

# 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)
```

berekenen chi-kwadraat

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

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

```
#sum(afwijking)
```

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

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

```
chi_sq
```

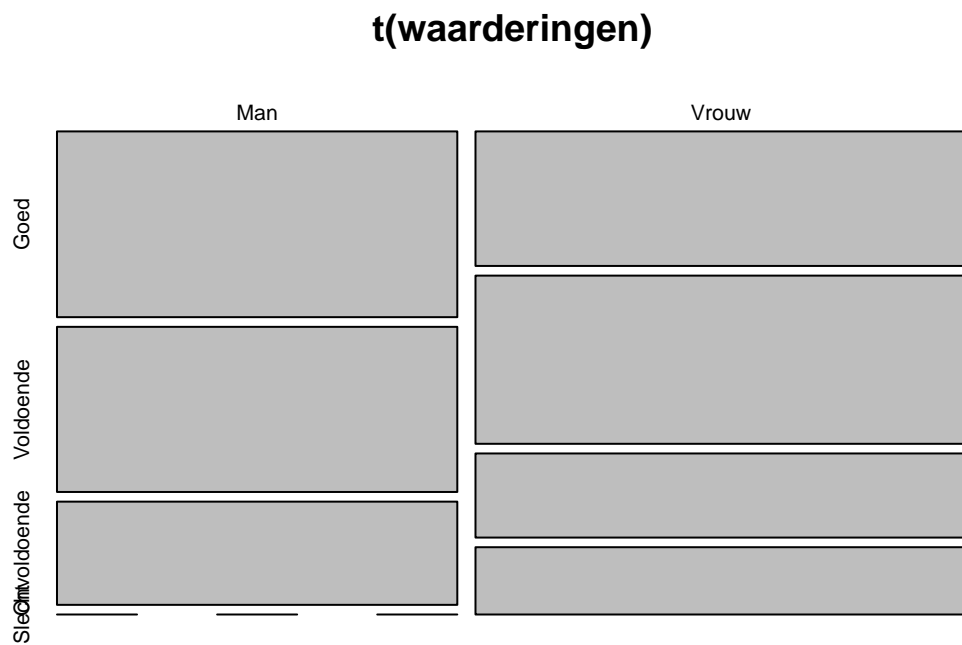
```
## [1] 3.810518
```

berekenen van Cramér's V

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

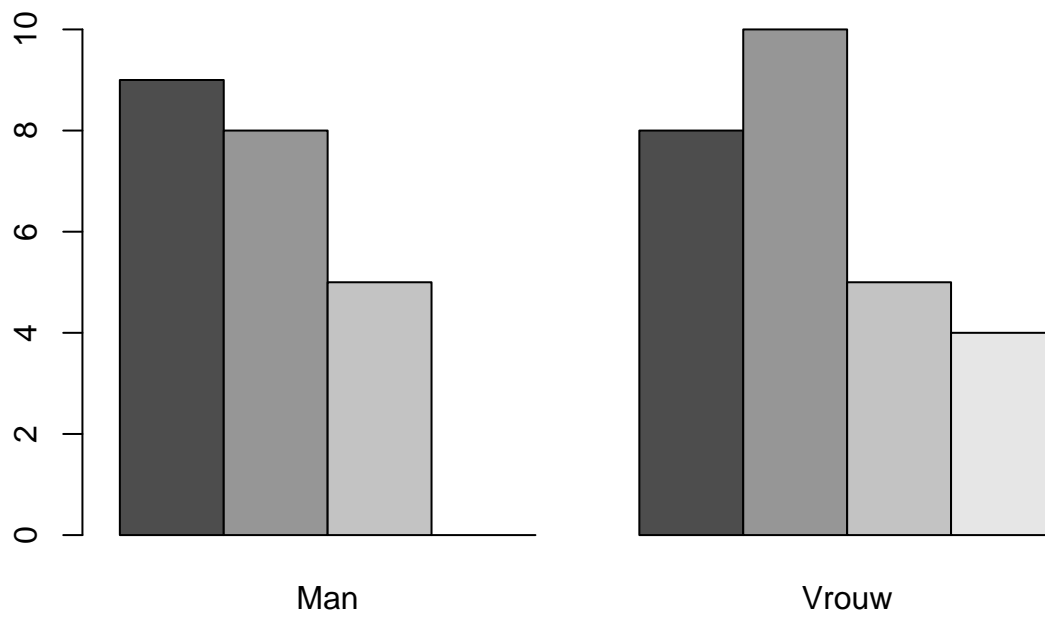
maken tabel

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



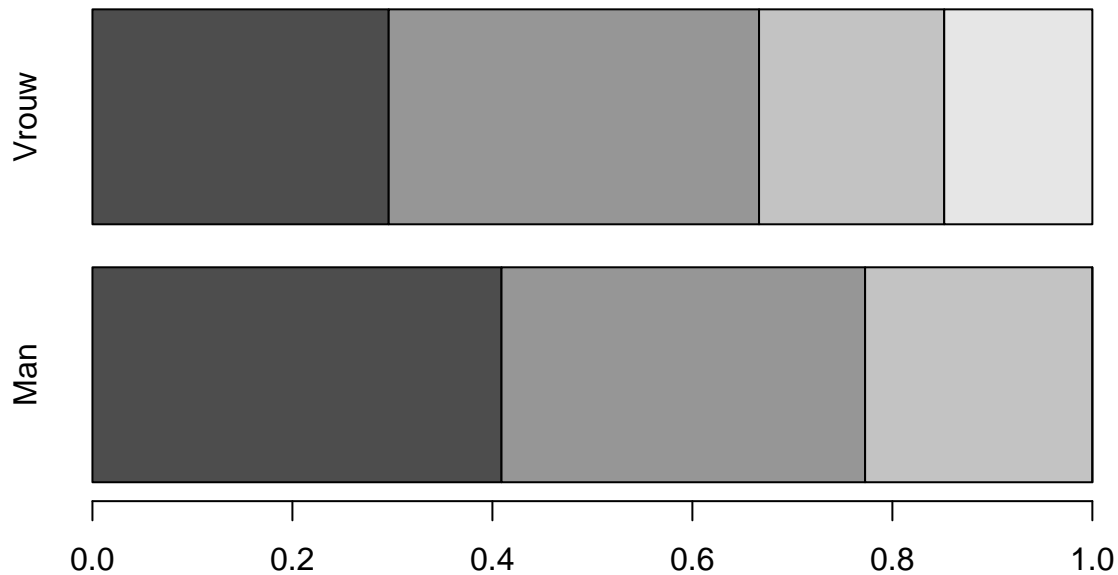
maken geclustere bar chart

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



maken van percentage chart

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



## 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 flessen 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  $\chi^2$  toetsingsgrootte.
5. Bereken manueel de Cramér's V. Wat kan je hieruit besluiten?

### oplossing

inlezen csv bestand

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

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

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

2.

