

Statistiek B – C3 - RESPONS

Oefening 1

Een deel van de leerlingen volgden een techniekproject. We vragen ons af of de percepties van leerlingen m.b.t. het project ('Projectbijgeleerd', 'Projectleuk', 'Projectinteressant', 'Projectmoeilijk') onderling samenhangen?

- Beantwoord de vraag a.d.h.v. de gepaste analyses.
- Visualiseer de resultaten.

OEFFENING 1 a

```
> Project <- data.frame(Techniek$Projectbijgeleerd,  
+                       Technik$Projectleuk,  
+                       Technik$Projectinteressant,  
+                       Technik$Projectmoeilijk)  
> Tabelproject <- cor.prob(Project)  
> write.csv2(Tabelproject, file="Tabel project.csv") # uitleg einde respons  
> Tabelproject
```

	Techniek.Projectbijgeleerd	Techniek.Projectleuk	Techniek.Projectinteressant	Techniek.Projectmoeilijk
Techniek.Projectbijgeleerd	1.0000000	0.0000000	0.0000000	0
Techniek.Projectleuk	0.8100783	1.0000000	0.0000000	0
Techniek.Projectinteressant	0.8483395	0.9010072	1.0000000	0
Techniek.Projectmoeilijk	0.4322983	0.3751562	0.3882242	1

→ $r =$ tussen 0.38 en 0.90: de samenhang varieert van middelmatig tot groot (groene getallen)

→ $p < 0.05$: kans dat H_0 opgaat is kleiner dan 5% (rode getallen)

We verwachten in de populatie WEL een samenhang tussen de verschillende percepties over het project.

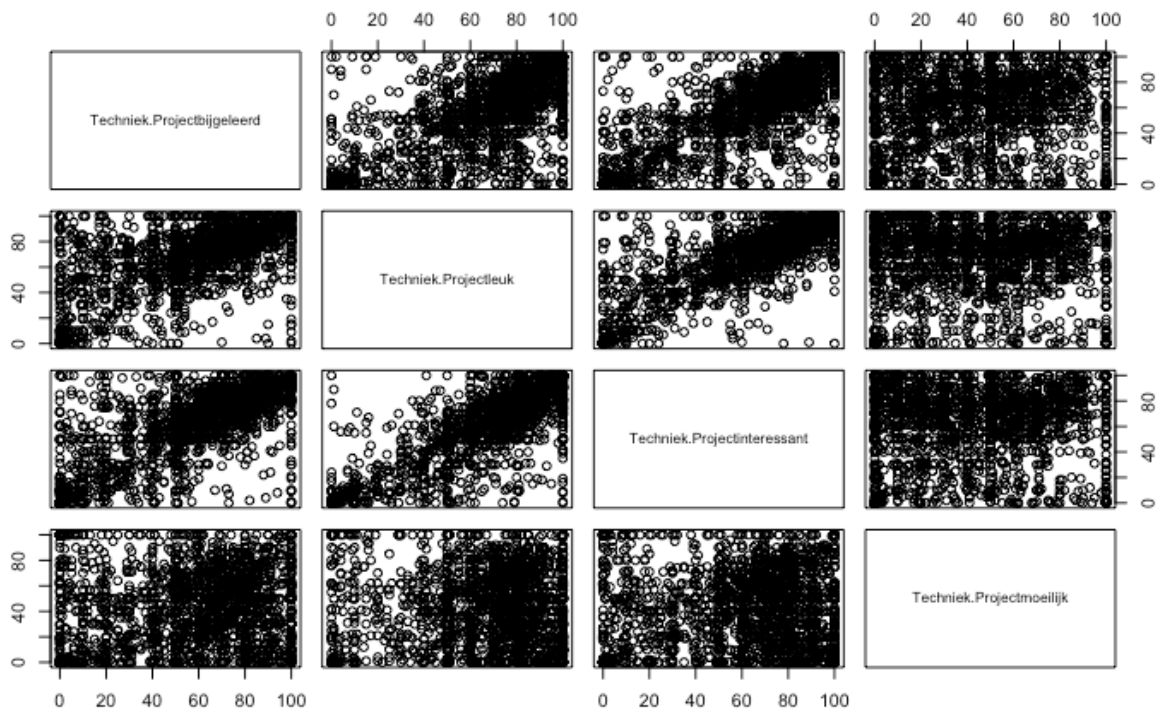
CONCLUSIE:

