

STATISTIEK VOOR DE PSYCHOLOGIE

DEEL 2: TOETSEN VOOR TWEE GEMIDDELDEN EN TOETSINGSTHEORIE

Statistiek voor de psychologie

Deel 2: Toetsen voor twee gemiddelden en toetsingstheorie

Jules L. Ellis

Derde druk

Boom Lemma uitgevers
Den Haag
2014

Inhoud

Voorwoord		11
Deel A	Toetsen voor twee gemiddelden	13
HOOFDSTUK 1	Elementair rapport van verschil in gemiddelden	14
1.1	Inleiding	14
1.2	Definitie van het elementaire rapport	14
HOOFDSTUK 2	Datadesign	16
2.1	Inleiding	16
2.2	Samenvatting	17
2.3	Het within-subjectdesign: het vergelijken van situaties	17
2.4	Het between-subjectdesign: het vergelijken van groepen personen	18
2.5	Vergelijking van de twee datadesigns	18
2.6	Afhankelijke en onafhankelijke variabele	21
2.6.1	Afhankelijke en onafhankelijke variabele in een between-subjectdesign	22
2.6.2	Afhankelijke en onafhankelijke variabele in een within-subjectdesign	22
HOOFDSTUK 3	Mate van controle	24
3.1	Inleiding	24
3.2	Samenvatting	24
3.3	Passief-observerend onderzoek	25
3.4	Actief-manipulerend onderzoek en experimenten	26
3.5	Counterbalancing	27
3.6	Randomiseren	28
3.7	Het belang van controlemaatregelen	29
3.8	Longitudinaal is niet experimenteel	30
3.9	Passief-observerend is niet slecht	32

HOOFDSTUK 4	De hypothesen	33
4.1	Inleiding	33
4.2	Samenvatting	34
4.3	Het formuleren van de hypothesen	34
4.3.1	Populatie versus steekproef	35
4.3.2	De rollen van H_0 en H_a	36
4.4	Hypothesen en design	37
HOOFDSTUK 5	De analyseprocedure	39
5.1	Inleiding	39
5.2	Het kiezen van de analyseprocedure	39
HOOFDSTUK 6	Analyse van within-subjectdesign: t -toets voor gekoppelde paren	40
6.1	Inleiding	40
6.2	Samenvatting	40
6.3	Berekenen van de geaggregeerde data	41
6.4	Het ruwe effect	46
6.5	Het relatieve effect	46
6.6	De t -grootheid	48
6.7	Het aantal vrijheidsgraden	49
6.8	De p -waarde	50
6.8.1	Bepaling van de p -waarde	50
6.9	Beslissing	51
6.10	Causale interpretatie	52
6.11	Controle op assumpties	59
6.12	Samenvatting	61
HOOFDSTUK 7	Analyse van between-subjectdesign: t -toets voor onafhankelijke steekproeven	63
7.1	Inleiding	63
7.2	Samenvatting	63
7.3	Berekenen van de geaggregeerde data	64
7.4	Het ruwe effect	64
7.5	Het relatieve effect	65
7.6	De t -grootheid	69
7.7	Het aantal vrijheidsgraden	71
7.8	De p -waarde	72
7.9	Beslissing	73
7.10	Causale interpretatie	73
7.11	Controle op assumpties	74
7.12	Samenvatting	76

HOOFDSTUK 8	Mixed designs en hun analyse	78
8.1	Inleiding	78
8.2	Samenvatting	78
8.3	Mixed designs	78
8.4	Analyses bij mixed designs	79
8.5	Elementair rapport van een toets voor interactie	80
HOOFDSTUK 9	Visualiseren: de t -toets voor onafhankelijke steekproeven	81
9.1	Inleiding	81
9.2	Samenvatting	81
9.3	Voorbeeld	81
HOOFDSTUK 10	De intuïtieve psycholoog en causaliteit	85
10.1	Inleiding	85
10.2	Het gelijkeniscriterium	86
10.3	De beschikbaarheid van oorzaken	87
10.3.1	De methodologische norm over causaliteit	87
10.3.2	Beperking van het intuïtief denken bij causale attributie	88
10.3.3	Onderzoeken over de invloed van beschikbaarheid op causale attributies	89
10.4	De vuistregel van het Mooie Verhaal	93
10.4.1	Onderzoek over de vuistregel van het Mooie Verhaal	93
10.4.2	Het dagelijks leven: voorbeelden van de vuistregel van het Mooie Verhaal	94
10.5	Causale verklaringen die men voor zichzelf geeft	95
10.5.1	Onderzoek over causale verklaringen voor zichzelf	95
HOOFDSTUK 11	Opgaven deel A	98
HOOFDSTUK 12	Leerdoelen en zelftoetsen deel A	114
12.1	Leerdoelen	114
12.2	Zelftoets A1	117
12.3	Uitwerkingen van zelftoets A1	122
12.4	Zelftoets A2	126
12.5	Uitwerkingen van zelftoets A2	131

