

Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

Samenvatting

In een profielwerkstuk worden leerlingen geacht verslag te doen van bijvoorbeeld een uitgewerkte scheikundige, biologische of natuurkundige proef. Vaak is de onderzoeksvraag op een dusdanige wijze geformuleerd, dat de leerling geïnteresseerd is in een systematisch verschil tussen groepen waarnemingen bijvoorbeeld door het toepassen van twee behandelingen, of in een samenhang tussen twee grootheden. Voor toetsen van zulke verschillen of samenhangen heeft Wageningen Universiteit drie lesbrieven ontworpen. Dit is de tweede van deze drie lesbrieven.

In deze lesbrief ga ik in op een situatie waarbij twee groepen aan verschillende 'omstandigheden' worden blootgesteld. De vraag is dan of het verschil in omstandigheden een verschillend effect heeft op een gemeten eigenschap. Het aantal waarnemingen in de twee groepen die vergeleken gaan worden hoeft voor het kunnen toepassen van deze toets niet gelijk te zijn. Eerst wordt aan de hand van voorbeelden een leidraad gegeven waarmee de leerling kan bepalen in welke situaties de Twee steekproeventoets van Wilcoxon kan worden gebruikt. Daarna volgt een strategie, waarin stapsgewijs duidelijk wordt gemaakt hoe uit de waarnemingen conclusies kunnen worden getrokken met betrekking tot het verschil tussen de twee groepen.

Inleiding

Voor een project dat uit moet monden in een profielwerkstuk is het raadzaam om de volgende vijf fasen te doorlopen:

1. Probleemstelling
2. Planning
3. Verkenning
4. Uitvoering
5. Conclusie.

Voor veel mensen komt de statistiek pas om de hoek kijken als een proef al is uitgevoerd en de waarnemingen klaar liggen om verwerkt te worden tot een conclusie. Dit uitgangspunt is de meest voorkomende beginnersfout. Een conclusie kan slechts op een statistisch verantwoorde manier op basis van de waarnemingen worden getrokken als al vanaf het begin van het project duidelijk is met welke statistische methode men de uiteindelijke waarnemingen gaat verwerken. Tevens is het dan van groot belang om de waarnemingen op een dusdanige manier te verzamelen dat de beoogde toets ook kan worden gebruikt.

In verband met onzekerheid in de waarnemingen is het niet verstandig om te volstaan met één waarneming. Zo'n onzekerheid in de waarnemingen wordt ook wel *stochasticiteit* genoemd. Wanneer een waarneming meerdere malen herhaald wordt, krijgt degene die



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

de experimenten uitvoert een steeds beter beeld van de onzekerheid in de waarnemingen. Als aan één kenmerk waarnemingen worden gedaan dan is het van belang om er op te letten dat al het overige gelijk blijft: het zogenaamde '*ceteris paribus*' principe. Omdat het meestal onmogelijk is om al het overige exact gelijk te houden, is het raadzaam om te loten wie in welke (behandel)groep terecht komt. Dit voorkomt een mogelijke verstrengeling van niet constante factoren. Bij moeilijke probleemstellingen worden om de nauwkeurigheid van de waarnemingen te verhogen soms bepaalde waarnemingen in groepen met een verschillende combinatie van behandelingen in gedeeld.

Een voorbeeld van zo'n proefopzet is het opkweken van planten. Men wil het effect van herbicide en van kunstmest weten. De waarnemingen worden dan gedaan aan vier groepen planten

1. zonder kunstmest en met herbicide
2. zonder kunstmest en zonder herbicide
3. met kunstmest en met herbicide
4. met kunstmest en zonder herbicide.

In dit voorbeeld zijn de vier combinaties de zogenaamde *blokken*.

In verslagen die op de middelbare school worden gemaakt over proeven is het veelal van belang om systematische verschillen tussen twee groepen waarnemingen bijvoorbeeld door het toepassen van twee behandelingen aan te tonen. In deze lesbrief is aangenomen dat een leerling is geïnteresseerd in het verschil tussen twee groepen die aan verschillende 'omstandigheden' worden blootgesteld. De vraag is dan of het verschil in omstandigheden op een gemeten kenmerk tot een aan te tonen systematisch verschil leidt. Dit houdt onder meer in dat het aantal waarnemingen in de twee groepen die vergeleken gaan worden niet gelijk hoeft te zijn.

We kunnen bijvoorbeeld kijken of twee meststoffen een verschillend effect hebben op de groei van een bepaald soort gras. Aan het begin van de proef waren de plantjes onder identieke omstandigheden al drie weken gekweekt. De helft van de plantjes krijgt nu gedurende 5 weken meststof A toegediend en de andere helft meststof B. Gedurende die 5 weken zijn er twee plantjes, die meststof B kregen aangevreten door een konijn en 1 die meststof A kreeg. Van de niet aangevreten plantjes wordt het bovengrondse gewicht bepaald en voor deze gegevens kan met de Wilcoxon twee steekproeventoets bekeken worden of er een systematisch verschil is in de bovengrondse productie bij meststof A en B.

Aan de hand van voorbeelden geef ik een leidraad waarmee de leerling kan bepalen in welke situaties de Twee steekproeventoets van Wilcoxon kan worden gebruikt. Daarna volgt een strategie, waarin stapsgewijs duidelijk wordt gemaakt hoe uit de waarnemingen conclusies kunnen worden getrokken met betrekking tot het verschil tussen de twee groepen waarnemingen die aan verschillende omstandigheden zijn blootgesteld.



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

Vraagstellingen waarbij het gebruik van de twee steekproeven toets gerechtvaardigd is

Is de ontwikkelingstijd van insecten verschillend voor mannelijke en vrouwelijke insecten?

Als een vlinder eitjes legt op een plant waarvan de rupsen kunnen eten, voordat zij zich gaan verpoppen, dan komt er uiteindelijk een mannelijke of vrouwelijke vlinder uit de pop. Van al de uitgekomen vlinders is het geslacht bepaald en de ontwikkelingstijd van ei tot volwassen vlinder. Als op deze manier van 25 volwassen vlinders het geslacht en de ontwikkelingstijd bekend zijn, dan kan met de Wilcoxon twee steekproeven toets worden geconcludeerd of de ontwikkelingstijden van de mannelijke en vrouwelijke vlinders systematisch van elkaar verschillen.

