



Universiteit Utrecht

Bachelorthesis

De toename van de kans op toevallige significante resultaten bij het toetsen van meerdere hypotheses

Faculteit Sociale Wetenschappen, Afdeling Methoden & Statistiek
Universiteit Utrecht

Student: Anouk Sprokel, 5775671

Woorden: 7192

1^e beoordelaar/begeleider: Fayette Klaassen

2^e begeleider: Herbert Hoijtink

2^e beoordelaar: Irene Klugkist

Datum: 29-06-2016

Samenvatting

Er is een hoge prevalentie van p-waardes net onder het significantieniveau (Brodeur, Lé, Sangnier & Zylberberg, 2016; De Winter & Dodou, 2014; Lakens, 2015; Masicampo & Lalande, 2012). Eerdere onderzoeken hebben geprobeerd dit fenomeen te verklaren (Gerber & Malhotra, 2008; Ioannidis, 2005; Simmons, Nelson & Simonsohn, 2011). In onderstaand onderzoek ligt de focus op de inflatie van de type 1 fout wanneer meerdere hypothesen gelijktijdig worden getoetst. Dit is onderzocht aan de hand van data gesimuleerd uit een populatie waar geen correlatie aanwezig is. Tevens is data gesimuleerd uit een populatie waar een kleine, irrelevante correlatie van .1 aanwezig is. Tijdens de analyses zijn verschillende aantallen respondenten gebruikt en een verschillend aantal variabelen. De resultaten die uit de analyses naar voren kwamen zijn dat de kans op ten minste één significant resultaat, wanneer geen correlatie in de populatie aanwezig is, groeit wanneer meer variabelen worden getoetst. Wanneer er een irrelevante correlatie van .1 aanwezig is, groeit deze kans sneller. Het aantal significante resultaten bij een correlatie van 0 blijft beperkt tot 1 à 2 op de 28 significante resultaten, terwijl het bij een populatie met een correlatie van .1 kan gaan om 1 op de 4 significante resultaten. Conclusie is dat een correctie, zoals de Bonferronicorrectie, gebruikt moet worden bij het toetsen van meerdere hypothesen om het onderzoek meer betrouwbaar en valide te maken.

Introductie

Het doel van de wetenschap is om de wereld om ons heen te verklaren, wetenschappers proberen dit te doen door middel van onderzoek (Tijmstra & Boeije, 2011). Als een onderzoeker een onderzoek opzet, stelt hij een hypothese op. Door middel van data probeert hij de nulhypothese te verwerpen, hij probeert een significant resultaat te behalen. Indien hij een significant resultaat vindt, wordt zijn artikel vaker geaccepteerd en gepubliceerd dan wanneer hij geen significant resultaat vindt (Gerber & Malhotra, 2008). In verschillende onderzoeken wordt gebruik gemaakt van het toetsen van correlaties. Een correlatie toetst of er samenhang aanwezig is tussen verschillende variabelen. Een aantal zaken zijn belangrijk bij het rapporteren van een correlatie, namelijk de power en het gekozen alpha-level. Power wordt beïnvloed door de grootte van de steekproef, de keuze voor één- of tweezijdig toetsen, de verwachte effectgrootte en de keuze van het alpha-level. Hoe groter de power, hoe kleiner de kans op een onterechte verwerping. Iedere onderzoeker kan een toepasselijk alpha-level kiezen. Vaak is dit een alpha-level van .05 of .01. Wanneer een statistisch onderzoek een correlatie rapporteert, zal dit vaak gebeuren aan de hand van de correlatiecoëfficiënt r en de p-waarde. Wanneer een p-waarde significant is, hangt af van het gekozen alpha-level. In de meeste onderzoeken wordt gekozen voor een alpha-level van .05. Als een p-waarde zich tussen de 0 en 0.5 bevindt, wordt hij als significant gezien. Je zou verwachten dat alle p-waardes tussen de 0 en .05 eenzelfde dalende lijn volgen, echter is dit niet het geval. Zowel Brodeur en collega's (2016) als Masicampo en Lalande (2012) hebben onderzoek gedaan naar de prevalentie van p-waardes. Net onder het meest voorkomende significantieniveau van .05, specifiek tussen de .041 en .049, is een piek te zien. We zien dat deze p-waardes beduidend vaker voorkomen dan dat we zouden verwachten (Brodeur et al., 2016; Masicampo & Lalande, 2012).

De hoge prevalentie is een reden om beter te kijken naar onderzoeken. Zo zijn er onderzoeken die veel hypothesen tegelijkertijd toetsen of alleen die resultaten rapporteren en beschrijven die significant zijn en zo minder de nadruk leggen op de resultaten die niet significant zijn. Zoals het artikel van Noom (1999) waar alleen iets gezegd wordt over de correlaties die significant zijn gebleken en niets gezegd wordt over mogelijke samenhang tussen de variabelen, over de toetsen die niet significant zijn en over het toetsen van meerdere hypothesen tegelijk. Als meerdere hypothesen tegelijk worden getoetst, dan kapitaliseert de onderzoeker de kans op het vinden van een toevallig significant resultaat. Hier is al een oplossing voor bedacht (de Bonferronicorrectie), maar deze wordt vaak niet toegepast zoals te

zien is in Noom (1999), Poortman en Voorpostel (2008) en Garnefski, Kraai en Spinhoven (2001). Men kan stellen dat de piek van p-waardes net onder de .05 alleen voorkomt in sociale studies. De Winter en Dodou (2015) hebben een onderzoek gedaan naar de prevalentie van p-waardes. Zij hebben gekeken of deze hoge prevalentie van p-waardes tussen de .041 en .049 voornamelijk voorkomen in een bepaald werkveld (sociaal, biologisch of natuurwetenschappelijk werkveld) en of deze hoge prevalentie iets is van de laatste jaren of dat dit al langer voorkomt. Zij bewezen dat zowel het aantal p-waardes tussen .041 en .049 als het aantal p-waardes tussen .051 en .059 in 2013 meer voorkwamen dan in 1990. Echter de hoeveelheid p-waardes onder dit significantieniveau is sneller gestegen dan boven het significantieniveau. Ook vonden zij dat in het biologisch werkveld het aantal p-waardes tussen .041 en .049 een exponentiële groei liet zien tussen 1990 en 2013. Hier kwamen deze waardes maar liefst 65,7 keer vaker voor dan in het natuurwetenschappelijk werkveld (De Winter & Dodou, 2015). Blijkbaar is het dus niet alleen aanwezig bij sociale studies, maar wordt deze prevalentie gevonden in meerdere werkvelden.