Deel B	Toetsingstheorie	135
HOOFDSTUK 13	De steekproevenverdeling van het gemiddelde	137
13.1	Inleiding	137
13.2	Belangrijkste begrippen	139
13.2.1	Formulering van de nulhypothese	142
13.2.2	Waarom willen we überhaupt een p -waarde?	145
13.2.3	De veronderstelling dat de nulhypothese waar is	146
13.3	Uitschrijven van de steekproevenverdeling van het gemiddelde	147
13.3.1	Aannamen over de populatie individuele subjecten	148
13.3.2	Data in de steekproef	149
13.3.3	Uitschrijven van alle mogelijke steekproeven	149
13.3.4	Beschrijving van de steekproevenverdeling van het gemiddelde	151
13.3.5	Berekening van de theoretische standard error	151
13.3.6	Positioneren van de geobserveerde steekproef in de steekproevenverdeling	152
13.3.7	De p -waarde	153
13.3.8	Beoordelen of de berekende kans exact is	156
13.3.9	Samenvatting voorbeeld	157
13.4	Simuleren van de steekproevenverdeling van het gemiddelde	159
13.4.1	Aannamen over de populatie individuele subjecten	160
13.4.2	Data in de steekproef	161
13.4.3	Simuleren van een groot aantal steekproeven	161
13.4.4	Beschrijving van de steekproevenverdeling van het gemiddelde	166
13.4.5	Berekening van de theoretische standard error	166
13.4.6	Positioneren van de geobserveerde steekproef in de steekproevenverdeling	167
13.4.7	De p -waarde	168
13.4.8	Beoordelen of de berekende kansen exact zijn	170
13.4.9	Samenvatting voorbeeld	170
13.5	Beredeneren van de steekproevenverdeling van het gemiddelde	172
13.5.1	Aannamen over de populatie individuele subjecten	173
13.5.2	Data in de steekproef	174
13.5.3	Beredeneren van de steekproevenverdeling van het gemiddelde	174
13.5.4	Beschrijving van de steekproevenverdeling van het gemiddelde	176

	13.5.5	Berekening van de theoretische standard error	176
	13.5.6	Positioneren van de geobserveerde steekproef in de steekproevenverdeling	178
	13.5.7	De p -waarde	179
	13.5.8	Beoordelen of de berekende kans exact is	181
	13.5.9	Samenvatting voorbeeld	181
	13.6	Vergelijking van de drie methoden om een p -waarde te bepalen	182
	13.7	De wortel- N -wet en de Wet van de Grote Aantallen	184
	13.8	Betrouwbaarheidsinterval met z	185
	13.9	Veel gemaakte fouten	187
	13.9.1	Standard error	187
	13.9.2	Wanneer moet je delen door \sqrt{N} ?	188
	13.9.3	Foute interpretaties van de CLS	189
HOOFDSTUK	14	De steekproevenverdeling van t	191
	14.1	Inleiding	191
	14.2	Samenvatting	191
	14.3	De geschatte standard error	192
	14.4	De variabiliteit van de geschatte standard error	193
	14.5	De t -waarde	195
	14.6	De referentieverdeling van t (uitschrijven)	196
	14.7	De referentieverdeling van t (simuleren)	199
	14.8	De referentieverdeling van t (redeneren)	202
	14.9	Betrouwbaarheidsinterval met t	204
	14.10	De verdeling van t als de nulhypothese onwaar is	205
	14.11	Power	206
	14.12	Power bij t -toetsen	212
	14.13	Veel gemaakte fouten	212
HOOFDSTUK	15	Algemene toetsingstheorie	214
	15.1	Inleiding	214
	15.2	Principes van statistische toetsen	217
	15.2.1	Wat is toetsen?	217
	15.2.2	Waarom is toetsen belangrijk?	217
	15.2.3	Wat is een statistische toets?	217
	15.2.4	De beperking van statistische toetsen	218
	15.2.5	De stappen van elke statistische toets	218
	15.2.6	Eisen waar een statistische toets aan moet voldoen	220
	15.2.7	Statistische conclusievaliditeit	220
	15.2.8	Questionable research practices	222
	15.2.9	De NHST-controverse	224
	15.2.10	Fisher versus Neyman-Pearson	228

