

# GKN - Contactmoment 5

## Logistische regressieanalyses (Deel 1)

Sven De Maeyer & Bea Mertens

09/12/2021

1 / 56

## Categorische afhankelijke variabelen

2 / 56

# Categorische afhankelijke variabelen



3 / 56

# Categorische afhankelijke variabelen

Enkele klassiekers:

- Al dan niet slagen;
- Al dan niet doorstromen;
- Al dan niet falen;
- Al dan niet nieuwe job;
- Al dan niet norm halen;
- ...

Maar soms ook handig bij numerieke variabelen ...

4 / 56

# Categorische afhankelijke variabelen

Variabelen met meerdere categorieën zijn herleidbaar tot een reeks van **dummyvariabelen**

Voor elke categorie kan je een variabele maken die aangeeft of een respondent al dan niet tot die categorie behoort

5 / 56

## Vb. Onderpresteerders... in PIRLS

Leesvaardigheid van Vlaamse lln. in 5 categorieën ingedeeld (variabele `ASRIBM01`):

1. BELOW 400 (code 1)
2. AT OR ABOVE 400 BUT BELOW 475 (code 2)
3. AT OR ABOVE 475 BUT BELOW 550 (code 3)
4. AT OR ABOVE 550 BUT BELOW 625 (code 4)
5. AT OR ABOVE 625 (code 5)

Code 1 en 2 samen vormen een groep van onderpresterende leerlingen

```
table(Vlaanderen_1_2_3$ASRIBM01)
```

1	2	3	4	5
135	917	2329	1613	204

6 / 56

# Onderpresteerders - afhankelijke var. hercoderen

Eerst **hercoderen**...

Dummyvariabele maken die aanstaat voor 'Onderpresteren' (dus voor leerlingen met code 1 of 2 voor de variabele ASRIBM01)

```
library(car)
Vlaanderen_1_2_3$Onderpresteren <- recode(Vlaanderen_1_2_3$ASRIBM01, "1=1;2=1;3=0;4=0;5=0")
table(Vlaanderen_1_2_3$ASRIBM01, Vlaanderen_1_2_3$Onderpresteren)
```

	0	1
1	0	135
2	0	917
3	2329	0
4	1613	0
5	204	0

7 / 56

# Onderpresteerders - voorspeller?

*Is het zo dat jongens een grotere kans hebben om onder te presteren dan meisjes?*

```
Vlaanderen_1_2_3$Gender <- recode(Vlaanderen_1_2_3$ASBG01, "1 = 'Girls'; 2 = 'Boys'")
kruistabel.kolom(Vlaanderen_1_2_3$Onderpresteren, Vlaanderen_1_2_3$Gender)
```

	x		
y	Boys	Girls	Sum
0	77.57266	82.04833	79.83699
1	22.42734	17.95167	20.16301
Sum	100.00000	100.00000	100.00000

8 / 56

# Onderpresteerders - kansen! (1)

**Wat is de kans op ....?**

1. onderpresteren?
2. onderpresteren voor jongens?
3. onderpresteren voor meisjes?
4. onderpresteren voor meisjes - onderpresteren voor jongens?

y	x		
	Boys	Girls	Sum
0	77.57266	82.04833	79.83699
1	22.42734	17.95167	20.16301
Sum	100.00000	100.00000	100.00000

9 / 56

# Onderpresteerders - kansen! (2)

**Wat is de kans op....?**

$$P(\text{Onderpr.}) = .202$$

$$P(\text{Onderpr.} | \text{Boys}) = .224$$

$$P(\text{Onderpr.} | \text{Girls}) = .180$$

$$P(\text{Onderpr.} | \text{Girls}) - P(\text{Onderpr.} | \text{Boys}) = .180 - .224 = -.044$$

10 / 56

# Dummy als afhankelijke variabele? (1)

## Waarom niet gewoon dummy als afhankelijke variabele?

```
Model_Dummy <- lm(Onderpresteren ~ Gender, data = Vlaanderen_1_2_3)
summary(Model_Dummy)
```

```
Call:
lm(formula = Onderpresteren ~ Gender, data = Vlaanderen_1_2_3)

Residuals:
    Min       1Q   Median       3Q      Max
-0.2243 -0.2243 -0.1795 -0.1795  0.8205

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.224273   0.007941  28.244 < 2e-16 ***
GenderGirls -0.044757   0.011164  -4.009 6.18e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4007 on 5151 degrees of freedom
(45 observations deleted due to missingness)
Multiple R-squared:  0.003111,    Adjusted R-squared:  0.002917
F-statistic: 16.07 on 1 and 5151 DF,  p-value: 6.183e-05
```

