

Statistiek B – C2 - RESPONS

Oefening 1

Zouden onze leerlingen uit de verschillende categorieën ('Richting5cat') anders scoren als het gaat over het begrijpen van techniek ('Begrijpen.voor')? Indien dit zo is, welke groepen verschillen dan van elkaar?

```
> tapply(Techniek$Begrijpen.voor, Techniek$Richting5cat, FUN=mean, na.rm=TRUE)
      1      2      3      4      5
0.6216667 0.6452206 0.7098642 0.6337891 0.5429025
```

→ in de steekproef scoren leerlingen uit categorie 3 gemiddeld het hoogst op 'Begrijpen.voor' ($\mu = 0.71$) en leerlingen uit categorie 5 gemiddeld het laagst ($\mu = 0.54$).

```
> leveneTest(Techniek$Begrijpen.voor, Techniek$Richting5cat)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  4  2.0016 0.09181 .
      1873
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

→ p is > 0.05 , dus WEL gelijke binnengroepvariantie

```
> Model.Begrijpen <- aov(Techniek$Begrijpen.voor ~ Techniek$Richting5cat)
> summary(Model.Begrijpen)
              Df Sum Sq Mean Sq F value Pr(>F)
Techniek$Richting5cat    4    5.27   1.3182   49.34 <2e-16 ***
Residuals              1873   50.04   0.0267
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
489 observations deleted due to missingness
```

→ $p < 0.05$: kans dat H_0 opgaat is kleiner dan 5%

⇒ We verwachten in de populatie WEL een verschil in begrijpen van naar studierichting.

```
> etasq(Model.Begrijpen)
              Partial eta^2
Techniek$Richting5cat    0.0953278
Residuals                NA
```

→ η^2 tussen 0.06 en 0.14: het gaat om een medium effect

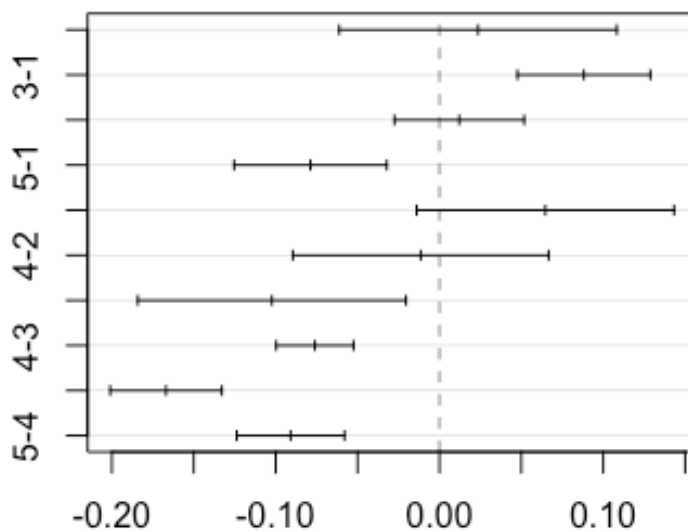
```
> Model.Begrijpen.Tukey <- TukeyHSD(Model.Begrijpen)
> Model.Begrijpen.Tukey
Tukey multiple comparisons of means
 95% family-wise confidence level
```

