Onderzoekspracticum 2, aantekeningen college 9

www.joho.org

Schending van normaliteit

Bij veel toetsen, zoals de t-toets, wordt de aanname gedaan dat er sprake is van een normaalverdeling in de populatie. Vaak is dit echter niet het geval. Er zijn verschillende methoden om te gebruiken wanneer de aanname van normaliteit geschonden wordt:

- Uitbijters moeten verwijderd worden wanneer ze niet tot de populatie behoren. Wanneer uitbijters wel tot de populatie behoren, moeten statistische technieken gebruikt worden die geen aanname van normaliteit doen. Dit worden non-parametrische technieken genoemd.
- Wanneer data scheef verdeeld zijn kunnen ze getransformeerd worden. Een voorbeeld hiervan is het gebruik van de logaritmen van de data.
- Soms kunnen verdelingen beter worden beschreven door andere standaardverdelingen. Een voorbeeld hiervan is de Weibullverdeling.
- Bootstrapmethoden en permutatietoetsen kunnen gebruikt worden, omdat deze geen normaliteit of een andere verdelingsvorm vereisen.
- Ook andere non-parametrische methoden kunnen gebruikt worden. Het verschil met bootstrapmethoden en permutatietoetsen is dat deze andere methoden geen gebruik maken van de werkelijke waarden van observaties.

Rangsomtoets van Wilcoxon

De rangsomtoets van Wilcoxon is de non-parametrische versie van de t-toets voor twee steekproeven, wanneer normaliteit geschonden is kan de Wilcoxon uitgevoerd worden (zie college 2 en 3). De aannamen zijn hetzelfde als voor deze t-toets, met uitzondering van de aanname van normaliteit. De rangsomtoets maakt gebruik van rangtransformatie. Dit houdt in dat alle waarnemingen gerangschikt worden van laag naar hoog. Vervolgens wordt aan elke waarneming een rangnummer toegekend. De laagste waarneming krijgt nummer 1, de daaropvolgende laagste waarneming krijgt nummer 2 enzovoorts. Met deze nummers wordt de toets uitgevoerd.

Bij de rangsomtoets wordt ervan uitgegaan dat er twee onafhankelijke steekproeven (simple random samples) zijn getrokken van grootte n_1 en n_2. Het totaal aantal observaties is dus $n_1 + n_2$, oftewel N. De som van de rangnummers van de eerste steekproef is W, de Mann-Witney rangsom statistiek. Welke van de twee steekproeven je hiervoor gebruikt maakt niet uit; het gaat erom dat één groep als uitgangspunt wordt gekozen. Wanneer beide populaties een gelijke continue verdeling hebben, dan geldt het volgende voor het gemiddelde van W:

$$m_W = \frac{n_1(N+1)}{2}$$

De standaarddeviatie van W is dan als volgt:

$$s_W = \sqrt{\frac{n_1 n_2 (N+1)}{12}}$$

De steekproevenverdeling van W kan worden benaderd met behulp van een standaardnormaalverdeling:

$$z = \frac{W - m_W}{s_W} = \frac{W - n_1(N+1)/2}{\sqrt{n_1 n_2 (N+1)/12}}$$

Aan de hand van deze z-waarde kan de p-waarde worden vastgesteld (Tabel A; *Introduction to the Practice of Statistics* van Moore, McCabe en Craig). De p-waarde wordt vervolgens vergeleken met de vastgestelde alpha α en zo kan er worden bepaald of er een significant effect is. In SPSS kunnen de z-score en de p-waarde afwijken, omdat SPSS gebruikmaakt van de Mann-Witney benadering in plaats van de rangsombenadering.

Eventueel kan er een continuïteitscorrectie toegepast worden op de rangsomtoets. Een continuïteitscorrectie voer je uit omdat de verdeling van W discreet is en niet continu. In praktijk is niets continu gemeten, dus zou je eigenlijk altijd een continuïteitscorrectie moeten toepassen. In werkelijkheid wordt dit vaak echter alleen gedaan als er weinig schaalpunten zijn. De continuïteitscorrectie uitvoeren doe je door 0.5 van W af te halen of er juist 0.5 bij op te tellen. Als W bijvoorbeeld 16 is, dan geldt het volgende:

- Als je wilt weten hoe groot de kans is op een W van 16 of *hoger*, oftewel P $(W \geq 16)$, dan haal je 0.5 van 16 af. Je bepaalt dan dus P $(W \geq 15.5)$. Dit gebruik je wanneer W groter is dan het gemiddelde van W (μ_W).
- Als je wilt weten hoe groot de kans is op een W van 16 of *lager*, oftewel P $(W \leq 16)$, dan tel je 0.5 bij 16 op. Je bepaalt dan dus P $(W \leq 16.5)$. Dit gebruik je wanneer W kleiner is dan het gemiddelde van W (μ_W).
- Bij beide manieren toets je dus eenzijdig, maar tweezijdig toetsen gaat heel gemakkelijk door éérst de continuïteitscorrectie toe te passen en vervolgens de p-waarde te vermenigvuldigen met 2.