11 / 56

# Dummy als afhankelijke variabele? (2)

## Kans op onderpresteren voor een meisje?

```
Call:
lm(formula = Onderpresteren ~ Gender, data = Vlaanderen_1_2_3)

Residuals:
    Min       1Q   Median       3Q      Max
-0.2243 -0.2243 -0.1795 -0.1795  0.8205

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.224273   0.007941  28.244 < 2e-16 ***
GenderGirls -0.044757   0.011164  -4.009 6.18e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4007 on 5151 degrees of freedom
(45 observations deleted due to missingness)
Multiple R-squared:  0.003111,    Adjusted R-squared:  0.002917
F-statistic: 16.07 on 1 and 5151 DF,  p-value: 6.183e-05
```

12 / 56

## Dummy als afhankelijke variabele? (3)

### Kans op onderpresteren voor een leerling wiens ouders 3 SD hoger dan gemiddeld scoren op Ouders\_GraagLezenZ?

```
Vlaanderen_1_2_3$Ouders_GraagLezenZ <- scale(Vlaanderen_1_2_3$ASBHPLR)
Model_Dummy2 <- lm(Onderpresteren ~ Ouders_GraagLezenZ, data = Vlaanderen_1_2_3)
summary(Model_Dummy2)
```

```
Call:
lm(formula = Onderpresteren ~ Ouders_GraagLezenZ, data = Vlaanderen_1_2_3)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.39762 -0.21144 -0.17131 -0.09858  0.95232
```

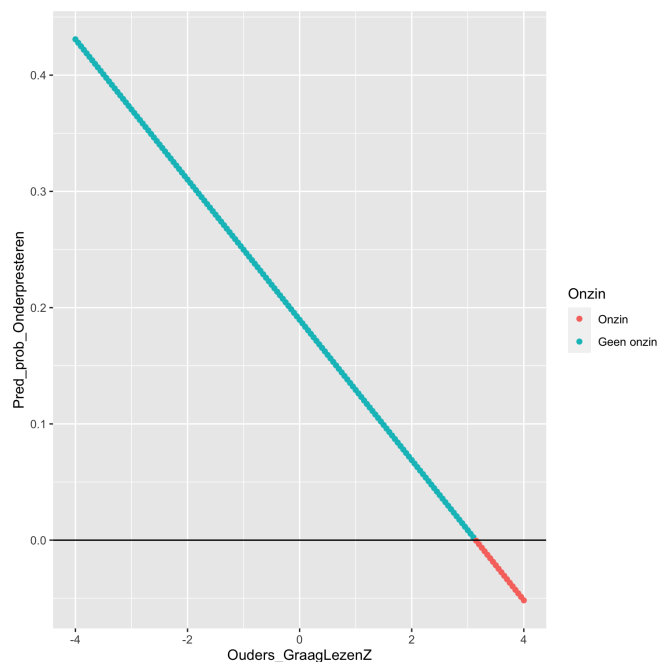
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.189537   0.005672   33.42  <2e-16 ***
Ouders_GraagLezenZ -0.060328   0.005672  -10.63  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.3873 on 4662 degrees of freedom
(534 observations deleted due to missingness)
Multiple R-squared:  0.02369,    Adjusted R-squared:  0.02348
F-statistic: 113.1 on 1 and 4662 DF,  p-value: < 2.2e-16
```

13 / 56

## Dummy als afhankelijke variabele? (4)