Waarom er een piek te zien is, kan aan verschillende aspecten toegeschreven worden. Hieronder zullen een aantal onderzoeken worden genoemd die mogelijke oorzaken hebben onderzocht wat betreft de hoge prevalentie van p-waardes. Gegeven is, dat het makkelijker is om een onderzoek te laten publiceren wanneer een significant resultaat gevonden is (Gerber & Malhotra, 2008). Hierdoor kunnen onderzoekers de drang krijgen om het onderzoek te manipuleren om een significant resultaat te verkrijgen. Fanelli (2009) heeft onderzoek gedaan naar het fabriceren en het vervalsen van data. Hierin gaf 2% toe dat ze ten minste één keer data hadden vervalst of veranderd en gaf 34% toe dat ze andere twijfelachtige dingen hadden gedaan met de resultaten of de data. Hoewel in dit onderzoek slechts gebruik wordt gemaakt van 21 ingevulde enquêtes, geeft dit wel een reden tot zorg. Fanelli (2009) geeft aan dat het erg lastig is om onderzoek te doen naar het fabriceren en vervalsen van data, omdat geen onderzoeker in aanraking wil komen met eventuele beschuldigingen of het krijgen van een slechte naam. Interessant is om te zien dat in ditzelfde onderzoek, 72% van de onderzoekers wel rapporteerde dat ze wisten dat een collega twijfelachtig was omgegaan met data. Met twijfelachtig omgaan bedoelt Fanelli (2009) alles wat een onderzoeker kan doen om een significant resultaat te behalen, behalve fabricatie (het zelf verzinnen van data) en falsificatie (het bewust vernietigen of weglaten van data). Een mogelijk ander aspect met betrekking tot de prevalentie van p-waardes kan zijn dat onderzoekers wel een significant resultaat vinden maar dat de power erg laag is (Lakens, 2015; Ioannidis, 2005). Hierdoor zijn de conclusies die aan deze data worden onttrokken, niet stabiel en kan het zijn dat dit niet representatief is voor

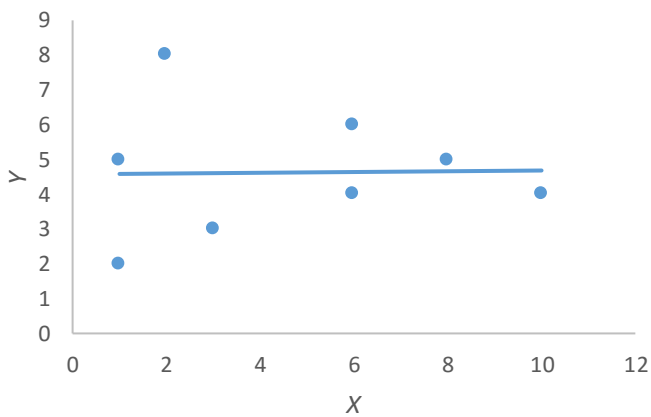
de populatie. Mogelijk komt de hoge prevalentie van p-waardes doordat onderzoekers verkeerde keuzes maken om maar een significant resultaat te behalen, zoals Simmons en collega's (2011) de 'researcher degrees of freedom' benoemen. Zij onderzochten een viertal factoren die eraan konden bijdragen het resultaat van een onderzoek te laten vertekenen. Ze onderzochten de flexibiliteit in de keuze van afhankelijke variabelen, de keuze voor de steekproefgrootte, de keuze voor het gebruik van covariaten en de keuze in welke informatie werd gegeven over de condities waarin het experiment werd uitgevoerd. Een laatste aspect wat in dit onderzoek behandeld wordt is dat de hoge prevalentie komt door de inflatie van een type 1 fout, die aanzienlijk groter wordt wanneer meerdere analyses tegelijk worden uitgevoerd. Zowel Noom (1999), als Poortman en Voorpostel (2008) als Garnefski en collega's (2001) toetsen meerdere hypothesen tegelijkertijd maar houden geen rekening met de inflatie van de type 1 fout. Door deze inflatie is het makkelijker om een significant resultaat te vinden wanneer meerdere hypothesen getoetst worden.

Samenvattend, we zien dat er een hoge prevalentie is van p-waardes tussen de 0.041 en 0.049. Verschillende mogelijk oorzaken zijn al onderzocht, zoals het manipuleren van data, de keuzes die onderzoekers maken ('research degrees of freedom'), het rapporteren van een groot aantal resultaten maar alleen ingaan op significante resultaten en het toetsen van meerdere hypothesen en de inflatie van de type 1 fout. In dit onderzoek zal de focus liggen op de inflatie van de type 1 fout bij het uitvoeren van meerdere analyses. Gekeken wordt naar hoe vaak de type 1 fout voorkomt wanneer de correlatie tussen 2 variabelen wordt getoetst. Hierna zal dit worden uitgebreid naar 4, 6, 8 en 10 variabelen, waar de correlaties tussen de variabelen tegelijkertijd getoetst worden. Dan kan er gekeken worden naar hoe vaak men een significant resultaat vindt. We weten immers hoe vaak we deze zouden mogen verwachten op basis van toeval, namelijk 5% van het totaal aantal steekproeven getrokken uit de populatie (uitgaande van een gekozen alpha-level van .05). Het doel van dit onderzoek is dan ook om te bekijken wat er met de toevallige kans op het onterecht verwerpen van de nulhypothese gebeurt wanneer meerdere correlaties tegelijk getoetst worden. Dit wordt gedaan aan de hand van data gesimuleerd uit een populatie waar de correlatie 0 is en uit een populatie waar een irrelevante correlatie van .1 is.