Alle percepties ten aanzien van het project hebben een verband met elkaar dat we niet enkel in deze steekproef vaststellen, maar ook kunnen doortrekken naar de populatie ($p < 0.05$).

Al deze verbanden zijn positief en de perceptie van bijleren en leuk vinden ($r = 0.81$), bijleren en interessant vinden ($r = 0.84$) en leuk en interessant vinden ($r = 0.90$) hangen bovendien sterk met elkaar samen. Ook als leerlingen het project moeilijk vinden, geven ze aan dat ze iets bijleren ($r = 0.43$) en het zelfs interessant ($r = 0.38$) en leuk vinden ($r = 0.38$). Bij deze drie laatste verbanden gaat het telkens om een middelmatig samenhang.

OEFENING 1 b

```
> plot(Project)
```



Oefening 2

We vermoeden dat lesstijl een rol zou kunnen spelen in de ontwikkeling van de technische geletterdheid ('TAC.na'). We hebben evidentie om aan te nemen dat het belangrijk is dat:

- *leerlingen zelf experimenteren ('PISA_Experimenteren')
- *leerlingen zelf onderwerpen kiezen ('PISA_EigenInbreng')
- *de leerkracht de leerstof in een ruimer perspectief zet ('PISA_Orientatie')
- *dit best kan d.m.v. interactie met de leerlingen ('PISA_Interactie').

Formuleer voor deze 4 hypothesen een afzonderlijk besluit o.b.v. de resultaten van je analyses.

```
> Model.Exp <- lm(TAC.na ~ PISA_Experimenteren, data=Techniek)
> summary(Model.Exp)
```

Call:

```
lm(formula = TAC.na ~ PISA_Experimenteren, data = Techniek)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.59061	-0.16357	-0.00047	0.18127	0.41600

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.598537	0.019153	31.250	<2e-16 ***
PISA_Experimenteren	-0.005287	0.008171	-0.647	0.518

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2087 on 1677 degrees of freedom

(688 observations deleted due to missingness)

Multiple R-squared: 0.0002496, Adjusted R-squared: **-0.0003466**

F-statistic: 0.4186 on 1 and 1677 DF, p-value: **0.5177**

→ $R^2 = -0.0003$: het gaat om een verwaarloosbaar effect (negatief % verklaarde variantie in 'TAC.na'!!)

Met $p = 0.52$: kans dat H_0 opgaat in de populatie is groter dan 5%

Dus we verwachten dat dit model in de populatie GEEN variantie verklaart in 'TAC.na'.

→ intercept = 0.60: een leerling die 0 scoort op 'PISA_Experimenteren' behaalt 0.60 op 'TAC.na'

Met $p < 0.05$: kans dat H_0 opgaat in de populatie is kleiner dan 5%

Dus we verwachten dit WEL in de populatie terug te vinden.

→ $\beta_{\text{PISA_Experimenteren}} = -0.005$; dus 1 punt hoger scoren op 'PISA_Experimenteren' leidt tot 0.005 punt lager scoren op 'TAC.na'

Met $p = 0.52$: kans dat H_0 opgaat in de populatie is groter dan 5%

Dus we verwachten dat 'PISA_Experimenteren' in de populatie GEEN invloed heeft op 'TAC.na'.

