

6. Niet-parametrische testen

Het gebruik van niet-parametrische testen

In het vorige hoofdstuk werd duidelijk dat er factoren kunnen zijn die een bias in de resultaten van statistische modellen veroorzaken. Soms kun je dit niet verhelpen. Zeker met kleine steekproeven, waarbij de centrale limietstelling niet opgaat, kan dit een probleem zijn. Een *niet-parametrische test* kan gebruikt worden als niet aan de aannames voldaan kan worden. Niet-parametrische testen hebben bijna geen assumpties en het principe is meestal het sorteren van de gegevens. Het sorteren is een volgorde geven aan de getallen (*ranken*). De nieuwe getallen die toegekend zijn aan de scores (de rangscores) worden dan gebruikt voor de analyse. Hierdoor heb je geen last meer van uitschieters en scheefheid.

Het vergelijken van twee onafhankelijke condities

Wanneer het een tussengroepsontwerp is met twee onafhankelijke condities kan je kiezen tussen de *Wilcoxon rank-sum test* en de *Mann-Whitney test*. Beide testen zijn de niet-parametrische versie de onafhankelijke t-toets. Er wordt een nummering aan de scores gegeven, de laagste score krijgt 1, de score daarna krijgt 2 enz. Bij deze rangorde behandel je het als één steekproef, de rangorde is ongeacht tot welke conditie een participant behoort. Dan worden van elke groep rangscores bij elkaar opgeteld. Als er geen verschil is tussen de condities, verwacht je evenveel hoge als lage rankings in beide condities.

De Wilcoxon rank-sum (W_s), de test statistiek van de Wilcoxon rank-sum methode, is simpelweg de kleinste opgetelde rangscore bij gelijke groepsgroottes en bij ongelijke groepsgrootte telt deze is het de opgetelde rangscore van de groep met de minste participanten.

Dezelfde waarden moeten hetzelfde rangnummer toegekend krijgen bij het ranken van de getallen. Dit doe je door het gemiddelde te nemen van de rangnummers die toegekend zouden worden. Om te bepalen of de Wilcoxon rank-sum significant is worden het gemiddelde ($\overline{W_s}$) en de standaard meetfout ($SE_{\overline{W_s}}$) berekend.

$$\text{Cohen's } d = \frac{\overline{X}_1 + \overline{X}_2}{s}$$

$$SE_{\overline{W_s}} = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

n is hier het aantal deelnemers in een conditie. Als het gemiddelde en de standaard meetfout bekend zijn kan de z-score berekend worden.

$$z = \frac{\overline{X} - \overline{X}}{s} = \frac{W_s - \overline{W_s}}{SE_{\overline{W_s}}}$$

De Mann-Whitney test (met U als test statistiek) lijkt erg op voorgaande test en is met de volgende formule te berekenen.

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

R_1 is hier het opgetelde totaal van de rangscores van groep 1. Er is een direct verband tussen deze twee testen, en SPSS geeft beide uitkomsten, dus het maakt niet uit welk van de twee toetsen je kiest.

De gegevens invoeren

In SPSS moet de onafhankelijke variabele gecodeerd worden. Eerst worden analyses uitgevoerd om te kijken of de gegevens normaal verdeeld zijn en er wordt gekeken naar de homogeniteit. Als de verdeling niet normaal is kan gekozen worden voor een niet-parametrische test.

Voor de analyse in SPSS ga je naar analyse – Nonparametric Tests – Independent Samples. Bij het tabblad ‘fields’ kun je de afhankelijke variabelen selecteren (zie blz. 224). De onafhankelijke variabele komt bij ‘groups’. Bij de settings kun je de toets selecteren. Je kunt SPSS de toets laten kiezen, of klikken op ‘customize tests’, waarbij je de toets zelf kunt kiezen.

De Mann-Whitney test kan op twee manieren gedaan worden. De een geeft een schatting van de p-waarde, die prima is voor grote steekproeven. Bij kleine steekproeven ($n < 50$) is het echter niet nauwkeurig genoeg, en moet de exacte methode worden gebruikt.

Bij opties kunnen de descriptives nog aangevinkt worden. Andere testen die geselecteerd kunnen worden zijn: Kolmogorov-Smirnov Z (deze test kijkt of groepen uit dezelfde populatie komen), Moses Extreme Reactions (het vergelijkt de variabiliteit van de scores in de twee groepen), Wald-Wolfowitz runs (deze test is een variant op de Mann-Whitney test).

