

4. De gegevens verkennen met grafieken

Presentatie van gegevens

Grafieken moeten:

- de gegevens weergeven,
- de lezer over de gegevens na laten denken,
- verstorende gegevens vermijden,
- veel nummers met zo min mogelijk inkt laten zien,
- veel gegevens als samenhangend geheel laten zien,
- de lezer aanmoedigen om verschillende gegevens te vergelijken,
- gegevens onthullen.

Overbodige dingen aan een grafiek wordt *chartjunk* genoemd. Bij een grafiek is de verticale as de y-as en de horizontale as de x-as. Creëer bij een grafiek geen verkeerde indrukken, gebruik zo min mogelijk afleidende dingen en zorg voor zo min mogelijk lijnen.

Chart Builder

Grafieken kun je maken in de *chart builder*. Ga naar graphs - chart builder. Op blz. 126 zie je een voorbeeld. Bij de gallery kan je kiezen hoe je grafiek eruit moet zien, in de variables list kan je de variabelen selecteren, de canvas laat zien hoe de grafiek eruit komt te zien en de geselecteerde variabelen kan je naar de juiste drop zones slepen.

Histogrammen

Als je bij de gallery voor een histogram kiest krijg je de keuze uit vier verschillende histogrammen (zie ook blz. 129).

Simple histogram: Deze optie is voor als je de frequenties van een simpele variabele wilt zien.

Stacked histogram: Hier zie je verschillende groepen in een histogram. Bijvoorbeeld man en vrouw.

Frequency histogram: Dit is hetzelfde als een simpel histogram alleen is het hier in een lijn te zien.

population pyramid: Hier zijn twee grafieken met de bodems tegen elkaar gezet, dus hier is de frequentie in de horizontale as te zien.

In een histogram en boxplot zie je snel wat de *uitbijters* of *uitschieters* zijn. Dit zijn scores die ver buiten alle andere scores vallen.

Boxplots

Een boxplot is een andere goede manier om de verdeling van de gegevens te bekijken. In het midden van de boxplot zit de mediaan. Voor een voorbeeld zie blz. 134. Er zijn drie verschillende boxplots waaruit je kan kiezen.

Simple boxplot: Deze optie is voor het gebruik van simpele variabelen. Je krijgt de categorieën in aparte boxplotjes naast elkaar te zien.

Clustered boxplot: Dit is hetzelfde als de eerste optie alleen kan je hier de gegevens in tweeën splitsen. De boxplots zijn dan in twee verschillende kleuren te zien.

1-D boxplot: Deze optie is voor het gebruik van één variabele zonder dat er een categorische variabele geselecteerd moet worden voor de x-as.

Een ster (*) in het plaatje is een uitbijter. De box zelf is de middelste 50%. Verder worden ook de laagste en hoogste scores weergegeven. Wanneer de armen vanaf de mediaan even lang zijn is de verdeling symmetrisch.

Uitbijters kunnen ook via SPSS met z-scores gevonden worden.

Grafische gemiddelden

De meest gebruikelijke manier om de gemiddelden weer te geven is met een *staafdiagram*. We gaan weer naar chart builder -> gallery -> bar. Dan komen er acht opties tevoorschijn:

- **Simple bar:** Deze optie gebruik je als je de gemiddelden van de verschillende groepen wilt zien.
- **Clustered bar:** Als je een tweede groepsvariabele eraan toe wilt voegen is deze optie handig. Hier worden twee staafdiagrammen in één plaatje weergegeven met twee verschillende kleuren.
- **Stacked bar:** Dit is ongeveer hetzelfde als de clustered bar alleen zijn hier de kleuren op elkaar gezet in plaats van aparte staafjes.
- **Simple 3-D bar:** Dit staafdiagram lijkt ook op de clustered bar alleen wordt hier de tweede groepsvariabele op een extra (derde) as gezet. En alles is één kleur.
- **Clustered 3-D bar:** Dit is de clustered bar alleen met drie assen. Het is heel moeilijk om hierin een gemiddelde te vinden en geen aanrader om te gebruiken.
- **Stacked 3-D bar:** Zie stacked bar. Drie assen en de kleuren op elkaar en ook geen aanrader om te gebruiken.
- **Simple error bar:** Hier wordt het gemiddelde met een punt aangegeven en de streep erdoorheen is meestal het betrouwbaarheidsinterval.
- **Clustered error bar:** Dit is hetzelfde als de simple error bar alleen hier weer meer groepsvariabelen en dus twee kleuren.

