

---

## 27. Beschrijvende statistiek

Wanneer je er zeker van bent dat er geen sprake is van error in je databestand dan kun je beginnen met de beschrijvende fase van je data-analyse. Dit noemde we beschrijvende statistiek ofwel *descriptive statistics*. Deze hebben als doel:

- het beschrijven van de kenmerken van je steekproef in de methodesectie van je artikel
- het checken van je variabelen om te onderzoeken of je aan bepaalde assumpties voldoet behorende bij de statistiekt technieken die je wilt uitvoeren om je onderzoeksvragen te beantwoorden
- het stellen van specifieke onderzoeksvragen

Wanneer het om onderzoek met menselijke proefpersonen gaat is het vrijwel noodzakelijk om algemene kenmerken te verzamelen. Denk hierbij aan het aantal mensen in de steekproef, het aantal of percentage mannen en vrouwen, de leeftijden, en opleidingsniveau.

Voorbeelden van beschrijvende statistieken zijn het gemiddelde, de standaarddeviatie en de spreiding van de scores.

### Procedure voor het maken van een codeboek

Als je alleen een snelle samenvatting wilt van de kenmerken van je variabelen in je databestand heb je waarschijnlijk genoeg aan een codeboek. Hier volgt de procedure om een codeboek te verkrijgen.

1. Klik op **Analyze** en ga naar **Reports** en kies **Codebook**.
2. Selecteer de variabelen die je wilt (bijvoorbeeld geslacht, leeftijd) en sleep deze variabelen naar het **Codebook Variables** venster.
3. Klik op het **Output** blad en vink alle **Opties** uit, behalve **Label**, **Value Labels** and **Missing Values**, uit.
4. Klik op **Statistics** en zorg ervoor dat alle opties in beide secties zijn aangevinkt.
5. Klik op **OK** (of op **Paste** om alles op te slaan in de Syntax Editor).

De syntax is dan als volgt:

```
DATASET ACTIVATE DataSet1.
```

```
CODEBOOK geslacht [n] leeftijd [s]
```

```
/VARINFO LABEL VALUELABELS MISSING
```

```
/OPTIONS VARORDER=VARLIST SORT=ASCENDING MAXCATS=200
```

```
/STATISTICS COUNT PERCENT MEAN STDDEV QUARTILES.
```

Hieronder is de output te zien:

---

**geslacht**

		Value	Count	Percent
Standard Attributes	Label	<none>		
Valid Values	0	man	12	40,0%
	1	vrouw	17	56,7%
	2		1	3,3%

**leeftijd**

		Value
Standard Attributes	Label	<none>
N	Valid	30
	Missing	0
Central Tendency and Dispersion	Mean	24,33
	Standard Deviation	1,470
	Percentile 25	23,00
	Percentile 50	24,00
	Percentile 75	25,00

Deze output geeft je een snelle samenvatting van de proefpersonen in je databestand. Als je meer gedetailleerde informatie wilt hebben kun je deze verkrijgen door middel van **Frequencies**, **Descriptive** of **Explore**. Om informatie te verkrijgen van categorische variabelen kun je gebruikmaken van **Frequencies** (zie het volgende hoofdstuk).

### Procedure voor het verkrijgen van beschrijvende statistiek voor continue variabelen

Voor continue variabelen (bijvoorbeeld leeftijd) is het makkelijker om **Descriptives** te gebruiken. Deze analyse verschaft de basis ‘samenvattende’ statistieken zoals het gemiddelde, de mediaan en de standaarddeviatie. Het betrouwbaarheidsinterval vind je door middel van **Explore**.

De procedure behorende bij het verkrijgen van beschrijvende statistiek voor continue variabelen is:

1. Klik op **Analyze** selecteer daarna **Descriptive Statistics** en vervolgens **Descriptives**.
2. Klik alle continue variabelen aan waarvan je graag beschrijvende statistieken voor wilt verkrijgen. Klik vervolgens op de pijltjestoets (wijzend naar rechts) om deze variabelen te verschuiven naar de **Variables** sectie.
3. Klik op **Options**. Zorg ervoor dat de volgende statistieken zijn aangevinkt: **mean**, **standard deviation**, **minimum**, **maximum** en klik vervolgens ook **skewness** en **kurtosis** aan.
4. Klik op **Continue** en vervolgens op **OK** (of op **Paste** om de analyse op te slaan in de Syntax Editor).

De syntax die bij deze procedure wordt gegenereerd is:

```
DESCRIPTIVES
```

```
VARIABLES = leeftijd
```

```
/STATISTICS=MEAN STDDEV MIN MAX KURTOSIS SKEWNESS
```

Hieronder vind je de bijbehorende output.

---

---

### Descriptive Statistics

	N Statistic	Minimum Statistic	Maximum Statistic	Mean Statistic	Std. Deviation Statistic	Skewness		Kurtosis	
						Statistic	Std. Error	Statistic	Std. Error
leeftijd	30	21	27	24,33	1,470	-,067	,427	-,434	,833
Valid N (listwise)	30								

De **Skewness** functie geeft informatie over de symmetrie van de verdeling van de scores. **Kurtosis** geeft informatie over de piek van de verdeling. Als de verdeling van de scores perfect normaal verdeeld zouden zijn zouden zowel de Skewness als de Kurtosis nagenoeg nul zijn. Een positieve waarde van skewness indiceert dat de scores zich met name aan de linkerkant bevinden. Negatieve waarden suggereren dat de scores zich met name aan de rechterkant begeven van het gemiddelde. Een Kurtosis van nagenoeg nul indiceert een verdeling die relaties vlak is (te veel proefpersonen in de extreme scores).

---

---