

---

## 33. Pearson correlatie

In het vorige hoofdstuk werd de *Pearson productmoment correlation coefficient* ( $r$ ), hierna verder afgekort met Pearson  $r$ , reeds geïntroduceerd. De Pearson  $r$  meet het verband of de relatie tussen twee variabelen van intervalniveau. De Pearson  $r$  heeft een waarde tussen de  $-1$  en  $+1$ . Het min- of plusteken geeft de richting aan van de correlatie. De grootte van de absolute waarde (dus door het signaal ervoor te negeren) geeft de sterkte aan van het verband. Immers: een correlatie verder weg van de nul staat gelijk aan een sterkere correlatie.

### Correlaties en spreidingsdiagrammen

Wanneer sprake is van een perfecte correlatie van  $-1$  of  $1$ , dan is in het bijbehorende spreidingsdiagram een rechte lijn te zien, ofwel naar links of wel naar rechts (afhankelijk van of de correlatie negatief of positief is). Wanneer sprake is van vrijwel geen correlatie (een coëfficiënt die nagenoeg nul is), dan is geen rechte lijn te zien. In het spreidingsdiagram zijn dan losse puntjes te zien zonder patroon.

### Pre-analyse voor de correlatieanalyse

Voordat de correlatieanalyse kan worden uitgevoerd dient eerst een aantal assumpties te worden gecheckt. Deze assumpties zijn als volgt:

- Er moet sprake zijn van een normale verdeling.
- Er moet sprake zijn van een lineair verband tussen de variabelen.
- De groepen moeten onafhankelijk zijn.

Deze assumpties worden met de volgende stappen gecheckt:

1. Het checken voor uitbijters.
2. Het inspecteren van de verdeling.
3. Het bepalen van de richting van de relatie tussen de variabelen.

#### **Stap 1. Het checken voor uitbijters**

Eerst dien je het spreidingsdiagram van de uitbijters te bekijken. Extreme uitbijters dienen nader te worden gecontroleerd. Als je een uitbijter hebt geïdentificeerd en je wil graag het nummer van deze proefpersoon terugvinden, dan kun je het **Data Label Model** icoon gebruiken in de Chart Editor. Klik tweemaal op het icoon dat een beetje op een schotroos lijkt (of kies **Data Label Mode** van het **Elements** menu) en beweeg de cursor naar een punt in de grafiek dat je wil identificeren. Klik hier eenmaal op en het nummer zal verschijnen.

#### **Stap 2. Het inspecteren van de verdeling van datapunten**

De verdeling van de datapunten kan verschillende dingen vertellen. Wanneer de datapunten bijvoorbeeld wijd verspreid zijn, dan suggereert dit een zeer lage correlatie. Als de punten vlakbij elkaar liggen in een rechte lijn dan suggereert dit een zeer hoge correlatie.

#### **Stap 3. Het bepalen van de richting van de relatie tussen de variabelenaam**

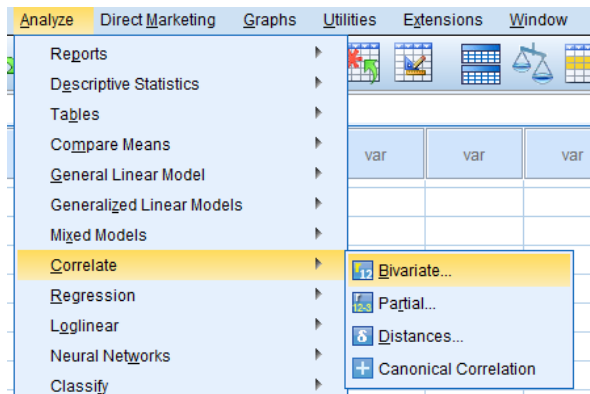
Het spreidingsdiagram kan informatie geven over of de relatie tussen de twee variabelen positief of negatief is. Een lijn die naar boven loopt suggereert een positieve relatie: hoge scores op  $X$  zijn geassocieerd met hoge scores op  $Y$  (of lage scores op  $X$  zijn geassocieerd met lage scores op  $Y$ ). Een lijn die naar beneden loopt suggereert een negatieve relatie: hoge scores op  $X$  zijn geassocieerd met lage scores op  $Y$  (of lage scores op  $X$  zijn geassocieerd met hoge scores op  $Y$ ).

Nadat je de verdeling van de scores in het spreidingsdiagram hebt bekeken dan kun je beginnen met de correlatieanalyse. Voordat je begint met de volgende procedure: kies **Edit** van het menu, selecteer **Options** en zorg ervoor dat in het **General** blad **No scientific notation for small numbers in tables** staat aangevinkt in de **Output** sectie.

---

## Procedure voor het opstellen van een Pearson r of Spearman rho

- Kies voor **Analyze**, selecteer daarna **Covariate** en vervolgens **Bivariate**.



- Selecteer de twee variabelen waarin je geïnteresseerd bent en sleep deze variabelen naar de **Variables box**.
- In **Correlation Coefficients** staat de **Pearson r** automatisch aangevinkt. Als je liever de Spearman rho wil uitrekenen vink je de Pearson r uit en kies je de **Spearman rho** aan.
- Klik op **Options**. Voor **Missing Values**, klik op **Exclude cases pairwise**. Onder **Options** kun je tevens gemiddelden en standaarddeviaties opvragen indien nodig.
- Klik op **Continue** en vervolgens op **OK** (of op **Paste** om de analyse op te slaan in de Syntax Editor).

### Descriptive Statistics

	Mean	Std. Deviation	N
tentamenstress	2,8000	1,24291	30
alcoholconsumptie	3,1000	1,39827	30

### Correlations

		tentamenstress	alcoholconsumptie
tentamenstress	Pearson Correlation	1	,647**
	Sig. (2-tailed)		,000
	N	30	30
alcoholconsumptie	Pearson Correlation	,647**	1
	Sig. (2-tailed)	,000	
	N	30	30

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

In de output hierboven is te zien dat de Pearson correlatie tussen de hoeveelheid tentamenstress en de hoeveelheid alcoholconsumptie ,647 is. Deze correlatie is significant. Dit betekent dat er een significant verband is tussen de hoeveelheid tentamenstress en de hoeveelheid alcoholconsumptie. De correlatie is positief. Dit betekent dat grotere hoeveelheden tentamenstress zijn geassocieerd met grotere hoeveelheden alcoholconsumptie (en lagere hoeveelheden tentamenstress zijn geassocieerd met kleinere hoeveelheden alcoholconsumptie).

De sterkte van de relatie kan klein, medium of groot zijn. Hiervoor gelden de volgende vuistregels:

- klein:  $r = .10$  tot  $.29$
- medium:  $r = .30$  tot  $.49$
- groot:  $r = .50$  tot  $1.0$

In het voorbeeld is dus sprake van een sterke relatie.