

Statistiek B – C5 – RESPONS

Voorbereidend werk

- Om alles eenvoudig interpreteerbaar te houden, maak je van alle kwantitatieve variabelen die je nodig hebt eerst een z-score. Het gaat om de variabelen 'TAC.na', 'Gender.voor', 'PISA_EigenInbreng' en 'PISA_Experimenteren'.
- Maak een dummy variabele voor 'Geslacht' die aanstaat voor "meisje". Zet de variabele 'Richting5cat' om in een reeks dummyvariabelen zodanig dat je in je analyses de volgende groepen van studierichtingen met elkaar kan vergelijken: de studierichting "Latijn", de studierichting "Moderne wetenschappen" en de studierichtingen "Overige". Deze laatste groep omvat de studierichtingen "Technische", "Kunst" en "STV/Handel".

VOORBEREIDEND WERK

```
> # Variabelen standaardiseren #
> Techniek$TAC.naZ <- scale(Techniek$TAC.na)
> Techniek$PISA_EigenInbrengZ <- scale(Techniek$PISA_EigenInbreng)
> Techniek$PISA_ExperimenterenZ <- scale(Techniek$PISA_Experimenteren)
> Techniek$Gender.voorZ <- scale(Techniek$Gender.voor)

> # Dummy variabelen maken #
> Techniek$GeslachtD <- (Techniek$Geslacht=="0")*1
> table(Techniek$GeslachtD, Techniek$Geslacht)

      0      1
0      0 1050
1 1317      0

> ## Controleer of je de dummyvariabele correct aanmaakte!

> Techniek$LatijnD <- (Techniek$Richting5cat=="3")*1
> Techniek$Mod_wetD <- (Techniek$Richting5cat=="4")*1
> Techniek$OverigeD <- (Techniek$Richting5cat=="1" |
+                       Techniek$Richting5cat=="2" |
+                       Techniek$Richting5cat=="5")*1
> ## | staat in R voor de logische operator OF...
> ## Als 'Richting5cat' gelijk is aan 1 OF aan 2 OF aan 5 geef die dan de
> ## waarde 1.

> table(Techniek$LatijnD, Techniek$Richting5cat)

      1      2      3      4      5
0  181    40      0 1040   316
1      0      0   742      0      0

> table(Techniek$Mod_wetD, Techniek$Richting5cat)

      1      2      3      4      5
0  181    40   742      0   316
1      0      0      0 1040      0

> table(Techniek$OverigeD, Techniek$Richting5cat)

      1      2      3      4      5
0      0      0   742 1040      0
1  181    40      0      0   316
```

Oefening 1

Ter voorbereiding maakte je een reeks dummy variabelen aan die aangeven in welke studierichting een leerling les volgt (Latijn, Moderne wetenschappen of Overige).

- Doe de nodige analyses om de volgende onderzoeksvraag te beantwoorden en bespreek zo grondig mogelijk de output: *Scoren de leerlingen uit de 3 studierichtingen verschillend op technische geletterdheid ('TAC.naZ') ongeacht de mate waarin leerlingen een eigen inbreng in de les krijgen ('PISA_EigenInbrengZ') of waarin er geëxperimenteerd wordt in de les ('PISA_ExperimenterenZ')?*
- Voortbouwend op het model dat je in a) hebt getest, vraagt een collega-onderzoeker aan jou of het niet zinvoller is om volgende onderzoeksvraag te onderzoeken: *Is het effect van eigen inbreng in de les ('PISA_EigenInbrengZ') op technische geletterdheid (TAC.naZ) wel identiek voor leerlingen uit Moderne wetenschappen?* Doe hiertoe de nodige analyses en bespreek kort de essentie om bovenstaande vraag te beantwoorden.
- Bereken de voorspelde score voor een leerling uit Moderne wetenschappen, die 2 SD hoger dan gemiddeld scoort op 'PISA_EigenInbrengZ' en 2.5 SD lager op 'PISA_ExperimenterenZ'. Bereken dit zowel voor de steekproef als voor de populatie. (Rond daarbij zowel de tussenstappen als de uitkomst af tot op 2 cijfers na de komma.)