```
Fit: aov(formula = Techniek$Begrijpen.voor ~ Techniek$Richting5cat)
```

```
$`Techniek$Richting5cat`
```

	diff	lwr	upr	p adj
2-1	0.02355392	-0.06121659	0.10832444	0.9422650
3-1	0.08819755	0.04762624	0.12876886	0.0000000
4-1	0.01212240	-0.02746614	0.05171093	0.9193976
5-1	-0.07876412	-0.12536703	-0.03216122	0.0000413
3-2	0.06464363	-0.01394620	0.14323346	0.1634082
4-2	-0.01143153	-0.08951854	0.06665549	0.9946245
5-2	-0.10231805	-0.18418477	-0.02045132	0.0059194
4-3	-0.07607515	-0.09968814	-0.05246217	0.0000000
5-3	-0.16696167	-0.20105203	-0.13287132	0.0000000
5-4	-0.09088652	-0.12380115	-0.05797189	0.0000000

95% family-wise confidence level



Differences in mean levels of Techniek\$Richting5ca

CONCLUSIE:

Afhankelijk van de studierichting die een leerling volgt, verschilt de mate waarin techniek wordt begrepen. Het gaat om een medium ($\eta^2 = 0.10$) en statistisch significant effect ($p < 0.05$). We verwachten dit effect dus ook in de populatie terug te vinden.

Uit de post-hoc analyse blijkt dat leerlingen uit categorie 5 (Handel/STV) het meest verschillen van de andere leerlingen. Hun score op 'Begrijpen.voor' is significant ($p < 0.05$) lager dan deze van leerlingen uit categorie 2 (Kunst), categorie 3 (Latijn) en categorie 4 (Moderne wetenschappen). Ook tussen de groepen Latijn (categorie 3) en Techniek (categorie 1) is een significant verschil ($p < 0.05$) in het begrijpen van techniek terug te vinden. De leerlingen uit de Latijnse scores significant ($p < 0.05$) beter dan de leerlingen uit de meer technische richtingen. De verschillen tussen leerlingen uit de technische (categorie 1), moderne (categorie 4) of kunstrichtingen (categorie 2) in onze steekproef kunnen we niet doortrekken naar de populatie ($p > 0.05$).

Oefening 2

We bekeken in de eerste oefening of er een verschil was in het begrijpen van techniek afhankelijk van de studierichting. Als we de afhankelijke variabele ('Begrijpen.voor') opdelen in vier categorieën op basis van de kwartielen kunnen we dit verband ook weergeven in een tabel.

- a) Hoe is de verdeling overheen de verschillende categorieën van begrijpen naar studierichting?
- b) Geef dit ook grafisch weer.

OE FENING 2 a

```
> quantile(Techniek$Begrijpen.voor, c(0.25,0.5,0.75), na.rm=TRUE)
  25%    50%    75%
0.5000 0.6875 0.7500
> Techniek$Begrijpen.voor.Cat <- recode(Techniek$Begrijpen.voor,
+                                     "0:0.5= 'Zeer
Laag';0.5:0.6875='Laag';0.6875:0.75='Hoog';0.75:1='Zeer Hoog'",
+                                     as.factor.result=TRUE,
+                                     levels=c("Zeer
Laag", "Laag", "Hoog", "Zeer Hoog"))
```

→ *variabele hercoderen*

```
> kruistabel.kolom(Techniek$Begrijpen.voor.Cat, Techniek$Richting5cat)
```

		x					
y		1	2	3	4	5	Sum
Zeer Laag		30.666667	29.411765	12.460064	28.485577	46.186441	25.559105
Laag		38.666667	29.411765	34.025559	38.100962	37.711864	36.581470
Hoog		13.333333	11.764706	17.731629	13.100962	9.322034	14.164004
Zeer Hoog		17.333333	29.411765	35.782748	20.312500	6.779661	23.695421
Sum		100.000000	100.000000	100.000000	100.000000	100.000000	100.000000

→ *kolompercentages*

```
> chi.kwadraat.test(table(Techniek$Richting5cat, Techniek$Begrijpen.voor.Cat))
```

Pearson's Chi-squared test

```
data: table(Techniek$Richting5cat, Techniek$Begrijpen.voor.Cat)
```

X-squared = 172.76, df = 12, p-value < 2.2e-16

95 percent confidence interval:

0.1428339 0.1959541

