

Oefening 1

Is er een verschil tussen eerste en tweedejaars in technische geletterdheid (TAC.na) en is het zo dat het effect van “eigen inbreng in de les” (‘PISA_EigenInbreng’) op de technische geletterdheid anders is voor eerste dan wel voor tweedejaars (‘Studiejaar’)? Vooraleer je deze analyse draait, maak je een dummyvariabele aan voor de variabele ‘Studiejaar’. Zorg ervoor dat deze dummy aanstaat voor de categorie “tweedejaars”. Hanteer deze dummyvariabele in je verdere analyses.

Dummyvariabele maken:

```
> levels(Techniek$Studiejaar)
[1] "1" "2"
> Techniek$Tweedejaars<- (Techniek$Studiejaar=='2')*1
> table(Techniek$Studiejaar,Techniek$Tweedejaars)
      0      1
1 1083      0
2      0 1284
```

PISA_EigenInbreng:

```
> Modella<-
lm(TAC.na~PISA_EigenInbreng+Tweedejaars+(PISA_EigenInbreng*Tweedejaars),data=Techniek)
> summary(Modella)

Call:
lm(formula = TAC.na ~ PISA_EigenInbreng + Tweedejaars + (PISA_EigenInbreng *
  Tweedejaars), data = Techniek)

Residuals:
    Min       1Q   Median       3Q      Max
-0.62846 -0.14602  0.01189  0.16344  0.47802

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.679370   0.018249  37.228 < 2e-16 ***
PISA_EigenInbreng -0.053860   0.009498  -5.671 1.67e-08 ***
Tweedejaars       0.010221   0.024576   0.416 0.678
PISA_EigenInbreng:Tweedejaars -0.007273  0.013114  -0.555 0.579
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2041 on 1675 degrees of freedom
(688 observations deleted due to missingness)
Multiple R-squared:  0.04447,    Adjusted R-squared:  0.04275
F-statistic: 25.98 on 3 and 1675 DF,  p-value: < 2.2e-16
```

De mate waarin leerlingen eigen inbreng mogen doen, het studiejaar waarin leerlingen zitten en de interactie tussen deze beide variabelen verklaren 4.28% van de variantie ($R^2 = 0.0428$) in technische geletterdheid (TAC.na). We verwachten dat deze variabelen ook in de populatie een invloed gaan hebben op technische geletterdheid ($F(3,1675) = 25.98, p < 0.001$).

Onze hypothesen met betrekking tot verschil in het effect van ‘PISA_EigenInbreng’ kloppen niet. Het is niet zo dat leerlingen in het tweede jaar meer baat hebben bij het hebben van eigen inbreng in de les ($\beta = 0.01, p = 0.678$). Ook het interactie-effect van de lesstijl met het jaar is niet significant ($\beta = -0.007, p = 0.579$). Anders gesteld: leerlingen uit het eerste of tweede jaar hebben dezelfde mate van eigen inbreng nodig in de les om technisch geletterder te worden.

Oefening 2

En hoe zit dat met de andere leerkrachtstijlen (PISA_Experimenteren; PISA_Oriëntatie)? Hebben die hetzelfde effect voor eerste en tweedejaars? Gebruik opnieuw de dummyvariabele ('Tweedejaars') om deze hypothesen te toetsen.

PISA_Experimenteren:

```
> Modellb<-
lm(TAC.na~PISA_Experimenteren+Tweedejaars+(PISA_Experimenteren*Tweedejaars),data=Techniek)
> summary(Modellb)
```

Call:

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

Residuals:

| | Min | 1Q | Median | 3Q | Max |
|--|----------|----------|---------|---------|---------|
| | -0.59012 | -0.16213 | 0.00697 | 0.18210 | 0.41744 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|---------------------------------|------------------|------------|---------|--------------|
| (Intercept) | 0.582276 | 0.029566 | 19.694 | <2e-16 *** |
| PISA_Experimenteren | 0.001171 | 0.012381 | 0.095 | 0.925 |
| Tweedejaars | 0.028006 | 0.038866 | 0.721 | 0.471 |
| PISA_Experimenteren:Tweedejaars | -0.011252 | 0.016513 | -0.681 | 0.496 |

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

Residual standard error: 0.2088 on 1675 degrees of freedom
(688 observations deleted due to missingness)
Multiple R-squared: 0.000561, Adjusted R-squared: **-0.001229**
F-statistic: **0.3134** on 3 and 1675 DF, p-value: **0.8157**