15.3	Voorbeeld I: enkelvoudige aselechte steekproef	231
15.4	Voorbeeld II: dementie	232
15.5	Voorbeeld III: computerratten	236
15.6	Voorbeeld IV: statistiek binnen één persoon	240
HOOFDSTUK 16	Opgaven deel B	244
HOOFDSTUK 17	Leerdoelen en zelftoetsen deel B	262
17.1	Leerdoelen	262
17.2	Zelftoets B1	264
17.3	Uitwerkingen zelftoets B1	266
17.4	Zelftoets B2	269
17.5	Uitwerkingen zelftoets B2	272
Appendix		276
Referenties		283
Register		286

Voorwoord

De serie *Statistiek voor de psychologie* is, zoals de titel al zegt, een inleiding in de statistiek toegespitst op opleidingen psychologie. De serie bestaat uit vijf delen. Het algemene uitgangspunt is dat studenten aan de opleidingen psychologie, die niet allereerst komen voor de statistiek, toch gemotiveerd worden om het vak statistiek met goed gevolg af te leggen. De serie doet dit door in ‘gewone taal’ statistische begrippen uit te leggen, maar ook door je met behulp van opdrachten te stimuleren zelfstandig met de stof om te gaan.

Meer concreet heeft de serie de volgende kenmerken. Elke analyse wordt behandeld volgens dezelfde structuur, het ‘elementaire rapport’. Dat maakt het leren veel gemakkelijker. Je wordt systematisch getraind op: datadesign en vraagstelling; betekenis en formulering van de conclusies; causale interpretaties en hun beperkingen. Je leert daarmee niet alleen het ‘hoe’ maar ook het ‘wanneer’, ‘welke’ en ‘waarom’ van de analyses op een praktische manier. Bij veel analyses wordt uitgebreid aandacht besteed aan visualiseren: het ‘lezen’ en inzichtsmatig redeneren met figuren. De stof begint extreem gemakkelijk (hoe bereken je een gemiddelde) maar eindigt op hoog niveau (dubbel multivariate repeated-measures ANOVA). In de eerste twee delen wordt uitvoerig stilgestaan bij de beperkingen van het intuïtieve statistische denken. Er is ook veel aandacht voor praktische regels. De stof wordt uitgelegd met veel gevarieerde voorbeelden, van het dagelijks leven tot gepubliceerd onderzoek. Met name onderzoeken die prototypisch zijn voor een bepaald gebied worden gebruikt. Tot slot is de stof vijf jaar lang getest en verbeterd naar aanleiding van vragen, problemen en suggesties van studenten van de opleiding Psychologie van de RU.

Naast de serie is in januari 2004 een site geopend waar docenten en studenten meer informatie kunnen vinden, zoals meer tentamenvragen, een informatiefunctie voor docenten, oplossingen van de opgaven in de boeken. Raadpleeg tegen die tijd www.uitgeverijboom.nl voor meer informatie.

Bij deze wil ik Maurits Geerts bedanken voor zijn intelligente bijdragen aan de eerste versies van de delen 1, 2 en 4 in 1996 / 1997. Tevens wil ik Dick Willems bedanken voor zijn gedegen bijdragen aan de delen 2 en 3 in dezelfde jaren. Mede dankzij jullie scherpe inzicht en gevoel voor taal zijn deze boeken prettig leesbaar geworden. En dan heb ik het nog niet over jullie praktische hulp bij het zoeken en beschrijven van voorbeelden en de gevechten met onwillige tekstverwerkers.

Voorts bedank ik de vele assistenten en studenten die met hun opmerkingen en suggesties in de afgelopen jaren hebben bijgedragen aan de ontwikkeling van deze serie.

Nijmegen, januari 2004

In deel A zijn enkele correcties verwerkt. In deel B zijn diverse gedeelten herschreven, toegevoegd, of verwijderd.

Verwijderd zijn de gedeelten over de kans dat een foutmarge overschreden wordt. Dit had weinig praktische toepassingen en het riep meer vragen op dan het aan inzicht verschaftte.