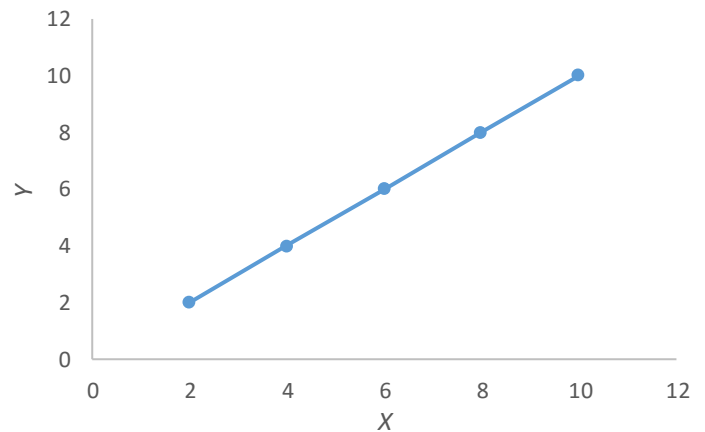
Theoretische achtergrond

Correlatie

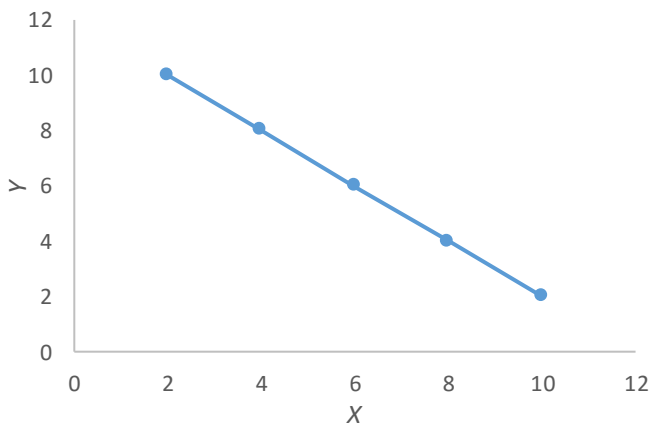
Een correlatie is een maat om te bekijken in hoeverre een variabele (X) samenhangt met een andere variabele (Y). De waarde van de correlatie ligt tussen de -1 en 1 . Men spreekt van een positieve relatie als de correlatie tussen de 0 en de 1 ligt, waar 0 geen correlatie is (zie Figuur 1) en 1 een perfecte positieve correlatie is (zie Figuur 2). Wanneer geen correlatie aanwezig is, kun je uitspraken doen als ‘Wanneer X toeneemt, kun je niks zeggen over Y ’. Wanneer een positieve relatie wordt gevonden, kun je uitspraken doen als ‘Wanneer X toeneemt, neemt Y ook toe’. Men spreekt van een negatieve correlatie wanneer de waarde van de correlatie tussen de -1 en 0 ligt, waar 0 geen correlatie is en -1 een perfecte negatieve correlatie is (zie Figuur 3). Wanneer een negatieve correlatie wordt gevonden, kun je uitspraken doen als ‘Wanneer X toeneemt, neemt Y af’.



Figuur 1: Geen correlatie, 0.



Figuur 2: Een positieve correlatie van 1.



Figuur 3: Een negatieve correlatie van -1 .

Er zijn 2 soorten correlaties, de bivariate en de partiële correlatie. Bij een bivariate correlatie wordt gekeken naar de samenhang tussen X en Y . Bij een partiële correlatie wordt gekeken naar de samenhang tussen X en Y terwijl rekening gehouden moet worden met een andere variabele (Z). Stel je hebt interesse in de samenhang tussen angst en leeftijd van kinderen. Onderzoek heeft uitgewezen dat over het algemeen meisjes bang zijn dan jongens. Je zou dan een partiële correlatie kunnen berekenen voor mate van angst en leeftijd terwijl je controleert voor sekse. In dit onderzoek wordt gekozen voor een bivariate correlatie. Dit omdat in de praktijk veelal de bivariate correlatie gebruikt wordt. Naast de keuze voor een bivariate of partiële correlatie, kies je ook nog de juiste correlatiemaat aan de hand van het meetniveau van de variabelen. Verschillende correlaties zijn de Pearson correlatie, de Spearman correlatie, Kendall's tau, de punt-biseriële correlatie en de Phi-correlatie (Gravetter, Wallnau, 2013; Field, 2014). In deze studie is gekozen om de Pearson correlatie te gebruiken want gezien is dat er vaak variabelen van interval- of ratiomeetniveau onderzocht worden. De formule die gebruikt wordt om de Pearson correlatie te berekenen is

$$\rho = \frac{cov(X,Y)}{SD(X)SD(Y)} \quad , \quad (1)$$

waarbij $X = [X_1, \dots, X_i, \dots, X_N]$ en $Y = [Y_1, \dots, Y_i, \dots, Y_N]$, N het aantal personen weergeeft, $cov(X, Y) = E[(X - E(X))(Y - E(Y))]$ de covariantie tussen X en Y weergeeft en

$$SD(X) = \sqrt{E[(X_i - E(X))^2]} \text{ en}$$

$$SD(Y) = \sqrt{E[(Y_i - E(Y))^2]} \text{ de standaardafwijking van } X \text{ en } Y \text{ weergeeft.}$$

Wanneer een correlatie getoetst wordt, is de nulhypothese: $H_0: \rho = 0$, waarin wordt verwacht dat er geen correlatie is en dus de correlatiecoëfficiënt (ρ) nul is. De alternatieve hypothese is $H_1: \rho \neq 0$, waarin wordt verwacht dat de correlatiecoëfficiënt (ρ) geen nul is. Er kan een uitspraak gedaan worden over de richting van de alternatieve hypothese. Deze kan namelijk positief ($H_1: \rho > 0$) of negatief ($H_1: \rho < 0$) zijn. Tijdens het toetsen van een correlatie, kun je toevallig een significant resultaat vinden, terwijl dit niet hoeft te betekenen dat de nulhypothese onjuist is. Dit wordt ook wel een fout positieve uitkomst genoemd. Gesproken wordt van een type 1 fout wanneer de nulhypothese onterecht verworpen wordt. De kans op deze fout is afhankelijk van het gekozen alpha-level. Wanneer een onderzoeker kiest voor een alpha-level van .05, dan is de kans op een type 1 fout 5%. Wanneer een onderzoeker kiest voor een alpha-level van .01, dan is de kans op een type 1 fout 1%. De tegenhanger van de

type 1 fout is de type 2 fout. Een type 2 fout is de kans dat de nulhypothese aangehouden wordt wanneer deze eigenlijk verworpen zou moeten worden. Wanneer meerdere toetsen achter elkaar worden uitgevoerd, stijgt de kans op ten minste één significant resultaat. Deze stijging wordt ook wel kanskapitalisatie genoemd. Hiermee wordt bedoeld dat bij elke volgende analyse de kans dat een bevinding per ongeluk significant is toeneemt. De formule die hiervoor gebruikt wordt is

$$P(I+) = 1 - (1 - \alpha)^p. \quad (2)$$

Waarin $P(I+)$ de kans op ten minste één significant resultaat weergeeft, p het aantal te toetsen correlaties weergeeft en α het gekozen alpha-level weergeeft. Wanneer je bijvoorbeeld 15 onafhankelijke hypothesen toetst, is de kans op ten minste één significant resultaat door toeval: $1 - (1 - 0.05)^{15} = 0.54$.

Oftewel wanneer je 15 hypothesen toetst, is de kans 54% dat je ten minste één toevallig significant resultaat vindt. In dit onderzoek zal gekeken worden naar hoe groot deze kans wordt als gevarieerd wordt in het aantal variabelen en respondenten. De onderzoeksvraag luidt dan ook als volgt: Verhoogt de toevallige kans op een significant resultaat wanneer er meerdere correlaties getoetst worden in een populatie waar geen significante resultaten verwacht worden? En wat gebeurt er met de kans op significante resultaten wanneer er meerdere correlaties getoetst worden in een populatie waar een irrelevante correlatie van .1 aanwezig is?

Methode

Datasimulatie

Er is gebruik gemaakt van het programma R (R Development Core Team, 2008). Met behulp van R is een code geschreven om data te simuleren en analyses uit te voeren. De volledige code staat in appendix A. De code volgt globaal een viertal stappen. Ten eerste is er een steekproef getrokken voor N respondenten en v variabelen uit zowel een populatie waar de correlatie 0 is als uit een populatie waar de correlatie .1 is. Ten tweede zijn al de bivariate correlaties binnen de steekproeven berekend en getoetst. Dit aantal is gegeven door de formule

$$(V \cdot (V - 1))/2.$$

Worden er bijvoorbeeld 2 variabelen gebruikt, betekent dit dat hierin $(2 \cdot (2-1))/2 = 1$ bivariate correlatie wordt getoetst. Ten derde is er bepaald of er ten minste één significante correlatie gevonden is en hoeveel significante correlaties aanwezig zijn. En ten vierde is dit proces 100.000 keer herhaald.

