

# Oefening 7.3

## Opgave

Laad de dataset Aids2 uit package MASS die informatie bevat over 2843 patiënten die voor 1991 in Australië met AIDS besmet werden. Onderzoek of er een relatie is tussen de variabele geslacht (Sex) en de manier van besmetting (T.categ).

1. Ga op de gebruikelijke manier te werk: visualiseren van de data,  $\chi^2$ ,  $g$  en  $p$ -waarde berekenen ( $\alpha = 0,05$ ), en tenslotte een conclusie formuleren.
2. Bepaal de gestandaardiseerde residuën om te bepalen welke categoriën extreme waarden bevatten.

## Gegeven

```
library(MASS)
attach(Aids2)

alpha <- 0.05
```

## Oplossing

1

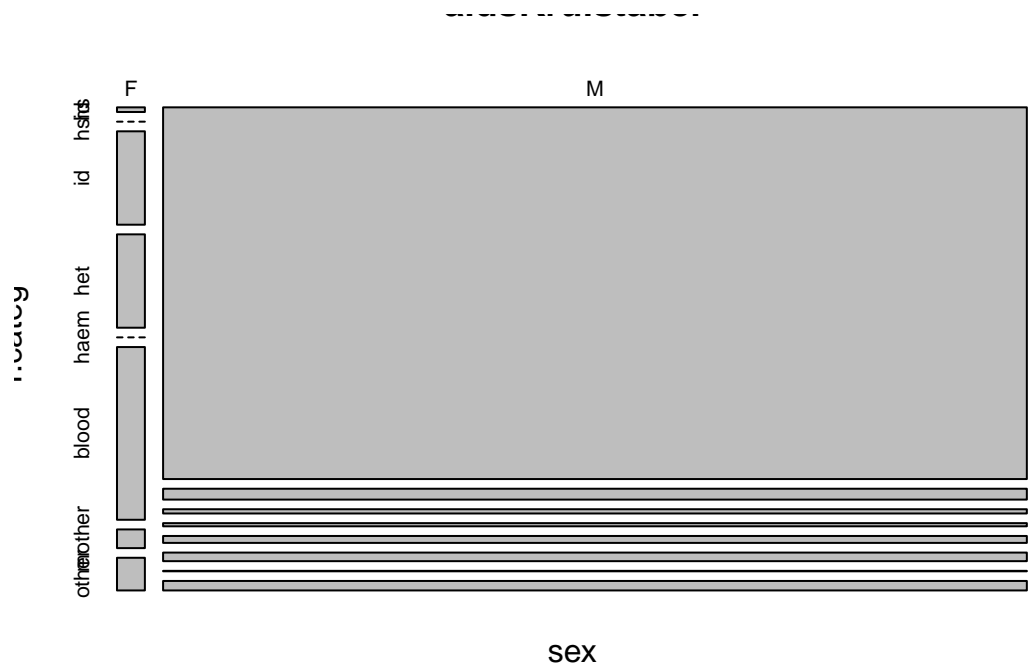
Ga op de gebruikelijke manier te werk: visualiseren van de data,  $\chi^2$ ,  $g$  en  $p$ -waarde berekenen ( $\alpha = 0,05$ ), en tenslotte een conclusie formuleren.

### Visualiseren van de data

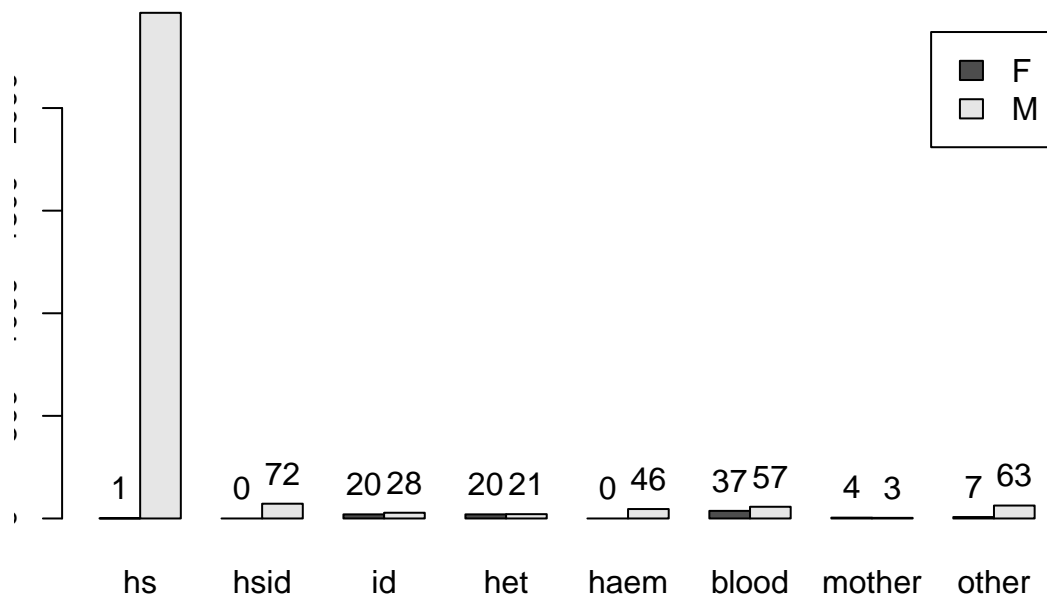
```
aidsKruistabel <- table(sex, T.categ)
aidsKruistabel # kruistabel

##      T.categ
## sex   hs hsid   id  het haem blood mother other
##   F     1   0   20   20    0   37      4      7
##   M 2464   72   28   21   46   57      3     63

plot(aidsKruistabel) # niet echt nuttig
```



```
bp <- barplot(aidsKruistabel, beside = TRUE, legend=rownames(aidsKruistabel))
text(bp, aidsKruistabel, aidsKruistabel, pos = '3') # text aan barplot toevoegen
```



$\chi^2$

```
model <- chisq.test(sex, T.categ)
```

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

```
chisq <- model$statistic
chisq
```

```
## X-squared
```

```
## 1083.373
```

$g$

```
g <- qchisq(1-alpha,df=model$parameter)
g
```

```
## [1] 14.06714
```

$p$ -waarde

```
p <- model$p.value
```

## Conclusie

Geen representatieve steekproef:  $\chi^2 (1083.3729135) > \text{grenswaarde } g (14.0671404) * p\text{-waarde } (1.1570708 \times 10^{-229}) < \alpha (0.05)$

## 2

Bepaal de gestandaardiseerde residuën om te bepalen welke categoriën extreme waarden bevatten.

```
model$stdres
```

```
##      T.categ
## sex      hs      hsid      id      het      haem      blood
##  F -24.160129 -1.545075  15.462745  16.907792 -1.229233  20.513873
##  M  24.160129  1.545075 -15.462745 -16.907792  1.229233 -20.513873
##      T.categ
## sex      mother      other
##  F   8.216321   3.341856
##  M  -8.216321  -3.341856
```

Gestandaardiseerde residuën duiden aan welke klassen de grootste bijdragen leveren aan de waarde van de grootheid. Algemene regel: waarden groter dan 2 of kleiner dan  $-2$  zijn “extreem”.