Output

Het eerste deel van de output geeft een samenvatting van de ranking die gegeven is aan de scores. De Mann-Whitney test heeft de volgorde gegeven van laag naar hoog. Dus de groep met het laagste gemiddelde is de groep met het grootste aantal lage scores. En de groep met het hoogste gemiddelde is de groep die meer hoge scores heeft. Het tweede deel van de output laat de test statistieken zien samen met de z-score. Wanneer geen voorspelling is gemaakt welke groep beter is, kan naar de significantie gekeken worden. Dit omdat de significantie ervan uit gaat dat er twee staarten in de grafiek zitten. Als de test significant is, concludeer je dat er een verschil is tussen de groepsgemiddelden.

De effectgrootte

Met de z-score kan ook de effectgrootte berekend worden. Het is handig om de effectgrootte te berekenen omdat het een standaardmaat is. Door het te berekenen kan het vergeleken worden met andere studies.

$$r = \frac{Z}{\sqrt{N}}$$

N is het totale aantal observaties.

Voor het rapporteren van de Mann-Whitney test moet je de test statistiek (U) rapporteren met de exacte significantie. Ook is het handig de effectgrootte, de z-score, de mediaan en de bijbehorende ranges te vermelden. De mediaan is geschikter dan het gemiddelde voor niet-parametrische testen.

Het vergelijken van twee gerelateerde condities

De *Wilcoxon signed-rank test* (de test statistiek is T) wordt gebruikt wanneer twee condities met dezelfde proefpersonen vergeleken moeten worden. Deze test kan vergeleken worden met de afhankelijke (paired-samples) t-toets en is net als de t-toets gebaseerd op het vergelijken van de scores tussen de twee condities.

De eerste stap is het berekenen van het verschil, met erbij vermeld of het verschil negatief of positief is. Als het verschil nul is, dus in de twee condities van de repeated measures wordt dezelfde waarde gemeten, dan worden deze scores uitgesloten van de ranking. De verschillscore is nu de ranking. Daarna worden alle positieve getallen bij elkaar opgeteld en alle negatieve getallen bij elkaar opgeteld. Voor het berekenen van het gemiddelde van de test statistiek is de volgende formule:

$$\bar{T} = \frac{n(n+1)}{4}$$

n is hierbij de steekproefgrootte, en omdat het een afhankelijke steekproef is, is er maar 1 sample.

$$SE\bar{T} = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Wanneer het verschil voor een persoon 0 is wordt deze niet meegenomen in de berekening. De groepsgrootte kan dus anders zijn dan in andere analyses. Ook hier kan de z-score berekend worden:

$$z = \frac{T - T_{\text{gem}}}{SE_{T_{\text{gem}}}}$$

De analyse in SPSS

Voor de Wilcoxon signed-rank test in SPSS ga je naar analyse – Nonparametric Tests – Related Samples. Selecteer de twee afhankelijke variabelen en sleep ze naar ‘test fields’, dan vormen ze aan paar dat vergeleken wordt. Bij customize tests kun je kiezen voor de Wilcoxon signed-rank test.

Andere testen die hier geselecteerd kunnen worden zijn: Sign (deze test is ongeveer hetzelfde als de Wilcoxon signed-rank test alleen wordt hier naar de richting gekeken, waardoor hij weinig power heeft in samples groter dan 6), McNemar (deze test is bruikbaar wanneer je nominale in plaats van ordinale gegevens hebt), Marginal Homogeneity (vergelijkbaar met Wilcoxon signed-rank test).

Output

Eerst wordt een samenvatting gegeven van de positieve en negatieve rankingen en in de volgende tabel wordt de significantie weergegeven (in de rij Asymptotic Sig. (2-tailed)). Als je dubbel klikt op deze tabel, zie je een histogram met de veranderingscores. De kleinste z-score wordt gebruikt, dit kan dus van de positieve of negatieve getallen zijn.

De effectgrootte

Bij twee condities van dezelfde personen kan de effectgrootte op dezelfde manier berekend worden als bij twee onafhankelijke condities, zoals de Mann-Whitney test. Ook het noteren van de resultaten is hetzelfde.