In dit onderzoek is ervoor gekozen om de studie 100.000 keer te repliceren. Dit is gebaseerd op de ‘Law of large numbers’. We zijn benieuwd naar de kans op ten minste één significant resultaat in een populatie waar geen of een irrelevante correlatie aanwezig is. De ‘law of large numbers’ geeft aan dat hoe vaker je iets herhaalt, hoe dichter je bij het gemiddelde van de populatie komt (Gravetter & Walnau, 2013; Wikipedia, n.d). Wanneer we een analyse slechts 1 of enkele keren uitvoeren, kan het zijn dat de steekproef niet representatief is voor de populatie. Omdat we een uitspraak willen doen over de populatie, nemen we een zeer groot aantal replicaties zodat we met ons gevonden gemiddelde dicht bij het ware gemiddelde van de populatie komen. De reden dat bij de tweede simulatie is gekozen voor een correlatie van .1, is omdat dit een kleine correlatie is en het interessant is om te kijken of een dergelijk kleine correlatie in de populatie effect heeft op de verdeling van significante resultaten. Wanneer twee variabelen een correlatie van .1 hebben, betekent dit dat deze variabelen slechts 1% variantie gemeenschappelijk hebben. Hierdoor wordt in dit onderzoek een correlatie van .1 als irrelevant en klein gelabeld.

Analysetechniek

Gekozen is voor het testen van bivariate correlaties, deze correlatie geeft weer of er een samenhang is tussen twee of meer variabelen. Om data te simuleren maken we gebruik van een multivariate normaalverdeling met een gemiddelde van 0, een variantie van 1 en covarianties van 0 wanneer we simuleren uit een populatie waar de correlatie 0 is. We maken gebruik van covarianties van .1 wanneer we simuleren uit een populatie waar de correlatie .1 is.

Respondenten en variabelen

Volgens Gravetter en Wallnau (2013) is de kans op het verkrijgen van een foutief significant resultaat bij een kleinere steekproef groter dan wanneer een grote steekproef gebruikt wordt. Mede hierdoor is ervoor gekozen om te variëren in de grootte van de groep respondenten. De ideale grootte van een steekproef hangt af van een drietal factoren, de effectgrootte, de power en de keuze van het alpha-level. Wanneer deze drie elementen vaststaan, kan een vierde (de ideale steekproefgrootte) worden berekend (Cohen, 1992). Wanneer een alpha-level van .05 gekozen wordt, de power 0.80 is en een effectgrootte van 0.30 verwacht wordt, is een minimale steekproefgrootte tussen de 80 en 100 personen een vereiste voor het toetsen van 1 correlatie (Cohen, 1992; "Sample size for bivariate correlation, Pearson correlation and Pearson product moment correlation", n.d.). Er worden ook studies met minder respondenten uitgevoerd, zoals de studie van Stewart en El-Mallakh (2007) waar slechts 25 respondenten aan het onderzoek meedoen. Echter benoemen zij geen effectgrootte en power. Mogelijk doen ze dit omdat ze weten dat de power klein zal zijn vanwege het kleine aantal respondenten. Ook geven zij aan bij de discussie dat het moeilijk is om conclusies te trekken voor de populatie vanwege de kleine steekproef. Onderzoeken die een groot aantal hypothesen toetsen, hebben vaker ook een grotere steekproef. Zoals Maheux en Price (2015) met een steekproefgrootte van 599 (45 hypothesen), Noom (1999) met een steekproefgrootte van 400 (36 hypothesen) en Garnefski en collega's (2001) met een steekproefgrootte van 547 (36 hypothesen). Uit de literatuur blijkt dat een minimale steekproefgrootte 80 tot 100 personen moet bevatten (voor een power van .80, verwachte effectwaarde van .30 en toetsen van 1 correlatie), dat onderzoeken die meerdere hypothesen testen vaak een grote steekproef hebben maar dat er ook nog steeds onderzoeken worden uitgevoerd met een kleine groep respondenten. Dat heeft geleid tot de verdeling die in dit onderzoek gebruikt wordt. De kleinste groep respondenten telt 25 fictieve personen, een tweede groep 50 fictieve personen en dit aantal loopt verder van 100, 250 tot en met 500. De

verdeling is zo gemaakt dat er mogelijk conclusies kunnen worden getrokken over de steekproefgrootte in verhouding tot de kans op een onterechte verwerping van de nulhypothese. Daarnaast is gekozen om maximaal 10 variabelen te gebruiken. Te beginnen bij 2 variabelen, waarmee 1 correlatie kan worden getoetst. Waarna 4, 6, 8 en 10 variabelen volgen. Met de formule $(V \cdot (V - 1))/2$ wordt bepaald hoeveel correlaties dit zijn. We gebruiken niet meer dan 10 variabelen omdat dit zou betekenen dat er meer dan 45 onafhankelijke hypothesen tegelijkertijd getoetst worden, dit komt in de praktijk weinig voor.

Resultaten

Alvorens de resultaten te bespreken, wordt eerst kort stilgestaan bij de te verwachte resultaten. Gezien is dat er een hogere prevalentie is van p-waardes net onder het significantieniveau. Waar in de inleiding meerdere mogelijke oorzaken besproken zijn, ligt in dit onderzoek de focus op de stijging van de kans op onterechte significante resultaten wanneer meerdere hypothesen tegelijk worden getoetst. We weten dat er sprake is van kanskapitalisatie wanneer we meerdere hypothesen tegelijk toetsen. Wanneer we de gegevens analyseren die onttrokken zijn uit een populatie waar de correlatie 0 is, verwachten we een stijging te zien wat betreft de kans op ten minste één onterechte significante resultaten. Deze stijging zal beïnvloedt worden door het aantal variabelen dat gebruikt wordt. De resultaten kunt u vinden in Tabel 1. Wanneer we kijken naar het aantal significante resultaten uit een populatie waar de correlatie 0 is, wordt verwacht dat het aantal significante resultaten af zal hangen van het aantal variabelen maar niet van het aantal respondenten. De resultaten kunt u vinden in Tabel 2 voor vier variabelen en in Tabel 3 voor acht variabelen. Wanneer we de gegevens analyseren die onttrokken zijn uit een populatie waar een irrelevante correlatie van .1 aanwezig is, verwachten we dat de stijging op ten minste één significant resultaat sneller stijgt dan wanneer geen correlatie aanwezig is in de populatie. Ook verwachten we dat de groei van de kans op ten minste één significant resultaat af zal hangen van zowel het aantal variabelen als het aantal respondenten. De resultaten kunt u vinden in Tabel 4. Wanneer we kijken naar het aantal significante resultaten uit een populatie met een irrelevante correlatie van .1, verwachten we dat het aantal af zal hangen van zowel het aantal variabelen als het aantal respondenten. De resultaten kunt u vinden in Tabel 5 voor vier variabelen, in Tabel 6 voor acht variabelen.

Tabel 1: De kans op 1 of meer onterechte significante resultaten als elke hypothese met een alpha-level van .05 wordt getoetst.