Heeft verhoging van het kooldioxide gehalte in de lucht een positief effect op de plantengroei?

Alleen als er slechts twee verschillende gehalten vergeleken worden dan kan de Wilcoxon twee steekproeven toets worden gebruikt. Enkele planten, zeg 15, groeien gedurende 4 weken op onder de normale in de buitenlucht heersende kooldioxide concentratie en een ander 15-tal planten groeit onder een verhoogde kooldioxide concentratie. Van de bovengrondse delen wordt na afloop het gewicht bepaald. Statistische analyse van deze waarnemingen geeft een aanwijzing of planten onder een verhoogd kooldioxide gehalte systematisch sneller groeien dan onder het heersende kooldioxide gehalte.

Verschilt de duur van de concentratie van mannen en vrouwen bij blootstelling aan geluid van een bepaalde frequentie?

Mannen worden in de loop van hun leven eerder doof voor hoge geluiden dan vrouwen. Geluidsgolven met relatief hoge frequenties leiden tot het horen van hoge tonen en relatief lage frequenties leiden tot het horen van lage tonen.

1. Als 10 mannen en 10 vrouwen van boven de 40 jaar blootgesteld worden aan een relatief hoogfrequent geluid, dan zijn vrouwen naar verwachting eerder door dit geluid gestoord dan mannen.
2. Als we daarentegen 10 mannen en 10 vrouwen aan een geluid met veel "bassen" erin blootstellen (=laagfrequent geluid), dan is vooraf niets bekend over hoe het verschil tussen beide geslachten zal uitpakken (als dat er al zou zijn).

In beide hierboven beschreven situaties bestaan de gegevens uit de duur van de concentratie na het begin van de blootstelling aan het geluid. Echter doordat we in het eerste geval de richting van het verschil denken te weten en in het tweede geval niet, pakt de aanpak van het toetsingsprobleem iets anders uit (zie onder "stappenplan")

Hebben twee verschillende mestsoorten effect op de productie van een bepaald gewas?

Vooraf is niet bekend of meststof A of B beter werkt. Met meststof A groeien 15 planten op en de overige 15 planten groeien op met meststof B. Tijdens het uitvoeren van de proef blijkt één of andere onverlaat 3 planten die met meststof A werden behandeld en 1 plant die met meststof B werd behandeld ontworteld te hebben. Na een bemesting van 3 weken zijn er dus 12 waarnemingen bij meststof A beschikbaar en 14 bij meststof B. Van elk plantje wordt het totale bovengrondse gewicht gemeten en op grond van deze gegevens kan met de Wilcoxon twee steekproeven toets worden bekeken of er een systematisch verschil is tussen de bovengrondse productie bij meststof A en B.



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

Is er een verschil in de leeftijd die mannelijke en vrouwelijke insecten bereiken?

Fruitvliegen zijn snel te kweken. Het verschil tussen mannelijke en vrouwelijke fruitvliegen is eenvoudig vast te stellen. De vraag is of er eventueel een verschil in leeftijd is tussen de seksen in het volwassen stadium (de fruitvlieg). Net nadat er vliegen uit poppen zijn gekropen wordt van de uitgekomen fruitvliegen het geslacht bepaald. Om de maximale leeftijd van vrouwtjes en mannetjes te bepalen worden 20 vrouwtjes en 20 mannetjes bij voldoende voedsel apart bewaard. Dagelijks wordt gekeken of ze nog leven en de dag waarop ze dood worden aangetroffen geldt als waargenomen leeftijd. Met de Twee steekproeventoets van Wilcoxon kan worden opgehelderd of er een systematisch verschil is tussen de maximale leeftijd van vrouwtjes en mannetjes.

Theorie

In het begin van deze paragraaf geef ik een overzicht van de begrippen die noodzakelijk zijn om een statistische toets goed uit te kunnen voeren. Voordat men een experiment uitvoert heeft men op grond van kennis of van een redenering soms al een idee welke van de twee waarnemingen systematisch groter of kleiner zal zijn.

In het boven beschreven voorbeeld over de duur van de concentratie bij mannen en vrouwen als ze aan een hoogfrequent geluid worden blootgesteld verwachten we dat de mannen van veertig jaar of ouder langer onverstoort doorwerken dan vrouwen van die leeftijd. Van de meststoffen in het voorbeeld uit de inleiding weten we niet vooraf welke de beste zal zijn. Om een statistische toets te kunnen uitvoeren is het voor beide genoemde experimenten noodzakelijk een veronderstelling te formuleren.

Met het formuleren van een zogenaamde *nulhypothese* wordt de collectie kansverdelingen voor het juiste onderliggende statistisch model ingeperkt. De nulhypothese sluit aan bij de tot nu toe aangenomen veronderstellingen (de traditie). Een nulhypothese heeft altijd een tegenhanger namelijk de *alternatieve hypothese*. Deze is altijd zodanig geformuleerd dat hij zegt dat het onderliggende statistische model niet beperkt is tot de collectie modellen onder de nulhypothese. Elke statistische toets geeft de mogelijkheid om op grond van de waarnemingen te besluiten of de nulhypothese al of niet verworpen dient te worden. Bij de duur van concentratie van mannen en vrouwen wanneer ze worden blootgesteld aan een hoog geluid is de (alternatieve) veronderstelling vooraf dat de gemeten duur van de concentratie bij mannen gemiddeld genomen groter zal zijn dan die bij vrouwen. Als in een hypothese het woord "groter" of "kleiner" voorkomt, dan heb je te maken met een *eenzijdig* te toetsen hypothese: afwijkingen naar een bepaalde kant wijzen namelijk op ondersteuning van de nulhypothese en afwijkingen de andere kant op leveren aanwijzingen voor het alternatief. Wanneer van tevoren niet duidelijk is naar welke kant een afwijking uit zou kunnen vallen, dan leveren afwijkingen naar beide kanten aanwijzingen voor het alternatief. In het geval van het effect van de meststoffen A en B is er vooraf geen enkele aanwijzing welke beter zou zijn dan de ander. Op deze manier wordt er *tweezijdig* getoetst.