```
library(ggplot2)
Ouders_GraagLezenZ <- seq(-4, 4, .05)
Pred_prob_Onderpresteren <- 0.189537 + (-0.060328 * Ouders_Graag
Onzin <- car::recode(
  Pred_prob_Onderpresteren,
  ' -2:-0.0000001 = "Onzin"; 0.0000001:1 = "Geen onzin" '
)
Onzin <- relevel(as.factor(Onzin), "Onzin")
qplot(x = Ouders_GraagLezenZ,
      y = Pred_prob_Onderpresteren,
      colour = Onzin) +
  geom_hline(yintercept = 0)
```



14 / 56

# Problemen met dummyvariabelen als afhankelijke variabele

Probabiliteiten zijn gelimiteerd tussen 0 en 1

- Regressielijn kan onmogelijke waarden bevatten
- Kan leiden tot onzinnige schattingen
- Hoe meer 'gemiddelde kans' afwijkt van 50%, hoe groter kans op onzinnige schattingen

15 / 56

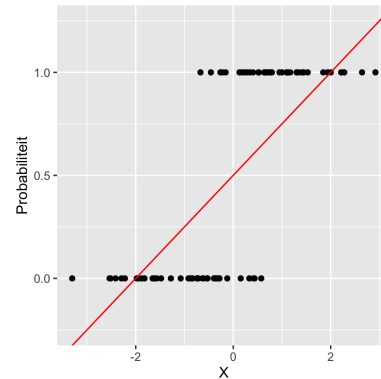
## Non-lineariteit

16 / 56



# Lineaire regressie is hier problematisch!

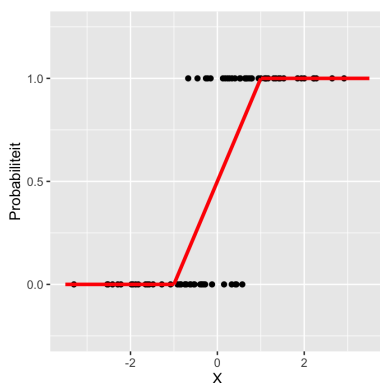
```
X <- c(rnorm( 40 , mean = -1 , sd = 1),  
      rnorm( 40, mean = 1 , sd = 1))  
Y <- c(rep( 0 , 40),  
      rep( 1 , 40))  
  
