Contactmoment 5: Respons computer lab

> **!** Important
>
> Vooraleer je de oefeningen kan oplossen is het belangrijk om zowel de dataset te laden, het pakket `car` te activeren en ook de OLP2 Functies te activeren.

Voorbereiding

> Voorbereiding
>
> 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 dummy variabelen 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".

Eerste stap: variabelen standaardiseren (door `scale()` te gebruiken of `zscore()`)

```r
 # 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)
```

Vervolgens maken we de gevraagde dummy variabelen aan

```r
# Dummy variabelen maken #
Techniek$GeslachtD <- (Techniek$Geslacht=="0")*1

# Nagaan of het goed is gelukt
table(Techniek$GeslachtD, Techniek$Geslacht)
```

```
       0    1
  0    0 1050
  1 1317    0
```

```r
## 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$Richting5cat=="2" | Techniek$Ric
## | 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.

## Controleer of je de dummyvariabele correct aanmaakte!
table(Techniek$LatijnD, Techniek$Richting5cat)
```

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

Oefening 1

Oefening 1

(a) 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`)?

(b) 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.

(c) 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.)

(a)

De eerste stap is een dataset aanmaken die geen NA's meer bevat voor alle variabelen die gehanteerd zullen worden in deze oefening 1. Deze stap is nodig omdat we verder in de oefening ook modellen met elkaar gaan vergelijken.

```
DataC5a <- na.omit(
  Techniek[
    c("TAC.naZ",
      "PISA_ExperimenterenZ",
      "PISA_EigenInbrengZ", "LatijnD", "Mod_wetD",
      "OverigeD"
      )
    ]
  )
```

Nu zijn we klaar om de modellen te schatten, gebruikmakend van de nieuwe dataset `DataC5a`.

```
# Model schatten #
Model1 <- lm(
  TAC.naZ ~ Mod_wetD + OverigeD + PISA_EigenInbrengZ + PISA_ExperimenterenZ,
  data=DataC5a)
```

In dit model is 'Latijn' 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.

Na het schatten kunnen we ook de output bestuderen.

```
summary(Model1)
```

```
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-kwadraat = 0.14: het gaat om een groot effect (14% verklaarde variantie in `TAC.naZ`). Met $p < 0.05$: kans dat H0 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 lager dan gemiddeld op `TAC.naZ` in de steekproef. Met $p < 0.05$: kans dat H0 opgaat in de populatie is kleiner dan 5%. Dus we hebben voldoende evidentie om te stellen dat in de populatie het intercept anders is dan 0.
- β_{Mod_wetD} = -0.52, dus een leerling uit de richting 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 H0 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 H0 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 H0 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 H0 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 (`Model1_alternatief`) 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.

```
Model1_alternatief <- lm(
  TAC.naZ ~ LatijnD + OverigeD + PISA_EigenInbrengZ + PISA_ExperimenterenZ,
  data=DataC5a)
summary(Model1_alternatief)
```

```

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
```

Uiteraard veranderen enkele parameters van waarde. Maar waar we nu voornamelijk naar willen kijken is het effect van de dummy variabele `OverigeD`.

- $\beta_{OverigeD}$ = -0.22, dus een leerling uit een overige studierichting scoort 0.22 SD (want z-score!) lager op `TAC.naZ` dan een leerling die Moderne Wetenschappen volgt (=nu de referentiecategorie). Met $p < 0.05$: kans dat H0 opgaat in de populatie is kleiner dan 5%. Dus we verwachten dit verschil in score op `TAC.naZ` tussen leerlingen die Moderne Wetenschappen en leerlingen die een overige studierichting volgen ook in de populatie terug te vinden.

(b)

Nu schatten we een model met daarin een extra parameter: het interactie-effect tussen `Mod_wetD` en `PISA_EigenInbrengZ`.

```
Model1b <- lm(
  TAC.naZ ~ Mod_wetD + OverigeD + Mod_wetD*PISA_EigenInbrengZ + PISA_EigenInbrengZ + PISA_Exp
  data=DataC5a)
```

We vergelijken dit model met het model uit deel (a) van de oefening via de functie `anova()`.