Het is altijd mogelijk dat de nulhypothese ten onrechte wordt verworpen. De nulhypothese is dan waar, maar gedurende het uitvoeren van de toets besluiten is toch het besluit



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

gevallen om hem te verwerpen. Naarmate de kans dat de nulhypothese onterecht wordt verworpen kleiner wordt is de uitkomst van een statistische toets betrouwbaarder. De *onbetrouwbaarheid* van een statistische toets is gelijk aan het maximum van de kans dat de nulhypothese onterecht wordt verworpen. De onbetrouwbaarheid wordt meestal aangeduid met de Griekse letter α . Veel voorkomende ingestelde waarden van α zijn 0,05 en 0,10. Als een toets wordt uitgevoerd met een onbetrouwbaarheid α van 5%, dan is de kans dat de nulhypothese onterecht wordt verworpen dus maximaal 0,05. Bij een tweezijdige toets wordt zowel links als rechts verworpen en de onbetrouwbaarheid waarbij in een tabel moet worden afgelezen is dan $\alpha/2$. Ik hoop dit duidelijk te maken in de uitgewerkte voorbeelden.

Voor het uitvoeren van een statistische toets wordt altijd een uit de waarnemingen afgeleide grootte gebruikt. Deze wordt de *toetsingsgrootte* genoemd.

Voor het meten van verschillen tussen verschillende behandelingen pas ik hier de twee steekproeventoets van Wilcoxon toe. Om te kunnen bepalen of de berekende waarde van de toetsingsgrootte al of niet leidt tot het verwerpen van de nulhypothese wordt het *kritieke gebied* bepaald. Dit kritieke gebied is de verzameling van alle mogelijke waarden van de toetsingsgrootte waarvoor de nulhypothese verworpen zal worden. Voor alle overige waarden wordt de nulhypothese niet verworpen.

Uitvoering

Stappenplan Wilcoxon twee steekproeventoets.

Voor het systematisch uitwerken van een toetsingsprocedure voor de twee steekproeventoets van Wilcoxon is het volgende schema van toepassing:

1. Formuleer de probleemstelling in woorden, waarbij de waarnemingen binnen de ene behandeling worden aangeduid met een letter (zeg x) en de waarnemingen van de andere behandeling met een andere letter (zeg y).
2. Formuleer de nulhypothese en de alternatieve hypothese in woorden. Op grond van de nulhypothese en de alternatieve hypothese bepaal je of je eenzijdig of tweezijdig gaat toetsen.
3. Bepaal de toetsingsgrootte W . Voor de Wilcoxon twee steekproeventoets is dat de som van de rangnummers in de groep met het kleinste aantal waarnemingen. Geef aan of je verwacht dat W grote of kleine waarden aanneemt als de alternatieve hypothese waar is. Bij een tweezijdige toets verwacht je dat W ofwel grotere ofwel kleinere waarden aanneemt onder de alternatieve hypothese. In dat laatste geval betekenen middelmatige waarden van W een ondersteuning van de nulhypothese.
4. Kies voor de onbetrouwbaarheid α een waarde waarmee je de Wilcoxon twee steekproeventoets gaat uitvoeren (veelal 0,05 of 0,10).
5. Lees in de tabel de kritieke waarde(n) af en bepaal het kritieke gebied.
6. Bepaal de waarde van de toetsingsgrootte W .
7. Trek een conclusie op een statistische verantwoorde manier en vertel het resultaat vervolgens in je eigen woorden.



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

Als bovenstaande procedure stap voor stap wordt gevolgd kan voor elk probleem waarbij het verschil tussen twee behandelingen wordt onderzocht een verantwoorde toetsing worden verkregen.

Uitgewerkte voorbeelden

Voorbeeld: het effect van twee meststoffen op de bovengrondse productie van gras.

De 13 grasplanten van Engels Raaigras die bemest waren gedurende 5 weken met meststof A hadden als spruitgewicht in grammen: 1,750; 5,620; 2,070; 4,370; 4,970; 6,850; 7,890; 0,860; 0,640; 0,360; 1,350; 5,970; 5,260. Voor de 12 plantjes die met meststof B waren bemest zijn de volgende spruitgewichten gemeten: 3,450; 5,220; 5,270; 3,520; 6,150; 7,800; 4,270; 1,260; 3,910; 5,700; 2,340; 5,460.

1. Probleemstelling: leidt het geven van meststof A of meststof B tot een systematisch verschil in de bovengrondse productie? Metingen zijn gedaan aan het spruitgewicht a van plantjes die met meststof A zijn gekweekt en aan het spruitgewicht b van plantjes die met meststof B zijn gekweekt.
2. Als nulhypothese nemen we aan dat er geen systematisch verschil is tussen de gewichten in de groepen planten die gekweekt zijn met meststof A en B. De bijbehorende alternatieve hypothese is dan dat er wel een systematische verschil is aan te tonen tussen de twee groepen. Aangezien we vooraf geen idee hebben hoe elke meststof de bovengrondse productie zal beïnvloeden gaan we een tweezijdige toets uitvoeren.
3. De toetsingsgrootte W is de som van de rangnummers in de steekproef met het kleinste aantal waarnemingen (groep B)

Meststof A	Meststof B	A van klein naar groot	Rangnummer s voor groep A	B van klein naar groot	Rangnummers voor groep B
1,750	3,450	0,360	1	1,260	4
5,620	5,220	0,640	2	2,340	8
2,070	5,270	0,860	3	3,450	9
4,370	3,520	1,350	5	3,520	10
4,970	6,150	1,750	6	3,910	11
6,850	7,800	2,070	7	4,270	12
7,890	4,270	4,370	13	5,220	15
0,860	1,260	4,970	14	5,270	17
0,640	3,910	5,260	16	5,460	18
0,360	5,700	5,620	19	5,700	20
1,350	2,340	5,970	21	6,150	22
5,970	5,460	6,850	23	7,800	24
5,260		7,890	25		

Onder de alternatieve hypothese heeft W de neiging om relatief kleine of relatief grote waarden aan te nemen.

