

ZSO 4

Voorbereiding

Voor deze ZSO ga je aan de slag met een nieuwe dataset: `Titanic2.RData`. De data kan je downloaden door op onderstaande link te klikken met een rechtermuisklik.

[DOWNLOAD DE DATA](#)

Tip

Een goede workflow in `RStudio` is van cruciaal belang op lange termijn. In dit vak is het geen verplichting, maar werken met projecten in `RStudio` kan je leven een pak makkelijker maken.

Daarom raden we je aan om een nieuw project te maken per ZSO. Concreet betekent dit dat je tevens op je pc een map maakt per ZSO met daarin alle nodige files voor die ZSO:

- databestanden
- `OLP2 functies.R` file
- Je script met analyses (of Quarto document)

Meer uitleg over werken met projecten en Quarto documenten kan je vinden onder Extra materiaal ([KLIK HIER](#))

OPDRACHTEN

Vraag 1

Na het zinken van de Titanic deden er verschillende verhalen de ronde. Na vele jaren onderzoek heeft men een nagenoeg volledig bestand kunnen aanleggen van eigenschappen van de passagiers. Nu zijn we in staat om enkele prangende vragen te onderzoeken. De eerste vraag

die mensen boeide was of men tijdens die ramp echt wel voorrang heeft gegeven aan vrouwen. Hadden vrouwen een grotere overlevingskans dan mannen?

Het bijhorende databestand (Titanic2.RData) bevat gegevens over een groot deel van de passagiers op de Titanic. De volgende drie variabelen zijn opgenomen in het bestand: - Survived (“Yes”, “No”); - Class (“1st”, “2nd”, “3rd”, “Crew”); - Gender (“Male”, “Female”).

1.1 Ga na of er een verband is tussen het geslacht en de overlevingskans. Voer de gepaste analyses uit om deze vraag te beantwoorden. Geef de relevante output van je analyses en bespreek deze.

1.2 Maak ook een gepaste visualisatie van het resultaat.

Vraag 2

Sommige onderzoekers van de ramp met de Titanic beweerden kort na de ramp dat de klasse waarin passagiers geboekt hadden meer invloed had op de kans op overleven dan kenmerken zoals het geslacht.

2.1 Is er een effect van de klasse waarin men boekte? Staaf je antwoord met output van je analyses.

2.2 Klopt de uitspraak dat de klasse meer invloed had op de overlevingskans dan het geslacht? Staaf je antwoord met gegevens uit de output van je analyses.

[RESPONS ACHTERAAN DOCUMENT]

RESPONSEN

Vooraleer we de analyses kunnen uitvoeren moeten we uiteraard de data inladen in R en tevens de OLP2 Functies activeren. Dit doen we bijvoorbeeld door gebruik te maken van de commando's `load()` en `source()`.

```
load("Titanic2.RData")
source("OLP2 Functies.R")
library(car)
```

Vraag 1

1.1

a) Analyses

Om het verband na te gaan maken we gebruik van een kruistabel, met in de kolommen 'Gender' (Eigenlijk kunnen we veronderstellen dat dit de onafhankelijke variabele is) en in de rijen 'Survived'. Bovendien vragen we kolompercentages op.

```
kruistabel.kolom(Titanic2$Survived, Titanic2$Gender)
```

| | x | | |
|-----|-----------|-----------|-----------|
| y | Female | Male | Sum |
| no | 26.80851 | 78.79838 | 67.69650 |
| yes | 73.19149 | 21.20162 | 32.30350 |
| Sum | 100.00000 | 100.00000 | 100.00000 |

```
chi.kwadraat.test(table(Titanic2$Survived, Titanic2$Gender))
```

Pearson's Chi-squared test

```
data: table(Titanic2$Survived, Titanic2$Gender)
X-squared = 456.87, df = 1, p-value < 2.2e-16
95 percent confidence interval:
 0.4138267 0.4973825
sample estimates:
Cramer's V
 0.4556048
```