qplot(  
  x = X,  
  y = Y,  
  ylab = 'Probabiliteit',  
  ylim = c(-0.25,1.25)) +  
  geom_abline(slope = 0.25,  
             intercept = 0.50,  
             col="red")
```



*We krijgen aan de uiteinden problemen met onrealistische schattingen...*

17 / 56

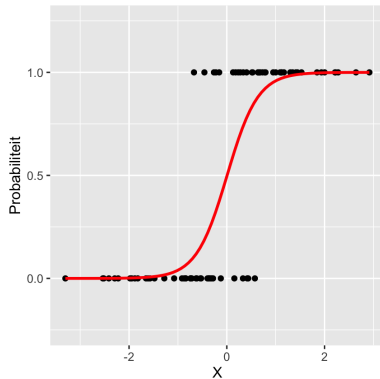
## Oplossing 1



Dit model is **DETERMINISTISCH**: vanaf een bepaalde waarde voor X is de kans gelijk aan 0% of 100%!

18 / 56

## Oplossing 2

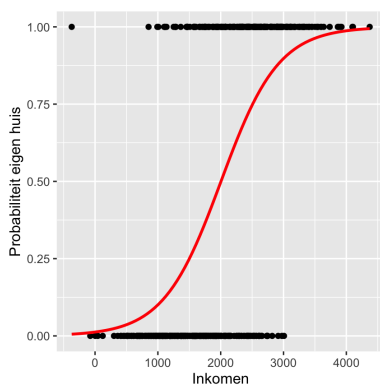


Dit model is

- **REALISTISCHER**: Vanaf een bepaalde waarde voor X benadert de kans 0% of 100%, maar is nooit exact 0% of 100%!
- **NIET-LINEAIR**: Kansen nemen niet lineair toe of af overheen de schaal van X

19 / 56

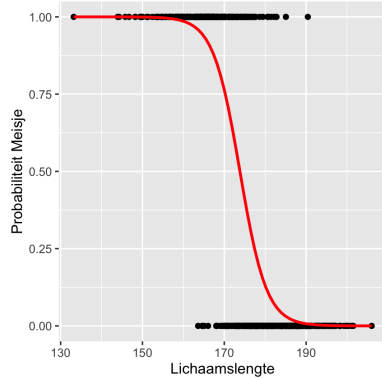
## Voorbeelden uit het 'dagelijkse leven' (1)



- Kans om een eigen huis te bezitten is hoger voor gezinnen met hoger inkomen
- Toch ook gezinnen met hoog inkomen die geen eigen huis bezitten en vice versa
- 200€ meer verdienen heeft geen gelijkaardige impact (tussen 1500€ en 2500€ is de impact het sterkst)

20 / 56

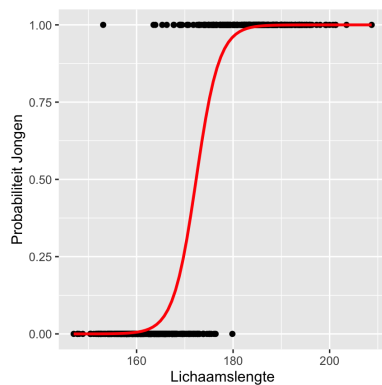
## Voorbeelden uit het 'dagelijkse leven' (2a)



- Hoe groter lichaamslengte, hoe kleiner de kans dat het een meisje is
- Toch ook grote meisjes en kleine jongens
- 1cm groter of kleiner zijn heeft geen gelijkaardige impact

21 / 56

## Voorbeelden uit het 'dagelijkse leven' (2b)



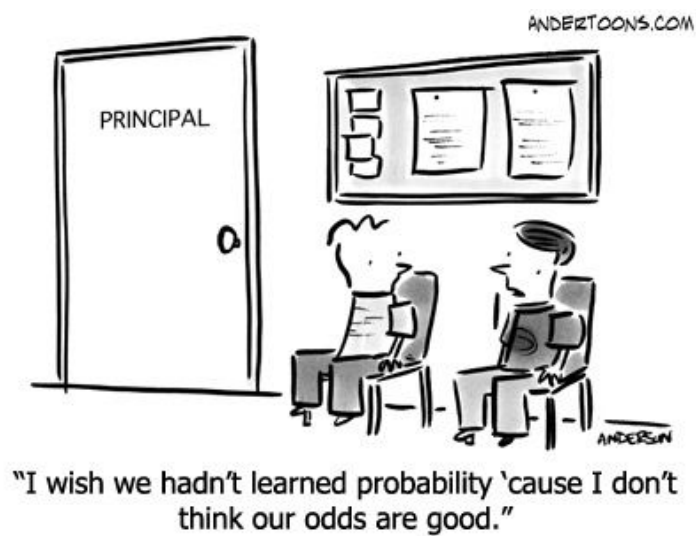
- Net hetzelfde model kan 'gespiegeld' worden!

22 / 56

# Logits

23 / 56

Let's talk in LOGITS...



24 / 56

# Kansen op een andere schaal uitgedrukt

Probabiliteiten → Odds → Logits

25 / 56

## Odds

De Odds voor een 'event' is:

$$Odds(Y = 1) = \frac{P(Y = 1)}{1 - P(Y = 1)} = \frac{P(Y = 1)}{P(Y \neq 1)}$$

26 / 56

# Odds - voorbeeld onderpresteren

Kans op onderpresteren is 0.25 keer groter dan de kans op niet onderpresteren

Kans op onderpresteren is 1/0.25 keer kleiner dan kans op niet onderpresteren

$$Odds(Onderpr. = 1) = \frac{P(Onderpr. = 1)}{1 - P(Onderpr. = 1)} = \frac{0.202}{0.798} = 0.253$$

Kans op niet onderpresteren is 4 keer groter dan kans op onderpresteren

Kans op niet onderpresteren is 1/4 keer kleiner dan kans op onderpresteren

$$Odds(Onderpr. = 0) = \frac{P(Onderpr. = 0)}{1 - P(Onderpr. = 0)} = \frac{0.798}{0.202} = 3.95$$

27 / 56

## Enkele odds...

Probabiliteit	Odds
0.10	(= 0.1/0.9) = 0.11
0.30	0.43
0.50	1.00
0.70	2.33
0.90	9.00

Loopt van 0 tot  $\infty$

28 / 56

## Odds zijn echter niet superhandig

- Nog steeds een nulwaarde
- Niet symmetrisch
- Dus niet handig voor lineaire modellen...

29 / 56

## Tijd voor logits

**Logit** is het natuurlijk logaritme van een Odds:

$$\text{Logits}(\text{Onderpr.} = 1) = \ln(\text{Odds}) = \ln(0.253) = -1.347$$

30 / 56

## Uitstapje naar ... logaritmes

$$\log_a(x) \rightarrow a^? = x$$

$$\log_{10}(10) \rightarrow 10^? = 10 \rightarrow 10^1 = 10$$

$$\log_{10}(1) \rightarrow 10^? = 1 \rightarrow 10^0 = 1$$