OEFFENING 1a

```
> # Dataset zonder NA's aanmaken voor oefening 1 #
> DataC5a <- na.omit(Techniek[c("TAC.naZ", "PISA_ExperimenterenZ",
+                               "PISA_EigenInbrengZ", "LatijnD", "Mod_wetD",
+                               "OverigeD")])
> ## Dit is nodig omdat we modellen moeten vergelijken voor oefening 1b

> # Modellen schatten #
> Modell1 <- lm(TAC.naZ ~ Mod_wetD + OverigeD + PISA_EigenInbrengZ +
+               PISA_ExperimenterenZ, data=DataC5a)
> Modell1b <- lm(TAC.naZ ~ LatijnD + OverigeD + PISA_EigenInbrengZ +
+                PISA_ExperimenterenZ, data=DataC5a)
> ## In model 1b is 'Moderne wetenschappen' de referentiecategorie.
> ## Dit maakt het mogelijk om ook na te gaan of leerlingen uit deze categorie
> ## significant verschillen m.b.t. 'TAC.naZ' deze uit de overige
> ## studierichtingen.
>
> summary(Modell1)
```

Call:

```
lm(formula = TAC.naZ ~ Mod_wetD + OverigeD + PISA_EigenInbrengZ +
    PISA_ExperimenterenZ, data = DataC5a)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.2091	-0.6044	0.0593	0.6900	2.4599

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.45339	0.03827	11.849	< 2e-16 ***
Mod_wetD	-0.51724	0.05122	-10.099	< 2e-16 ***
OverigeD	-0.73258	0.06293	-11.642	< 2e-16 ***

```

PISA_EigenInbrengZ    -0.22603    0.02628    -8.601    < 2e-16 ***
PISA_ExperimenterenZ  0.12492    0.02592     4.819    1.57e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 0.9126 on 1635 degrees of freedom
Multiple R-squared:  0.1401, Adjusted R-squared:  0.138
F-statistic: 66.58 on 4 and 1635 DF, p-value: < 2.2e-16

```

→ $R^2 = 0.14$: het gaat om een groot effect (14% verklaarde variantie in 'TAC.naZ')

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.naZ'.

→ intercept = 0.45: een leerling uit de studierichting Latijn die gemiddeld scoort op 'PISA_EigenInbrengZ' en 'PISA_ExperimenterenZ' (want allemaal z-scores) scoort 0.45 SD hoger dan gemiddeld op 'TAC.naZ'

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.

→ $\beta_{Mod_wetD} = -0.52$, dus een leerling uit moderne wetenschappen scoort 0.52 SD (want z-score!) lager op 'TAC.naZ' dan een leerling die Latijn volgt (=referentiecategorie).

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

Dus we verwachten dit verschil in score op 'TAC.naZ' tussen leerlingen die Latijn en leerlingen die moderne wetenschappen volgen ook in de populatie terug te vinden.

→ $\beta_{OverigeD} = -0.73$, dus een leerling uit een overige studierichting scoort 0.73 SD (want z-score!) lager op 'TAC.naZ' dan een leerling die Latijn volgt (=referentiecategorie).

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

Dus we verwachten dit verschil in score op 'TAC.naZ' tussen leerlingen die Latijn en leerlingen die een overige studierichting volgen ook in de populatie terug te vinden.

→ $\beta_{PISA_EigenInbrengZ} = -0.23$, dus 1 SD (want z-score!) hoger scoren op 'PISA_EigenInbrengZ' leidt tot 0.23 SD (want z-score!) lager scoren op 'TAC.naZ'

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

Dus we verwachten dat 'PISA_EigenInbrengZ' in de populatie WEL invloed heeft op 'TAC.naZ'. Bovendien is dit effect sterker dan dat van 'PISA_ExperimenterenZ'. (Je mag de sterkte van deze effecten met elkaar vergelijken, omdat beide variabelen gestandaardiseerd zijn en dus op dezelfde schaal staan.)

→ $\beta_{PISA_ExperimenterenZ} = 0.12$, dus 1 SD (want z-score!) hoger scoren op 'PISA_ExperimenterenZ' leidt tot 0.12 SD (want z-score!) hoger scoren op 'TAC.naZ'

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

Dus we verwachten dat 'PISA_ExperimenterenZ' in de populatie WEL invloed heeft op 'TAC.naZ'.

In de analyse in Model1 vormen de leerlingen die Latijn volgen de referentiecategorie. Op basis van bovenstaande analyse kunnen we dus geen uitspraken doen over het verschil in score op 'TAC.naZ' tussen leerlingen die moderne wetenschappen volgen en leerlingen uit de overige studierichtingen. Om hier een zicht op te krijgen, schatten we hetzelfde model (Model1b) en nemen de dummyvariabele die aanstaat voor "Latijn" op in het model en laten een andere dummyvariabele weg. De output van deze analyse vind je hieronder.