4. We kiezen als waarde voor de onbetrouwbaarheid $\alpha = 0,10$.



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

5. De kritieke waarde bij een tweezijdige toets met 12 ($=n$) en 13 ($=m$) waarnemingen in de twee steekproeven en $\alpha = 0,10$ is te vinden door in de tabel aan het eind van deze lesbrief bij $\alpha = 0,05$ af te lezen, omdat je de linker kritieke waarde wilt bepalen van een tweezijdige toets. Het getal dat daar staat is 125. Het kritieke gebied bestaat uit het stuk waar $W \leq 125$ en uit het stuk waar geldt dat $W \geq 12(13+12+1) - 125 = 187$ (algemene formule voor rechterstuk van het kritieke gebied $W \geq n(n+m+1)$ – "afgelezen getal voor linker kritieke waarde").
6. In de bovenstaande tabel is te zien hoe de toetsingsgrootte W berekend wordt uit de waarnemingen. In dit voorbeeld is $W=170$ (de som van de rangnummers in groep B).
7. De waarde van W ligt niet in het kritieke gebied en dan verwerpen we de nulhypothese niet. De conclusie luidt: er is geen systematisch verschil in bovengrondse productie bij bemesting met meststof A of meststof B.

Voorbeeld: de duur van de concentratie van mannen en vrouwen bij blootstelling aan geluid van een bepaald frequentie

Mannen worden in de loop van hun leven eerder doof voor hoge geluiden dan vrouwen. We stellen 12 mannen en 12 vrouwen van boven de 40 jaar bloot aan een redelijk hoog geluid. De verwachting is dat vrouwen eerder door dit geluid worden gestoord dan mannen.

1. Probleemstelling: worden in de categorie "mensen boven de 40" vrouwen eerder door een bepaald hoog geluid afgeleid dan mannen? Hoelang het duurt voordat 12 mannen en 12 vrouwen gestoord worden is gemeten met behulp van een stopwatch. De tijden in de onderstaande tabel zijn dus tijden in seconden.
2. Als nulhypothese nemen we aan dat er geen systematisch verschil is tussen de tijden van doorwerken voor mannen en vrouwen. De bijbehorende alternatieve hypothese is dat vrouwen aantoonbaar eerder gestoord worden door het geluid. We hebben nu dus vooraf een idee over hoe de tijden verschillen en we voeren dus een eenzijdige toets uit.
3. De toetsingsgrootte W is de som van de rangnummers in de steekproef met het kleinste aantal waarnemingen. Hier hebben beide groepen 12 waarnemingen en is er dus geen kleinste groep. We nemen voor W de som van de rangnummers voor de vrouwen. Onder de alternatieve hypothese heeft W de neiging om relatief kleine waarden aan te nemen.



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

vrouwen	mannen	Vrouw van laag naar hoog	Rangnummers voor vrouwen	man van laag naar hoog	Rangnummers voor mannen
528	805	76	1	130	2
293	364	194	3	364	6
659	997	293	4	805	11
906	1250	339	5	881	13
888	130	434	7	928	16
1016	928	528	8	955	17
76	1133	544	9	997	18
544	881	659	10	1074	20
863	955	863	12	1094	21
339	1094	888	14	1133	22
434	1274	906	15	1250	23
194	1074	1016	19	1274	24

4. We kiezen als waarde voor de onbetrouwbaarheid $\alpha = 0,05$.
5. De linker-kritieke waarde bij een eenzijdige toets met 12 ($=n$) en 12 ($=m$) waarnemingen in de twee steekproeven en $\alpha = 0,05$ is 120 (zie tabel aan eind van deze lesbrief bij $\alpha = 0,05$). Het kritieke gebied bestaat uit het stuk waar $W \leq 120$.
6. In de bovenstaande tabel is te zien hoe de toetsingsgrootte W berekend wordt uit de waarnemingen. In dit voorbeeld is $W=103$ (de som van de rangnummers in de groep vrouwen).
7. De waarde van W ligt dus in het kritieke gebied en dan verwerpen we de nulhypothese. Voor de categorie "mensen van boven de 40" luidt de conclusie: vrouwen worden systematisch eerder gestoord door een hoge toon dan mannen.



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

Vragen en opdrachten

Opdracht: het effect een nieuw dienstrooster bij de Nederlandse Spoorwegen

Op 10 juni 2001 is er een nieuw dienstrooster ingevoerd bij de Nederlandse Spoorwegen. Daarbij werden conducteurs wat vaker op hetzelfde traject ingezet (het zogenaamde "rondje om de kerk"). De spoorwegvakbonden zijn steeds van mening geweest dat een dienstrooster met daarin het rondje om de kerk tot méér vertragingen zou leiden dan voorheen. Een reizigersorganisatie wil nagaan of de bewering van de spoorwegvakbonden juist is. Om dit aan te tonen zijn er in de maand mei van 2001 acht treinreizen aselekt gekozen en daarvan is de vertraging vastgesteld. Na ingang van het nieuwe dienstrooster heeft men dit voor weer acht aselekt gekozen treinreizen herhaald. Voor de acht aselekt gekozen treinreizen in Mei 2002 was de vertraging (in minuten) 1; 6; 9; 2; 18; 0; 5; 6 en voor die in Juli 2001 was de gemeten vertraging achtereenvolgens 12; 7; 3; 26; 14; 4; 35 en 61 minuten. Toets met onbetrouwbaarheid $\alpha = 0,10$ of de invoering van het nieuwe dienstrooster tot een systematisch grotere vertraging heeft geleid.

Opdracht: het effect van geslacht op de Quetelet index

De Quetelet index is het getal dat verkregen wordt door het gewicht van een persoon (in kg) te delen door het kwadraat van zijn lengte (in m). Het is een veelgebruikte index om overgewicht en ondergewicht in beeld te brengen. Men wil weten of er bij aanvang van een studie aan de Wageningen Universiteit een systematisch verschil is in de Quetelet Index van mannelijke en vrouwelijke studenten. Hiertoe is een aselechte steekproef van 15 vrouwelijke en 15 mannelijke eerstejaars getrokken. Drie vrouwelijke studenten bleken niet tot dezelfde leeftijdsgroep te behoren als alle overige studenten en zijn daarom buiten beschouwing gelaten. De berekende Quetelet Indices voor vrouwen zijn 15,6; 16,8; 18,9; 19,6; 19,6; 19,7; 20,7; 21,1; 21,2; 22,0; 22,3 en 23,1. Voor de mannelijke studenten zijn de volgende indices berekend: 17,5; 18,4; 18,8; 19,8; 20,3; 20,6; 21,1; 21,3; 22,2; 22,4; 23,6; 25,7; 26,0; 26,2 en 27,3. Toets met onbetrouwbaarheid $\alpha = 0,10$ of er een systematisch verschil is in Quetelet index bij mannelijke en vrouwelijke eerstejaars bij aanvang van hun studie.