Verschillen tussen meerdere onafhankelijke groepen

De niet-parametrische test die vergelijkbaar is met de eenweg onafhankelijke ANOVA, is de *Kruskal-Wallis test* (H). Wanneer niet aan de aannames voldaan kan worden is het handig om deze test te gebruiken. De Kruskal-Wallis test lijkt op de Mann-Whitney test. Ook hier worden de getallen geordend en rangnummers gegeven van laag naar hoog. Daarna worden de toegekende rangscores van alle groepen bij elkaar opgeteld (R_i). De Kruskal-Wallis test kan als volgt berekend worden.

$$H = (12/N(N-1)) (\sum R_i^2 / n_i) - 3(N+1)$$

N is het totale aantal deelnemers en n is het aantal deelnemers in de groep. Het aantal vrijheidsgraden is hier k (het aantal groepen) - 1. Deze test statistiek heeft een chi-square distributie.

Als de Kruskal-Wallis test significant blijkt te zijn, weet je dat je groepen uit verschillende populaties komen. Je weet alleen niet precies welke groepen van elkaar verschillen. Je kunt dit oplossen door *paarsgewijze vergelijkingen*, waarbij je elke groep los met elkaar vergelijkt. Omdat dit meerdere toetsen zijn, kun je hierbij een familywise error rate krijgen, waarbij de kans op een type I fout erg toeneemt. Een correctie hiervoor zorgt echter voor vermindering van de power.

Een alternatief voor paarsgewijze vergelijkingen is een 'stepped procedure'. Hierbij kijk je naar de som van de rangscores in de groepen, en zet je die op volgorde van laag naar hoog.

Vervolgens vergelijk je de eerste twee groepen in deze rangorde met elkaar. Als er geen significant verschil is, voeg je de derde groep toe, en kijk je of die drie groepen allemaal hetzelfde zijn. Als je een significant verschil vindt, voeg je de groep die je het laatst hebt toegevoegd bij de volgende stap.

In stap 2 vergelijk je de significant verschillende groep uit de vorige stap met de overgebleven groep. Als dit verschil niet significant is, komen ze samen in een subgroep. Als het verschil wel significant is, heb je dus verschillende subgroepen.

De gegevens invoeren in SPSS

Voor deze analyse wordt eerst een gecodeerde variabele gemaakt van de onafhankelijke variabele (de groepen). Als de gegevens ingevoerd zijn kunnen ze getest worden op normaliteit en homogeniteit. Om naar het verschil tussen groepen te kijken moeten deze assumpties per groep getest worden.

Voor de Kruskal-Wallis test ga je naar analyse – Nonparametric Tests – K independent Samples. De afhankelijke variabele gaat in het vak Test Fields (Test Variable List in oudere versies van SPSS). De onafhankelijke variabele gaat bij Group (Grouping Variable). Selecteer customize tests, Kruskal-Wallis 1-way ANOVA (k samples). Hiernaast staat de knop Multiple comparisons, waarbij je paarsgewijze (all pairwise) of stepped (stepwise step-down) vergelijkingen kunt kiezen (zie bladzijde 242). Wanneer je naar de lineaire trend wil kijken, kan je de Jonckheere-Terpstra aanvinken.

Output

De eerste tabel geeft een samenvatting van de geordende gegevens die nodig zijn voor het interpreteren van de effecten. De volgende tabel laat de statistieken zien. Chi-square is de statistiek van de Kruskal-Wallis test. Als de asymp. sig. kleiner is dan .05 is er een effect.

De significantie laat zien dat er een verschil is. Het laat niet zien waar het verschil zit. Een manier om dat wel te bekijken is met boxplots. Dit is echter een subjectief oordeel, je kunt niet zien of de boxplots significant verschillen. Daarom doe je een follow-up analyse.

Follow-up analyse

Zoals al eerder gezegd, zijn er twee manieren voor een follow-up analyse. In SPSS heb je een van beide geselecteerd (all pairwise of stepwise step-down). De output van deze tests komt niet gelijk in je viewer. Je kunt dit zichtbaar maken via View. Je klikt dan op pairwise comparisons of op homogeneous subsets (voor de stepped manier).

Op bladzijde 245 zie je de beide outputs. Bij de pairwise comparisons moet je kijken bij de Adj. Sig., dit is de p-waarde die gecorrigeerd is voor het feit dat er meerdere tests uitgevoerd worden, om een familywise error rate te voorkomen.

Bij de homogeneous subsets worden subsets gemaakt van groepen die niet significant van elkaar verschillen. De subsets zelf verschillen wel significant van elkaar. De groepen die bij elkaar in een subset horen, worden met dezelfde kleur aangegeven.

