

Oplossingen hoofdstuk 11

1. Hierbij vind je de resultaten van het onderzoek naar de relatie tussen een leestest en een schoolrapport voor lezen. Deze gegevens hebben betrekking op een regressieanalyse bij 12 leerlingen.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,855 ^a	,731	,704	6,781

a. Predictors: (Constant), leestest

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1249,864	1	1249,864	27,183	,000 ^a
	Residual	459,803	10	45,980		
	Total	1709,667	11			

a. Predictors: (Constant), leestest

b. Dependent Variable: schoolrapport

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1,692	5,871		-,288	,779
	leestest	1,232	,236	,855	5,214	,000

a. Dependent Variable: schoolrapport

Tevens de SD van de beide variabelen:

Statistics

		leestest	schoolrapport
N	Valid	12	12
	Missing	0	0
Std. Deviation		8,649	12,467

Stel de regressielijn op van het model: $Y = a + bX$. (Y is het schoolrapport en X de test)

$$Y = -1,69 + 1,23X$$

Wat is de concrete betekenis van deze b-waarde?

Deze b-waarde betekent dat wanneer de variabele X met één eenheid toeneemt de variabele Y met een waarde b zal toenemen.

Hoeveel bedraagt de Pearson correlatie tussen beide variabelen?

Deze bedraagt .86

Welk is de determinatiecoëfficiënt?

Deze bedraagt .73. Dat wil zeggen dat 73% van de verschillen in de afhankelijke variabele verklaard kunnen worden door dit model.

Hoe groot is de standaardfout van estimatie?

Deze bedraagt 6,78.

Dit wil zeggen dat indien de uitslag voor het schoolrapport voorspeld zal worden op basis van deze regressielijn, we in tweederde van de gevallen een fout zullen maken die kleiner zal zijn dan 6,78.

Kunnen deze resultaten iets zeggen over de betrouwbaarheid, validiteit of ijking van deze leestest?

Ze hebben enkel betrekking op de validiteit.

Indien we de criteriumuitslag met twee zouden vermenigvuldigen (aldus ontstaat Y2).

Wat is daarvan de betekenis voor de nieuwe regressielijn, waarbij we deze Y2 trachten te voorspellen op grond van de testuitslag X?

Ofschoon de correlatie tussen X en Y2 dezelfde is als tussen X en Y1, zal er een nieuw regressiemodel geformuleerd dienen te worden. Met andere woorden de constante en de regressiecoëfficiënt zullen veranderen.

Zal de determinatiecoëfficiënt veranderen?

Neen, omdat de correlatie niet zal veranderen ten gevolge van deze lineaire transformatie. En bijgevolg de determinatiecoëfficiënt eveneens niet.

Zal de standaardfout van estimatie veranderen?

Deze zal wel veranderen.

Omdat deze standaardfout recht evenredig is met de standaarddeviatie van de afhankelijke variabele, zal deze standaardfout van estimatie veranderen. De standaarddeviatie van de afhankelijke variabele zal met twee vermenigvuldigd worden. Dus...

Zal de Pearson correlatie veranderen?

Neen. Deze lineaire transformatie heeft geen impact op de correlatie.

Zal de bèta-coëfficiënt veranderen?

Neen, omdat deze bij een enkelvoudige regressie gelijk is aan de correlatie van Pearson.

2. Een logopediste onderzoekt het verband tussen de scores van een leesvoorwaardentest (X) en het resultaat van een leesniveautest (Y). Van 6 leerlingen werden volgende gegevens verzameld.

Leerling	Score leesvoorwaarden	Score leesniveautest
1	3	6
2	5	8
3	0	3
4	2	5
5	2	6
6	3	5

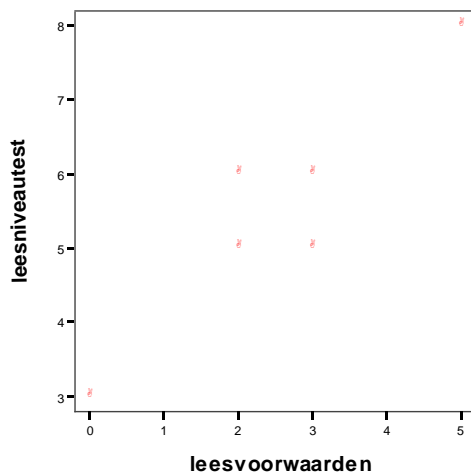
Kan men zeggen dat er een verband bestaat tussen beide variabelen?

Maak hierbij gebruik van SPSS.

Interpreteer dit resultaat.

Kunnen we op grond van de score voor de leesvoorwaardentest iets voorspellen over de leesniveautest? Hoe goed is deze voorspelling?

De scatterplot toont dat er een positief rechtlijnig verband bestaat tussen de beide variabelen. Zie hiervoor de output in de Viewer.



De Pearson correlatie bedraagt .93, zoals onderstaande output aantoont.

Correlations

		leesvoorwaarden	leesniveautest
leesvoorwaarden	Pearson Correlation	1	,926**
	Sig. (2-tailed)		,008
	N	6	6
leesniveautest	Pearson Correlation	,926**	1
	Sig. (2-tailed)	,008	
	N	6	6

** . Correlation is significant at the 0.01 level (2-tailed).

Vanuit dergelijke hoge correlatie is het zinvol een regressielijn te tekenen.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,926 ^a	,857	,822	,694

a. Predictors: (Constant), leesvoorwaarden

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11,574	1	11,574	24,038	,008 ^a
	Residual	1,926	4	,481		
	Total	13,500	5			

a. Predictors: (Constant), leesvoorwaarden

b. Dependent Variable: leesniveautest

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3,185	,551		5,785	,004
	leesvoorwaarden	,926	,189	,926	4,903	,008

a. Dependent Variable: leesniveautest

We kunnen dus schrijven dat het $Y = 3,18 + 0,93X$, waarbij Y het leesniveau is en X het resultaat van de leesvoorwaardentest.