Opdracht: effect van woonwijk op verdeling van huis-, tuin- en keukenwerk over mannen en vrouwen

In twee wijken in een stad wil men het verschil van het huis-, tuin- en keukenwerk uitgevoerd door vrouwen en mannen bij standaardgezinnen (2 ouders van verschillend geslacht, 2 kinderen in de leeftijd van 4-12 jaar, geen extra huis-, tuin- en keukenhulp) vergelijken. Het te onderzoeken verschil is dus gedefinieerd als het aantal uren dat de vrouw des huizes per week aan huis-, tuin- en keukenwerk besteed min het aantal uren dat de man des huizes per week aan dezelfde soort zaken besteed. Men vermoedt dat dit verschil in wijk 2 systematisch groter is dan in wijk 1. Uit beide wijken wordt een aselechte steekproef van 10 gezinnen genomen. Uit wijk 1 blijken bij nader onderzoek 2 van de 10 geselecteerde gezinnen verhuisd te zijn en uit wijk 2 één gezin. Van de acht gezinnen in wijk 1 zijn de gerapporteerde verschillen 4,8; 3,4; 8,1; -6,0; 0,0; 2,8; 1,2 en -1,8 en die



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

van de negen gezinnen uit wijk 2 zijn 3,3; 10,3; 9,7; 2,3; 3,4; 8,3; 3,4; 3,1 en 8,5. Toets met een betrouwbaarheid $\alpha = 0,10$ of er een systematisch verschil is tussen de wijken in besteding van uren aan huis-, tuin- en keukenwerk door vrouwen en mannen

Suggesties voor verder onderzoek

Documentatie

Kuipers F.F. (1998) Voor de variatie: inleiding statistiek. Wageningen Pers, Wageningen

Staal H., Alten T. van, Spijkers F., Janssen C., Beusekom P van, Swaan M., Haven A., Lorist P., Kuijk L., Essers J., Evers F. en Ament J. (1999) Pascal wiskunde voor de tweede fase VWO informatieboek CM&EM, Thieme, Zutphen

Oriëntatie op vervolgonderwijs

Het onderwerp van deze lesmodule kom je ook tegen bij de opleiding Biologie van Wageningen Universiteit.

Kijk voor meer informatie op www.wageningenuniversiteit.nl.



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

Bijlage 1 Tabel voor de Wilcoxon twee steekproeventoets

Kritieke waarden voor de twee-steekproeventoets van Wilcoxon. Linker kritieke waarden voor de toetsingsgrootheid W bij steekproefomvang n en m ($n \leq m$) en onbetrouwbaarheidsdrempel α . Rechter kritieke waarde = $n(N + 1)$ - linker kritieke waarde, met $N = n + m$.

n	m	α					$n(N + 1)$
		0,005	0,010	0,025	0,050	0,100	
2	3	-	-	-	-	3	12
3	3	-	-	-	6	7	21
2	4	-	-	-	-	3	14
3	4	-	-	-	6	7	24
4	4	-	-	10	11	13	36
2	5	-	-	-	3	4	16
3	5	-	-	6	7	8	27
4	5	-	10	11	12	14	40
5	5	15	16	17	19	20	55
2	6	-	-	-	3	4	18
3	6	-	-	7	8	9	30
4	6	10	11	12	13	15	44
5	6	16	17	18	20	22	60
6	6	23	24	26	28	30	78
2	7	-	-	-	3	4	20
3	7	-	6	7	8	10	33
4	7	10	11	13	14	16	48
5	7	16	18	20	21	23	65
6	7	24	25	27	29	32	84
7	7	32	34	36	39	41	105
2	8	-	-	3	4	5	22
3	8	-	6	8	9	11	36
4	8	11	12	14	15	17	52
5	8	17	19	21	23	25	70
6	8	25	27	29	31	34	90
7	8	34	35	38	41	44	112
8	8	43	45	49	51	55	136
1	9	-	-	-	-	1	11
2	9	-	-	3	4	5	24
3	9	6	7	8	10	11	39
4	9	11	13	14	16	19	56
5	9	18	20	22	24	27	75
6	9	26	28	31	33	36	96
7	9	35	37	40	43	46	119
8	9	45	47	51	54	58	144
9	9	56	59	62	66	70	171
1	10	-	-	-	-	1	12
2	10	-	-	3	4	6	26
3	10	6	7	9	10	12	42
4	10	12	13	15	17	20	60
5	10	19	21	23	26	28	80
6	10	27	29	32	35	38	102
7	10	37	39	42	45	49	126
8	10	47	49	53	56	60	152
9	10	58	61	65	69	73	180
10	10	71	74	78	82	87	210



