De drie mooiste modellen deel 3

Medische statistiek
Logistische regressie

Hans Burgerhof
h.burgerhof@home.nl

Wat staat in deze tabel?

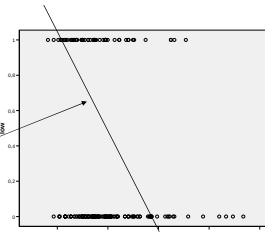
Table 5. Predictors of LBW risk: univariate analysis

Predictor variable	Unit	Odds ratio (95% CI)[a]	AUC ± SE ROC curve
Age	+5 years	1.1 (0.9–1.2)	0.54 ± 0.04
Parity	+1	1.0 (0.0–1.2)	0.51 ± 0.04
Intake of cigarettes[b]	1	1.7 (1.4–2.1)	0.86 ± 0.03
Smoking during pregnancy	Yes	4.1 (2.4–6.9)	0.68 ± 0.03
Marital status	Single	1.6 (0.8–3.0)	0.53 ± 0.05
Country of birth	Asia	1.3 (0.3–4.1)	0.52 ± 0.04
Economic quintile	+1	0.9 (0.7–1.1)	0.55 ± 0.04
Educational quintile	+1	0.9 (0.7–1.1)	0.54 ± 0.04

AUC = area under the receiver operating characteristic curve.
[a] Values are odds ratio and 95% confidence interval. Statistical significance at $p < 0.05$ level is indicated in bold-faced figures.
[b] Intake of cigarettes was coded as follows: 1 = 1–4 cigarettes per day; 2 = 5–10 cigarettes per day; and 3 = 11 or more cigarettes per day.

Een responsievariabele en meerdere verklarende variabelen

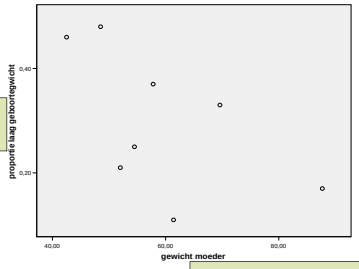
- Als Y continu, lineaire regressie
- Als Y dichotoom?



Interpretatie?

Proportie "laag geboortegewicht" per gewichtsklasse van de moeder

0≤ prop ≤ 1



Prop(Y=1) = β0 + β1*gewicht?

Logistische regressie

- De (natuurlijke) logaritme van de odds wordt als een lineaire functie van X gemodelleerd

$$odds = \frac{p}{1-p}$$

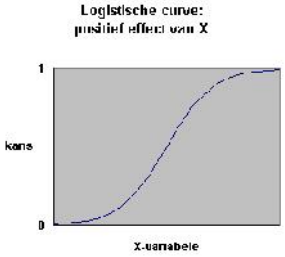
"kansverhouding"

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X \Leftrightarrow p = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

"logit"

Logistische functie

Logistische curve



Een dichotome verklarende X

smoking status during pregnancy * low birth weight Crosstabulation

			low birth weight		
			no (bwt >= 2500 g)	yes (bwt < 2500)	Total
smoking status during pregnancy	no	Count	86	29	115
		% within smoking status during pregnancy	74,8%	25,2%	100,0%
	yes	Count	44	30	74
		% within smoking status during pregnancy	59,5%	40,5%	100,0%
Total		Count	130	59	189
		% within smoking status during pregnancy	68,8%	31,2%	100,0%

$$odds_{rooksters} = \frac{0{,}405}{0{,}595} = 0{,}681$$

$$odds_{niet-rooksters} = \frac{0{,}252}{0{,}748} = 0{,}337$$

$$OR = \frac{odds_{rooksters}}{odds_{niet-rooksters}} = \frac{0{,}681}{0{,}337} = 2{,}02$$

Logistische regressie in SPSS

Odds Ratio

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	smoke	,704	,320	4,852	1	,028	2,022
	Constant	-1,087	,215	25,627	1	,000	,337

a. Variable(s) entered on step 1: smoke.