Toegevoegd zijn de paragrafen over questionable research practices, de NHST-controverse, en Fisher versus Neyman-Pearson. In de paragraaf over de NHST-controverse wordt ook het verschil tussen $P(D|H)$ en $P(H|D)$ uitgelegd. Opgaven 26 tot en met 32 zijn toegevoegd.

Herschreven zijn de volgende gedeelten. De inleiding op steekproeventheorie in 13.2 is uitgebreid. De gedeelten over betrouwbaarheidsintervallen hebben aparte paragrafen gekregen, met meer uitleg en meer voorbeelden, vooral voor t . De paragraaf over power is aanzienlijk uitgebreid, met meer uitleg en meer voorbeelden. De eerste zelftoets is vereenvoudigd.

Nijmegen, augustus 2013

Deel A Toetsen voor twee gemiddelden

Leerdoelen deel A

- Onderscheid kunnen maken tussen een within-subject- en een between-subjectdesign, alsmede de daarvoor geëigende controlemaatregelen en analyses.
- Systematisch kunnen rapporteren over het verschil tussen twee gemiddelden bij een within-subjectdesign, inclusief een causale interpretatie.
- Systematisch kunnen rapporteren over het verschil tussen twee gemiddelden bij een between-subjectdesign, inclusief een causale interpretatie.
- Systematisch kunnen rapporteren over het verschil tussen twee gemiddelden bij een mixed design, inclusief de primaire causale verklaring.
- Op grond van een grafische weergave (spreidingsdiagram) conclusies kunnen trekken over diverse statistische maten.
- De bekende beperkingen van het intuïtief menselijk denken bij causale attributie kunnen onderbouwen en toepassen.

1 Elementair rapport van verschil in gemiddelden

1.1 Inleiding

Achtergrond

Veel onderzoeksvragen gaan over de vergelijking van twee gemiddelden. Gaan vrouwen gemiddeld vaker naar de dokter dan mannen? Hebben mensen bij volle maan gemiddeld vaker seks dan bij nieuwe maan? Dat zijn vragen die veel interessanter zijn dan het vaststellen van een enkel gemiddelde. Zeg nou zelf.

Het vergelijken van gemiddelden kan op **twee manieren**. Dat hangt af van de manier waarop de data zijn verzameld, zoals we zullen zien. Er bestaan dan ook twee verschillende statistische procedures voor het vergelijken van gemiddelden. Je moet kunnen beargumenteren of je de ene, dan wel de andere procedure gebruikt.

Voorts zullen we ons in dit onderdeel voor het eerst bezig gaan houden met de vraag of de conclusies generaliseerbaar zijn naar een **populatie**. Dit betekent dat we rekening zullen houden met de grootte van de steekproef. De procedures waarmee je dat doet, heten significantietoetsen.

Ook zullen we ons meer bezighouden met de interpretatie van gegevens. De vraag is daarbij door welke **oorzaak-gevolgrelaties** de resultaten kunnen worden verklaard.

Doel

Na bestudering van hoofdstuk 1 tot en met 8 kun je een elementair rapport maken voor de vergelijking van twee gemiddelden.

1.2 Definitie van het elementaire rapport

Een elementair rapport van het vergelijken van twee gemiddelden dient het volgende in te houden.

Typering van het onderzoek

- *datadesign*: within-subjectdesign of between-subjectdesign; afhankelijke variabele en onafhankelijke variabele; hun meetniveaus (kwantitatief of kwalitatief)
- *mate van controle*: experiment of niet-experiment (passief-observerend)

- *nulhypothese en alternatieve hypothese*: eenzijdig of tweezijdig; geformuleerd in statistische termen (met gebruik van ' μ ')

Analyse van de data

- *de gekozen analyseprocedure*: *t*-procedure voor gekoppelde paren of *t*-procedure voor onafhankelijke steekproeven
- *geaggregeerde data*: gemiddelde, standaardafwijking en *N*; van de verschillcores of per groep
- *schatting van het effect*: ruw effect en relatief effect
- *houdbaarheid van de nulhypothese*: *t*-waarde, aantal vrijheidsgraden en *p*-waarde

Conclusies

- *beslissing omtrent de hypothesen*: de nulhypothese behouden of de nulhypothese verwerpen; formulering in termen van het onderzoek
- *causale interpretatie*: exclusiviteit van de verklaring voor het gevonden resultaat; de primaire verklaring; zo mogelijk een alternatieve verklaring

Indien gevraagd ook nog

- *controle op assumpties*: beoordeling of aan de voorwaarden van de toets is voldaan, en eventuele consequenties voor de correctheid van de *p*-waarde

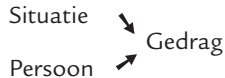
Deze zaken zullen nu worden behandeld.