$$\log_{10}(100) \rightarrow 10^? = 100 \rightarrow 10^2 = 100$$

31 / 56

## Natuurlijk logaritme

$$\ln(x) = \log_e(x) \rightarrow e^? = x$$

met:

$$e = 2,718281828$$

$$\ln(1) \rightarrow e^? = 1 \rightarrow e^0 = 1$$

$$\ln(10) \rightarrow e^? = 10 \rightarrow e^{2.303} = 10$$

32 / 56



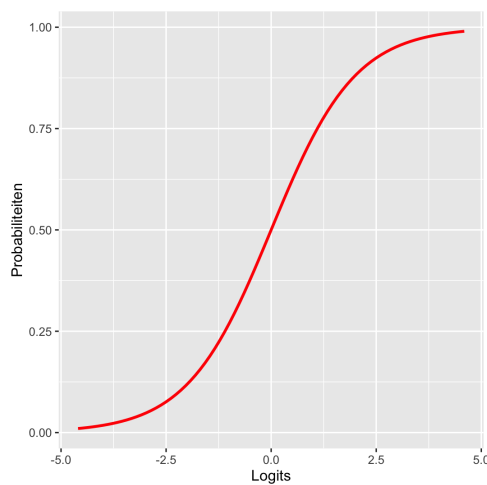
## Enkele logits...

Probabiliteit	Odds	Logits
0.10	0.11	-2.20
0.30	0.43	-0.85
0.50	1.00	0.00
0.70	2.33	0.85
0.90	9.00	2.20

- Loopt van  $-\infty$  tot  $\infty$
- 0 is midden (=50% kans)
- Symmetrisch!

33 / 56

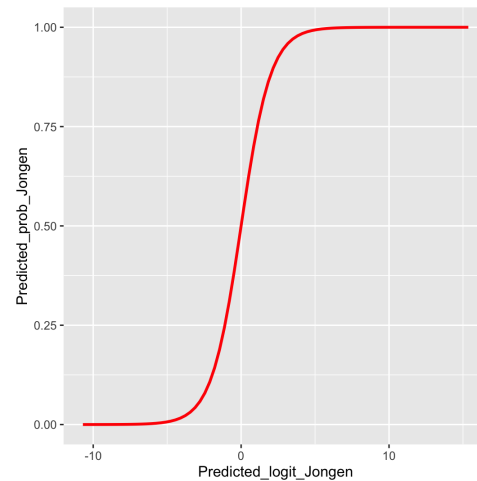
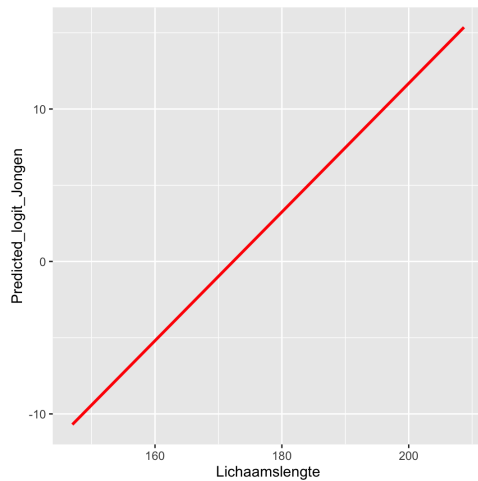
## Verhouding tussen probabiliteiten en logits



- Elke toename van 1 logit is niet lineair gerelateerd aan toename in probabiliteiten!
- Lineair model in logits gaat gepaard met non-lineair model in probabiliteiten

34 / 56

## Voorbeeld uit 'dagelijkse leven'



35 / 56

## Voorbeeld uit 'dagelijkse leven'

Het statistisch model achter het verband tussen lengte en kans dat het om een jongen gaat:

$$\text{Logit}(\text{Jongen} = 1) = -72.717 + 0.422 * \text{Lichaamslengte}$$

waarbij

**-72.717** = intercept = Voorspelde **kans in logits** dat het een jongen is als iemand 0 cm lang is

**0.422** = slope = Per cm dat een persoon langer is, neemt **kans in logits** dat het om een jongen gaat met 0.422 logits toe

36 / 56

## Voorbeeld uit 'dagelijkse leven'

5 fictieve personen + wat voorspelt ons model

$$\text{Logit}(\text{Jongen} = 1) = -72.717 + 0.422 * \text{Lichaamslengte}$$

Lengte	Voorspelde logit Jongen
155	$-72.717 + 0.422 * 155 = -7.307$
165	$-72.717 + 0.422 * 165 = -3.087$
175	$-72.717 + 0.422 * 175 = 1.133$
185	$-72.717 + 0.422 * 185 = 5.353$
195	$-72.717 + 0.422 * 195 = 9.573$

37 / 56

## Voorbeeld uit 'dagelijkse leven'

5 fictieve personen + wat voorspelt ons model in logits

$$\text{Logit}(\text{Jongen} = 1) = -72.717 + 0.422 * \text{Lichaamslengte}$$

Lengte	Voorspelde logit Jongen	$\Delta$ Logit
155	$-72.717 + 0.422 * 155 = -7.307$	
		4.22
165	$-72.717 + 0.422 * 165 = -3.087$	
		4.22
175	$-72.717 + 0.422 * 175 = 1.133$	
		4.22
185	$-72.717 + 0.422 * 185 = 5.353$	
		4.22
195	$-72.717 + 0.422 * 195 = 9.573$	

38 / 56

## Voorbeeld uit 'dagelijkse leven'

5 fictieve personen + wat voorspelt ons model in probabiliteiten

$$\text{Logit}(\text{Jongen} = 1) = -72.717 + 0.422 * \text{Lichaamslengte}$$

Lengte	Voorspelde logit Jongen	$\Delta$ Logit	Voorspelde Prob. Jongen	$\Delta$ Prob.
155	$-72.717 + 0.422 * 155 = -7.307$		0.0007	
		4.22		0.0429
165	$-72.717 + 0.422 * 165 = -3.087$		0.0436	
		4.22		0.7128
175	$-72.717 + 0.422 * 175 = 1.133$		0.7564	
		4.22		0.2388
185	$-72.717 + 0.422 * 185 = 5.353$		0.9952	
		4.22		0.0047
195	$-72.717 + 0.422 * 195 = 9.573$		0.9999	

39 / 56

## Logistische regressie in R

40 / 56

# Functie `glm( )`

Werkwijze zeer gelijkaardig aan lineaire regressieanalyse

Functie `glm( )`

