

---

## 40. Meerdere regressieanalyse

### Introductie

De multiple regressieanalyse kijkt of er een (voorspellend) verband is op basis van de correlatie van meerdere onafhankelijke variabelen met de afhankelijke variabelen. De multiple regressieanalyse maakt gebruik van continue of ordinale data maar kan daarnaast ook één of meerdere categorische variabelen meenemen als onafhankelijke variabelen.

In principe kunnen factorial ANOVA en meerdere regressie hetzelfde berekenen. Factorial ANOVA wordt in de praktijk vaker gebruikt voor experimenteel onderzoek en de multiple regressie meestal voor niet experimenteel onderzoek.

Er bestaan drie soorten meerdere regressie: standaard, hiërarchisch of stapsgewijs.

### Assumpties

De assumpties die horen bij een multiple regressieanalyse zijn:

- normale verdeling (de residuen zijn perfect normaal verdeeld rond de voorspelde afhankelijke variabele scores)
- lineaire relatie tussen onafhankelijke en afhankelijke variabele
- homogeniteit van varianties (Homoscedasticiteit). Dit kun je checken door te kijken naar de standaarddeviaties. Hierbij is de regel dat de grootste SD kleiner moet zijn dan 2x de kleinste SD
- onafhankelijke groepen
- geen perfecte multicollineariteit. Er is sprake van multicollineariteit wanneer de onafhankelijke variabelen sterk met elkaar zijn gecorreleerd (wanneer de correlatiecoëfficiënt .9 is of groter).

### Voorbeeld onderzoeksvraag

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

**Onderzoeksvraag:** In hoeverre voorspellen tentamenstress en leeftijd de scores op de depressietest?

**Wat heb je nodig:**

- één continue afhankelijke variabele (in dit geval de depressiescores)
- twee of meer continue of dichotome onafhankelijke variabelen (in dit geval tentamenstress en leeftijd)

### Procedure

Nu volgt de procedure voor het uitvoeren van een multiple regressieanalyse.

1. Kies **Analyze** en selecteer daarna **Regression** en **Linear**.
  2. Selecteer je afhankelijke variabele (depressiescores) en sleep deze naar de **Dependent** sectie.
  3. Selecteer je onafhankelijke variabelen (tentamenstress en leeftijd) en sleep deze naar de **Independent** sectie.
  4. Kijk bij **Method** of **Enter** is aangevinkt (dit zorgt voor een standaard multiple regressie).
  5. Klik op **Statistics**. Selecteer de volgende onderdelen: **Estimates, Confidence Intervals, Model fit, Descriptives, Part and partial correlation, and Collinearity diagnostics**. Bij de **Residuals** sectie selecteer je **Casewise diagnostics** en **Outliers outside 3 standard deviations**. Klik op **Continue**.
-

6. Klik op **Options**. In de **Missing values** sectie selecteer je **Exclude cases pairwise**. Klik op **Continue**.
7. Klik op **Plots**. Klik op **\*ZRESID** en de pijltjestoets om deze te verplaatsen naar de Y-sectie. Klik op **\*ZPRED** en de pijltjestoets om deze te verplaatsen naar de X-sectie. In de sectie genaamd **Standardized Residual Plots** vink je **Normal probability plot** aan. Klik op **Continue**.
8. Klik op **Save**. In de sectie genaamd **Distances** selecteer je **Mahalanobis** en **Cook's**. Klik op **Continue** en dan op **OK** (of op **Paste** als je de analyse wil opslaan in de Syntax Editor).

## De output interpreteren

### Descriptive Statistics

	Mean	Std. Deviation	N
depressiescores	2,87	1,332	30
tentamenstress	2,80	1,243	30
leeftijd	24,33	1,470	30

### Correlations

		depressiescores	tentamenstress	leeftijd
Pearson Correlation	depressiescores	1,000	,462	,323
	tentamenstress	,462	1,000	,075
	leeftijd	,323	,075	1,000
Sig. (1-tailed)	depressiescores	.	,005	,041
	tentamenstress	,005	.	,346
	leeftijd	,041	,346	.
N	depressiescores	30	30	30
	tentamenstress	30	30	30
	leeftijd	30	30	30

In de Correlations tabel is te zien hoe de verschillende variabelen aan elkaar zijn gerelateerd. Je kunt nu de multicollineariteit checken (onderzoeken of de onafhankelijke variabelen zeer sterk aan elkaar zijn gerelateerd).

### Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,545 <sup>a</sup>	,297	,245	1,158

a. Predictors: (Constant), leeftijd, tentamenstress

b. Dependent Variable: depressiescores

### ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	15,292	2	7,646	5,707	,009 <sup>b</sup>
	Residual	36,175	27	1,340		
	Total	51,467	29			

a. Dependent Variable: depressiescores

b. Predictors: (Constant), leeftijd, tentamenstress

Kijk vervolgens in de **Model Summary** tabel naar de waarde van de **R Square**. Deze waarde vertelt in hoeverre de variantie van de afhankelijke variabele wordt voorspeld door het model (in dit geval leeftijd en tentamenstress). In het voorbeeld is deze waarde .245. Dit betekent dat het model 24.5% van de variantie in de afhankelijke variabele (depressiescores) voorspelt. In de **ANOVA** tabel is te zien dat het resultaat significant is ( $p < .05$ ).

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics		
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF	
1	(Constant)	-4,841	3,571		-1,356	,186	-12,167	2,486						
	tentamenstress	,472	,173	,440	2,722	,011	,116	,828	,462	,464	,439	,994	1,006	
	leeftijd	,262	,147	,290	1,790	,085	-,038	,563	,323	,326	,289	,994	1,006	

a. Dependent Variable: depressiescores

Vervolgens kun je in de **Coefficients** tabel kijken welke variabelen in het model bijdragen aan de voorspelling van de afhankelijke variabele. Kijk hiervoor in de kolom **Beta** onder **Standardized Coefficients**. Kijk welke Beta het grootst is (in dit geval die van tentamenstress, namelijk .440). Dit betekent dat tentamenstress een grotere bijdrage heeft aan de voorspelling van depressiescores dan leeftijd. Kijk vervolgens naar **Sig.** Te zien is dat tentamenstress een significante p-waarde heeft, leeftijd heeft dit echter niet. Dit betekent dat tentamenstress een *unieke* bijdrage heeft aan de voorspelling. Als de p-waarde groter is dan .05 (zoals het geval is bij leeftijd) dan betekent dit dat deze variabele geen significante unieke bijdrage levert aan de voorspelling van je afhankelijke variabele.