Merk op dat deze oefening een louter didactische functie heeft, en dat voor het toepassen van een regressielijn het aantal respondenten in feite te weinig is.

3. Een kinderpsycholoog wenst het IQ van kinderen te voorspellen bij een leeftijd van 10 jaar aan de hand van de leeftijd (in maanden) waarop het kind zijn eerste woordjes sprak. Voor een groepje van 5 kinderen vond hij volgende waarden op deze variabelen:

Leeftijd eerste woordjes	IQ op latere leeftijd
15	95
26	71
10	83
9	91
15	102

Bereken de correlatie via SPSS.

Eerder hebben we (hoofdstuk IX) aangetoond dat er een negatieve (en rechtlijnige) samenhang bestaat tussen de beide variabelen. De Pearson correlatie bedroeg $-.56$. We stellen de regressielijn op via PASW.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,560 ^a	,314	,085	11,390

a. Predictors: (Constant), leeftijdwoordjes

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	178,022	1	178,022	1,372	,326 ^a
	Residual	389,178	3	129,726		
	Total	567,200	4			

a. Predictors: (Constant), leeftijdwoordjes

b. Dependent Variable: IQ

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	103,235	13,650		7,563	,005
	leeftijdwoordjes	-,989	,844	-,560	-1,171	,326

a. Dependent Variable: IQ

Op grond van deze output kunnen we schrijven dat het IQ gelijk is aan 103,24 verminderd met 0,99 keer de leeftijd in maanden waarop het kind de eerste woordjes kon spreken. $Y = 103,24 - 0,99 * X$.

Let op in deze output staat dat de correlatie gelijk is aan .56, maar omdat de b waarde negatief is weten we dat deze correlatie negatief moet zijn. Dat bleek ook eerder bij de berekening van de Pearson correlatie.

Indien we een regressielijn zouden maken hoe groot is dan de standaardfout van estimatie? De standaardfout van estimatie is gelijk aan 11,39. Dat wil zeggen dat indien we het IQ van een kind willen voorspellen op grond van de leeftijd dat het kind kan spreken we in tweederde van de gevallen een fout zullen maken die kleiner is dan 11,39.

Hoe groot is de determinatiecoëfficiënt?

Deze is .31. Dat wil zeggen dat we 31% van de verschillen van de IQ's van kinderen kunnen verklaren op grond van de leeftijd in maanden waarop het kind de eerste woordjes kon zeggen. Ook kunnen we stellen dat 69% van deze verschillen in IQ niet verklaard worden door dit model.

Indien een kind op 8 maand de eerste woordjes zou spreken, welk IQ verwacht je dan?

We zouden hierbij de volgende regressielijn kunnen invullen.

$$IQ = 103,24 - 0,99 * 8 = 95,32.$$

Maar aangezien het model van deze regressieanalyse niet significant is, kunnen we in feite geen uitspraken doen over het IQ van een kind dat de eerste woordjes zegt op 8 maanden.

4. De fouten in het dictee.

Ofschoon de correlatie tussen beide variabelen niet erg hoog is en het aantal respondenten eerder gering, kunnen we onderzoeken of we een regressielijn kunnen tekenen.

We starten in SPSS met Analyze - Regression - Lineair. Let hierbij op in het volgende keuzemenu dat fouten in het dictee de dependent variabele vormt en het aantal boeken de independent. Na dit commando verschijnt volgende output in de Viewer.

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	aantal gelezen boeken ^b		Enter

a. Dependent Variable: aantal fouten in dictee

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,362 ^a	,131	-,086	1,864

a. Predictors: (Constant), aantal gelezen boeken

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2,100	1	2,100	,604	,480 ^b
	Residual	13,900	4	3,475		
	Total	16,000	5			

a. Dependent Variable: aantal fouten in dictee

b. Predictors: (Constant), aantal gelezen boeken

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5,600	2,194		2,552	,063
	aantal gelezen boeken	-,300	,386	-,362	-,777	,480

a. Dependent Variable: aantal fouten in dictee

In deze output lezen we dat de Pearson correlatie tussen de twee variabelen .36 bedraagt en dat de determinatiecoëfficiënt slechts 13% is. Dat wil zeggen dat 87% van de verschillen in het aantal fouten in het dictee niet verklaard kan worden door de hoeveelheid boeken dat gelezen wordt. Dit is dus een pover resultaat.

In de ANOVA tabel lezen we dat het model niet significant is. Dat wil zeggen dat de samenhang in de steekproef niet veralgemeend kan worden naar de populatie. Tot slot vinden we in de derde tabel de coëfficiënten van de regressie. Deze regressie zouden we kunnen schrijven als: Fouten in dictee = 5,60 - 0,30 * aantal boeken.

Gezien echter de matige correlatie en het niet significant zijn van het model is het niet toegelaten om deze regressielijn te hanteren om toekomstige resultaten te voorspellen. Dus het is in dit geval niet zinvol om de uitkomst van de regressie te gebruiken.

5. Lees het betreffende artikel en ga na of je tot dezelfde uitkomst kunt komen.

6. Bestudeer het artikel en controleer de uitkomsten via SPSS.