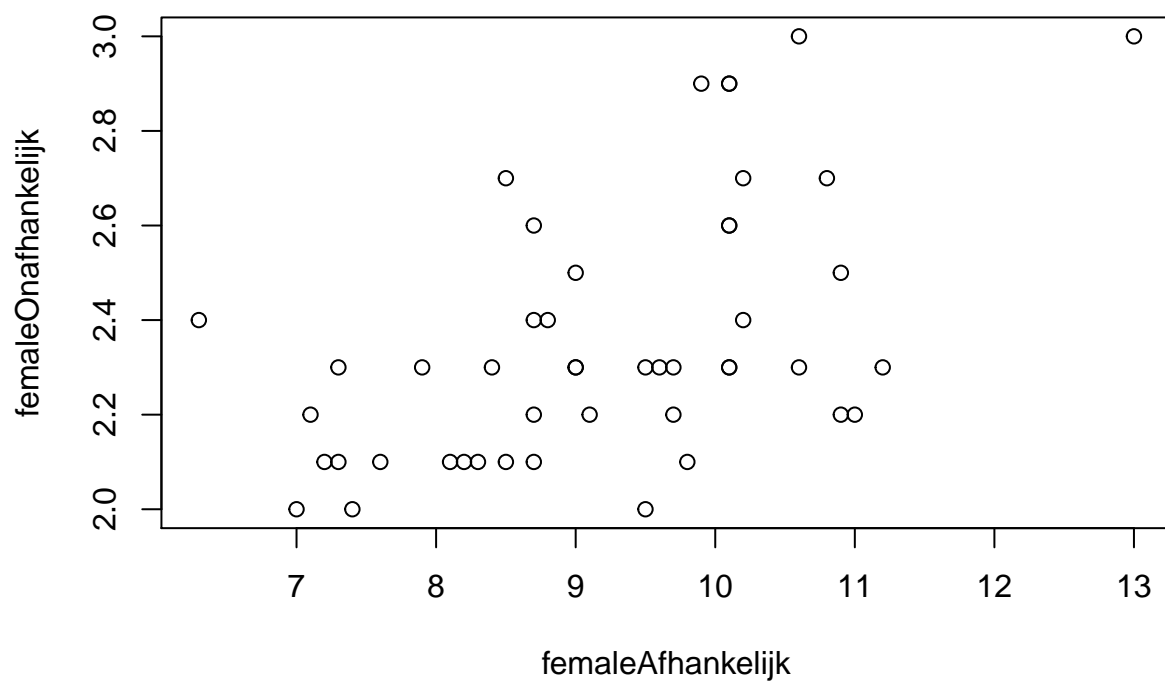
```
plot(maleAfhankelijk, maleOnafhankelijk)
```