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

n	m	α					$n(N+1)$
		0,005	0,010	0,025	0,050	0,100	
1	11	-	-	-	-	1	13
2	11	-	-	3	4	6	28
3	11	6	7	9	11	13	45
4	11	12	14	16	18	21	64
5	11	20	22	24	27	30	85
6	11	28	30	34	37	40	108
7	11	38	40	44	47	51	133
8	11	49	51	55	59	63	160
9	11	61	63	68	72	76	189
10	11	73	77	81	86	91	220
11	11	87	91	96	100	106	253
1	12	-	-	-	-	1	14
2	12	-	-	4	5	7	30
3	12	7	8	10	11	14	48
4	12	13	15	17	19	22	68
5	12	21	23	26	28	32	90
6	12	30	32	35	38	42	114
7	12	40	42	46	49	54	140
8	12	51	53	58	62	66	168
9	12	63	66	71	75	80	198
10	12	76	79	84	89	94	230
11	12	90	94	99	104	110	264
12	12	105	109	115	120	127	300
1	13	-	-	-	-	1	15
2	13	-	3	4	5	7	32
3	13	7	8	10	12	15	51
4	13	13	15	18	20	23	72
5	13	22	24	27	30	33	95
6	13	31	33	37	40	44	120
7	13	41	44	48	52	56	147
8	13	53	56	60	64	69	176
9	13	65	68	73	78	83	207
10	13	79	82	88	92	98	240
11	13	93	97	103	108	114	275
12	13	109	113	119	125	131	312
13	13	125	130	136	142	149	351
1	14	-	-	-	-	1	16
2	14	-	3	4	6	8	34
3	14	7	8	11	13	16	54
4	14	14	16	19	21	25	76
5	14	22	25	28	31	35	100
6	14	32	34	38	42	46	126
7	14	43	45	50	54	59	154
8	14	54	58	62	67	72	184
9	14	67	71	76	81	86	216
10	14	81	85	91	96	102	250
11	14	96	100	106	112	118	286
12	14	112	116	123	129	136	324
13	14	129	134	141	147	154	364
14	14	147	152	160	166	174	406



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

n	m	α					$n(N+1)$
		0,005	0,010	0,025	0,050	0,100	
1	15	-	-	-	-	1	17
2	15	-	3	4	6	8	36
3	15	8	9	11	13	16	57
4	15	15	17	20	22	26	80
5	15	23	26	29	33	37	105
6	15	33	36	40	44	48	132
7	15	44	47	52	56	61	161
8	15	56	60	65	69	75	192
9	15	69	73	79	84	90	225
10	15	84	88	94	99	106	260
11	15	99	103	110	116	123	297
12	15	115	120	127	133	141	336
13	15	133	138	145	152	159	377
14	15	151	156	164	171	179	420
15	15	171	176	184	192	200	465
1	16	-	-	-	-	1	18
2	16	-	3	4	6	8	38
3	16	8	9	12	14	17	60
4	16	15	17	21	24	27	84
5	16	24	27	30	34	38	110
6	16	34	37	42	46	50	138
7	16	46	49	54	58	64	168
8	16	58	62	67	72	78	200
9	16	72	76	82	87	93	234
10	16	86	91	97	103	109	270
11	16	102	107	113	120	127	308
12	16	119	124	131	138	145	348
13	16	136	142	150	156	165	390
14	16	155	161	169	176	185	434
15	16	175	181	190	197	206	480
16	16	196	202	211	219	229	528
1	17	-	-	-	-	1	19
2	17	-	3	5	6	9	40
3	17	8	10	12	15	18	63
4	17	16	18	21	25	28	88
5	17	25	28	32	35	40	115
6	17	36	39	43	47	52	144
7	17	47	51	56	61	66	175
8	17	60	64	70	75	81	208
9	17	74	78	84	90	97	243
10	17	89	93	100	106	113	280
11	17	105	110	117	123	131	319
12	17	122	127	135	142	150	360
13	17	140	146	154	161	170	403
14	17	159	165	174	182	190	448
15	17	180	186	195	203	212	495
16	17	201	207	217	225	235	544
17	17	223	230	240	249	259	595



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

n	m	α					$n(N+1)$
		0,005	0,010	0,025	0,050	0,100	
1	18	-	-	-	-	1	20
2	18	-	3	5	7	9	42
3	18	8	10	13	15	19	66
4	18	16	19	22	26	30	92
5	18	26	29	33	37	42	120
6	18	37	40	45	49	55	150
7	18	49	52	58	63	69	182
8	18	62	66	72	77	84	216
9	18	76	81	87	93	100	252
10	18	92	96	103	110	117	290
11	18	108	113	121	127	135	330
12	18	125	131	139	146	155	372
13	18	144	150	158	166	175	416
14	18	163	170	179	187	196	462
15	18	184	190	200	208	218	510
16	18	206	212	222	231	242	560
17	18	228	235	246	255	266	612
18	18	252	259	270	280	291	666
1	19	-	-	-	1	2	21
2	19	3	4	5	7	10	44
3	19	9	10	13	16	20	69
4	19	17	19	23	27	31	96
5	19	27	30	34	38	43	125
6	19	38	41	46	51	57	156
7	19	50	54	60	65	71	189
8	19	64	68	74	80	87	224
9	19	78	83	90	96	103	261
10	19	94	99	107	113	121	300
11	19	111	116	124	131	139	341
12	19	129	134	143	150	159	384
13	19	148	154	163	171	180	429
14	19	168	174	183	192	202	476
15	19	189	195	205	214	224	525
16	19	210	218	228	237	248	576
17	19	234	241	252	262	273	629
18	19	258	265	277	287	299	684
19	19	283	291	303	313	325	741
1	20	-	-	-	1	2	22
2	20	3	4	5	7	10	46
3	20	9	11	14	17	21	72
4	20	18	20	24	28	32	100
5	20	28	31	35	40	45	130
6	20	39	43	48	53	59	162
7	20	52	56	62	67	74	196
8	20	66	70	77	83	90	232
9	20	81	85	93	99	107	270
10	20	97	102	110	117	125	310
11	20	114	119	128	135	144	352
12	20	132	138	147	155	164	396
13	20	151	158	167	175	185	442
14	20	172	178	188	197	207	490
15	20	193	200	210	220	230	540
16	20	215	223	234	243	255	592
17	20	239	246	258	268	280	646
18	20	263	271	283	294	306	702
19	20	289	297	309	320	333	760
20	20	315	324	337	348	361	820