```
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.

```
#library(gmodels)
#library(lsr)

#crosstable(kruistabel)
```

4.

```
#cramersV(kruistabel)
```

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

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

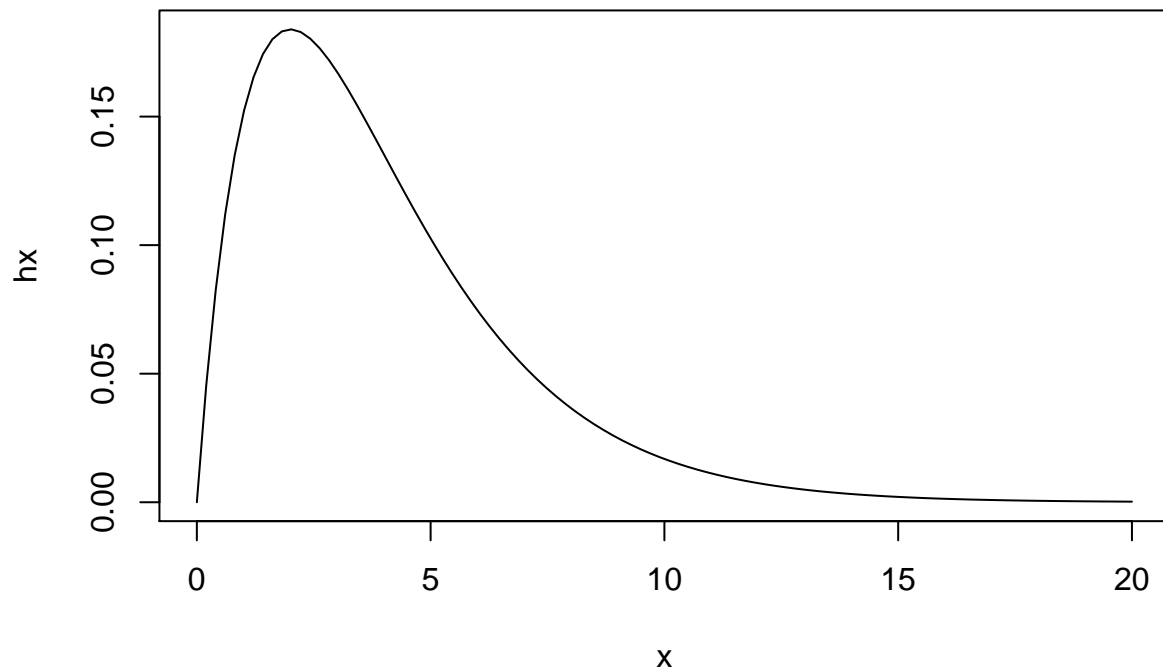
```
qchisq(0.95, df = 4)
```

```
## [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).
2. Stel de percentages verkochte wijnen voor in rependiagram (stacked bar chart) ## oplossing