PISA_Oriëntatie:

```
> Modellc <- lm(TAC.na~PISA_Oriëntatie+Tweedejaars+(PISA_Oriëntatie*Tweedejaars),
data=Techniek)
> summary(Modellc)
```

Call:

```
lm(formula = TAC.na ~ PISA_Oriëntatie + Tweedejaars + (PISA_Oriëntatie *
  Tweedejaars), data = Techniek)
```

Residuals:

| | Min | 1Q | Median | 3Q | Max |
|--|----------|----------|---------|---------|---------|
| | -0.60970 | -0.16449 | 0.00952 | 0.17363 | 0.42555 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-----------------------------|------------------|------------|---------|--------------|
| (Intercept) | 0.544560 | 0.028204 | 19.308 | <2e-16 *** |
| PISA_Oriëntatie | 0.016128 | 0.010834 | 1.489 | 0.137 |
| Tweedejaars | -0.005354 | 0.035844 | -0.149 | 0.881 |
| PISA_Oriëntatie:Tweedejaars | 0.005019 | 0.014177 | 0.354 | 0.723 |

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

Residual standard error: 0.2083 on 1675 degrees of freedom
(688 observations deleted due to missingness)
Multiple R-squared: 0.004543, Adjusted R-squared: **0.002761**
F-statistic: **2.548** on 3 and 1675 DF, p-value: **0.05431**

Wat beide andere lesstijlen betreft, blijkt dat de verklaarde variantie in technische geletterdheid (TAC.na) telkens verwaarloosbaar is ($R^2_{\text{Experimenteren}} = -0.0012$, $R^2_{\text{Orientatie}} = 0.0028$) en bovendien niet statistisch significant ($p_{\text{Experimenteren}} = 0.816$, $p_{\text{Orientatie}} = 0.054$). De onafhankelijke variabelen hebben dus geen effect op technische geletterdheid in de populatie. Bovendien is voor beide modellen zowel 'Tweedejaars' ($\beta_{\text{Experimenteren}} = -0.007$, $p = 0.471$; $\beta_{\text{Orientatie}} = 0.016$, $p = 0.137$) als de interactie tussen 'Tweedejaars' en lesstijl ($\beta_{\text{Experimenteren}*\text{Tweedejaars}} = -0.011$, $p = 0.496$; $\beta_{\text{Orientatie}*\text{Tweedejaars}} = 0.005$, $p = 0.723$) niet statistisch significant. In de populatie vinden we deze effecten dus niet terug.

Oefening 3

Tijdens een vorig contactmoment gingen we al na in welke mate de interesse van leerlingen in het project (Projectinteressant.z) en de mate van interesse in techniek voor het project (Interest.voor.z) een invloed hebben op de interesse na het project ('Interest.na.z'). Op basis van een uitgebreide literatuurstudie vermoeden we echter dat de interesse in techniek na het project niet voor alle studierichtingen hetzelfde is.

a) Ga deze hypothese na. Zet de variabele 'Richting5cat' eerst om in de nodige dummyvariabelen. Hanteer vervolgens de studierichting "Technisch" als referentiecategorie in je analyses. Maak ook de nodige gestandaardiseerde variabelen aan.

b) Interpreteer vervolgens ook de andere parameterschattingen.

Dummyvariabelen maken:

```
> levels(Techniek$Richting5cat)
[1] "1" "2" "3" "4" "5"
> Techniek$Technisch<-(Techniek$Richting5cat=='1')*1
> Techniek$Kunst<-(Techniek$Richting5cat=='2')*1
> Techniek$Latijn<-(Techniek$Richting5cat=='3')*1
> Techniek$ModWet<-(Techniek$Richting5cat=='4')*1
> Techniek$Handel<-(Techniek$Richting5cat=='5')*1
> table(Techniek$Richting5cat,Techniek$Technisch)
      0      1
1     0  181
2    40     0
3   742     0
4  1040     0
5   316     0
> table(Techniek$Richting5cat,Techniek$Kunst)
      0      1
1  181     0
2     0   40
3   742     0
4  1040     0
5   316     0
> table(Techniek$Richting5cat,Techniek$Latijn)
      0      1
1  181     0
2   40     0
3     0  742
4  1040     0
5   316     0
> table(Techniek$Richting5cat,Techniek$ModWet)
      0      1
1  181     0
2   40     0
3   742     0
4     0  1040
5   316     0
```

```
> table(Techniek$Richting5cat, Techniek$Handel)
```

```

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

```

Variabelen standaardiseren:

```

> Techniek$Projectinteressant.z<-scale(Techniek$Projectinteressant)
> Techniek$Interest.voor.z<-scale(Techniek$Interest.voor)
> Techniek$Interest.na.z<-scale(Techniek$Interest.na)

