

---

## 44. Logistische regressie

Logistische regressie (*logistic regression*) wordt gebruikt wanneer je categorische uitkomsten wil voorspellen met twee of meer categorieën. De voorspellende (onafhankelijke) variabele kan categorisch of continue zijn, de afhankelijke variabele categorisch (dit in tegenstelling tot multiële regressie waarbij de afhankelijke variabele continue is). In dit hoofdstuk wordt een voorbeeld gegeven van binaire logistische regressie. Hierbij is de afhankelijke variabele dichotoom (twee categorieën of waarden). Als je afhankelijke variabele meer dan twee categorieën of waarden heeft dien je gebruik te maken van de Multinomial Logistic set.

### Assumpties

De assumpties behorende bij een logistische regressieanalyse zijn:

- er dient geen sprake te zijn van multicollineariteit
- check de steekproefgrootte (is deze groot genoeg?) en kijk of er sprake is van uitbijters

### Voorbeeld onderzoeksvraag

Nu volgt een voorbeeld van een onderzoeksvraag die we gaan onderzoeken met logistische regressie.

**Onderzoeksvraag:** Welke factoren vergroten de kans dat de proefpersoon aangeeft problemen te hebben met zijn of haar eetlust?

**Wat heb je nodig:**

- een categorische (dichotome) afhankelijke variabele (problemen met eetlust: Nee/Ja, gecodeerd met 0/1)
- twee of meer continue of categorische (onafhankelijke) variabelen (in dit geval geslacht en leeftijd)

**Wat doet het:** Logistische regressie vertelt je in welke mate je set van voorspellende (onafhankelijk) variabelen je categorische afhankelijke variabele verklaren.

### Procedure

Nu volgt de procedure voor het uitvoeren van een binaire logistische regressie.

1. Kies voor **Analyze**, klik op **Regression** en vervolgens op **Binary Logistic**.
  2. Kies je categorische afhankelijke variabele (in dit geval eetlust) en sleep deze naar de **Dependent** sectie. Klik dan op je voorspellende onafhankelijke variabelen (in dit geval geslacht en leeftijd) en sleep deze variabelen naar de sectie **Covariates**. Zorg ervoor dat bij de sectie **Methode** de optie **Enter** staat aangevinkt.
  3. Als je categorische voorspellers hebt (van nominaal of ordinaal meetniveau) dan moet je op **Categorical** klikken. Selecteer je categorische variabelen en sleep deze naar de **Categorical covariates** box. In dit geval dus geslacht. Selecteer elke categorische variabele en klik op de knop **First** in de **Change contrast** sectie. Klik op **Change** en je ziet 'first' verschijnen achter de variabelenaam. Herhaal dit voor de andere categorische variabelen. Klik op **Continue**.
  4. Klik op **Options**. Selecteer **Classification plots**, **Hosmer-Lemeshow**, **goodness of fit**, **Casewise listing of residuals** en **CI for Exp(B)**.
  5. Klik op **Continue** en dan op **OK** (of op **Paste** als je de analyse wil opslaan in de Syntax Editor).
-

---

## Interpretatie van de output

### Case Processing Summary

Unweighted Cases <sup>a</sup>		N	Percent
Selected Cases	Included in Analysis	30	100,0
	Missing Cases	0	,0
	Total	30	100,0
Unselected Cases		0	,0
Total		30	100,0

a. If weight is in effect, see classification table for the total number of cases.

### Dependent Variable Encoding

Original Value	Internal Value
geen problemen	0
wel problemen	1

### Categorical Variables Codings

		Frequency	Parameter coding	
			(1)	(2)
geslacht	man	12	,000	,000
	vrouw	17	1,000	,000
2		1	,000	1,000

## Block 0: Beginning Block

### Classification Table<sup>a,b</sup>

Observed		Predicted		Percentage Correct
		geen problemen	wel problemen	
Step 0	eetlust	0	15	,0
		0	15	100,0
Overall Percentage				50,0

a. Constant is included in the model.

b. The cut value is ,500

Hierboven zie je bij **Block 0** de resultaten van de analyse zonder dat één van de onafhankelijke variabelen is gebruikt in het model. Dit is een baseline ter vergelijking met de rest van het model.

---

---

## Block 1: Method = Enter

### Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	5,042	3	,169
	Block	5,042	3	,169
	Model	5,042	3	,169

### Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	36,547 <sup>a</sup>	,155	,206

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

### Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	11,839	6	,066

Hierboven zie je in Blok 1 dat de onafhankelijke variabelen zijn toegevoegd in het model. De **Omnibus Tests of Model Coefficients** geeft een algemene indicatie over goed het model presteert in vergelijking met Blok 0. Dit is de 'goodness of fit' test. Je wil graag een significant resultaat ( $p < .05$ ). In het voorbeeld hierboven is het resultaat helaas niet significant ( $p = .169$ ). Dit betekent dat het tweede model (waaraan de onafhankelijke variabelen zijn toegevoegd) niet beter is dan het baseline model.

---