2 Datadesign

2.1 Inleiding

Achtergrond

Menselijk gedrag wordt in belangrijke mate bepaald door zowel de **situatie** als de **persoon** (zie figuur 2.1).



Figuur 2.1

Om te beginnen wordt gedrag dus in hoge mate beïnvloed door de situatie waarin de persoon verkeert. Daarom wil men vaak twee situaties met elkaar vergelijken, bijvoorbeeld volle maan en nieuwe maan. De vraag is daarbij of het gedrag in de ene situatie systematisch verschilt van het gedrag in de andere situatie.

Voorts wordt gedrag in belangrijke mate beïnvloed door de personeigenschappen van de mensen die het gedrag vertonen. Daarom wil men vaak twee groepen personen met elkaar vergelijken, bijvoorbeeld mannen en vrouwen. De vraag is daarbij of het gedrag van de ene groep systematisch verschilt van het gedrag van de andere groep.

Deze twee typen vragen leiden tot twee verschillende manieren om data te verzamelen. Het within-subjectdesign gebruiken we om verschillende situaties met elkaar te vergelijken. Het between-subjectdesign gebruiken we om verschillende groepen personen te vergelijken.

Doel

Na bestudering van dit hoofdstuk kun je uit de beschrijving van een onderzoek opmaken of het een within- dan wel een between-subjectdesign heeft. In beide gevallen kun je aangeven wat de afhankelijke en de onafhankelijke variabele zijn, welk meetniveau zij hebben, en welke niveaus zij bevatten.

2.2 Samenvatting

Het datadesign heeft betrekking op de manier waarop de data worden verzameld. Bij de specificatie van het design geef je aan:

- of het een within-subjectdesign dan wel een between-subjectdesign is;
- wat de afhankelijke variabele is, en welk meetniveau die heeft;
- wat de onafhankelijke variabele is, alsmede het meetniveau en de niveaus daarvan.

Bij een within-subjectdesign wordt ieder subject herhaaldelijk gemeten, in verschillende situaties. Bij een between-subjectdesign worden de subjecten in meerdere groepen verdeeld en ieder subject wordt eenmaal gemeten. Het within-subjectdesign heeft in de eerste plaats betrekking op verschillen tussen situaties, het between-subjectdesign heeft in de eerste plaats betrekking op verschillen tussen personen. In de praktijk kan dit laatste onderscheid echter door elkaar lopen. De afhankelijke variabele bevat de scores waarvan het gemiddelde moet worden berekend. De onafhankelijke variabele geeft de conditie of groep aan.

2.3 Het within-subjectdesign: het vergelijken van situaties

In een within-subjectdesign wordt iedere persoon op twee tijdstippen geobserveerd of getest. Eerst in de ene situatie en dan in de andere situatie. De data bestaan dus uit **twee scores** per persoon. Daarna vergelijken we het gemiddelde van de ene situatie met het gemiddelde van de andere situatie. Hier is ‘binnen’ de gegevens van een persoon sprake van twee verschillende scores. Andere veelgebruikte namen voor dit design zijn: gepaarde waarnemingen, gekoppelde waarnemingen, gekoppelde paren, gepaarde steekproeven, gecorreleerde steekproeven, afhankelijke steekproeven.

Voorbeelden

1 Stemmingmakerij

In het stemmingmakerijonderzoek zijn voor iedere persoon twee scores voor depressiviteit (BDI) verkregen. De eerste score is de depressiviteit vóór de cursus (Bdv). De tweede score is de depressiviteit na de cursus (Bdn). Je kan je nu afvragen of het gemiddelde van Bdn kleiner is dan het gemiddelde van Bdv. Dat zou betekenen dat de proefpersonen in de loop van de cursus minder depressief zijn geworden.