```

Model schatten:

```

> Model2a<-lm(Interest.na.z~Projectinteressant.z+Interest.voor.z+Kunst+Latijn+
ModWet+Handel, data=Techniek)
> summary(Model2a)

```

Call:

```
lm(formula = Interest.na.z ~ Projectinteressant.z + Interest.voor.z +
    Kunst + Latijn + ModWet + Handel, data = Techniek)
```

Residuals:

```

      Min       1Q   Median       3Q      Max
-3.9210 -0.4538  0.0149  0.4781  2.7577

```

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|----------------------|-----------------|------------|---------|-----------------------|
| (Intercept) | 0.28057 | 0.05886 | 4.767 | 1.99e-06 *** |
| Projectinteressant.z | 0.18773 | 0.01643 | 11.423 | < 2e-16 *** |
| Interest.voor.z | 0.56743 | 0.01706 | 33.259 | < 2e-16 *** |
| Kunst | -0.02814 | 0.13274 | -0.212 | 0.832 |
| Latijn | -0.33973 | 0.06588 | -5.157 | 2.73e-07 *** |
| ModWet | -0.31074 | 0.06355 | -4.890 | 1.08e-06 *** |
| Handel | -0.30135 | 0.07370 | -4.089 | 4.49e-05 *** |

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

Residual standard error: 0.7497 on 2236 degrees of freedom

(124 observations deleted due to missingness)

Multiple R-squared: 0.4403, Adjusted R-squared: **0.4388**

F-statistic: **293.2** on **6** and **2236** DF, p-value: **< 2.2e-16**

a) De hypothese klopt! Uit de analyses blijkt dat leerlingen die gemiddeld scoren op 'Projectinteressant.z' en 'Interest.voor.z' maar les volgen in de studierichtingen Latijn ($\beta = -0.34$, $p < 0.001$), Moderne Wetenschappen ($\beta = -0.311$, $p < 0.001$) of Handel ($\beta = -0.301$, $p < 0.001$) significant lager scoren op 'Interest.na.z' in vergelijking met leerlingen uit een technische richting. Het verschil in score op 'Interest.na.z' tussen leerlingen uit Kunst en leerlingen uit een technische richting is niet significant ($\beta = -0.028$, $p = 0.832$).

b) Een leerling uit een technische richting die gemiddeld scoort op 'Projectinteressant.z' en op 'Interest.voor.z' behaalt een score van 0.281 SD boven het gemiddelde ($\beta_0 = 0.281$, $p < 0.001$). De kans dat het intercept in de populatie 0 bedraagt, is zeer klein ($p < 0.001$). We gaan er dus vanuit dat dit ook voor de populatie geldt. Als we het effect van Projectinteressant.z op 'Interest.na.z' vergelijken met dat van 'Interest.voor.z' op 'Interest.na.z', dan blijkt dat het effect van de laatste variabele groter is. We mogen beide effecten vergelijken, omdat het in beide gevallen om gestandaardiseerde variabelen gaat. 'Interest.voor.z' heeft een positief effect op 'Interest.na.z' ($\beta = 0.567$, $p < 0.001$). Voor elke SD die een leerling hoger scoort op 'Interest.voor.z' stijgt zijn score op 'Interest.na.z' met 0.567 SD, ongeacht de studierichting waarin de leerling lesvolgt en zijn score op 'Project.interestant.z'. (Ook de toename in 'Interest.na.z' mogen we in SD's interpreteren aangezien ook deze

variabele is gestandaardiseerd.) Aangezien p kleiner is dan 0.05 verwachten we dit effect ook in de populatie terug te vinden. Hetzelfde geldt voor het effect van 'Projectinteressant.z'. Eén SD hoger score op deze variabele leidt tot een toename van 0.188 SD op 'Interest.na.z' ($\beta = 0.188, p < 0.001$), ongeacht de studierichting waarin een leerling les volgt en zijn score op 'Interest.voor.z'.

Ten slotte blijkt uit de adjusted R^2 dat dit model 43.9% van de variantie in 'Interest.na.z' verklaart ($adj. R^2 = 0.439$). De F-toets leert dat we dit ook in de populatie mogen verwachten ($F(6,2236) = 293.2, p < 0.001$). Dus ook de in populatie verklaren de onafhankelijke variabelen een deel van de verschillen in interesse in techniek na het project.

Oefening 4

Ten slotte verwachten we dat het effect van 'Interest.voor.z' op interesse in techniek mogelijk anders is voor leerlingen uit de studierichting "Technisch".