```
anova(Model1, Model1b)
```

```
Analysis of Variance Table

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

We kunnen in deze output de *Residuals Sum of Squares (RSS)* (eigenlijk de SSE uit het OLP) voor beide modellen aflezen:

- RSS_Model1 = 1361.8,
- RSS_Model2 = 1361.2, p = 0.41

Model1b heeft wel een lagere RSS, maar dat verschil in RSS (RSS = 0.69) is te klein om te kunnen doortrekken naar de populatie ($p > 0.05$). Model1b is dus NIET statistisch significant beter dan Model1. We verwachten bijgevolg GEEN verschil in RSS in de populatie.

We kunnen desalniettemin de output van dit tweede model wel bekijken.

```r
summary(Model1b)
```

```
Call:
lm(formula = TAC.naZ ~ Mod_wetD + OverigeD + Mod_wetD * PISA_EigenInbrengZ +
    PISA_EigenInbrengZ + 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
```

- $\beta_{Mod\_wetD*PISA\_EigenInbrengZ} = 0.04$, 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-*0.24) + (*20.04) = -0.46)

Met $p > 0.05$: kans dat H0 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 samen de technische geletterdheid van leerlingen (`TAC.naZ`). Een model waarin bovendien de interactie tussen `PISA_EigenInbrengZ` en `Mod_wetD` is opgenomen is geen beter model dan een model met enkel hoofdeffecten (RSS = 0.69, p = 0.41). Bovendien is de 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, statistisch significant effect (R-kwadraat = 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 (ß = 0.12): 1 SD hoger scoren op `PISA_ExperimenterenZ` leidt tot een toename van 0.12 SD in technische geletterdheid.

(c)

Tot slot gaan we voorspelde scores berekenen aan de hand van `Model1`.

Hernemen we de resultaten:

```r
summary(Model1)
```

```

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
```

We starten met nog een keer de regressievergelijking op te stellen:

$\text{TAC.naZ} = \beta_0 + \beta_1 \cdot \text{Mod\_wetD} + \beta_2 \cdot \text{OverigeD} + \beta_3 \cdot \text{PISA\_EigenInbrengZ} + \beta_4 \cdot \text{PISA\_ExperimenterenZ}$

Vullen we de parameters in dan krijgen we dit:

$\text{TAC.naZ} = 0.45 + (-0.52 \cdot \text{Mod\_wetD}) + (-0.73 \cdot \text{OverigeD}) + (-0.23 \cdot \text{PISA\_EigenInbrengZ}) + (0.12 \cdot \text{PISA\_ExperimenterenZ})$

Nu kunnen we de score berekenen voor **de steekproef** door de waarden in te vullen ipv de namen van de variabelen. We zijn geïnteresseerd in een leerling die les volgt in moderne wetenschappen (`Mod_wetD = 1` & `OvergeD = 0`), 2 SD hoger scoort op `PISA_EigenInbrengZ` en 2.5 SD lager op `PISA_ExperimenterenZ` in de steekproef:

$\text{TAC.naZ} = 0.45 + (-0.52 \cdot 1) + (-0.73 \cdot 0) + (-0.23 \cdot 2) + (0.12 \cdot -2.5)$

Rekenen we dit uit dan komen we op -0.83 .

Voor **de populatie** is het net dezelfde werkwijze. Immers, alle parameterschattingen zijn statistisch signifcant ($p < 0.05$).

$\text{TAC.naZ} = 0.45 + (-0.52 \cdot 1) + (-0.73 \cdot 0) + (-0.23 \cdot 2) + (0.12 \cdot -2.5)$

Rekenen we dit uit dan komen we op -0.83 .

Oefening 2

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.

(a) 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 technische richting volgen (`Richting2cat`)?

(b) Hoeveel scoren jongens/meisjes die al dan niet techniek volgen op `Gender.voorZ`?

Vul o.b.v. je output onderstaande gegevens in. (Rond daarbij zowel de tussenstappen als de uitkomst af tot op 2 cijfers na de komma.)
Voorspelde scores **voor de steekproef**:

- Jongen - Geen techniek = ...
- Jongen - Wel techniek = ...
- Meisje - Geen techniek = ...
- Meisje - Wel tecchniek = ...

Voorspelde scores **voor de populatie**:

- Jongen - Geen techniek = ...
- Jongen - Wel techniek = ...
- Meisje - Geen techniek = ...
- Meisje - Wel techniek = ...