2 Maakt het weekend werkende mensen gelukkig?

Je laat een stel personen die allen een baan hebben, twee keer een testje invullen. De test meet hoe prettig zij zich voelen op de dag van afname. De eerste keer moeten zij de test invullen op een woensdag, de tweede keer op een zaterdag. De con-

clusie zou kunnen zijn dat mensen met een baan zich op zaterdag gemiddeld prettiger voelen dan op woensdag. (Wat volgens mij niet vanzelfsprekend is, gezien het Nederlandse arbeidsethos).

2.4 Het between-subjectdesign: het vergelijken van groepen personen

In een between-subjectdesign wordt de totale groep personen in twee groepen gesplitst. De personen van beide groepen worden getest of geobserveerd. De data bestaan uit een groepsnummer en één score per persoon. Daarna vergelijken we het gemiddelde van de ene groep met het gemiddelde van de andere groep. Dit betreft een vergelijking ‘tussen’ verschillende personen. Dit design noemt men ook wel onafhankelijke steekproeven of ongecorrleerde steekproeven.

Voorbeelden

1 Stemmingmakerij

In het stemmingmakerijonderzoek kun je een onderscheid maken tussen ‘Deelnemers’ en ‘Niet-deelnemers’. Dat zijn dus twee groepen. Voor al deze personen is de depressiviteit na de cursus (Bdn) gemeten. Je kan je nu afvragen of het gemiddelde van de Deelnemers op Bdn kleiner is dan het gemiddelde van de Niet-deelnemers. Dat zou betekenen dat Deelnemers na de cursus minder depressief zijn dan Niet-deelnemers.

2 Gaan vrouwen vaker naar de film dan mannen?

Je vraagt een groep personen:

- 1 of zij man of vrouw zijn, en
- 2 hoeveel films zij het afgelopen jaar in de bioscoop hebben gezien.

De conclusie zou kunnen zijn dat vrouwen gemiddeld vaker naar de film gaan dan mannen.

2.5 Vergelijking van de twee datadesigns

Veel studenten hebben in het begin moeite om de twee datadesigns – within-subject en between-subject – uit elkaar te houden. Toch is dat hier, en ook in de volgende boeken van deze serie, erg belangrijk. Het datadesign heeft namelijk gevolgen voor de statistische procedure die je dient te volgen, zoals we zullen zien. Het helpt je misschien om de datamatrices te visualiseren. Bestudeer daarom nu het verschil in structuur tussen tabel 2.1 en tabel 2.2.

Tabel 2.1 Datamatrix bij een within-subject onderzoek

<i>persoon</i>	<i>meting in situatie A</i>	<i>meting in situatie B</i>
1	66	25
2	82	36
3	73	84
4	19	52
5	22	16

Tabel 2.2 Datamatrix bij een between-subject onderzoek

<i>persoon</i>	<i>groep</i>	<i>meting</i>
1	A	66
2	A	82
3	A	73
4	A	19
5	A	22
6	B	25
7	B	36
8	B	84
9	B	52
10	B	16

Er zijn drie klassieke gevallen waarbij de twee datadesigns vaak worden verward. In het eerste geval wordt men door de **presentatie** van de data op het verkeerde been gezet. Zo is het bij een between-subjectdesign niet ongebruikelijk om de scores van de verschillende groepen naast elkaar te zetten. De laatste datamatrix, tabel 2.2, wordt dan gepresenteerd zoals in tabel 2.3.

Tabel 2.3

<i>groep A</i>	<i>groep B</i>
66	25
82	36
73	84
19	52
22	16

Oppervlakkig gezien lijkt dit op een within-subjectdatamatrix, maar in feite is het between-subject. De verwarring ontstaat doordat de data niet in de standaardvorm van een datamatrix (subjecten bij variabelen) zijn opgeschreven.

In het tweede geval wordt men door het begrip ‘subject’ op het verkeerde been gezet. Je moet je realiseren dat het woord subject in principe een algemene betekenis heeft. Een subject hoeft niet per se een persoon te zijn. Stel bijvoorbeeld dat we onderzoek doen naar huwelijksatisfactie, en bij een steekproef van echtparen aan zowel de man als de vrouw vragen in hoeverre de betrokkene tevreden is met de huishoudelijke verrichtingen van de partner. We krijgen dan voor elk echtpaar twee scores: één van de man, en één van de vrouw. Je zou nu kunnen denken dat dit een between-subjectdesign is, omdat er twee groepen personen zijn: mannen en vrouwen. Deze scores zijn echter paarsgewijs aan elkaar gekoppeld. De twee scores van een echtpaar horen steeds bij elkaar. We doen er dan beter aan hier het echtpaar als ‘subject’ te beschouwen. Er is daarom sprake van een within-subjectdesign. Kortom, je moet twee dingen goed onderscheiden:

- 1 Een datadesign met mannen en vrouwen, waarbij de mannen niets met de vrouwen hebben te maken. Dat is een between-subjectdesign.
- 2 Een datadesign met mannen en vrouwen waarbij iedere man aan precies één vrouw is gekoppeld in het selectieproces. Dat is een within-subjectdesign.