female

```
plot(femaleAfhankelijk, femaleOnafhankelijk)
```

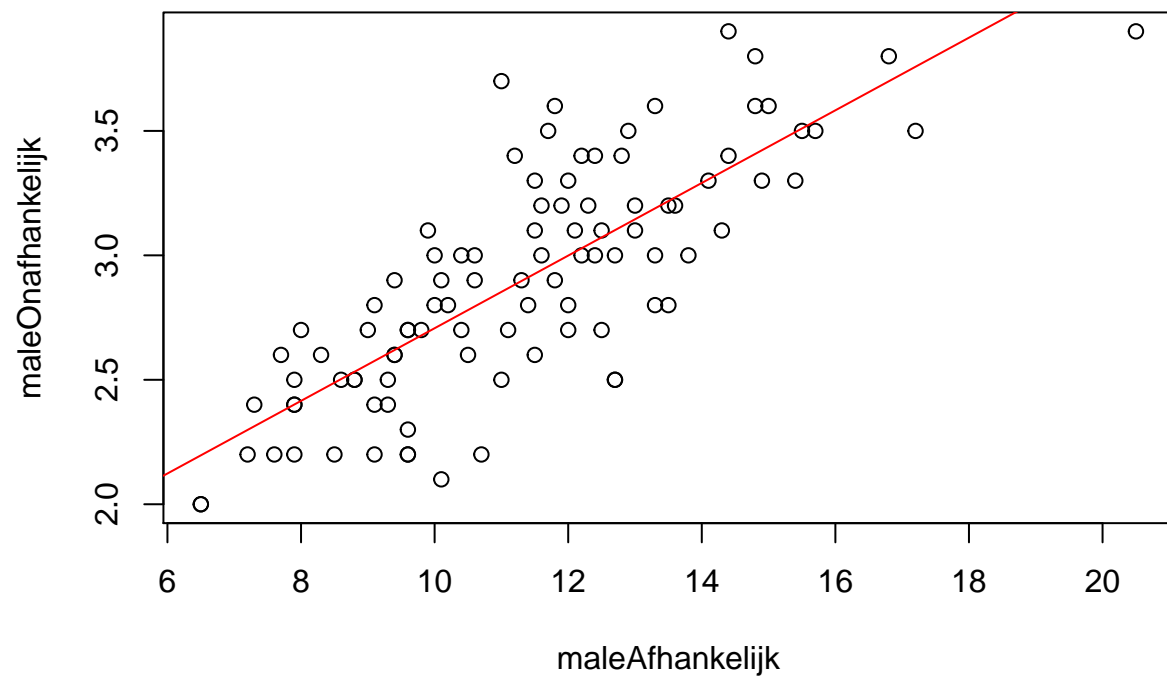




### 3. regressierechte

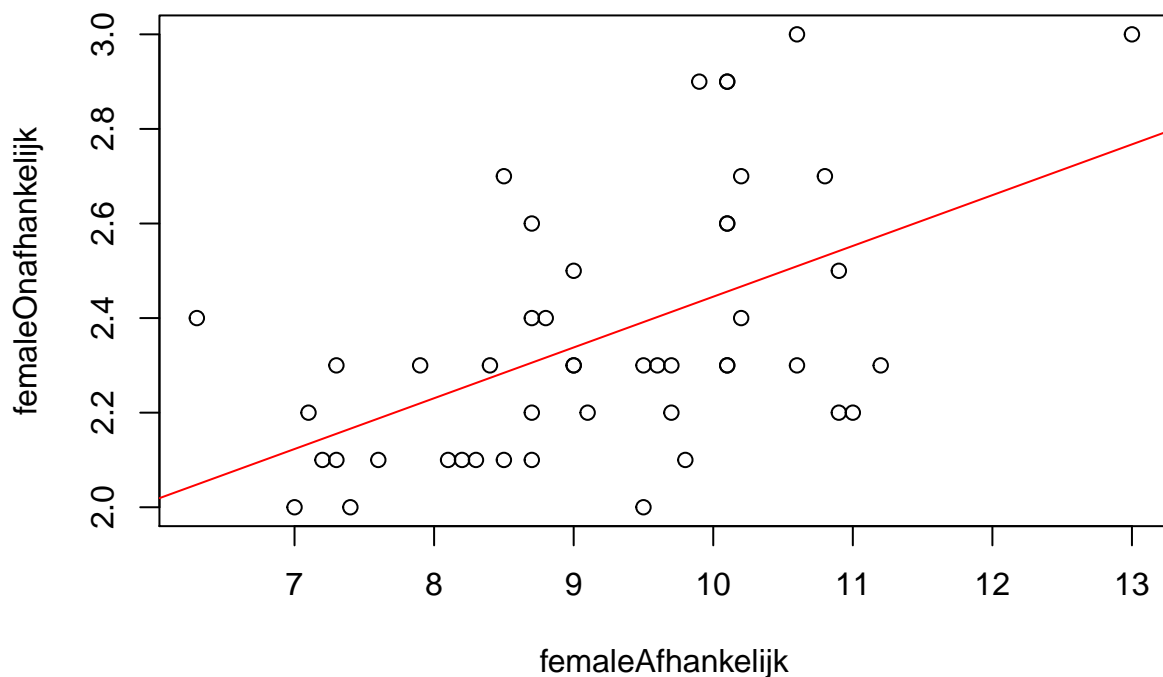
male:

```
plot(maleAfhankelijk, maleOnafhankelijk)
abline(maleRegressie, col = "red")
```



female:

```
plot(femaleAfhankelijk, femaleOnafhankelijk)  
abline(femaleRegressie, col = "red")
```



#### 4. correlatie en determinantcoefficient

male

```
maleCo <- cor(maleOnafhankelijk,maleAfhankelijk, method = 'pearson')
maleCo
```

```
## [1] 0.7930443
```

```
maleDeterminantcoefficient <-maleCo * maleCo
maleDeterminantcoefficient
```

```
## [1] 0.6289193
```

female

```
femaleCo <- cor(femaleOnafhankelijk,femaleAfhankelijk, method = 'pearson')
femaleCo
```

```
## [1] 0.5320497
```

```
femaleDeterminantcoefficient <-femaleCo * femaleCo
femaleDeterminantcoefficient
```

```
## [1] 0.2830768
```

#### 5. interpretatie

male

- sterk sijgend lineair verband

female

- zwak stijgend lineair verband

## oefening 6.8.

*(oefening zelf gemaakt, geen oplossing)*

### opgave

Lees het databestand “Pizza.csv” in.

