9. Vergelijken van twee gemiddelden

Verschillen bekijken

De simpelste vorm van experimenten zijn experimenten waarbij er slechts één onafhankelijke variabele is die op twee manieren gemanipuleerd wordt, en waarbij er ook maar één uitkomst gemeten wordt. Vaak bestaat de onafhankelijke variabele uit een experimentele conditie en een controle conditie. Je hebt hierbij twee groepen. Dit hoofdstuk gaat over onderzoek waarbij twee groepen, met twee gemiddelden worden vergeleken.

Er zijn twee manieren waarop je zo'n onderzoek kunt uitvoeren. Je kunt twee groepen participanten hebben die je blootstelt aan verschillende manipulaties (onafhankelijk of tussen-groep design), of je hebt dezelfde groep participanten die de verschillende manipulaties op verschillende tijdstippen krijgen (herhaalde metingen design).

Als je de verschillen tussen twee groepsgemiddelden wil vergelijken, is dat hetzelfde als een uitkomst voorspellen op basis van het lidmaatschap van die twee groepen. Dit is dus een regressie met één dichotome predictor. Ook het vergelijken van gemiddelden doe je op basis van een lineair model. Je kunt dus ook het standaard model hierbij gebruiken:

Uitkomst = (b_0 +b_1X)+ error.

Bij twee groepen heb je echter een nominale variabele. Deze moet je omzetten in de cijfers 0 en 1. 0 staat daarbij voor de 'baseline', dus de controleconditie. De controleconditie krijgt dus het cijfer 0, de experimentele groep krijgt het cijfer 1. De beste gok voor de uitkomst van iemand in een bepaalde conditie, is het gemiddelde in die groep. In de controleconditie wordt de formule dan (als je de errorterm weglaat):

$$\overline{X}_{\text{controle}} = b_0 + (b_1 \text{ x } 0)$$

$$\overline{X}_{\text{controle}} = b_0$$

Voor de experimentele groep wordt de formule:

$$\overline{X}_{\text{experimenteel}} = b_0 + (b_1 \text{ x } 1)$$

$$\overline{X}_{\text{experimenteel}} = b_0 + b_1$$

Uit de formule voor de controle groep bleek dat het gemiddelde daarvan gelijk is aan b_0, dus:

$$\overline{X}_{\text{experimenteel}} = \overline{X}_{\text{controle}} + b_1$$

$$b_1 = \overline{X}_{\text{experimenteel}} - \overline{X}_{\text{controle}}$$

b_0 staat in dit model dus voor het gemiddelde van de groep die met nul is gecodeerd en b_1 staat voor het verschil tussen de twee groepsgemiddelden. Een t-toets toetst of dit verschil significant afwijkt van 0, wat dus een significant verschil tussen de twee groepsgemiddelden betekent. Dit bewijst dat ook een t-toets op te vatten is als een lineair model.

T-toets

Hoewel je met de methode hiervoor twee groepsgemiddelden kunt vergelijken, wordt de t-toets toch meestal apart gezien van lineaire modellen. De t-toets kan gebruikt worden om te kijken of twee groepsgemiddelden verschillend van elkaar zijn. Er zijn twee verschillende t-toetsen, namelijk de onafhankelijke t-toets (independent-samples t-test) en de afhankelijke t-toets (paired-samples t-test). De onafhankelijke t-toets wordt gebruikt bij twee verschillende experimentele condities waarbij verschillende proefpersonen bij een conditie worden geplaatst. De afhankelijke t-toets wordt gebruikt bij twee experimentele condities waarbij dezelfde proefpersonen deelnemen in beide condities.

Hoofdgedachtes

Beide t-toetsen hebben een aantal punten gemeen. Bij beide worden gegevens verzameld en de gemiddeldes van de steekproeven vergeleken. We verwachten dat beide gemiddeldes ongeveer gelijk zijn aan elkaar als ze uit dezelfde populatie komen. Ook wordt bij beide t-toetsen het geobserveerde verschil vergeleken met het verwachte verschil. De standaard meetfout wordt gebruikt als maat voor de variabiliteit tussen de steekproefgemiddelden. Wanneer de nulhypothese verworpen wordt, kunnen we zeggen dat er een verschil tussen de gemiddelden is door de experimentele manipulatie. De t-toets kan gezien worden als:

$$t = \frac{\text{Geobserveerd verschil - verwacht verschil in de populatie als } H_0 \text{ waar is}}{\text{Standaardfout}}$$

De onafhankelijke t-toets

Bij een onafhankelijke t-toets is het geobserveerde verschil het ene groepsgemiddelde min het andere. Als de nulhypothese waar is, verwacht je geen verschil in de populatie. Dus de formule voor de onafhankelijke t-toets is:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\text{Standard error}}$$

Om de standaarddeviatie van de steekproefverdeling (=standard error) van de verschillen tussen de steekproefgemiddelden te krijgen, gebruiken we de *wet van de variantiesom*. Deze wet zegt dat de variantie van het verschil tussen twee onafhankelijke variabelen gelijk is aan de som van de varianties. Je kan de standaard deviaties van de steekproeven gebruiken om de standaard meetfout te berekenen.