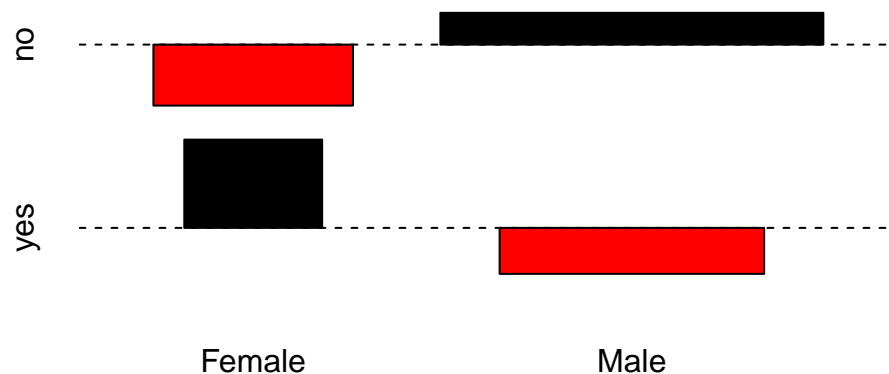
b) Resultaten bespreken met verwijzing naar output

Uit de kruistabel blijkt dat vrouwen (73,19%) duidelijk een grotere overlevingskans hebben dan mannen (21,20%). Uit de Chi-kwadraattoets ($\chi^2(1) = 456.874$, $p < 0.001$) kunnen we bovendien afleiden dat de kans dat de nulhypothese opgaat zeer klein is. M.a.w. de kans dat in de populatie van Titanic-passagiers geen verschil in overlevingskans was tussen vrouwen en mannen is zo klein dat we deze verwerpen. (We verwerpen H_0 !) Het verband tussen ‘Survival’ en ‘Gender’ is dus statistisch significant. De sterkte van het verband is af te leiden uit de Cramer’s V (0.456) en dit geldt ook voor de populatie, 95% BI [0.414,0.497].

1.2

Om een grafiek te maken, gebruiken we `assocplot()`. Let op de volgorde van de variabelen binnen dit commando!

```
assocplot(table(Titanic2$Gender,Titanic2$Survived))
```



Vraag 2

2.1

a) Analyses

```
kruistabel.kolom(Titanic2$Survived, Titanic2$Class)
```

| | | x | | | | |
|-----|--|-----------|-----------|-----------|-----------|-----------|
| y | | 1st | 2nd | 3rd | Crew | Sum |
| no | | 37.53846 | 58.59649 | 74.78754 | 76.04520 | 67.69650 |
| yes | | 62.46154 | 41.40351 | 25.21246 | 23.95480 | 32.30350 |
| Sum | | 100.00000 | 100.00000 | 100.00000 | 100.00000 | 100.00000 |

```
chi.kwadraat.test(table(Titanic2$Survived, Titanic2$Class))
```

Pearson's Chi-squared test

```
data: table(Titanic2$Survived, Titanic2$Class)
X-squared = 190.4, df = 3, p-value < 2.2e-16
95 percent confidence interval:
 0.2506724 0.3344483
sample estimates:
Cramer's V
 0.2941201
```

b) Resultaten bespreken met verwijzing naar output

Uit de kruistabel kan je aflezen dat de overlevingskans het grootst is voor passagiers uit 1ste klasse (62,462%), het tweede grootst voor passagiers uit 2de klasse (41,404%) en het laagst voor bemanningsleden (23,955%). Dit wijst duidelijk op een samenhang tussen 'Survived' en 'Class'. Deze samenhang is ook statistisch significant ($\chi^2(3) = 190.401$, $p < 0.001$). We kunnen dus stellen dat deze samenhang zich ook in de populatie van Titanic-passagiers zal voordoen. De sterkte van verband wordt uitgedrukt door Cramer's V (0.294). Uit het 95% BI [0.251,0.334] rond Cramer's V kunnen we afleiden dat de Cramer's V ergens tussen 0.251 en 0.334 zal bedragen voor de populatie.

2.2

Om na te gaan of ‘Class’ inderdaad meer invloed had op de overlevingskans (‘Survived’) dan geslacht(‘Gender’) vergelijken we de Cramer’s V voor beide kruistabellen. De Cramer’s V voor de associatie tussen geslacht en de ramp overleven bedraagt 0.456. Voor het verband tussen klasse en de ramp overleven is de Cramer’s V 0.294. De uitspraak dat klasse meer invloed had op de overlevingskans dan geslacht klopt dus niet.