N	V(P)				
	2(1)	4(6)	6(15)	8(28)	10(45)
25	0.050	0.265	0.539	0.765	0.903
50	0.050	0.268	0.540	0.765	0.902
100	0.049	0.264	0.536	0.761	0.899
250	0.049	0.266	0.536	0.762	0.900
500	0.050	0.266	0.536	0.764	0.900

Tabel 1 toont de resultaten van de analyses van data onttrokken uit een populatie met een correlatie van 0. In de meest linkse kolom zien we het aantal respondenten staan (N), in de bovenste rij zien we het aantal variabelen staan (V) en tussen de haakjes staan het aantal correlaties wat getoetst is (P). Er is een groei te zien wanneer gekeken wordt naar het percentage significante resultaten. Wanneer we slechts twee variabelen nemen, zien we dat we gemiddeld genomen in 5% van de gevallen een significant resultaat vinden. Deze 5% is te verklaren door de type 1 fout, ofwel de kans op een toevallig significant resultaat dat samenhangt met het gekozen alpha-level. Echter wanneer we vier variabelen nemen zien we dat deze kans groter wordt. De kans op een significant resultaat wanneer we zes correlaties toetsen is 27%. Deze kans neemt alleen nog maar toe wanneer we meerdere correlaties toetsen. Wanneer er 45 onafhankelijke hypothesen worden getoetst, is de kans 90% dat er een significant resultaat gevonden wordt wat berust op toeval. De verwachting wat betreft de groei van de kans op ten minste één toevallig significant resultaat wordt bevestigd. Te zien is dat de resultaten in Tabel 1 een directe consequentie zijn van Formule 2. Als volgens de formule de kanskapitalisatie wordt uitgerekend voor bijvoorbeeld 28 hypothesen (8 variabelen) dan zien we dat hier 76% uitkomt. Want, $1-(1-0.05)^{28} = 0.76$.

Nadat gezien is dat de kans op ten minste één toevallig significant resultaat stijgt wanneer meerdere onafhankelijke hypothesen worden getoetst, zijn we nu benieuwd naar hoe ‘erg’ deze stijging is. Daarvoor wordt gekeken naar of dat het telkens slechts 1 significante correlatie is, of dat er vaker 2, 3, 4, 5 of 6 significante correlaties zijn. De verwachting is dat het aantal respondenten geen verschil zal maken, het aantal variabelen daarentegen wel.

Tabel 2: Percentage aantal significante resultaten bij 4 variabelen uit een populatie met een correlatie van 0.

N	0	1	2	3	4	5	6
25	0.732	0.237	0.028	0.003	0	0	0
50	0.734	0.234	0.030	0.002	0	0	0
100	0.736	0.231	0.030	0.002	0	0	0
250	0.737	0.231	0.030	0.002	0	0	0
500	0.733	0.233	0.031	0.002	0	0	0

Kijkende naar het aantal toevallige significante resultaten, zien we dat ook hier de verwachting waar blijkt te zijn. Het aantal significante resultaten hangt niet af van het aantal respondenten. Voor elke grootte van de respondentengroepen vinden we in 23% van de gevallen slechts één toevallig significant resultaat, terwijl in slechts 3% van de gevallen twee toevallige significante resultaten wordt gevonden en in minder dan 1% van de gevallen drie

toevallige significante resultaten worden gevonden. In geen van de replicaties vonden we vier of meer significante resultaten. Hieruit kunnen we concluderen dat het aantal toevallige significante resultaten onafhankelijk is van het aantal respondenten in de steekproef.

Tabel 3: Percentage aantal significante resultaten bij 8 variabelen uit een populatie met een correlatie van 0.

<i>N</i>	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6-28</i>
25	0.236	0.357	0.249	0.110	0.037	0.010	0*
50	0.237	0.351	0.250	0.112	0.037	0.010	0*
100	0.237	0.350	0.250	0.114	0.037	0.009	0*
250	0.238	0.352	0.250	0.112	0.037	0.009	0*
500	0.238	0.352	0.248	0.113	0.038	0.009	0*

Noot. *: < 0.001

Ook wanneer we kijken naar het aantal significante resultaten bij het toetsen van 28 correlaties (acht variabelen) zien we geen invloed van het aantal respondenten wat betreft het percentage toevallige significante resultaten. In slechts 23% van de gevallen vinden we geen significante resultaten, in 35% van de gevallen slechts 1 significant resultaat en in 25% van de gevallen slechts 2 significante resultaten. In 76% van de gevallen vinden we dus 1 tot 5 significante resultaten van de 28 correlaties die getoetst worden.

Zoals in bovenstaande resultaten te zien is, groeit de kans op een toevallig significant resultaat exponentieel wanneer meerdere correlaties tegelijkertijd getoetst worden. Hoewel de verwachting was dat het aantal respondenten niet van invloed zou zijn op de analyses, zijn we toch benieuwd naar of er mogelijk wel een invloed wordt uitgeoefend. Mogelijk kan deze rol beter onderzocht worden wanneer we hetzelfde onderzoek herhalen, maar nu ervoor kiezen dat er wel een correlatie aanwezig is in de populatie waaruit we een steekproef trekken. Gekozen is voor een correlatie van .1. We kiezen voor een correlatie van .1 omdat dit een erg kleine en irrelevante correlatie is. Wanneer een correlatie van .1 gevonden worden dan wil dat zeggen dat beide variabelen slechts 1% variantie gemeenschappelijk hebben. Zoals eerder genoemd, wordt verwacht dat zowel het aantal variabelen als het aantal respondenten invloed zal hebben op de resultaten.

Tabel 4: De kans op 1 of meer significante resultaten als elke hypothese met een alpha-level van .05 wordt getoetst uit een populatie waar de correlatie .1 is.

N	V(P)				
	2(1)	4(6)	6(15)	8(28)	10(45)
25	0.076	0.370	0.674	0.870	0.959
50	0.106	0.472	0.782	0.934	0.985
100	0.167	0.639	0.906	0.984	0.998
250	0.349	0.901	0.994	0.999	1.000
500	0.608	0.992	0.999	1.000	1.000

Ook wanneer we steekproeven trekken uit een populatie waar de correlatie .1 is, zien we dat de kans op een toevallig significant resultaat vele malen groter wordt wanneer meerdere hypothesen tegelijk worden getoetst (zie Tabel 4). Wat echter ook te zien is, is dat het aantal respondenten in de steekproef invloed heeft op de kans op ten minste één significant resultaat. Hoe groter de steekproef, hoe groter de kans is op ten minste één significant resultaat. Waar de kans op ten minste één significant resultaat bij twee variabelen en 250 respondenten 35% is, is de kans op ten minste één significant resultaat bij acht en tien variabelen al gegroeid naar 100%. Ook het aantal respondenten is van invloed. Bij vier variabelen en 25 respondenten is 37% kans op een toevallig significant resultaat, terwijl dit bij 500 respondenten (en vier variabelen) al 99% is. Dit wil dus zeggen dat hoe groter de steekproef en het aantal respondenten, wanneer er daadwerkelijk een correlatie aanwezig is in de populatie, hoe groter de kans op ten minste één significant resultaat.