In het derde geval wordt men door de **formulering** van de vraagstelling op het verkeerde been gezet. Dat kan bijvoorbeeld gebeuren als de steekproef bestaat uit mannen en vrouwen die met elkaar getrouwd zijn, en als daarbij de vraag is: ‘Onderzoek of getrouwde mannen gemiddeld meer tevreden zijn dan getrouwde vrouwen.’ Sommige mensen denken nu dat uit deze vraagstelling volgt dat je de data als een between-subjectdesign moet behandelen, omdat er in de vraag niet wordt gezegd dat je moet onderzoeken of de mannen tevredener zijn dan *hun* vrouwen. Maar dat maakt niet uit. Bij deze data is het gemiddelde van *de* vrouwen gelijk aan het gemiddelde van *hun* vrouwen, dus beide manieren om de vraag te formuleren betekenen hetzelfde. Het gaat om de manier waarop de data verkregen zijn: zodra vaststaat welke mannen in de steekproef zitten, staat ook vast welke vrouwen erin zitten. Dat is een within-subjectdesign, ongeacht de formulering van de vraag.

Tot slot: De bovenstaande inleiding heeft wellicht nog een derde bron van verwarring gecreëerd. Daar werd het within-subjectdesign gekoppeld aan verschillen tussen situaties, en het between-subjectdesign werd gekoppeld aan verschillen tussen personen. Dit was slechts ter inleiding. In de praktijk loopt het onderscheid tussen ‘situaties’ en ‘personen’ vaak *door elkaar*. In twee verschillende situaties vind je namelijk ook vaak verschillende personen. In die gevallen kan een between-subjectdesign soms door de het woordgebruik worden gepresenteerd als een onderzoek naar het verschil tussen situaties. Soms is dat terecht, soms niet.

Voorbeelden

1 Ratten

Stel, je doet een experiment waarbij je één groep ratten beloont in 80% van de keren dat zij een correcte respons vertonen. Een andere groep ratten beloon je in

slechts 50% van de keren dat zij een correcte respons vertonen. De ratten worden at random in de ene, dan wel in de andere groep geplaatst. Na verloop van tijd meet je in beide groepen het gemiddeld aantal correcte responsen. Je zult vinden dat dit gemiddelde groter is in de groep die voor 80% werd beloond. Dit verschil mag je dan toeschrijven aan het verschil in situatie (het beloningspercentage). Immers, door de random toewijzing van de ratten kunnen er geen systematische ‘persoonsverschillen’ bestaan tussen de ratten van de ene en de andere groep. In dit geval is het dus gerechtvaardigd om uit een between-subjectdesign een conclusie te trekken over een verschil tussen situaties.

2 Televisieprogramma’s

Stel, je wilt weten welk programma meer gewaardeerd wordt: *GTST* of *EenVandaag*. Daartoe vraag je een steekproef mensen die in de afgelopen maand *GTST* hebben gezien om hun waardering voor dit programma te geven met een rapportcijfer van 1 tot 10. Tevens vraag je een steekproef mensen die in de afgelopen maand *EenVandaag* hebben gezien om hun waardering voor dat programma te geven. Vervolgens vergelijk je de gemiddelde cijfers met elkaar. Veel mensen zullen geneigd zijn dit onderzoek te beschrijven als een onderzoek naar een verschil tussen twee situaties: het ene versus het andere programma. In feite vergelijk je echter twee groepen personen met elkaar: de *GTST*-kijkers met de *EenVandaag*-kijkers. Er is dus sprake van een between-subjectdesign.

En laten we wel zijn, het is natuurlijk best mogelijk dat een eventueel verschil in gemiddelde niet door het verschil in programma wordt veroorzaakt, maar doordat bijvoorbeeld de *EenVandaag*-kijkers kritischer zijn en daarom sowieso aan elk programma een lager cijfer geven dan de *GTST*-kijkers. In dit geval is het dus niet zonder meer gerechtvaardigd om uit het between-subjectdesign een conclusie te trekken over het verschil tussen de situaties.