```
Model1 <- glm(Jongen ~ Lichaamslengte, data = DataGender, family = binomial())
```

Vervolgens

```
summary(Model1)
```

41 / 56

## Voorbeeld uit 'dagelijkse leven'

Eerst maken we data aan om mee te werken:

```
set.seed(1975) # Maakt voorbeeld reproduceerbaar
Lichaamslengte <- c(rnorm( 300 , mean = 162 , sd = 6.2), #Genereer 300 waarden met mean 162 en SD 6.2
                   rnorm( 300, mean = 184 , sd = 7))    #Genereer 300 waarden met mean 184 en SD 7
Jongen          <- c(rep( 0 , 300), rep ( 1 , 300))
DataGender <- data_frame(Lichaamslengte, Jongen)
```

Dan de analyse:

```
Model1 <- glm(Jongen ~ Lichaamslengte, data = DataGender, family = binomial())
```

42 / 56

# De output

```
summary(Model1)

Call:
glm(formula = Jongen ~ Lichaamslengte, family = binomial(), data = DataGender)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5329  -0.1701  -0.0030   0.1110   4.0325

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -72.71668     6.93641  -10.48  <2e-16 ***
Lichaamslengte  0.42193     0.04031   10.47  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 831.78  on 599  degrees of freedom
Residual deviance: 188.99  on 598  degrees of freedom
AIC: 192.99

Number of Fisher Scoring iterations: 7
```

43 / 56

## Welke kans ben je aan het voorspellen?

*Kans dat het een jongen is of kans dat het een meisje is?*

Hier: dummy die aanstaat voor jongen!

Standaard:

- Hoogste categorie (bv. waarde 1 bij een dummy)
- of laatste categorie alfabetisch (bv. categorie "Meisje" indien de variabele categorieën "Meisje" en "Jongen" heeft)

44 / 56

# De voorbeelddata

We maken enkele varianten van de variabele geslacht.  
Vervolgens herschatten we model met elke variant als afhankelijke variabele.

```
DataGender$Meisje <- c(rep(1, 300), rep(0, 300))
DataGender$Geslacht <- as.factor(c(rep("Meisje", 300), rep("Jongen", 300)))

M1_Meisje <- glm(Meisje ~ Lichaamslengte, data = DataGender, family = binomial())
M1_Geslacht <- glm(Geslacht ~ Lichaamslengte, data = DataGender, family = binomial())

coefficients(Model1)
```

```
(Intercept) Lichaamslengte
-72.7166813    0.4219255
```

```
coefficients(M1_Meisje)
```

```
(Intercept) Lichaamslengte
 72.7166813   -0.4219255
```

```
coefficients(M1_Geslacht)
```

```
(Intercept) Lichaamslengte
 72.7166813   -0.4219255
```

45 / 56

## Veranderen van referentiecategorie

De functie `relevel( )`

```
DataGender$Geslacht_x <- relevel(DataGender$Geslacht, 2)
M1_Geslacht_x <- glm(Geslacht_x ~ Lichaamslengte, data = DataGender, family = binomial())
summary(M1_Geslacht_x)
```

```
Call:
glm(formula = Geslacht_x ~ Lichaamslengte, family = binomial(),
    data = DataGender)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5329  -0.1701  -0.0030   0.1110   4.0325
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -72.71668    6.93641  -10.48  <2e-16 ***
Lichaamslengte  0.42193    0.04031   10.47  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 831.78  on 599  degrees of freedom
Residual deviance: 188.99  on 598  degrees of freedom
AIC: 192.99
```

```
Number of Fisher Scoring iterations: 7
```

46 / 56

# Onderpresteerders - meerdere voorspellers

Effect van 'Gender' en 'Ouders\_GraagLezenZ' op kans op onderpresteren ?