Onder element properties kan je voor een I-bar kiezen waarbij alleen de lijnen zichtbaar zijn. Als je een staafdiagram wilt zien met daarbij de error bar krijg je het *meetfout staafdiagram*. Je krijgt ze allebei als je het knopje Display error bars aanvinkt.

Wanneer herhaalde metingen bij dezelfde personen wordt gedaan is er geen groepsvariabele en kan die niet op de x-as ingevuld worden.

Om toch een staafdiagram te krijgen selecteer je alle variabelen die je op de y-as wilt zien met de ctrl knop. Hiermee kan je meerdere variabelen tegelijk selecteren. Alle variabelen sleep je dan naar de y-as en dan komt er een nieuwe grafiek uit. SPSS maakt er dan van dat de index (stelt de onafhankelijke variabele voor) op de x-as komt en summary (stelt de groepsvariabele voor) op de y-as. Bij element properties kan je alles juist instellen voor de goede titel enz.

We hebben nu gezien hoe je een plot maakt met gemiddelden die aan elkaar gerelateerd zijn. Maar hoe doe je het als je nog een tweede onafhankelijke variabele hebt? Dit is niet mogelijk via SPSS.

De chart builder kan wel grafieken maken van mixed designs. Als je één variabele als herhaalde meting hebt kan je grafieken maken.

Lijngrafieken

Lijngrafieken zijn hetzelfde als staafdiagrammen alleen wordt hier een lijn gegeven. Voor een staafdiagram kan ook een lijndiagram gebruikt worden. Het startpunt is weer de chart builder. Daarna klik je op gallery - line en dan krijg je twee opties te kiezen:

- Simple line: Deze optie kan je kiezen als je de gemiddelden van de scores van de verschillende groepen wilt zien.
- Multiple line: Hier kan je ook het gemiddelde zien van een bepaalde variabele en ook meerdere lijnen met verschillende kleuren krijgen voor meerdere variabelen.

Scatterplot

Een *scatterplot* is handig als je naar de relatie tussen twee variabelen wilt kijken. Een scatterplot laat de score op de ene variabele en de score op de andere variabele zien. Bij een scatterplot kan je ook zien wat voor soort relatie er is tussen twee variabelen. Ook zijn de uitbijters makkelijk in de grafiek te vinden. Om een scatterplot in SPSS te maken gaan we naar chart builder -> scatter/dot en dan krijg je acht opties te zien.

- Simple scatter: In deze grafiek kan je een continue variabele tegen een andere variabele uitzetten.
- Grouped scatter: Dit is hetzelfde als de simple scatter alleen kan je hier onderscheid maken tussen groepen door middel van verschillende kleuren.
- Simple 3-D scatter: Deze opties kan je gebruiken als je een continue variabele tegen twee andere variabelen uit wilt zetten.
- grouped 3-D scatter: Dit is hetzelfde als de grouped scatter alleen kan je hier nog een variabele toevoegen.
- Summary point plot: Deze plot is hetzelfde als de staafdiagram alleen is het hier een punt in plaats van een staaf.
- Simple dot plot: Dit wordt ook wel de *density plot* genoemd. Het laat de frequenties zien en elke punt is één score. Het is een manier om de verdeling van de scores te zien.
- Scatterplot matrix: Deze laat meerdere plotjes zien tussen paren van variabelen.

- Drop-line: Een punt laat hier het gemiddelde zien en de lijn maakt de verbinding tussen de gemiddelden van de groepen. Dit is handig voor het vergelijken van de statistieken.

Bij het plotten van een scatterplot is het handig om de *regressielijn* erin zichtbaar te maken. Dit is een lijn die de relatie tussen de twee variabelen samenvat. Je krijgt de lijn door de dubbelklikken op de grafiek en dan bij properties voor de optie lineair te kiezen. (Als je een lineaire lijn wilt).

Voor een grouped scatterplot wordt er nog een derde categorische variabele aan het geheel toegevoegd. En kan je ook meerdere regressielijnen zichtbaar maken.

Het kan soms onoverzichtelijk zijn als je een 3-D scatterplot maakt. Met de optie rotation kan je de grafiek draaien om een duidelijker beeld te krijgen.

Een matrix scatterplot laat alle mogelijke combinaties zien van alle variabelen die je invoert. Selecteer eerst alle variabelen die je vergeleken wilt hebben en sleep ze dan naar scattermatrix.

De opmaak van grafieken

Dubbelklikken op de grafiek zorgt ervoor dat de *chart editor* wordt geopend. Hier kan je de grafiek opmaken. Voor uitleg van de iconen zie blz. 159 en 160 van het boek.