SE van steekproefverdeling van populatie 1 = $\frac{s_1}{\sqrt{N_1}}$

SE van steekproefverdeling van populatie 2 = $\frac{s_2}{\sqrt{N_2}}$

De variantie is het kwadraat van de standard error.

$$\left(\frac{s_1}{\sqrt{N_1}}\right)^2 = \frac{s_1^2}{N_1}$$

Om de verschillen te vinden geldt volgens de wet van de variantiesom het volgende:

Variantie van de steekproefverdeling van verschillen = $\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}$

Om van de variantie weer naar de standard error te gaan, moeten we de wortel hieruit trekken:

SE van de steekproefverdeling van verschillen = $\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}$

Als we alles met de t-toets verwerken komt de volgende formule eruit:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

Deze formule kan alleen gebruikt worden als de steekproefgroottes van de twee groepen gelijk zijn. Wanneer dat niet zo is, kan de gepoolde variantie (s_p^2) geschat worden. Hierbij wordt de variantie meegenomen relatief aan zijn vrijheidsgraden (n-1):

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Als we dit verwerken in de formule voor de t-toets krijgen we het volgende:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{N_1} + \frac{s_p^2}{N_2}}}$$

De afhankelijke t-toets (paired-sample t-test)

De afhankelijke t-toets vergelijkt de verschillen in groepsgemiddelden van de steekproef () het verschil in populatiegemiddelden (μ_D).

De formule voor deze t-toets is:

$$t = \frac{\bar{D} - \mu_D}{s_D \sqrt{N}}$$

$s_D / \sqrt{N}$ is de standaard meetfout van de verschillen.

Er kan gekeken worden naar het verschil tussen steekproefgemiddelden. Dit verschil is meestal nul of dicht bij de nul omdat de steekproeven vaak hetzelfde gemiddelde hebben als ze uit dezelfde populatie komen. De *standaard meetfout van verschillen* is de standaardafwijking van de steekproefverdeling. Gelijke gemiddelden betekent een kleine standaard meetfout. Dus, als de H_0 waar is, is μ_D nul.

De standaardafwijking van deze steekproefvergeling is de standaard meetfout van de verschillen. Een grote standaardafwijking betekent dat steekproeven erg kunnen afwijken van het populatiegemiddelde, dus dat er veel verschillen tussen paren kunnen zijn op basis van toeval. Er is dan een groter verschil nodig om een significant resultaat te krijgen.

$\bar{D}$ is een schatting van de systematische variantie en laat het experimentele effect zien. De standaard meetfout laat ongeveer zien hoe goed de steekproeven overeenkomen met de populatie en is een meting van de niet-systematische variantie. Als het gemiddelde verschil groot is en de standaard meetfout klein kunnen we zeggen dat het verschil niet door toeval komt. De standaard meetfout voor verschillen is:

$$\sigma_{\bar{D}} = \frac{s_D}{\sqrt{N}}$$

We verwachten dat de systematische variantie groter is dan de niet-systematische variantie als de experimentele condities een effect heeft. t is dan groter dan 1. Wanneer er geen effect is verwachten we dat de variantie veroorzaakt wordt door individuele verschillen groter is dan de variantie veroorzaakt door de experimentele conditie. t is dan niet significant groter dan 1. Je kan de verkregen t-waarde vergelijken met de t-waarde die je zou krijgen door alleen toeval erin te betrekken bij dezelfde vrijheidsgraden. Wanneer de kritieke waarde bereikt wordt, concluderen we dat de onafhankelijke variabele een effect heeft.

Assumpties

De *onafhankelijke t-toets* en de *afhankelijke t-toets* zijn beide parametrische toetsen gebaseerd op de normaalverdeling. In hoofdstuk 5 is te lezen over de bias die ontstaat als deze assumptie geschonden wordt. Bij afhankelijke t-toetsen is het belangrijk om te weten dat de verschillen tussen scores normaal verdeeld moeten zijn, niet de scores zelf.