1. Voer een volledige lineaire regressieanalyse uit op de variabelen Rating en CostPerSlice. Trek hieruit de juiste conclusies en ga deze ook grafisch na.
2. Onderzoek een mogelijk verband tussen Rating en Neighbourhood. Welke methode kan je hiervoor gebruiken? Kan je deze gegevens van Rating hiervoor in dezelfde vorm gebruiken?
3. Geef een interpretatie aan deze resultaten.
4. Stel de kruistabel grafisch voor met een staafdiagram. Voorzie een legende.

### oplossing

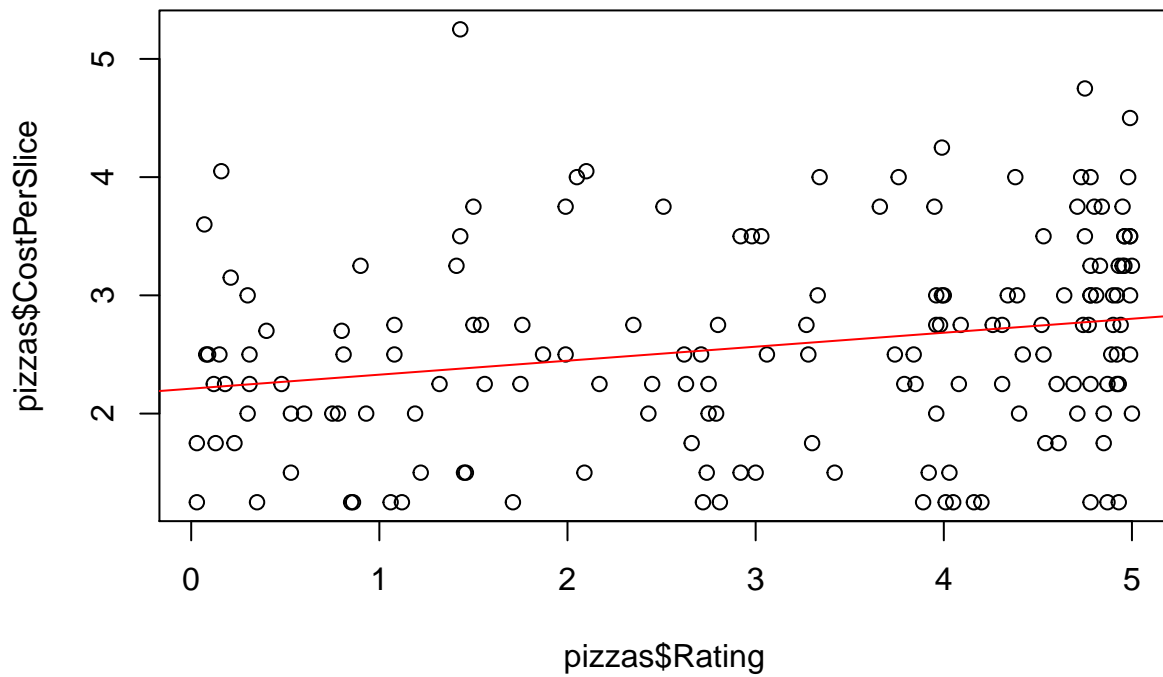
inlezen data

```
pizzas <- read.csv("C:\\Users\\tijsm\\Google Drive\\HoGent 2018-2019\\2e semester\\Onderzoekstechnieken\\data\\Pizza.csv")
View(pizzas)
```

#### 1. regressie

```
plot(pizzas$Rating, pizzas$CostPerSlice)

regr <- lm(pizzas$CostPerSlice ~ pizzas$Rating)
abline(regr, col='red')
```



```
regr
```

```
##
## Call:
## lm(formula = pizzas$CostPerSlice ~ pizzas$Rating)
##
## Coefficients:
## (Intercept) pizzas$Rating
##      2.2107      0.1183
```

er is weinig samenhang

## 2. relatie tussen rating en neighbourhood