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

n	m	α					$n(N+1)$
		0,005	0,010	0,025	0,050	0,100	
1	21	-	-	-	1	2	23
2	21	3	4	6	8	11	48
3	21	9	11	14	17	21	75
4	21	18	21	25	29	33	104
5	21	29	32	37	41	47	135
6	21	40	44	50	55	61	168
7	21	53	58	64	69	76	203
8	21	68	72	79	85	92	240
9	21	83	88	95	102	110	279
10	21	99	105	113	120	128	320
11	21	117	123	131	139	148	363
12	21	136	142	151	159	169	408
13	21	155	162	171	180	190	455
14	21	176	183	193	202	213	504
15	21	198	205	216	225	236	555
16	21	220	228	239	249	261	608
17	21	244	252	264	274	287	663
18	21	269	277	290	301	313	720
19	21	295	303	316	328	341	779
20	21	322	331	344	356	370	840
21	21	349	359	373	385	399	903
<hr/>							
1	22	-	-	-	1	2	24
2	22	3	4	6	8	11	50
3	22	10	12	15	18	22	78
4	22	19	21	26	30	35	108
5	22	29	33	38	43	48	140
6	22	42	45	51	57	63	174
7	22	55	59	66	72	79	210
8	22	70	74	81	88	95	248
9	22	85	90	98	105	113	288
10	22	102	108	116	123	132	330
11	22	120	126	135	143	152	374
12	22	139	145	155	163	173	420
13	22	159	166	176	185	195	468
14	22	180	187	198	207	218	518
15	22	202	210	221	231	242	570
16	22	225	233	245	255	267	624
17	22	249	258	270	281	294	680
18	22	275	283	296	307	321	738
19	22	301	310	323	335	349	798
20	22	328	337	351	364	378	860
21	22	356	366	381	393	408	924
22	22	386	396	411	424	439	990



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

n	m	α					$n(N+1)$
		0,005	0,010	0,025	0,050	0,100	
1	23	-	-	-	1	2	25
2	23	3	4	6	8	12	52
3	23	10	12	15	19	23	81
4	23	19	22	27	31	36	112
5	23	30	34	39	44	50	145
6	23	43	47	53	58	65	180
7	23	57	61	68	74	81	217
8	23	71	76	84	90	98	256
9	23	88	93	101	108	117	297
10	23	105	110	119	127	136	340
11	23	123	129	139	147	156	385
12	23	142	149	159	168	178	432
13	23	163	170	180	189	200	481
14	23	184	192	203	212	224	532
15	23	207	214	226	236	248	585
16	23	230	238	251	261	274	640
17	23	255	263	276	287	300	697
18	23	280	289	303	314	328	756
19	23	307	316	330	342	357	817
20	23	335	344	359	371	386	880
21	23	363	373	388	401	417	945
22	23	393	403	419	432	448	1012
23	23	424	434	451	465	481	1081
1	24	-	-	-	1	2	26
2	24	3	4	6	9	12	54
3	24	10	12	16	19	24	84
4	24	20	23	27	32	37	116
5	24	31	35	40	45	51	150
6	24	44	48	54	60	67	186
7	24	58	63	70	76	84	224
8	24	73	78	86	93	101	264
9	24	90	95	104	111	120	306
10	24	107	113	122	130	140	350
11	24	126	132	142	151	161	396
12	24	146	153	163	172	183	444
13	24	166	174	185	194	205	494
14	24	188	196	207	218	229	546
15	24	211	219	231	242	254	600
16	24	235	244	256	267	280	656
17	24	260	269	282	294	307	714
18	24	286	295	309	321	335	774
19	24	313	323	337	350	364	836
20	24	341	351	366	379	394	900
21	24	370	381	396	410	425	966
22	24	400	411	427	441	457	1034
23	24	431	443	459	474	491	1104
24	24	464	475	492	507	525	1176



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

n	m	α					$n(N+1)$
		0,005	0,010	0,025	0,050	0,100	
1	25	-	-	-	1	2	27
2	25	3	4	6	9	12	56
3	25	11	13	16	20	25	87
4	25	20	23	28	33	38	120
5	25	32	36	42	47	53	155
6	25	45	50	56	62	69	192
7	25	60	64	72	78	86	231
8	25	75	81	89	96	104	272
9	25	92	98	107	114	123	315
10	25	110	116	126	134	144	360
11	25	129	136	146	155	165	407
12	25	149	156	167	176	187	456
13	25	170	178	189	199	211	507
14	25	192	200	212	223	235	560
15	25	216	224	237	248	260	615
16	25	240	249	262	273	287	672
17	25	265	275	288	300	314	731
18	25	292	301	316	328	343	792
19	25	319	329	344	357	372	855
20	25	348	358	373	387	403	920
21	25	377	388	404	418	434	987
22	25	408	419	435	450	467	1056
23	25	439	451	468	483	500	1127
24	25	472	484	501	517	535	1200
25	25	505	517	536	552	570	1275



Voor de docent

Uitwerking van de vragen

Opdracht: het effect een nieuw dienstrooster bij de Nederlandse Spoorwegen

1. Probleemstelling: leidt het invoeren van een nieuw dienstrooster tot systematisch meer vertragingen? Metingen zijn gedaan aan de vertragingstijd in de maand Mei 2001 m en aan de vertragingstijd in de maand Juli 2001 j .
2. Als nulhypothese nemen we aan dat er geen systematisch verschil is tussen de vertraging in de maanden Mei en Juli 2001. De bijbehorende alternatieve hypothese is dan dat er systematische meer vertraging optrad in de maand Juli 2001. Aangezien we vooraf het idee hebben dat de vertraging in de maand Juli groter zou zijn, een eenzijdige toets uitvoeren.
3. De toetsingsgrootte W is de som van de rangnummers in de steekproef met het kleinste aantal waarnemingen. De waarnemingsgroepen zijn echter even groot en dus kiezen we er één, zeg de groep gemeten in Mei. Onder de alternatieve hypothese heeft W de neiging om relatief kleine waarden aan te nemen.
4. We kiezen als waarde voor de onbetrouwbaarheidsdrempel $\alpha = 0,10$.
5. De kritieke waarde bij een eenzijdige toets met 8 ($=n$) en 8 ($=m$) waarnemingen in de twee steekproeven en $\alpha = 0,10$ is 55 (zie tabel aan eind van deze lesbrief bij $\alpha = 0,10$ aflezen) Het kritieke gebied bestaat uit het stuk waar $W \leq 55$.