Ook in het aantal significante resultaten zien we de invloed van zowel het aantal variabelen als het aantal respondenten. Ook zie je een verschuiving van het aantal significante resultaten naarmate het aantal variabelen toeneemt en het aantal respondenten toeneemt.

Tabel 5: Percentage aantal significante resultaten bij 4 variabelen uit een populatie met een correlatie van .1.

N	0	1	2	3	4	5	6
25	0.630	0.297	0.061	0.010	0.002	0*	0*
50	0.528	0.341	0.103	0.023	0.004	0*	0*
100	0.361	0.376	0.185	0.059	0.015	0.003	0*
250	0.099	0.248	0.292	0.214	0.107	0.035	0.006
500	0.008	0.049	0.142	0.243	0.277	0.202	0.080

Noot. *: < 0.001

Tabel 6: Percentage aantal significante resultaten bij 8 variabelen uit een populatie met een correlatie van .1.

<i>N</i>	<i>0</i>	<i>1-5</i>	<i>6-10</i>	<i>11-15</i>	<i>16-20</i>	<i>21-24</i>	<i>24-28</i>
25	0.130	0.838	0.032	0*	0	0	0
50	0.066	0.832	0.100	0.002	0*	0	0
100	0.016	0.652	0.309	0.023	0*	0*	0
250	0*	0.094	0.501	0.345	0.057	0*	0*
500	0*	0*	0.033	0.292	0.504	0.158	0.013

Noot. *: < 0.001

Je ziet dit bij zowel vier variabelen als acht variabelen. Kijkend naar Tabel 5, zien we dat bij de eerste 3 rijen (25, 50 en 100 respondenten) het aantal significante resultaten grotendeels bij de 0, 1 en 2 ligt. Terwijl dit bij 250 en 500 opschuift naar 1, 2, 3 en 4. Ook in Tabel 6 is deze verschuiving te zien.

Conclusie

In dit onderzoek is de inflatie van de type 1 fout onderzocht. Als er geen correlatie aanwezig is in de populatie, stijgt de kans op een type 1 fout wanneer meerdere hypothesen worden getoetst. De kans op ten minste één onterecht significant resultaat stijgt, dit stijgt tot 90% wanneer 45 hypothesen getoetst worden. Echter het aantal onterechte significante resultaten blijft beperkt. Dit betekent dat wanneer 28 hypothesen tegelijk worden getoetst, het aantal onterechte significante resultaten zich beperkt tot slechts enkele. Echter blijft de kans op ten minste één onterecht significant resultaat wel groot, namelijk 76% bij het toetsen van 28 hypothesen. Dit betekent dat onderzoeken die meerdere hypothesen tegelijk toetsen mogelijk minder betrouwbaar en valide zijn, omdat ze resultaten hebben gevonden die niet aanwezig zijn in de populatie. Hoe meer hypothesen de onderzoeker toetst, hoe groter de kans wordt op onterechte significante resultaten.

Vervolgens is onderzocht welk effect kanskapitalisatie heeft wanneer een irrelevante kleine correlatie aanwezig is in de populatie. In dat geval blijkt dat de kans op ten minste één toevallig significant resultaat nog sterker stijgt. Hierin speelt niet alleen het aantal variabelen een rol, maar stijgt de kans op toevallige significante resultaten ook met het aantal respondenten. Ook wanneer gekeken wordt naar het aantal significante resultaten, blijft dit niet slechts bij enkele maar neemt dit aantal sterk toe naarmate er meer respondenten zijn. Dit betekent dat veel correlaties significant bevonden worden, ondanks dat ze slechts een zeer kleine en irrelevante samenhang hebben met de andere variabelen. Ook worden bij een groter aantal variabelen onttrokken uit een populatie met een correlatie van .1 meer significante resultaten gevonden dan bij eenzelfde aantal variabelen onttrokken uit een populatie met een correlatie van 0. We kunnen dus concluderen dat de kans op toevallige significante resultaten stijgt naarmate meer correlaties worden getoetst. Hoe snel deze stijging verloopt, hangt af van de aanwezige correlatie. Het aantal toevallige significante resultaten hangt af van het aantal respondenten, het aantal variabelen en de aanwezige correlatie in de populatie.

Hoewel het in een populatie waar geen correlatie aanwezig is het aantal significante resultaten beperkt blijft, zal het advies luiden om een correctie zoals de Bonferronicorrectie toe te passen bij het toetsen van meerdere hypothesen. Bij deze correctie wordt het alpha-level gekozen aan de hand van het aantal te toetsen hypothesen. Zo kan het onderzoek beschermd worden tegen zowel onterechte als irrelevante resultaten. In de volgende sectie is een voorbeeld opgenomen waar de Bonferronicorrectie wordt toegepast.

Heranalyse van bestaande data

Zoals eerder benoemd, bestaat er al een oplossing voor de inflatie van de type 1 fout bij het toetsen van meerdere hypothesen, de Bonferronicorrectie. De Bonferronicorrectie stelt het alpha-level op aan de hand van het aantal getoetste hypothesen. Bijvoorbeeld als 9 hypothesen getoetst worden, zal het alpha-level $.05/9 = .0056$ worden. De kans op een type 1 fout blijft dan in principe hetzelfde, terwijl je meerdere hypothesen toetst (Wikipedia, n.d.).

Om te laten zien wat deze correctie met de resultaten kan doen, wordt gebruik gemaakt van data uit het artikel van Noom (1999). Tabel 4.4 uit het artikel wordt gebruikt waarin 36 correlaties worden weergegeven. Deze 36 hypothesen zijn getoetst met een eenzijdige t-toets en een alpha-level van .01. De keuze voor dit alpha-level kan erop duiden dat Noom (1999) zich al bewust was van de mogelijke gevolgen van het feit dat hij meerdere hypothesen toetst. Echter is het alpha-level van .01 niet streng genoeg gezien de gesimuleerde resultaten in het onderzoek. Omdat het een gebruikelijke uitgangspositie is om de Bonferronicorrectie toe te passen bij een alpha-level van .05, kies ik ervoor om zowel de correctie toe te passen op .05 als op .01. De Bonferronicorrectie wordt bij een alpha-level van .05, $.05/36 = .00139$. De Bonferronicorrectie wordt bij een alpha-level van .01, $.01/36 = .00028$. In het artikel van Noom (1999) worden 14 correlaties significant gevonden. De p-waarden worden opnieuw bekeken met beide correcties. Opvallend is dat Noom (1999) twee correlatiecoëfficiënten als niet significant ziet ($p = .12$), terwijl dit aan de hand van de p-waarde en het alpha-level wel te stellen is. Hierdoor zien we in onderstaande tabel 16 significante resultaten.

Tabel 7: Een voorbeeld waarin de Bonferronicorrectie wordt toegepast op data van Noom (1999).