```
> summary(Model1b)

Call:
lm(formula = TAC.naZ ~ LatijnD + OverigeD + PISA_EigenInbrengZ +
    PISA_ExperimenterenZ, data = DataC5a)

Residuals:
    Min       1Q   Median       3Q      Max
-3.2091 -0.6044  0.0593  0.6900  2.4599

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.06385    0.03399   -1.879  0.060465 .
LatijnD         0.51724    0.05122   10.099 < 2e-16 ***
OverigeD       -0.21534    0.06013   -3.581  0.000352 ***
PISA_EigenInbrengZ -0.22603    0.02628   -8.601 < 2e-16 ***
PISA_ExperimenterenZ 0.12492    0.02592    4.819 1.57e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9126 on 1635 degrees of freedom
Multiple R-squared:  0.1401, Adjusted R-squared:  0.138
F-statistic: 66.58 on 4 and 1635 DF, p-value: < 2.2e-16
```

Uit deze analyse blijkt dat leerlingen uit de overige studierichtingen 0.22 SD lager score op 'TAC.naZ' ($\beta_{OverigeD} = -0.22$) dan leerlingen uit moderne wetenschappen. Bovendien verwachten we dit verschil ook terug te vinden in de populatie ($p < 0.05$).

OEFENING 1b

```
> # Model schatten #
> Model1c <- lm(TAC.naZ ~ Mod_wetD + OverigeD + PISA_EigenInbrengZ +
+               PISA_EigenInbrengZ*Mod_wetD + PISA_ExperimenterenZ,
+               data=DataC5a)

> # Modellen vergelijken #
> anova(Model1, Model1c)
Analysis of Variance Table

Model 1:  TAC.naZ ~ Mod_wetD + OverigeD + PISA_EigenInbrengZ +
PISA_ExperimenterenZ
Model 2:  TAC.naZ ~ Mod_wetD + OverigeD + PISA_EigenInbrengZ +
PISA_EigenInbrengZ * Mod_wetD + PISA_ExperimenterenZ

  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1   1635 1361.8
2   1634 1361.2  1   0.57441 0.6895 0.4064
```

→ $RSS_{Model1} = 1361.8$, $RSS_{Model2} = 1361.2$, $p = 0.41$

Model1c heeft wel een lagere RSS, maar dat verschil in RSS ($\Delta RSS = 0.69$) is te klein om te kunnen doortrekken naar de populatie ($p > 0.05$). Model1c is dus NIET statistisch significant beter dan Model1. We verwachten bijgevolg GEEN verschil in RSS in de populatie.

```
> # Parameter van interactie-effect checken #
> summary(Model1c)

Call:
lm(formula = TAC.naZ ~ Mod_wetD + OverigeD + PISA_EigenInbrengZ +
    PISA_EigenInbrengZ * Mod_wetD + PISA_ExperimenterenZ, data = DataC5a)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.2211	-0.6101	0.0641	0.6846	2.4938

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.45035	0.03844	11.715	< 2e-16	***
Mod_wetD	-0.51439	0.05134	-10.020	< 2e-16	***
OverigeD	-0.72602	0.06343	-11.446	< 2e-16	***
PISA_EigenInbrengZ	-0.24338	0.03357	-7.249	6.44e-13	***
PISA_ExperimenterenZ	0.12568	0.02594	4.845	1.39e-06	***
Mod_wetD:PISA_EigenInbrengZ	0.03843	0.04629	0.830	0.406	