```
sample estimates:
```

Cramer's V

0.1751129

→ $p < 0.05$: kans dat H_0 opgaat is kleiner dan 5%

⇒ We verwachten in de populatie WEL een verschil in begrip van techniek naar studierichting.

→ Cramer's $V = 0.18$: een klein effect

CONCLUSIE:

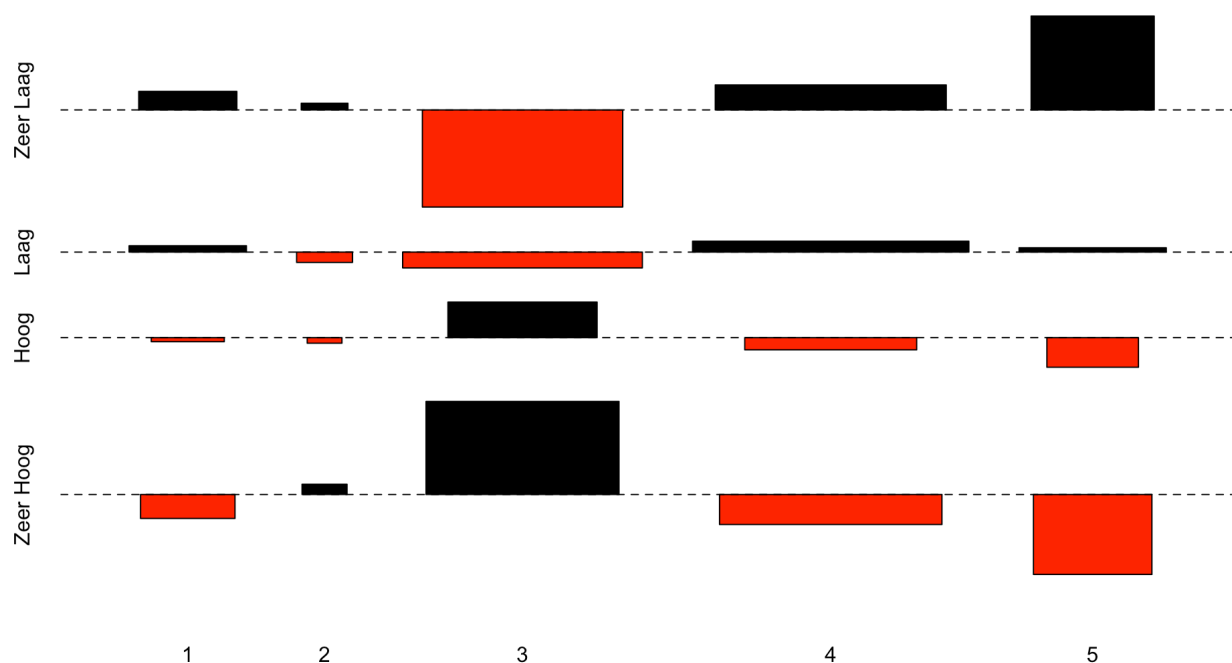
De kolompercentages geven weer hoe binnen de verschillende clusters van studierichtingen wordt gescoord op het begrijpen van techniek. We bespreken enkele 'opvallende' cijfers...

Bij de leerlingen uit technische studierichtingen zit de grootste groep in de categorie "Laag" (38.67%). Het percentage van deze leerlingen in de categorie "Zeet hoog" bedraagt slechts 17.33%. De grootste groep leerlingen die Latijn volgen is terug te vinden in de categorie "Zeet hoog" (35.78%). Deze groep leerlingen scoort dus in het algemeen beter dan de leerlingen uit de andere studierichtingen. Handel/STV-leerlingen scoren dan weer sterk lager. 46.19% van deze leerlingen is terug te vinden in de categorie "Zeet Laag" en slechts 6.78% in de categorie "Zeet Hoog".

De χ^2 -analyse leert ons dat het verband statistisch significant is ($p < 0.05$), maar eerder beperkt (Cramer's $V = 0.18$).

OEFENING 2 b