	<i>Attitudinal autonomy</i>	<i>Emotional autonomy</i>	<i>Functional autonomy</i>
<i>Individual characteristics</i>			
<i>Pubertal timing</i>	.11 (0.014)	.05 (0.159)	.13*(0.005)
<i>Temperament</i>	.03 (0.275)	.11 (0.014)	.27***(0.000#)
<i>Familial characteristics</i>			
<i>Monitoring father</i>	.14* (0.003)	.14* (0.003)	.16***(0.001)
<i>Monitoring mother</i>	.07 (0.081)	.07 (0.081)	.08 (0.055)
<i>Emotional support father</i>	.12* (0.008)	.14* (0.003)	.11 (0.014)
<i>Emotional support mother</i>	.03 (0.275)	.07 (0.081)	.05 (0.159)
<i>Stimulating autonomy father</i>	.07 (0.081)	.13* (0.005)	.18*** (0.000#)
<i>Stimulating autonomy mother</i>	.02 (0.345)	.05 (0.159)	.18*** (0.000#)
<i>Attachment to father</i>	.21*** (0.000#)	.25*** (0.000#)	.12* (0.008)
<i>Attachment to mother</i>	.09 (0.036)	.16** (0.001)	.01 (0.421)
<i>Cohesion</i>	.09 (0.036)	.05 (0.159)	.01 (0.421)
<i>Extra-familial characteristics</i>			
<i>Attachment to peers</i>	.10 (0.023)	.14* (0.003)	.20*** (0.000#)

Noot. Getoetst met 1-tailed .01, p-waardes staan tussen haakjes achter ρ . *significant bij .01, **significant na correctie bij .05/36=0.00139, *** significant na correctie bij .01/36=0.00028, # <0.001. Aangepast van "Adolescent Autonomy: Characteristics and Correlates" door M Noom, 1999, hoofdstuk 4, p. 62.

In eerste instantie, voordat de Bonferronicorrectie was toegepast ($\alpha=.01$), werden 16 correlaties significant bevonden. Nadat de correctie is toegepast op een alpha-level van .05, blijven er slechts 8 significante resultaten over. Nadat de correctie is toegepast op een alpha-level van .01, blijven er slechts 6 significante resultaten over. In dit artikel betekent dit dus dat meer dan de helft van de significante resultaten 'verdwijnen'. We zien dat er weinig verschil zit tussen de Bonferronicorrectie op .05 en de Bonferronicorrectie op .01, het scheidt 2 significante resultaten. Zonder de Bonferronicorrectie wordt toegepast, is de kans dat een willekeurige hypothese significant is veel groter. Dit was ook te verwachten gezien de resultaten van de simulatie (zie Tabel 1 en 4). Het is dus van belang om een correctie toe te passen wanneer meerdere hypothesen tegelijk worden getoetst.

Discussie

Er zijn een aantal kanttekeningen te plaatsen bij bovenstaand onderzoek. Wanneer gekeken wordt naar de populatie met een correlatie van .1, wordt deze als irrelevant en klein bestempeld. Hoewel een correlatie van .1 betekent dat de variabelen slechts 1% variantie gemeenschappelijk hebben, kan het toch zijn dat in sommige gevallen een correlatie van .1 wel van waarde is. Bijvoorbeeld wanneer onderzoek gedaan wordt naar een zeldzame ziekte. Daarnaast wordt, wanneer steekproeven getrokken worden uit een populatie met een correlatie van .1, de steekproef getoetst met de nulhypothese dat er geen correlatie aanwezig is, ofwel $H_0: \rho=0$. We weten zo dus niet welke van de significante resultaten daadwerkelijk significant zijn bevonden vanwege de correlatie van .1 en welke een toevallig gevolg zijn van de kanskapitalisatie. Dit probleem wordt relevanter naarmate het aantal respondenten toeneemt. Het onderzoek zou herhaald kunnen worden met $H_0: \rho=.1$ en $H_1: \rho \neq .1$.

Hoewel in dit onderzoek de Bonferronicorrectie wordt geïllustreerd, zijn er meer correcties die de type 1 fout kunnen minimaliseren in gevallen waarin meerdere hypothesen worden getoetst. Een aanbeveling voor verder onderzoek is dan ook een vergelijking van de verschillende beschikbare correcties. Wat zijn de voor- en nadelen van deze correcties en in hoeverre zijn ze toepasbaar in alle werkvelden/onderzoeksterreinen? Een tweede aanbeveling voor verder onderzoek is om onderzoek te verrichten naar de mate waarin de Bonferronicorrectie wordt toegepast in de praktijk. Dit omdat in vele artikelen nog geen correctie wordt toegepast en het mogelijk interessant is om de beweegredenen van onderzoekers te achterhalen. Zo kan mogelijk een vertaalslag worden gemaakt naar de vaste implementatie van een correctie indien meerdere hypothesen getoetst worden.

Referenties

- Brodeur, A., Lé, M., Sangnier, M., & Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal*, 8, 1-32. doi: 10.1257/app.20150044
- Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, 1, 98-101. <http://dx.doi.org/10.1111/1467-8721.ep10768783>
- De Winter, J. C., & Dodou, D. (2015). A surge of p-values between 0.041 and 0.049 in recent decades. *PeerJ* 3:e733. doi: 10.7717/peerj.733
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*, 4, e5738. doi: 10.1371/journal.pone.0005738
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics (4th edition)*. London: Sage.
- Garnefski, N., Kraaij, V., & Spinhoven, P. (2001). Negative life events, cognitive emotion regulation and emotional problems. *Personality and Individual Differences*, 30, 1311-1327. doi:10.1016/s0191-8869(00)00113-6
- Gerber, A. S., & Malhotra, N. (2008). Publication bias in empirical sociological research. *Sociological Methods & Research*, 37, 3-30. doi: 10.1177/0049124108318973
- Gravetter, F. J., & Wallnau, L. B. (2013). *Statistics for the Behavioral Sciences*. London: Thomson Wadsworth.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2, 696-701. doi: 10.1371/journal.pmed.0020124
- Lakens, D. (2015). On the challenges of drawing conclusions from p-values just below 0.5. *PeerJ*, 3, e1142. doi: 10.7717/peerj.1142

- Maheux, A., & Price, M. (2015). The indirect effect of social support on post-trauma psychopathology via self-compassion. *Personality and Individual Differences*, 88, 102-107. <http://dx.doi.org/10.1016/j.paid.2015.08.051>
- Masicampo, E. J. & Lalande, D. R. (2012). A peculiar prevalence of p values just below .05. *The Quarterly Journal of Experimental Psychology*.
<http://dx.doi.org/10.1080/17470218.2012.711335>
- Noom, M. (1999). Adolescent Autonomy: Characteristics and Correlates. Proefschrift ter verkrijging van de graad van doctor aan de Universiteit Utrecht. Hoofdstuk 4
- Poortman, A., & Voorpostel, M. (2009). Parental divorce and sibling relationships: A research note. *Journal of Family Issues*, 30, 74-91. doi: 10.1177/0192513X08322782
- R Development Core Team. (2008). R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Sample size for bivariate correlation, Pearson correlation and Pearson product moment correlation. (n.d.) Retrieved from <https://www.statisticssolutions.com/sample-size-for-bivariate-correlation-pearson-correlation-and-pearson-product-moment-correlation/>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359-1366. doi: 10.1177/0956797611417632
- Stewart, C., & El-Mallakh, R.S. (2007). Is bipolar disorder overdiagnosed among patients with substance abuse? *Bipolar Disorders*, 9, 646-648. doi: 10.1111/j.1399-5618.2007.00465.x
- Tijmstra, J., & Boeijs, H. (2011). *Wetenschapsfilosofie in de context van de sociale wetenschappen*. Den Haag: Boom/Lemma