Kijken naar trends

De Jonckheere-Terpstra test kijkt naar het patroon van de medianen van de groepen die je met elkaar wilt vergelijken. Het doet hetzelfde als de Kruskal-Wallis test, maar het geeft daarnaast ook informatie over of de volgorde van de groepen van belang is. In het voorbeeld over of soja slecht is voor het aantal zaadcellen in sperma, is er de verwachting dat hoe meer soja een persoon eet, hoe minder zaadcellen hij heeft. Hiervoor kun je dus de Jonckheere-Terpstra test gebruiken. Je kunt in SPSS kiezen of je van de kleinste naar de grootste groep wil vergelijken of andersom.

Net als met de Kruskal-Wallis test, komt er hier ook alleen een samenvattende tabel in de output. Door hierop te dubbelklikken, verschijnen meer resultaten. De plus of min van de z-score vertelt je of de medianen een toenemende of afnemende trend hebben. Een positieve z-score betekent dat de mediaan groter wordt bij hogere nummers van de coding variable, een negatieve z-score betekent dus een kleiner wordende mediaan.

De effectgrootte

De algemene effectgrootte is niet van belang omdat je dan het algemene effect berekend. De effectgrootte per vergelijking kan beter berekend worden. Dit kun je doen met dezelfde formule als eerder is gezegd, de z-score delen door de wortel van de steekproefgrootte. De steekproefgrootte is dan het aantal participanten in de vergeleken groepen. Een paarsgewijze vergelijking tussen twee groepen met 20 participanten heeft dus $N=40$. Het rapporteren van de Kruskal-Wallis test gaat hetzelfde als voorgaande testen.

Verschillen tussen meerdere gerelateerde groepen

Friedman's ANOVA (F_r) kan gebruikt worden wanneer dezelfde proefpersonen in meer dan twee verschillende condities gebruikt zijn. Deze test wordt gebruikt wanneer niet aan de assumpties voor eenweg herhaalde metingen ANOVA voldaan kan worden. Friedman's ANOVA is net als de andere testen gebaseerd op het ranken.

Elke repeated measures conditie krijgt een kolom. Geef dan een rangvolgorde aan bij elk persoon. Dus geef bij elk persoon de laagste waarde over de repeated measures een 1, de op een na laagste een 2, enzovoort. In het voorbeeld van gewicht krijgt het tijdstip waarop een bepaalde vrouw het laagste gewicht heeft een 1, het hoogste gewicht van die vrouw krijgt een 3, en het gewicht ertussen een 2. Tel dan de condities bij elkaar op om de Friedman's ANOVA test statistiek F_r te berekenen.

$$F_r = [(12/Nk(k+1))\sum R_i^2] - 3N(k+1)$$

Het aantal vrijheidsgraden bij deze verdeling is $k-1$. Net als de Kruskal-Wallis test, heeft deze test een chi-square distributie.

De gegevens invoeren

Als de gegevens in de data editor staan wordt eerst gekeken naar de normaliteit en homogeniteit. Voor de analyse in SPSS ga je naar analyse – Nonparametric Tests – K Related Samples. Sleep de afhankelijke variabelen in Test Fields (Test Variables bij oude versies van SPSS). Bij customize tests kies je voor Friedman's 2-way ANOVA by ranks (k samples). Net als bij de Kruskal-Wallis test, kun je naast deze knop de optie Multiple comparisons kiezen, all pairwise of stepwise step-down, als een follow-up analyse.

Output

In SPSS is de F_1 aangegeven met Chi-square. Verder is de output hetzelfde ingedeeld als bij de voorgaande test, er is een samenvattende tabel, maar meer output kan verkregen worden door dubbelklik op de tabel.

De effectgrootte

Ook hier is het moeilijk om van de chi-square statistiek de effectgrootte te berekenen. Het is beter de effectgroottes van de vergelijkingen te berekenen. De formule is weer hetzelfde. Het enige dat belangrijk is hierbij is dat N staat voor het aantal observaties in de vergelijking en niet voor het aantal proefpersonen. Je hebt immers een herhaalde metingen design, dus meerdere observaties komen van dezelfde proefpersonen. Als je het gewicht van de 10 vrouwen over twee maanden gaat vergelijken, heb je dus $N=20$.

Het rapporteren van de gegevens is hetzelfde als voorgaande testen. Vermeld de chi-square statistiek, het aantal vrijheidsgraden en de significantie. Rapporteer een effectgrote van de follow-up tests en de medianen en de reikwijdte daarvan (of een boxplot).