SPSS

Onafhankelijke t-toets
Voor de onafhankelijke t-toets ga je naar analyze – compare means – independent-samples t test. De onafhankelijke variabele voer je in bij grouping variabele en bij define groups kan de codes van de condities invullen. Ook hier kun je kiezen voor bootstrap als je mogelijk bias in je data hebt.

Bij de onafhankelijke t-toets heeft de output twee tabellen (drie als je kiest voor bootstrap). De eerste tabel laat een samenvatting van de statistieken zien (zie blz. 374). De tweede tabel laat de belangrijkste statistieken zien en is in twee rijen opgebouwd. De eerste rij is voor als je kunt aannemen dat je gelijke varianties hebt en de tweede rij is voor als je niet kunt aannemen dat je gelijke varianties hebt. Welke rij je gebruikt is afhankelijk van Levene's test. Bij een significante Levene's test kunnen we aannemen dat de varianties verschillend zijn en er geen homoscedasticiteit is.

In dat geval kijk je dus in de tweede rij.

De effectgrootte is te berekenen met deze formule:

r = $\sqrt{\frac{t^2}{t^2 + df}}$

Afhankelijke t-toets

Bij de afhankelijke t-toets heb je te maken met de assumptie dat de verschillen tussen scores normaal verdeeld moeten zijn. Bij een grote steekproef hoef je je hier geen zorgen over te maken, maar bij een kleine steekproef moet dat wel gecheckt worden.

Error bar charts problemen

Een probleem is dat SPSS in een error bar chart de data behandeld alsof het van verschillende participanten komt, in plaats van van dezelfde participanten. Hiervoor moet je handmatig corrigeren. Eerst moet je het gemiddelde voor elke participant berekenen. Dit doe je via Transform – Compute variable. Selecteer bij Function Group 'Statistical'. Bij Functions and Special variables kun je dan voor Mean kiezen. Vul variabelen in op de plek van de vraagtekens (zie bladzijde 380).

Vervolgens moet je het overkoepelende gemiddelde berekenen, het gemiddelde voor alle scores, ongeacht uit welke conditie de score komt. Dit is gewoon het gemiddelde van alle individuele gemiddelden. Ga via analyze naar explore, frequencies of descriptives. Sleep de variabele 'Mean' naar het veld Variable(s). Bij options kun je nu Mean aanklikken.

Stap drie is het berekenen van de aanpassingsfactor. Sommige participanten scoren hoger of lager dan anderen, los van de conditie. De error bar chart moet hierop aangepast worden. De aanpassingsfactor bereken je door het overkoepelende gemiddelde min het persoonlijke gemiddelde te doen. Dit voer je door in SPSS via Transform – Compute variable (zie bladzijde 382).

De laatste stap is een aangepaste variabele te maken voor de condities. Dit doe je ook via Transform – Compute variable. Je maakt een nieuwe aangepaste variabele voor een conditie door de oude variabele te nemen plus te aanpassingsfactor. Zie bladzijde 383.

De afhankelijke t-toets kan je vinden bij analyze – compare means – paired-samples t test. Daarna selecteer je de paren die geanalyseerd moeten worden.

In de output komen drie tabellen (vier als je bootstrap erbij doet). Als eerste komt een samenvatting van de statistieken in de twee experimentele condities. Deze bevatten het gemiddelde, de standaarddeviatie, de N en de standaard meetfout. Er wordt ook een tabel met de Pearson correlatie gegeven. De volgende tabel vertelt ons of het verschil in gemiddelden significant is. De laatste tabel is de bootstrap tabel.

De effectgrootte r wordt op dezelfde manier berekend als bij een independent t-toets. Cohen's d kan soms echter de voorkeur hebben, omdat die een effectgrootte heeft die onafhankelijk is van het design. Ongeacht of je dus een onafhankelijke of afhankelijke t-toets hebt, de Cohen's d blijft even groot. Bij de r is dit niet zo.

Tussengroepontwerp of herhaalde metingen?

Bij herhaalde metingen (dezelfde proefpersonen) is de niet-systematische variantie een stuk kleiner dan bij een tussengroepdesign. Dit betekent dat je sneller een significant effect vindt bij een herhaalde metingen design, dan wanneer je verschillende proefpersonen gebruikt.

Schenden van assumpties

Als je de assumpties van de t-toets schendt, is het mogelijk om non-parametrische testen te gebruiken (zie hoofdstuk 6), maar je kunt beter de stappen nemen die in hoofdstuk 5 zijn uitgelegd om uitschieters te corrigeren en een bootstrap betrouwbaarheidsinterval gebruiken.