Idealiter zou je eigenlijk altijd een continuïteitscorrectie moeten uitvoeren want in de praktijk is niets continu gemeten. In de praktijk wordt het alleen gedaan wanneer er weini g schaalpunten zijn, anders is het effect van de continuïteitscorrectie niet effectief (bv. IQ heeft geen zin).

Hypothesen bij de rangsomtoets van Wilcoxon
Omdat we bij de rangsomtoets geen gemiddelden toetsen maar gebruikmaken van rangnummers, toetsen we eigenlijk de gelijkheid van medianen. De nulhypothese en alternatieve hypothese moeten dan ook op de volgende manier geformuleerd worden:

H_0: $\text{mediaan}_1 = \text{mediaan}_2$
H_a: $\text{mediaan}_1 \neq \text{mediaan}_2$

De hypothesen mogen echter alleen op deze manier geformuleerd worden als beide verdelingen een identieke vorm hebben. Als dit niet het geval is moet er gebruik gemaakt worden van een volgende soort formulering:

H_0: De twee verdeling en zijn gelijk.
H_a: Eén verdeling heeft waarden die systematisch hoger zijn.

De alternatieve hypothese is tweezijdig. Er wordt weliswaar een richting aangegeven, maar er wordt niet vermeld *welke* van de twee verdelingen systematisch hogere waarden heeft.

Knopen (ties)
Het kan zo zijn dat bepaalde waarden die je hebt gevonden meerdere keren voorkomen, bijvoorbeeld als drie proefpersonen hetzelfde cijfer hebben gehaald voor een toets. In zo'n geval spreken we van knopen of ties. Het is dan onduidelijk welk rangnummer deze waarden krijgen. De oplossing hiervoor is om deze waarden het gemiddelde rangnummer te geven van de rangen die ze bezetten.

Als het cijfer 6 bijvoorbeeld drie keer voorkomt en rang 8, 9 en 10 bezet, dan krijgt elk cijfer 6 het rangnummer (8+9+10)/3 = 9. Omdat de exacte verdeling van de rangsom verandert bij knopen moet de standaarddeviatie van W aangepast worden. Statistische software doet hiervoor de nodige aanpassingen. Verder is het belangrijk dat er bij knopen nooit een continuïteitscorrectie gedaan wordt.

Toetsinsschema rangsomtoets

Begin zoals bij (samengestelde) t-toets voor twee steekproeven (je wilt in principe gemiddelden toetsen op gelijkheid).

Stap 2 (aannamen): Zelfde aannamen afgaan als bij (samengestelde) t-toets. Bij ernstige schending normaliteit:

- formuleer hypothesen niet in gemiddelden bij stap 3.
- kies rangsomtoets bij stap 4.

Stap 3 (hypothesen): Vanwege schending van aannamen toetsing van gemiddelden niet mogelijk, dus formuleer H0 en Ha in medianen of in woorden!

Stap 4 (toetskeuze):

- Rangsomtoets van Wilcoxon.
- Kies α; zeg α = 0.05.

Stap 8 (effectgrote) $r = \frac{|z|}{\sqrt{N}}$

0.10 < r < 0.20	klein effect
0.20 < r < 0.50	medium effect
r > 0.50	groot effect

De Wilcoxon rangtekentoets

De rangtekentoets is de non-parametrische versie van de gepaarde t-toets, dus deze wordt gebruikt wanneer er geen normaliteit is en wanneer er tegelijkertijd sprake is van gematchte paren of herhaalde metingen (zie college 2 en 3). De rangtekentoets maakt gebruik van rangnummers van de absolute verschillen binnen de paren. Het absolute verschil wordt aangegeven met $|d|$. Een voorbeeld hiervan is het verschil tussen voor- en nameting. Bij de rangtekentoets wordt uitgegaan van W^+, oftewel de som van alle rangnummers die bij een positief verschil horen. Het maakt hierbij niet uit of je de verschillen berekent van voormeting min nameting of andersom. Als de verdelingen van responsen of scores niet worden veroorzaakt door verschillen in behandeling binnen paren, dan heeft W^+ het volgende gemiddelde:

$$m_{W^+} = \frac{n(n+1)}{4}$$

In de formule is n het aantal paren. De standaarddeviatie van W^+ is als volgt:

$$s_{W^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

De steekproevenverdeling van W^+ kan worden benaderd met behulp van een standaardnormaalverdeling:

$$z = \frac{W^+ - m_{W^+}}{s_{W^+}}$$

Voor de hypothesen geldt hetzelfde als voor de rangsomtoets van Wilcoxon: hypothesen worden niet in gemiddelden geformuleerd, maar in medianen of in woorden. Verder kun je evenals bij de rangsomtoets een continuïteitscorrectie toepassen, maar alleen als er geen knopen in je data zitten. Bij afhankelijke steekproeven kunnen er twee soorten knopen zijn:

- Knopen tussen absolute verschillen. Twee proefpersonen komen bijvoorbeeld uit op hetzelfde verschil tussen de voor- en nameting. In dat geval wordt het gemiddelde genomen van de rangnummers.
- Knopen binnen een paar. Een proefpersoon scoort bijvoorbeeld op de voor- en nameting even hoog. Het absolute verschil is dan nul. Paren waarbij dit het geval is moet je buiten beschouwing laten. Dit betekent echter wel dat de test vertekend raakt richting de alternatieve hypothese, waardoor de standaarddeviatie van W^+ aangepast moet worden. Statistische software doet hiervoor de nodige aanpassingen.

Toetsingsschema rangtekentoets
Je begint alsof het een gepaarde t-toets is, je wilt gemiddelden op gelijkheid toetsen. Je formuleert een onderzoeksvraag in gemiddelden.

Bij stap 2 (aannamen) : zelfde aannamen afgaan als bij gepaarde t-toets. Bij ernstige schending van normaliteit : formuleer de hypothesen niet in gemiddelden bij stap drie.

Stap 3 (hypothesen) : vanwege schending van aannamen toetsing van het gemiddelde niet mogelijk, dus formuleer H0 en Ha in medianen of in woorden.

Stap 4 (toetskeuze) : Rangstekentoets van Wilcoxon, kies α; zeg α = 0.05

Bij stap 8, de effectgrote hebben we dezelfde formule als bij de rangsomtoets. N is nu de som van de voormeting + nameting. $r = \frac{|z|}{\sqrt{N}}$

De Kruskall-Wallis toets
De Kruskall-Wallis toets is de non-parametrische versie voor variantieanalyse (ANOVA). Bij de Kruskall-Wallis toets worden alle data gerangschikt en op de rangnummers wordt vervolgens enkelvoudige variantieanalyse toegepast. De SSG van de rangnummers, oftewel de kwadratensom tussen groepen, is de statistische maat van Kruskall-Wallis. Deze wordt H genoemd.

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

In deze formule staat n_i voor de grootte van de i'de steekproef. Het totaal aantal observaties is N. R_i is de som van de rangnummers van de i'de steekproef. Wanneer alle populaties, aangeduid met I, dezelfde continue verdeling hebben, dan heeft H bij benadering een chi-kwadraat verdeling. Deze verdeling heeft I – 1 vrijheidsgraden. De nulhypothese is dat alle populaties dezelfde verdeling hebben. Als de waarde van H groot is, dan wordt de nulhypothese verworpen. De gevonden waarde van H kan worden opgezocht in Tabel F (*Introduction to the Practice of Statistics* van Moore, McCabe en Craig) en vervolgens kan de bijbehorende p-waarde worden bepaald. Omdat het gaat om een chi-kwadraat verdeling, hoef je de p-waarde niet meer met 2 te vermenigvuldigen. De toets is automatisch tweezijdig. Ook voer je geen continuïteitscorrectie uit.

Belangrijk om bij het gebruik van SPSS in de gaten te houden, is het feit dat SPSS niet de waarden geeft van R_1, R_2,…, R_i, maar van R_1/n_1, R_2/n_2,…, R_i/n_i. Om de waarden van R_i te verkrijgen moeten de waarden die SPSS geeft dus nog worden vermenigvuldigd met de steekproefgrootte n_i.