```
M1_PIRLS <- glm(Onderpresteren ~ Gender + Ouders_GraagLezenZ,
  data = Vlaanderen_1_2_3, family = binomial())
summary(M1_PIRLS)
```

Call:  
glm(formula = Onderpresteren ~ Gender + Ouders\_GraagLezenZ, family = binomial(),  
data = Vlaanderen\_1\_2\_3)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.1701	-0.6869	-0.5991	-0.4531	2.2943

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.37641	0.05319	-25.88	< 2e-16 ***
GenderGirls	-0.25424	0.07611	-3.34	0.000836 ***
Ouders_GraagLezenZ	-0.39405	0.03856	-10.22	< 2e-16 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4496.7 on 4634 degrees of freedom  
Residual deviance: 4377.4 on 4632 degrees of freedom  
(563 observations deleted due to missingness)  
AIC: 4383.4

Number of Fisher Scoring iterations: 4

47 / 56

## `alog( )` functie - PIRLS voorbeeld

Voor interpretatie handiger om in probabiliteiten te spreken

$$P(x = 1) = \frac{e^{\text{logit}(x=1)}}{1 + e^{\text{logit}(x=1)}}$$

Bv. Kans op onderpresteren voor een jongen is -1.376 logits (~ intercept)

In R

```
exp(-1.376)/(1+exp(-1.376))
```

```
[1] 0.2016522
```

Handiger: functie `alog( )` uit "OLP2 Functies.R"

```
alog(-1.376)
```

```
[1] 0.2016522
```

48 / 56



## alog( ) functie - PIRLS voorbeeld

```
# Probabiliteit tot onderpresteren voor jongens:  
alog(-1.376)
```

```
[1] 0.2016522
```

```
# Probabiliteit tot onderpresteren voor meisjes:  
alog(-1.376 - 0.254)
```

```
[1] 0.1638304
```

```
# Verschil in probabiliteit tot onderpresteren tussen jongens en meisjes:  
alog(-1.376) - alog(-1.376 - 0.254)
```

```
[1] 0.03782182
```

49 / 56

## alog( ) functie - PIRLS voorbeeld

Bij kwantitatieve voorspellers (bv. Ouders\_GraagLezenZ) fictieve respondenten als voorbeeld hanteren:

```
# Probabiliteit tot onderpresteren voor een jongen die gemiddeld scoort:  
alog(-1.376)
```

```
[1] 0.2016522
```

```
# Probabiliteit tot onderpresteren voor een jongen die 2 st.dev. hoger dan gemiddeld scoort:  
alog(-1.376 + 2 * -0.394)
```

```
[1] 0.1030302
```

```
# Probabiliteit tot onderpresteren voor een jongen die 2 st.dev. lager dan gemiddeld scoort:  
alog(-1.376 - 2 * -0.394)
```

```
[1] 0.3570939
```

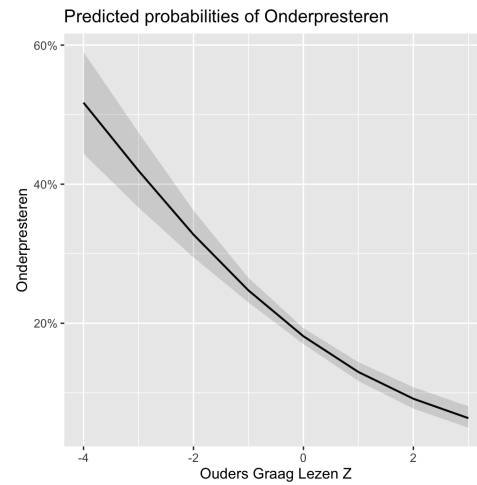
50 / 56

# Grafisch samenvatten van model

Functie `plot_model()` uit het pakket `sjPlot` (niet in OLP!)