```
> Tabel_asso <- table(Techniek$Richting5cat, Techniek$Begrijpen.voor.Cat)
> assocplot(Tabel_asso)
```



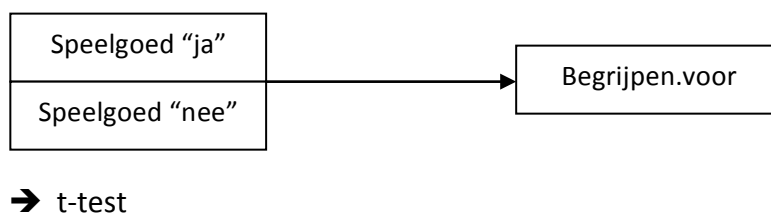
Oefening 3

We zijn ook geïnteresseerd in welke mate de aanwezigheid van technisch speelgoed ('Speelgoed') een invloed heeft op het begrijpen van techniek ('Begrijpen.voor'). We kunnen dit op 2 manieren analyseren nu we beschikken over een kwalitatieve en een kwantitatieve variabele die iets zeggen over het begrijpen van techniek.

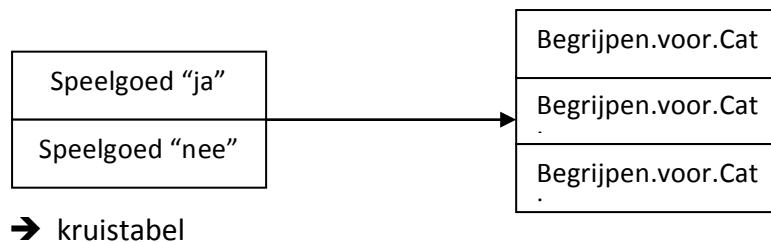
- a) Teken beide modellen en vermeld de analysetechniek.
- b) Maak beide oefeningen en vergelijk de resultaten.

OEFFENING 3 a

Optie 1:



Optie 2:



OEFFENING 3 b

Optie 1:

```
> leveneTest(Techniek$Begrijpen.voor, Techniek$Speelgoed)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  0.1089  0.7415
1890
```

→ p is > 0.05 , dus WEL gelijke binnengroepvariantie

```
> t.test(Techniek$Begrijpen.voor~Techniek$Speelgoed, var.equal=TRUE)
```

Two Sample t-test

```
data:  Techniek$Begrijpen.voor by Techniek$Speelgoed
t = -3.6108, df = 1890, p-value = 0.0003132
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.04621105 -0.01368049
sample estimates:
mean in group 0 mean in group 1
```

0.6251900

0.6551357

→ $p < 0.05$: kans dat H_0 opgaat is kleiner dan 5%

⇒ We verwachten in de populatie WEL een verschil in begrijpen van techniek afhankelijk van al dan niet technisch speelgoed hebben.

→ gemiddeld begrijpen groep 0 (geen technisch speelgoed) = 0.63

gemiddeld begrijpen groep 1 (wel technisch speelgoed) = 0.66

```
> d(Techniek$Begrijpen.voor, Techniek$Speelgoed)
[1] -0.1737467
```

→ $d < 0.2$: het gaat om een verwaarloosbaar effect

CONCLUSIE:

In de steekproef scoren leerlingen gemiddeld 0.66 op begrijpen van techniek, terwijl leerlingen die thuis geen technisch speelgoed hebben iets lager scoren ($\mu = 0.63$). Het effect van al dan niet technisch speelgoed hebben, is statistisch significant ($p < 0.05$). We kunnen het dus doortrekken naar de populatie. Het effect is echter zo klein ($d = -0.17$) dat het van weinig praktische waarde is.

Optie 2:

```
> kruistabel.kolom(Techniek$Begrijpen.voor.Cat, Techniek$Speelgoed)
```

		x		
y		0	1	Sum
Zeer	Laag	29.63526	24.06807	26.00423
	Laag	38.44985	35.57536	36.57505
	Hoog	12.15805	15.07293	14.05920
Zeer	Hoog	19.75684	25.28363	23.36152
Sum		100.00000	100.00000	100.00000

→ kolompercentages

```
> chi.kwadraat.test(table(Techniek$Begrijpen.voor.Cat, Techniek$Speelgoed))
```

Pearson's Chi-squared test