```
> Model.EI <- lm(TAC.na ~ PISA_EigenInbreng, data=Techniek)
> summary(Model.EI)
```

Call:

```
lm(formula = TAC.na ~ PISA_EigenInbreng, data = Techniek)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.62709	-0.14642	0.00749	0.16134	0.46876

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.684679	0.012193	56.152	<2e-16 ***
PISA_EigenInbreng	-0.057591	0.006535	-8.813	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.204 on 1677 degrees of freedom

(688 observations deleted due to missingness)

Multiple R-squared: 0.04426, Adjusted R-squared: **0.04369**

F-statistic: 77.66 on 1 and 1677 DF, p-value: **< 2.2e-16**

→ $R^2 = 0.04$: het gaat om een klein effect (4% verklaarde variantie in 'TAC.na')

Met $p < 0.05$: kans dat H_0 opgaat in de populatie is kleiner dan 5%

Dus we verwachten dat dit model in de populatie WEL variantie verklaart in 'TAC.na'.

→ intercept = 0.68: een leerling die 0 scoort op 'PISA_EigenInbreng' scoort 0.68 op 'TAC.na'

Met $p < 0.05$: kans dat H_0 opgaat in de populatie is kleiner dan 5%

Dus we verwachten dit WEL in de populatie terug te vinden.

→ $\beta_{PISA_EigenInbreng} = -0.06$, dus 1 punt hoger scoren op 'PISA_EigenInbreng' leidt tot 0.06 punten lager scoren op 'TAC.na'

Met $p < 0.05$: kans dat H_0 opgaat in de populatie is kleiner dan 5%

Dus we verwachten dat 'PISA_EigenInbreng' in de populatie WEL invloed heeft op 'TAC.na'.

```
> Model.Or <- lm(TAC.na ~ PISA_Orientatie, data=Techniek)
```

```
> summary(Model.Or)
```

Call:

```
lm(formula = TAC.na ~ PISA_Orientatie, data = Techniek)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.61621	-0.16243	0.00984	0.17759	0.42678

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.542515	0.017312	31.338	< 2e-16 ***
PISA_Orientatie	0.018424	0.006918	2.663	0.00781 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2082 on 1677 degrees of freedom

(688 observations deleted due to missingness)

Multiple R-squared: 0.004212, Adjusted R-squared: **0.003618**

F-statistic: 7.093 on 1 and 1677 DF, p-value: **0.007812**

→ $R^2 = 0.003$: het gaat om een verwaarloosbaar effect (0.3% verklaarde variantie in 'TAC.na')

Met $p < 0.05$: kans dat H_0 opgaat in de populatie is kleiner dan 5%

Dus we verwachten dat dit model in de populatie WEL variantie verklaart in 'TAC.na'.

→ intercept = 0.54: een leerling die 0 scoort op 'PISA_Orientatie' scoort 0.54 op 'TAC.na' in de steekproef

Met $p < 0.05$: kans dat H_0 opgaat in de populatie is kleiner dan 5%

Dus we verwachten dit WEL in de populatie terug te vinden.

→ $\beta_{PISA_Orientatie} = 0.02$; dus 1 punt hoger scoren op 'PISA_Orientatie' leidt tot 0.02 punten hoger scoren op 'TAC.na'

Met $p < 0.05$: kans dat H_0 opgaat in de populatie is kleiner dan 5%

Dus we verwachten dat 'PISA_Orientatie' in de populatie WEL invloed heeft op 'TAC.na'.

```
> Model.Int <- lm(TAC.na ~ PISA_Interactie, data=Techniek)
> summary(Model.Int)
```

Call:

```
lm(formula = TAC.na ~ PISA_Interactie, data = Techniek)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.58999	-0.16337	-0.00362	0.18043	0.41709

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.575826	0.017589	32.738	<2e-16 ***
PISA_Interactie	0.004720	0.007382	0.639	0.523

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2087 on 1677 degrees of freedom
(688 observations deleted due to missingness)

Multiple R-squared: 0.0002437, Adjusted R-squared: **-0.0003524**

F-statistic: 0.4088 on 1 and 1677 DF, p-value: **0.5227**

→ $R^2 = -0.0003$: het gaat om een verwaarloosbaar effect (negatief % verklaarde variantie in 'TAC.na'!!)