- Ga ook deze hypothese na. Rapporteer je bevindingen a.d.h.v. de output van je analyses.
- Bereken de scores van volgende leerlingen in de steekproef en in de populatie.
 - een leerling die 0 scoort op alle onafhankelijke variabelen
 - een leerling uit een technische richting die 1SD hoger scoort op 'Interest.voor.z' en 2 SD hoger scoort op 'Projectinteressant.z'

```
> Model2b<-
  lm(Interest.na.z~Projectinteressant.z+Interest.voor.z+Technisch+Latijn+ModWet+Handel+(Inter
    est.voor.z*Technisch),data=Techniek)
> summary(Model2b)
```

```
Call:
lm(formula = Interest.na.z ~ Projectinteressant.z + Interest.voor.z +
    Technisch + Latijn + ModWet + Handel + (Interest.voor.z *
    Technisch), data = Techniek)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-3.9124 -0.4541  0.0083  0.4787  2.7545
```

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.25216    0.11865   2.125  0.0337 *
Projectinteressant.z  0.18798    0.01644  11.437 <2e-16 ***
Interest.voor.z    0.56381    0.01752  32.183 <2e-16 ***
Technisch        -0.02903    0.14693  -0.198  0.8434
Latijn           -0.31190    0.12190  -2.559  0.0106 *
ModWet           -0.28240    0.12092  -2.335  0.0196 *
Handel           -0.27353    0.12642  -2.164  0.0306 *
Interest.voor.z:Technisch  0.06698    0.07379   0.908  0.3642
---
```

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

```
Residual standard error: 0.7497 on 2235 degrees of freedom
(124 observations deleted due to missingness)
Multiple R-squared:  0.4405, Adjusted R-squared:  0.4388
F-statistic: 251.4 on 7 and 2235 DF, p-value: < 2.2e-16
```

- Deze hypothese klopt niet! Uit de analyse blijkt dat het interactie-effect tussen 'Interest.voor.z' en een technische studierichting volgen niet significant is ($\beta = 0.067, p = 0.364$). Het effect van 'Interest.voor.z' op 'Interest.na.z' is in de populatie dus niet anders voor leerlingen uit een technische studierichting.

(Als dit wel statistisch significant zou zijn, dan zou dat betekenen dat voor leerlingen uit een technische studierichting het effect van 'Interest.voor.z' op 'Interest.na.z' groter zou zijn. 1 SD hoger scoren op 'Interest.voor.z' leidt dan voor een leerling uit een technische studierichting tot een toename van 0.631 SD (0.564+0.067) op 'Interest.na.z'.)

b)

- Een leerling die gemiddeld scoort op alle onafhankelijke variabelen behaalt zowel in de steekproef als in de populatie een score op 'Interest.na.z' van 0.252 (= het intercept!).

| | Intercept | Project Interessant.z | Interest. voor.z | Technisch | Latijn | ModWet | Handel | Interest.voor.z: Technisch |
|------------|-----------|-----------------------|------------------|-----------|----------|----------|----------|----------------------------|
| Steekproef | 0.252 | 0*0.188 | 0*0.564 | 0*-0.029 | 0*-0.312 | 0*-0.282 | 0*-0.274 | 0*0*0.067 |
| populatie | 0.252 | 0*0.188 | 0*0.564 | 0*0 | 0*-0.312 | 0*-0.282 | 0*-0.274 | 0*0*0 |

In de steekproef: 0.252

In de populatie: 0.252

- Een leerling uit een technische richting die 1 SD hoger scoort op 'Interest.voor.z' en 2 SD hoger scoort op 'Projectinteressant.z' behaalt een score van 1.23 op 'Interest.na.z' in de steekproef en van 1.192 in de populatie.

| | Intercept | Project Interessant.z | Interest. voor.z | Technisch | Latijn | ModWet | Handel | Interest.voor.z: Technisch |
|------------|-----------|-----------------------|------------------|-----------|----------|----------|----------|----------------------------|
| Steekproef | 0.252 | 2*0.188 | 1*0.564 | 1*-0.029 | 0*-0.312 | 0*-0.282 | 0*-0.274 | 1*1*0.067 |
| populatie | 0.252 | 2*0.188 | 1*0.564 | 1*0 | 0*-0.312 | 0*-0.282 | 0*-0.274 | 1*1*0 |

In de steekproef: $0.252 + 0.376 + 0.564 + (-0.029) + 0.067 = 1.23$

In de populatie: $0.252 + 0.376 + 0.564 = 1.192$