```
library(sjPlot)

plot_model(M1_PIRLS,
  transform = NULL,
  type = "eff",
  terms = c("Ouders_GraagLezenZ")
)
```

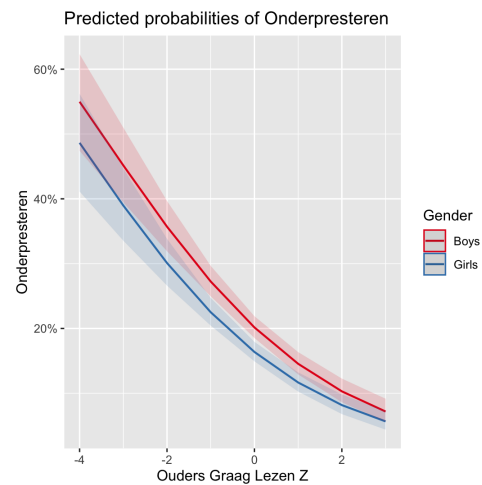


51 / 56

# Grafisch samenvatten van model

```
library(sjPlot)

plot_model(M1_PIRLS,
  transform = NULL,
  type = "eff",
  terms = c("Ouders_GraagLezenZ", "Gender")
)
```



52 / 56

# Interactie-effecten - PIRLS voorbeeld

## Interactie-effect van 'Gender' en 'Ouders\_GraagLezenZ' op kans op onderpresteren?

```
M2_PIRLS <- glm(Onderpresteren ~ Gender + Ouders_GraagLezenZ + Gender*Ouders_GraagLezenZ,  
  data = Vlaanderen_1_2_3, family = binomial())  
summary(M2_PIRLS)
```

```
Call:  
glm(formula = Onderpresteren ~ Gender + Ouders_GraagLezenZ +  
  Gender * Ouders_GraagLezenZ, family = binomial(), data = Vlaanderen_1_2_3)
```

```
Deviance Residuals:  
    Min       1Q   Median       3Q      Max  
-1.1653  -0.6870  -0.5989  -0.4517   2.2982
```

```
Coefficients:  
                Estimate Std. Error z value Pr(>|z|)  
(Intercept)      -1.375627   0.053804  -25.567  < 2e-16 ***  
GenderGirls       -0.255926   0.078173   -3.274  0.00106 **  
Ouders_GraagLezenZ -0.390525   0.053658   -7.278 3.39e-13 ***  
GenderGirls:Ouders_GraagLezenZ -0.007286   0.077165   -0.094  0.92477  
---
```

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

```
(Dispersion parameter for binomial family taken to be 1)
```

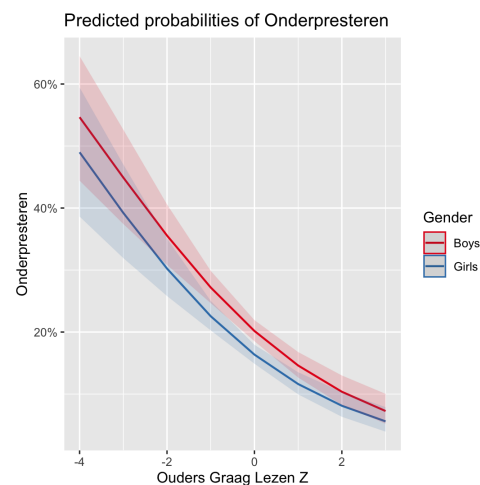
```
Null deviance: 4496.7 on 4634 degrees of freedom  
Residual deviance: 4377.4 on 4631 degrees of freedom  
(563 observations deleted due to missingness)  
AIC: 4385.4
```

```
Number of Fisher Scoring iterations: 4
```

53 / 56

## Grafisch samenvatten van model

```
library(sjPlot)  
  
plot_model(M2_PIRLS,  
  transform = NULL,  
  type = "eff",  
  terms = c("Ouders_GraagLezenZ", "Gender")  
)
```



54 / 56

# Time to pRactice!

## Instructies:

- Laat deze sessie open staan
- Open Blackboard opnieuw in een ander venster
- Ga naar de cursus GKN
- Ga naar de Blackboard Collaborate omgeving van je groep
- Zet je microfoon/video aan

Eén van ons maakt zo meteen een ronde langs de groepen!

55 / 56

## ExtRa's

Zit je te popelen om te weten ...

- hoe je toch random slopes voor kwalitatieve variabelen kan plotten zonder foutmelding?
- hoe je multilevel analyse gebruikt om longitudinale analyses te doen?

Check Blackboard (Leermateriaal > Materiaal per week)

56 / 56