Met $p = 0.52$: kans dat H_0 opgaat in de populatie is groter dan 5%

Dus we verwachten dat dit model in de populatie GEEN variantie verklaart in 'TAC.na'.

→ intercept = 0.58: een leerling die 0 scoort op 'PISA_Interactie' scoort 0.58 op 'TAC.na';

Met $p < 0.05$: kans dat H_0 opgaat in de populatie is kleiner dan 5%

Dus we verwachten dit WEL in de populatie terug te vinden.

→ $\beta_{PISA_Interactie} = 0.005$; dus 1 punt hoger scoren op 'PISA_Interactie' leidt tot 0.005 punten hoger scoren op 'TAC.na'

Met $p = 0.52$: kans dat H_0 opgaat in de populatie is groter dan 5%

Dus we verwachten dat 'PISA_Orientatie' in de populatie GEEN invloed heeft op 'TAC.na'.

CONCLUSIE:

Wat betreft het effect van 'PISA_Experimenteren' en 'PISA_Interactie' kunnen we erg kort zijn. Het intercept is telkens significant ($p < 0.05$), maar de hellingsgraad is niet statistisch significant ($p = 0.52$). Ook het model zelf is niet statistisch significant ($p = 0.52$). M.a.w. of leerlingen al dan niet mogen experimenteren in de lessen techniek of in interactie gaan met elkaar, doet er niet toe als het gaat over technische geletterdheid ('TAC.na').

Als leerlingen zelf een inbreng mogen hebben in de lessen techniek stellen we wel vast dat dit een klein significant effect heeft ($R^2 = 0.04$; $p < 0.05$). Het effect is hier negatief ($\beta = -0.06$)! Voor elk punt dat een leerling meer scoort op de de schaal 'PISA_EigenInbreng' zal de technische geletterdheid met 0.06 punten afnemen. Wanneer we weten dat het hier ('TAC.na') over een schaal gaat van nul tot een kunnen we stellen dat leerlingen dan 6% slechter scoren dan wanneer ze een punt minder halen op 'PISA_EigenInbreng'.

(Om dit eigenlijk volledig goed te begrijpen had het zinvol geweest eerst z-scores te maken van de lesstijlen. Op die manier zou het 1 punt meer of minder behalen meer betekenis hebben gekregen en bovendien vergelijkbaar zijn tussen de analyses.)

Als we kijken naar de omkadering ('PISA_Orientatie') die de leerkracht geeft bij de leerstof dan stellen we vast dat we hier wel een significant effect vinden ($p < 0.05$). Leerlingen die 1

punt meer scoren op 'PISA_Orientatie' zullen 0.02 punten meer scoren op technische geletterdheid ('TAC.na'). Dit model is statistisch significant ($p < 0.05$) en we kunnen dus aannemen dat we dergelijke trends ook in de populatie zullen aantreffen. Maar het effect is echter wel verwaarloosbaar klein! Met andere woorden, ook hier stellen we vast dat de invloed van lesstijl op de ontwikkeling van de technische geletterdheid van de leerling van weinig praktische waarde is.

UITSMIJTER!

```
> write.csv2(Tabelproject, file="Tabel project.csv")
```

Omdat je tabellen niet altijd even makkelijk van R naar Word kan overzetten, is het commando `write.csv2(x = naam object, file = "Naamtabel.csv")` erg handig. Hiermee kan je eenvoudig een R object (tabel of databestand) omzetten naar een csv-bestand dat je vervolgens kan openen in Excel. Onder `x` verwijst je naar het R object dat je wil wegschrijven als csv-bestand. In het argument `file` geef je de naam van het csv-bestand op. Let op! Deze naam moet eindigen op `.csv` en tussen aanhalingstekens worden geplaatst.

R schrijft dit csv-bestand dan weg naar je working directory (= de locatie waarin je bestand wordt weggeschreven). Je working directory vind je door het commando `getwd()` te laten lopen.