```
data: table(Techniek$Begrijpen.voor.Cat, Techniek$Speelgoed)
```

```
X-squared = 14.2894, df = 3, p-value = 0.002537
```

```
95 percent confidence interval:
```

```
0.03199597 0.12691805
```

```
sample estimates:
```

```
Cramer's V
```

```
0.08690546
```

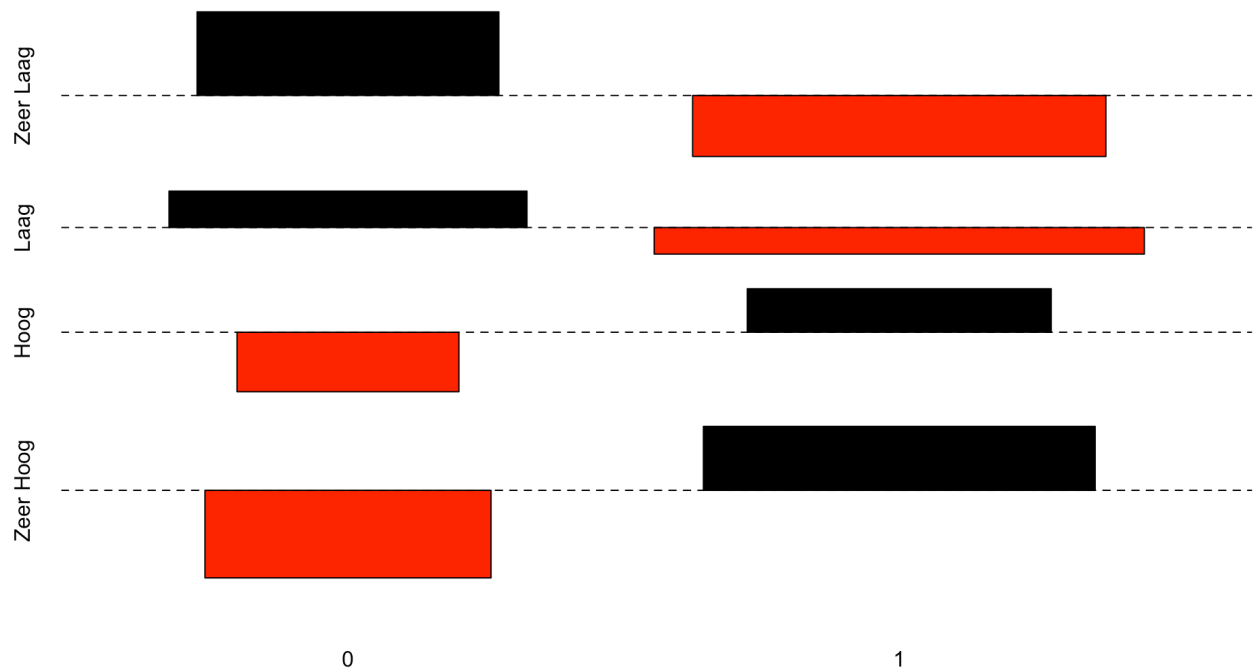
→ $p < 0.05$: kans dat H_0 opgaat is kleiner dan 5%

⇒ We verwachten in de populatie WEL een verschil in begrijpen van techniek naar al dan niet technisch speelgoed bezitten.

→ Cramer's $V = 0.09$: een klein effect

```
> Tabel_asso <- table(Techniek$Speelgoed, Techniek$Begrijpen.voor.Cat)
```

```
> assocplot(Tabel_asso)
```



CONCLUSIE:

De kolompercentages geven weer of leerlingen die al dan niet technisch speelgoed bezitten anders scoren op het begrijpen van techniek. Uit de kruistabel blijkt dat de verschillen eerder klein zijn. Beide groepen leerlingen hebben gelijkaardige percentages binnen de verschillende categorie van 'Begrijpen.voor.Cat'.

Hoewel de χ^2 -analyse aangeeft dat het om een statistisch significant verband gaat ($p < 0.05$), wijst de Cramer's V op een klein effect (Cramer's V = 0.09).