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

Residual standard error: 0.9127 on 1634 degrees of freedom

Multiple R-squared: 0.1404, Adjusted R-squared: 0.1378

F-statistic: 53.39 on 5 and 1634 DF, p-value: < 2.2e-16

→ $\beta_{\text{Mod_wetD} \times \text{PISA_EigenInbrengZ}} = 0.04$, dus leerlingen die moderne wetenschappen volgen scoren voor elke 1 SD hoger op 'PISA_EigenInbrengZ' nog eens 0.04 SD (want z-score!) hoger op 'TAC.naZ'. (Een voorbeeld: een leerling die moderne wetenschappen volgt en 2 SD hoger scoort op 'PISA_EigenInbrengZ' scoort -0.46 op 'TAC.naZ' in de steekproef: $0.45 + (-0.51) + (2 \times -0.24) + (2 \times 0.04) = -0.46$)

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

Dus we verwachten het interactie-effect tussen 'PISA_EigenInbrengZ' en 'Mod_wetD' NIET terug te vinden in de populatie.

CONCLUSIE:

De studierichting die leerlingen volgen ('Mod_wetD', 'OverigeD'), de mate waarin leerlingen een eigen inbreng hebben ('PISA_EigenInbrengZ') en de mate waarin leerlingen mogen experimenteren tijdens de lessen ('PISA_ExperimenterenZ') verklaren de technische geletterdheid van leerlingen ('TAC.naZ') beter als het model waarin ook de interactie tussen 'PISA_EigenInbrengZ' en 'Mod_wetD' is opgenomen ($\Delta\text{RSS} = 0.69$, $p = 0.41$). Bovendien is deze interactieterm ook niet statistisch significant ($p > 0.05$).

Het model (Model1) zonder interactieterm verklaart 14% van de variantie in 'TAC.naZ'. Het gaat dus om een sterk, significant effect ($R^2 = 0.14$, $p < 0.05$). Leerlingen die moderne wetenschappen of een overige studierichting volgen, scoren respectievelijk 0.52 en 0.73 SD lager op 'TAC.naZ' dan leerlingen die Latijn volgen. Leerlingen uit de overige studierichtingen scoren op hun beurt 0.22 SD lager op technische geletterdheid dan leerlingen die moderne wetenschappen volgen. Al deze verschillen verwachten we bovendien ook in de populatie ($p < 0.05$).

Zowel 'PISA_EigenInbrengZ' als 'PISA_ExperimenterenZ' hebben een statistisch significant ($p < 0.05$) effect op 'TAC.naZ'. Een toename van 1 SD in 'PISA_EigenInbrengZ' leidt tot een afname van 0.22 SD in 'TAC.naZ'. Het effect van 'PISA_ExperimenterenZ' is kleiner ($\beta = 0.12$). 1 SD hoger scoren op 'PISA_ExperimenterenZ' leidt tot een toename van 0.12 SD in technische geletterdheid.

OEFENING 1c

Om onderstaande scores te berekenen, gebruiken we de regressievergelijking en de output van Model1 of Model1b. We berekenen het telkens a.d.h.v. de output van beide modellen.

Regressievergelijking:

$$\text{TAC.naZ} = \text{Intercept} + \beta_1 * \text{Mod_wetD} + \beta_2 * \text{OverigeD} + \beta_3 * \text{PISA_EigenInbrengZ} + \beta_4 * \text{PISA_ExperimenterenZ}$$

Output model 1:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.45339	0.03827	11.849	< 2e-16	***
Mod_wetD	-0.51724	0.05122	-10.099	< 2e-16	***
OverigeD	-0.73258	0.06293	-11.642	< 2e-16	***
PISA_EigenInbrengZ	-0.22603	0.02628	-8.601	< 2e-16	***
PISA_ExperimenterenZ	0.12492	0.02592	4.819	1.57e-06	***

Output model 1b:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-0.06385	0.03399	-1.879	0.060465	.
LatijnD	0.51724	0.05122	10.099	< 2e-16	***
OverigeD	-0.21534	0.06013	-3.581	0.000352	***
PISA_EigenInbrengZ	-0.22603	0.02628	-8.601	< 2e-16	***
PISA_ExperimenterenZ	0.12492	0.02592	4.819	1.57e-06	***

Welke score op 'TAC.naZ' behaalt een leerling die:

a) les volgt in moderne wetenschappen, 2 SD hoger scoort op 'PISA_EigenInbrengZ' en 2.5 SD lager op 'PISA_ExperimenterenZ' *in de steekproef*:

	Model1	Model1b
TAC.naZ	= 0.45 + -0.52*1 + -0.73*0 + -0.22*2 + 0.12*-2.5	= -0.06 + 0.51*0 + -0.22*0 + -0.22*2 + 0.12*-2.5
	= -0.81	= -0.80

b) les volgt in moderne wetenschappen, 2 SD hoger scoort op 'PISA_EigenInbrengZ' en 2.5 SD lager op 'PISA_ExperimenterenZ' *in de populatie*:

	Model1	Model1b
TAC.naZ	= 0.45 + 1*-0.52 + 0*-0.73 + 2*-0.22 + -2.5*0.12	= 0 + 0*0.51 + 0*-0.22 + 2*-0.22 + -2.5*0.12
	= -0.81	= -0.74

Voor Model1 wijken alle parameters statistisch significant af van 0 (en we kunnen deze dus doortrekken naar de populatie). De score op 'TAC.naZ' is dus hetzelfde in de steekproef als in de populatie. In Model1b is het intercept niet statistisch significant ($p > 0.05$). Daarom verschilt de score op 'TAC.naZ' hier wel tussen steekproef en populatie.

Oefening 2

We richten onze aandacht nu op een andere afhankelijke variabele: 'Gender.voorZ'. Deze variabele meet de mate waarin de respondenten vinden dat het onderwerp techniek iets is dat gepast is voor zowel jongens als meisjes. Hoe hoger de score, hoe meer de respondent daarmee akkoord gaat.

- Doe de nodige analyses om de volgende onderzoeksvraag te beantwoorden en bespreek zo grondig mogelijk de output: *Is er een verschil tussen jongens en meisjes in de mate van techniek iets vinden voor beide geslachten ('Gender.voorZ') en is dit verschil afhankelijk van al dan niet een richting met techniek volgen (Richting2cat)?*
- Hoeveel scoren jongens/meisjes die al dan niet techniek volgen op 'Gender.voorZ'? Vul o.b.v. je output onderstaande tabel in. (Rond daarbij zowel de tussenstappen als de uitkomst af tot op 2 cijfers na de komma.)

	Jongen		Meisje	
	Steekproef	Populatie	Steekproef	Populatie
Geen techniek				
Wel techniek				

OEFFENING 2a

```
> # Model schatten #
> Model2 <- lm(Gender.voorZ ~ GeslachtD + Richting2cat +
+               GeslachtD*Richting2cat, data=Techniek)
> summary(Model2)
```

Call:

```
lm(formula = Gender.voorZ ~ GeslachtD + Richting2cat + GeslachtD *
    Richting2cat, data = Techniek)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-2.41961 -0.65682 -0.04615  0.84391  1.78556
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.31804    0.03294  -9.654   <2e-16 ***
GeslachtD       0.57606    0.04241  13.582   <2e-16 ***
Richting2cat1   -0.06891    0.08160   -0.844    0.3985
GeslachtD:Richting2cat1  0.51142    0.24659   2.074    0.0382 *
```

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

Residual standard error: 0.9531 on 2286 degrees of freedom
(77 observations deleted due to missingness)

Multiple R-squared: 0.08869, Adjusted R-squared: 0.0875

F-statistic: 74.16 on 3 and 2286 DF, p-value: < 2.2e-16

→ $R^2 = 0.09$: het gaat om een medium effect (9% verklaarde variantie in 'Gender.voorZ')

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 'Gender.voorZ'.

→ intercept = -0.32: een jongen die geen techniek volgt, scoort 0.32 SD (want z-score) lager dan gemiddeld op 'Gender.voorZ'

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.

→ $\beta_{\text{GeslachtD}} = 0.58$, dus een meisje dat geen techniek volgt, scoort 0.58 SD (want z-score!) hoger op 'Gender.voorZ' dan een jongen die geen techniek volgt.

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

Dus we verwachten dit verschil in score op 'Gender.voorZ' tussen jongens en meisjes die geen techniek volgen ook in de populatie terug te vinden.

→ $\beta_{\text{Richting2cat1}} = -0.07$, dus een jongen die wel techniek volgt, scoort 0.07 SD (want z-score!) lager op 'Gender.voorZ' dan een jongen die geen techniek volgt.