(a)

We starten met het schatten van het model.

```
# 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
```

We overlopen de verschillende relevante delen uit de output.

- R-kwadraat = 0.09: het gaat om een medium effect (9% verklaarde variantie in `Gender.voorZ`). Met $p < 0.05$: kans dat H0 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 H0 opgaat in de populatie is kleiner dan 5%. Dus we verwachten dit WEL in de populatie terug te vinden.
- $\beta_{GeslachtD} = 0.58$, 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 H0 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.
- $\beta_{Richting2cat1}$ = -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 H0 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.
- $\beta_{GeslachtD:Richting2cat1} = 0.51$, 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 H0 opgaat in de populatie is kleiner dan 5%. Dus we verwachten de interactie tussen `GeslachtD` en `Richting2cat` ook terug te vinden in de populatie.

(b)

Vooraleer we de berekeningen doen, schrijven we de regressievergelijking opnieuw op.

$$\text{Gender.voorZ} = \beta_0 + \beta_1 \cdot \text{GeslachtD} + \beta_2 \cdot \text{Richting2cat1} + \beta_3 \cdot \text{GelachtD:Richting2cat1}$$

Voorspelde scores **voor de steekproef**:

Om de scores te berekenen voor de steekproef vullen we alle parameters in uit de output.

Gender.voorZ $= -0.32 + 0.58 \cdot$ GeslachtD $+ -0.07 \cdot$ Richting2cat1 $+ 0.51 \cdot$ GelachtD:Richting2cat1

- Jongen (`GenderD` = 0) - Geen techniek (`Richting2cat1`= 0)

Gender.voorZ $= -0.32 + 0.58 \cdot 0 + -0.07 \cdot 0 + 0.51 \cdot 0 \cdot 0$

Resultaat: -0.32

- Jongen (`GenderD` = 0) - Wel techniek (`Richting2cat1`= 1)

Gender.voorZ $= -0.32 + 0.58 \cdot 0 + -0.07 \cdot 1 + 0.51 \cdot 0 \cdot 1$

Resultaat: -0.39

- Meisje (`GenderD` = 1) - Geen techniek (`Richting2cat1`= 0)

Gender.voorZ $= -0.32 + 0.58 \cdot 1 + -0.07 \cdot 0 + 0.51 \cdot 1 \cdot 0$

Resultaat: 0.26

- Meisje (`GenderD` = 1) - Wel tecchniek (`Richting2cat1`= 1)

Gender.voorZ $= -0.32 + 0.58 \cdot 1 + -0.07 \cdot 1 + 0.51 \cdot 1 \cdot 1$

Resultaat: 0.7

Voorspelde scores **voor de populatie**:

Om de scores te berekenen voor de populatie vullen we enkel de statistisch significante parameters uit de output in onze vergelijking. Parameters die niet statistisch significant zijn vervangen we door de waarde 0.

Gender.voorZ $= -0.32 + 0.58 \cdot \text{GeslachtD} + 0 \cdot \text{Richting2cat1} + 0.51 \cdot \text{GelachtD:Richting2cat1}$

- Jongen (`GenderD` = 0) - Geen techniek (`Richting2cat1`= 0)

Gender.voorZ $= -0.32 + 0.58 \cdot 0 + 0 \cdot 0 + 0.51 \cdot 0 \cdot 0$

Resultaat: -0.32

- Jongen (`GenderD` = 0) - Wel techniek (`Richting2cat1`= 1)

Gender.voorZ $= -0.32 + 0.58 \cdot 0 + 0 \cdot 1 + 0.51 \cdot 0 \cdot 1$

Resultaat: -0.32

- Meisje (`GenderD` = 1) - Geen techniek (`Richting2cat1`= 0)

$\text{Gender.voorZ} = -0.32 + 0.58 \cdot 1 + 0 \cdot 0 + 0.51 \cdot 1 \cdot 0$

Resultaat: 0.26

- Meisje (`GenderD` = 1) - Wel tecchniek (`Richting2cat1`= 1)

$\text{Gender.voorZ} = -0.32 + 0.58 \cdot 1 + 0 \cdot 1 + 0.51 \cdot 1 \cdot 1$

Resultaat: 0.77