Logit(laag geboortegewicht=1) = -1,087 + 0,704*smoke

$$e^{\ smoke} = 2{,}022 = OR = \frac{odds_{rooksters}}{odds_{niet-rooksters}}$$

Geschatte proporties (mbv de logistische regressie)

$$\hat{\pi}(low) = \frac{e^{-1{,}087+0{,}704*smoke}}{1+e^{-1{,}087+0{,}704*smoke}}$$

$$\hat{\pi}(niet-rookster) = \frac{e^{-1{,}087}}{1+e^{-1{,}087}} = \frac{0{,}337}{1+0{,}337} = 0{,}252$$

$$\hat{\pi}(rookster) = \frac{e^{-1{,}087+0{,}704}}{1+e^{-1{,}087+0{,}704}} = \frac{0{,}682}{1+0{,}682} = 0{,}405$$

Drie toetsen

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	smoke	,704	,320	4,852	1	,028	2,022
	Constant	-1,087	,215	25,627	1	,000	,337

a. Variable(s) entered on step 1: smoke.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	4,867	1	,027
	Block	4,867	1	,027
	Model	4,867	1	,027

Logistische regressie

Likelihood ratio test

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4,924[a]	1	,026		
Continuity Correction[b]	4,236	1	,040		
Likelihood Ratio	4,867	1	,027		
Fisher's Exact Test				,036	,020
Linear-by-Linear Association	4,898	1	,027		
N of Valid Cases	189				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is

Chi-kwadraat uit de kruistabel

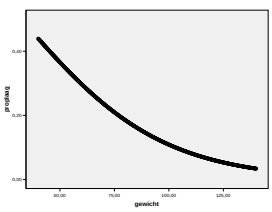
Een continue verklarende variabele

Variables in the Equation

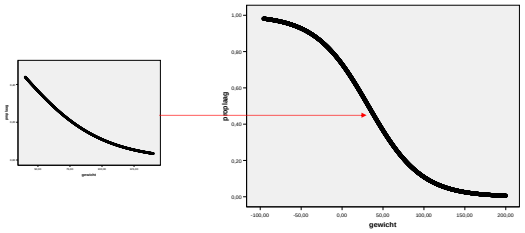
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	gewicht	-,031	,014	5,192	1	,023	,970
	Constant	,998	,785	1,616	1	,204	2,714

a. Variable(s) entered on step 1: gewicht.

$$\hat{\pi}(gewicht) = \frac{e^{0{,}998-0{,}031*gewicht}}{1+e^{0{,}998-0{,}031*gewicht}}$$



Uitvergroot op een kunstmatig domein



Interpretatie β1

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	gewicht	-,031	,014	5,192	1	,023	,970
	Constant	,998	,785	1,616	1	,204	2,714

a. Variable(s) entered on step 1: gewicht.

Odds(Moeder met gewicht x) $e^{\beta_0+\beta_1 x}$

Odds (moeder met gewicht x + 1) $e^{\beta_0+\beta_1(x+1)}$

Odds Ratio $\frac{e^{\beta_0+\beta_1(x+1)}}{e^{\beta_0+\beta_1 x}} = e^{\beta_0+\beta_1(x+1)-(\beta_0+\beta_1 x)} = e^{\beta_1}$

Voor twee moeders die 10 kg verschillen: $(0{,}97)^{10}$

Meervoudige logistische regressie

$$\pi(x_1,\ldots,x_k) = \frac{e^{\beta_0+\beta_1 x_1+\ldots+\beta_k x_k}}{1+e^{\beta_0+\beta_1 x_1+\ldots+\beta_k x_k}}$$

Kans op laag geboorte-gewicht

Ln(odds) = $\beta_0+\beta_1 x_1+\ldots+\beta_k x_k$

Vergelijken van modellen

Model 1
Gewicht en rookgedrag

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	224,341[a]	,053	,075

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Verschil 1,46 met 1df

Model 2
Gewicht, rookgedrag en leeftijd

Model Summary

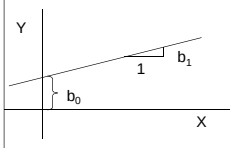
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	222,879[a]	,060	,085

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Summary regression models

Linear regression

Y continuous
Estimated Model: $Y = b_0 + b_1X$

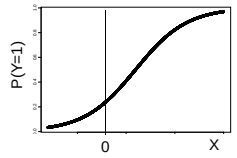


b_0 is the intercept
b_1 is the slope

t-test for individual variables
F-test for comparing models

Logistic regression

Y binary
Estimated model: $P(Y=1) = \frac{e^{b_0+b_1x}}{1+e^{b_0+b_1x}}$



e^{b_0} is the odds for Y=1 for a person with X =0
e^{b_1} is the odds ratio (for X = x +1 to X = x)

Wald test for individual variables
Likelihood ratio test for comparing models