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

Dus we verwachten dit verschil in score op 'Gender.voorZ' tussen een jongen die geen techniek en een jongen die wel techniek volgt NIET in de populatie terug te vinden.

→ $\beta_{\text{GeslachtD:Richting2cat1}} = 0.51$, dus een meisje dat wel techniek volgt (=scoort 1 op 'GeslachtD' en op 'Richting2cat') scoort nog eens 0.51 SD (want z-score!) hoger op 'Gender.voorZ'. (Dus, een meisje dat techniek volgt scoort 0.77 op 'Gender.voorZ' in de steekproef: $-0.32 + 0.58*1 + (-0.07)*1 + 0.51*1*1 = 0.7$)

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

Dus we verwachten de interactie tussen 'GeslachtD' en 'Richting2cat' ook terug te vinden in de populatie.

OEFFENING 2b

Om onderstaande scores te berekenen, gebruiken we de output van Model2 en de regressievergelijking.

Output model 2:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.31804	0.03294	-9.654	<2e-16 ***
GeslachtD	0.57606	0.04241	13.582	<2e-16 ***
Richting2cat1	-0.06891	0.08160	-0.844	0.3985
GeslachtD:Richting2cat1	0.51142	0.24659	2.074	0.0382 *

Regressievergelijking:

Gender.voorZ = Intercept + β_1 *GeslachtD + β_2 *Richting2cat + β_3 *GeslachtD*Richting2cat

a) Voor een jongen die geen techniek volgt:

	In de steekproef	In de populatie
Gender.voorZ =	-0.32	-0.32

Omdat "jongen" en "geen techniek volgen" de referentiecategorie zijn van 'GeslachtD' en 'Richting2cat' is het intercept hier gelijk aan de score op 'Gender.voorZ' voor een jongen die geen techniek volgt.

b) Voor een meisje dat geen techniek volgt:

	<i>In de steekproef</i>	<i>In de populatie</i>
Gender.voorZ =	$-0.32 + 0.58*1$	$-0.32 + 0.58*1$
=	0.26	0.26

Omdat “geen techniek volgen” de referentiecategorie is van ‘Richting2cat’ moeten we bij het intercept enkel nog het verschil tussen jongens en meisjes (‘GeslachtD’) optellen.

c) Voor een jongen die wel techniek volgt:

	<i>In de steekproef</i>	<i>In de populatie</i>
Gender.voorZ =	$-0.32 + (-0.07)*1$	$-0.32 + 0*1$
=	-0.39	-0.32

Omdat “jongen” de referentiecategorie is van ‘GeslachtD’ moeten we bij het intercept enkel nog het verschil tussen al dan niet techniek volgen (‘Richting2cat’) optellen. Dit verschil is echter niet statistisch significant ($p > 0.05$) en kunnen we dus niet doortrekken naar de populatie. Daarom verwachten we in de populatie dat een jongen die techniek volgt hetzelfde scoort op ‘Gender.voorZ’ als een jongen die geen techniek volgt (=het intercept).

d) Voor een meisje dat wel techniek volgt:

	<i>In de steekproef</i>	<i>In de populatie</i>
Gender.voorZ =	$-0.32 + 0.58*1 + (-0.07)*1 + 0.51*1$	$-0.32 + 0.58*1 + 0*1 + 0.51*1$
=	0.70	0.77

Om de score voor een meisje te berekenen, moeten we bij het intercept zowel het effect van geslacht (‘GeslachtD’), al dan niet techniek volgen (‘Richting2cat’) als het interactie-effect optellen. Het interactie-effect geeft immers aan hoeveel meer een meisje (GeslachtD = 1) dat techniek volgt (Richting2cat = 1) scoort op ‘Gender.voorZ’. Omdat het verschil in ‘Gender.voorZ’ voor ‘Richting2cat’ niet statistisch significant ($p > 0.05$) is, kunnen we het niet doortrekken naar de populatie.

Hieronder vind je de aangevulde tabel terug.

	Jongen		Meisje	
	<i>Steekproef</i>	<i>Populatie</i>	<i>Steekproef</i>	<i>Populatie</i>
Geen technische richting	-0.32	-0.32	0.26	0.26
Wel een technische richting	-0.39	-0.32	0.70	0.77