Wikipedia. (n.d.) Bonferroni correction. Retrieved on 18-04-2016 from
https://en.wikipedia.org/wiki/Bonferroni_correction#Alternatives

Wikipedia. (n.d.). Law of large numbers. Retrieved on 18-04-2016 from
https://en.wikipedia.org/wiki/Law_of_large_numbers

Appendix A. Code datasimulatie

```
require("psych")
require("MASS")

set.seed(88,NULL,NULL)
AlleRespondenten <- c(25, 50, 100, 250, 500)
AlleVariabelen <- c(2, 4, 6, 8, 10)
Replicatie <- c(100000)
output <- outCorperc <- matrix(0, length(AlleRespondenten), length(AlleVariabelen))
## Hiermee worden twee matrixen aangemaakt om de data op te slaan en makkelijker te
## kunnen bekijken, eenmaal voor de data zonder correlatie en eenmaal voor de data met
## correlaite
outputaantal <- array(0, dim=c(Replicatie, length(AlleRespondenten),
length(AlleVariabelen)))
## Hiermee wordt een array aangemaakt waarin het aantal significante resultaten op te zoeken
## is, waarin de rijen het aantal replicaties zijn, de kolommen het aantal respondenten en de
## slises het aantal variabelen
populatie_correlatie <- 0
## Hiermee wordt de correlatie aanwezig in de populatie op een getal gezet tussen de 0 en de
## 1.

for(r in 1:length(AlleRespondenten)) {
  for(v in 1:length(AlleVariabelen)) {
    means <- rep(0, AlleVariabelen[v])
    sigma <- diag(AlleVariabelen[v])
    sigma[col(sigma)!= row(sigma)] <- populatie.correlatie
    for(i in 1:Replicatie){
      datamatrixcor <- mvrnorm(AlleRespondenten[r], means, sigma)
      colnames(datamatrixcor) <- paste0("variabele", 1:AlleVariabelen[v])
      datamatrix <- matrix (0, AlleRespondenten[r],AlleVariabelen[v])
      ## Hiermee wordt een matrix gemaakt met de lengte van het aantal respondenten op dat
      ## moment(een eerste keer is dat 25) en met het aantal variabelen op dat moment (een ##eerste
      ## keer is dat 2)
```



```

colnames(datamatrix) <- paste0("variabele", 1:AlleVariabelen[v])
for(k in 1:AlleVariabelen[v]){
  datamatrix[,k] <- rnorm(AlleRespondenten[r], 0, 1)
}

```

Met deze loop wordt de datamatrix gevuld met waarden die uit een normaalverdeling komen met een grootte van [r], een gemiddelde van 0 en een standaardafwijking van 1.

```

correlaties <- corr.test(datamatrix)

```

Hiermee worden alle mogelijke correlaties berekend in de datamatrix

```

correlaties$p

```

Hiermee worden de p-waarden van de correlaties opgevraagd

```

lower.tri(correlaties$p)

```

Hiermee wordt gevraagd om alleen het onderste gedeelte van de tabel te laten zien, deze worden weergegeven als TRUE en FALSE

```

correlaties$p < .05

```

Hiermee wordt gevraagd welke p-waarden kleiner zijn dan .05.

Deze worden weergegeven als TRUE en FALSE

```

lower.tri(correlaties$p)*(correlaties$p<.05)

```

Hiermee wordt gevraagd om de onderste tabel keer de significante resultaten te doen zodat je alleen de waarden krijgt die significant zijn met een andere variabelen en dus niet met zichzelf

```

type1fout <- sum(lower.tri(correlaties$p)*(correlaties$p<.05)) > 0

```

Hiermee zorg je ervoor dat hij deze fouten opslaat, iedere keer dat hij de loop herhaalt (oftewel 100.000 keer)

```

output[r,v] <- output[r,v] + type1fout

```

Hiermee zorg je ervoor dat deze waarde wordt opgeslagen en telkens erbij opgeteld wordt, zodat we een aantal krijgen van hoe vaak er tenminste 1 significant resultaat is gevonden

```

correlatiesaantal <- corr.test(datamatrixcor)

```

Hiermee worden alle mogelijke correlaties berekend in de datamatrixcor

```

correlatiesaantal$p

```

Hiermee worden de p-waarden van de correlaties opgevraagd

```

lower.tri(correlatiesaantal$p)

```

```

## Hiermee wordt gevraagd om alleen het onderste gedeelte van de tabel te laten zien, ##
deze worden weergegeven als TRUE en FALSE
correlatiesaantal$p < .05
## Hiermee wordt gevraagd welke p-waardes kleiner zijn dan .05.
Deze worden weergegeven als TRUE en FALSE
lower.tri(correlatiesaantal$p)*(correlatiesaantal$p<.05)
## Hiermee wordt gevraagd om de onderste tabel keer de significante resultaten te doen
zodat je alleen de waardes krijgt die significant zijn met een andere variabelen en dus niet
met zichzelf
outputaantal[i,r,v] <- sum(lower.tri(correlatiesaantal$p)*(correlatiesaantal$p<.05))
## Hiermee zorg je ervoor dat de resultaten bij elkaar komen te staan in een array, zodat
er per replicatie, per aantal variabelen en per aantal respondenten je kan zien hoeveel
significante resultaten er zijn.
type1foutcor <- sum(lower.tri(correlatiesaantal$p)*(correlatiesaantal$p<.05)) > 0
## Hiermee zorg je ervoor dat hij deze fouten opslaat, iedere keer dat hij de loop herhaalt
(ofte wel 100.000 keer)
outCorperc [r,v] <- output[r,v] + type1foutcor
## Hiermee zorg je ervoor dat deze waarde wordt opgeslagen en telkens erbij opgeteld ##
wordt, zodat we een aantal krijgen van hoe vaak er tenminste 1 significant resultaat ## is
gevonden bij een steekproef uit een populatie waar de correlatie .1 is.
}
}
}
output/Replicatie
## Hiermee zorg je ervoor dat de output in proporties gegeven wordt
Outputaantal
## Hiermee zien we de array waarin je kan opzoeken hoe vaak er significante resultaten zijn
gevonden.
outCorperc
## Hiermee zien we hoe vaak er uberhaupt 1 significant resultaat gevonden is uit de populatie
waar de correlatie .1 is.
outCorperc/Replicatie
## Hiermee zorg je ervoor dat de output, van de resultaten uit een populatie met een correlatie
van .1, in proporties gegeven wordt

```