```
library(vcd)
```

```
assocstats(data)
```

## 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

/

## oefening 6.4.

*(oefening zelf gemaakt, geen oplossing)*

### opgave

- Bepaal handmatig de regressierechte  $y = \beta_0 + \beta_1 x$ .
- 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)

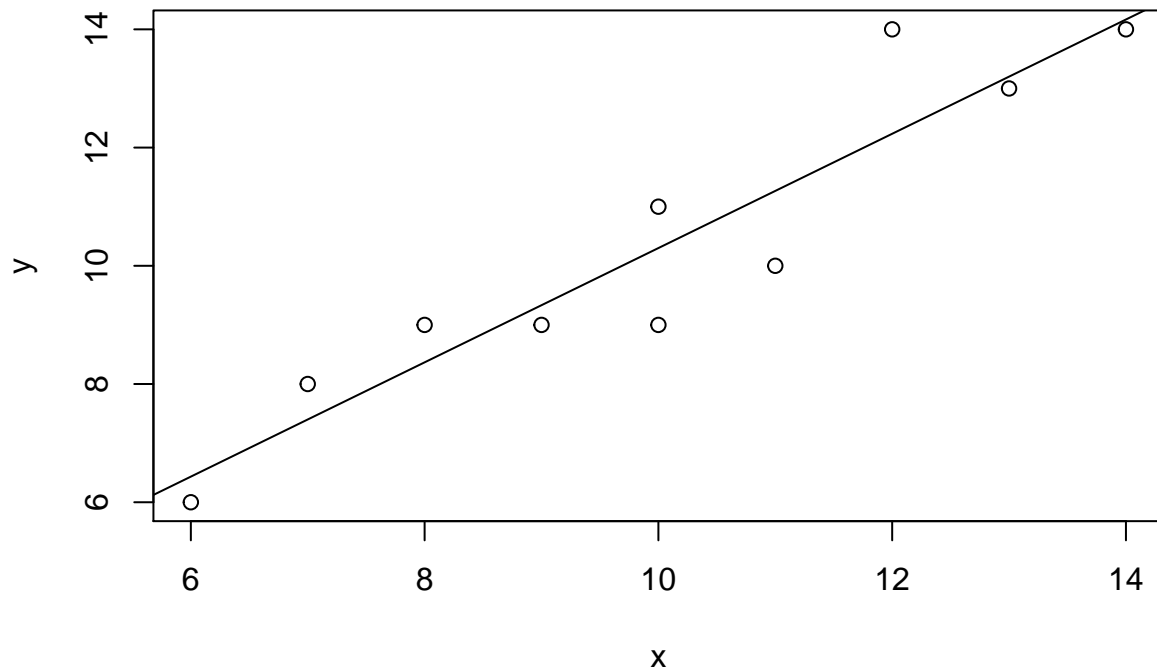
mx <- mean(x)
my <- mean(y)

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)
```



### 3. uitleg

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

### 4. bepaal handmatig r en r kwadraat

bepale 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

## 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 determinatiecoëf??ciënt. 5. Geef een interpretatie van deze resultaten.

### oplossing

/

## oefening 6.7.

*(oefening zelf gemaakt, geen oplossing)*

### opgave

1. Voer een lineaire regressieanalyse uit op de variabelen Lichaamsgewicht (Bwt) en Gewicht 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 determinatiecoëf??ciënt.
5. Geef een interpretatie aan deze resultaten.

### oplossing

## 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 de zelfde vorm gebruiken?
3. Geef een interpretatie aan deze resultaten.
4. Stel de kruistabel grafisch voor met een staafdiagram. Voorzie een legende.

### oplossing

/