Mei 2001	Juli 2001	m van klein naar groot	Rangnummers voor groep m	j van klein naar groot	Rangnummers voor groep j
1	12	0	1	3	4
6	7	1	2	4	5
9	3	2	3	7	9
2	26	5	6	12	11
18	14	6	7,5	14	12
0	4	6	8,5	26	14
5	35	9	10	35	15
6	61	18	13	61	16

6. In de bovenstaande tabel is te zien hoe de toetsingsgrootte W berekend wordt uit de waarnemingen. In dit voorbeeld is $W=51$ (de som van de rangnummers in groep m).
7. De waarde van W ligt in het kritieke gebied en dan verwerpen we de nulhypothese. De conclusie luidt: er is op basis van deze gegevens aangetoond dat er systematisch meer vertraging ontstaan is door het invoeren van het nieuwe dienstrooster.

Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

Opdracht: het effect van geslacht op de Quetelet index

1. Probleemstelling: Is er een systematisch verschil in Quetelet index tussen mannelijke en vrouwelijke studenten bij aanvang van de studie? Metingen zijn gedaan aan gewichten (in kg) en lengten (in m) en Quetelet index is berekend voor mannelijke (m) en vrouwelijke studenten (v).
2. Als nulhypothese nemen we aan dat er geen systematisch verschil is tussen de indices van mannelijke en vrouwelijke studenten. De bijbehorende alternatieve hypothese is dat er wel een systematisch verschil is in de Quetelet index van mannelijke en vrouwelijke studenten. Aangezien we vooraf geen idee hebben of de mannen of vrouwen een grotere Quetelet index zullen hebben, gaan we een tweezijdige toets uitvoeren.
3. De toetsingsgrootheid W is de som van de rangnummers in de steekproef met het kleinste aantal waarnemingen. Dat is hier door uitval van drie vrouwen de steekproef met vrouwen. Onder de alternatieve hypothese heeft W de neiging om relatief kleine of relatief grote waarden aan te nemen.
4. We kiezen als waarde voor de onbetrouwbaarheidsdrempel $\alpha 0,10$.
5. De kritieke waarde bij een tweezijdige toets met 12 ($=n$) en 15 ($=m$) waarnemingen in de twee steekproeven en $\alpha = 0,10$ is 133 (zie tabel aan eind van deze lesbrief bij $\alpha = 0,05$ aflezen omdat je de linker kritieke waarde wilt bepalen) Het kritieke gebied bestaat uit het stuk waar $W \leq 133$ en uit het stuk waar geldt dat $W \geq 12(15+12+1)-133=203$.

man	vrouw	Rangnummers voor groep m	Rangnummers voor groep v
17,5	15,6	3	1
18,4	16,8	4	2
18,8	18,9	5	6
19,8	19,6	10	7,5
20,3	19,6	11	7,5
20,6	19,7	12	9
21,1	20,7	14,5	13
21,3	21,1	17	14,5
22,2	21,2	19	16
22,4	22,0	21	18
23,6	22,3	23	20
25,7	23,1	24	22
26,0		25	
26,2		26	
27,3		27	

6. In de bovenstaande tabel is te zien hoe de toetsingsgrootheid W berekend wordt uit de waarnemingen. In dit voorbeeld is $W=136,5$ (de som van de rangnummers in groep v).



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

7. De waarde van W ligt niet in het kritieke gebied en dan verwerpen we de nulhypothese. De conclusie luidt: er is geen systematisch verschil aangetoond tussen de Quetelet index van mannelijke en vrouwelijke eerstejaarsstudenten.

Opdracht: effect van woonwijk op verdeling van huis-, tuin- en keukenwerk over mannen en vrouwen

1. Probleemstelling: Is er een systematisch groter verschil in wijk 2 in uren besteed aan huis-, tuin- en keukenwerk door vrouwen ten opzichte van mannen? Metingen zijn gedaan aan uren die de vrouw des huizes aan zulk werk besteed en aan uren die de man des huizes aan zulk werk besteed. Het verschil, het aantal uren besteed door de vrouw min het aantal uren besteed door de man, is het te onderzoeken kenmerk.
2. Als nulhypothese nemen we aan dat er geen systematisch verschil is tussen het aantal uren besteed door de vrouw min het aantal uren besteed door de man in de twee wijken. De bijbehorende alternatieve hypothese is dat er een systematisch groter verschil is in het aantal uren besteed door de vrouw min het aantal uren besteed door de man. Aangezien we vooraf het idee hebben dat het aantal uren besteed door de vrouw min het aantal uren besteed door de man groter is in wijk 2, gaan we een eenzijdige toets uitvoeren.
3. De toetsingsgrootheid W is de som van de rangnummers in de steekproef met het kleinste aantal waarnemingen. Dat is hier wijk 1 met 8 waarnemingen. Onder de alternatieve hypothese heeft W de neiging om relatief kleine waarden aan te nemen.
4. We kiezen als waarde voor de onbetrouwbaarheidsdrempel $\alpha = 0,10$.
5. De kritieke waarde bij een tweezijdige toets met 8 ($=n$) en 9 ($=n$) waarnemingen in de twee steekproeven en $\alpha = 0,10$ is 58 (zie tabel aan eind van deze lesbrief bij $\alpha = 0,10$ aflezen omdat je alleen de linker kritieke waarde wilt bepalen) Het kritieke gebied bestaat uit het stuk waar $W \leq 58$.

Wijk 1	Wijk 2	1 van klein naar groot	Rangnummers voor Wijk 1	2 van klein naar groot	Rangnummers voor Wijk 2
4,8	3,3	-6,0	1	2,3	5
3,4	10,3	-1,8	2	3,1	7
8,1	9,7	0,0	3	3,3	8
-6,0	2,3	1,2	4	3,4	10
0,0	3,4	2,8	6	3,4	10
2,8	8,3	3,4	10	8,3	14
1,2	3,4	4,8	12	8,5	15
-1,8	3,1	8,1	13	9,7	16
	8,5			10,3	17

6. In de bovenstaande tabel is te zien hoe de toetsingsgrootheid W berekend wordt uit de waarnemingen. In dit voorbeeld is $W=51$ (de som van de rangnummers in groep 1).
7. De waarde van W ligt in het kritieke gebied en dus verwerpen we de nulhypothese. De conclusie luidt: er is aangetoond dat er in wijk 2 een systematisch groter verschil



Wilcoxon twee steekproeven toets

-Het toetsen van verschillen-

bestaat tussen het aantal uren besteed door de vrouw en het aantal uren besteed door de man aan huis-, tuin- en keukenwerk.