2.6 Afhankelijke en onafhankelijke variabele

In deel 1B van deze serie hebben we gesteld: De afhankelijke variabele is de variabele waarvan wordt vermoed dat hij het ‘gevolg’ is. De onafhankelijke variabele is de variabele waarvan wordt vermoed dat hij de ‘oorzaak’ is. Dat is nog steeds geldig. De analyseprocedures van het huidige deel (2A) zijn echter alleen geschikt voor een afhankelijke variabele die kwantitatief is en een onafhankelijke variabele die kwalitatief is.

Wat moet je doen in de omgekeerde situatie? De analyseprocedure daarvoor is logistische regressie, maar in situaties met een enkele onafhankelijke variabele is het gebruikelijk (en toegestaan) om de afhankelijke en onafhankelijke variabele om te wisselen. Dat betekent: de afhankelijke variabele moet kwantitatief zijn en de onafhankelijke variabele moet kwalitatief zijn.

Beschrijf de variabelen met **duidelijke namen** (dus niet ‘score’ of zoiets). Beschrijf daarom bij een kwalitatieve variabele ook altijd de **niveaus** (categorieën). Bijvoorbeeld, de variabele Geslacht heeft de niveaus ‘mannelijk’ en ‘vrouwelijk’. Als je dat er niet bijschrijft, zou de lezer kunnen denken dat het gaat om de geslachten *Panthera* en *Leopardus* in de familie der katachtigen.

2.6.1 Afhankelijke en onafhankelijke variabele in een between-subjectdesign

Een variabele is een kolom (loopt verticaal) in de datamatrix. Bij een between-subjectdesign zijn er twee variabelen in het spel (bijvoorbeeld Groep en Bdn). De ene variabele bevat de scores waarvan het gemiddelde moet worden berekend (Bdn). Dit is de afhankelijke variabele en hij is kwantitatief. De tweede variabele geeft aan in welke groep elk subject zit (Groep). Dit is de onafhankelijke variabele en hij is kwalitatief. Dat noemt men een **between-subjectfactor**.

2.6.2 Afhankelijke en onafhankelijke variabele in een within-subjectdesign

Bij een within-subjectdesign zijn er twee kwantitatieve variabelen waarvan het gemiddelde wordt berekend (bijvoorbeeld Bdv en Bdn). Beide dienen te worden gezien als een meting van de afhankelijke variabele (depressiviteit-volgens-Beck) en deze is kwantitatief. De onafhankelijke variabele geeft aan in welke conditie elke score is verkregen (conditie: voor of na de behandeling). Dit is de onafhankelijke variabele en hij is kwalitatief. Dat noemt men een **within-subjectfactor**.

We wijken hier op twee punten af van de eerder gemaakte afspraak dat een variabele een kolom in de datamatrix is. Ten eerste bevat ‘de afhankelijke variabele’ nu twee kolommen (Bdv en Bdn). Oké, bij deze breiden we de definitie gewoon uit: een variabele mag ook een verzameling kolommen zijn. Ten tweede is ‘de onafhankelijke variabele’ niet als een kolom scores in de datamatrix te vinden. Het is eerder een *rij* die je boven de datamatrix kan schrijven met de kopjes ‘voor’ en ‘na’.

Voorbeelden

1 Stemmingmakerij (1)

Als we in het onderzoek ‘Stemmingmakerij’ de Deelnemers met de Niet-deelnemers vergelijken op Bdn, dan krijgen we het volgende design:

Design:	between-subject	
	afhankelijke variabele	= Bdn (kwantitatief)
	onafhankelijke variabele	= Groep (kwalitatief: deelnemer, niet-deelnemer)

2 Stemmingmakerij (2)

Als we in het onderzoek ‘Stemmingmakerij’ de voormeting met de nameting vergelijken, dan is het design:

Design: within-subject
afhankelijke variabele = Depressiviteit = {Bdv, Bdn}
(kwantitatief)
onafhankelijke variabele = Tijd (kwalitatief: voor, na)

3 Maakt het weekend werkende mensen gelukkig?

Als je mensen op woensdag en zaterdag laat invullen hoe prettig zij zich voelen, resulterend in de variabelen Wpret en Zpret:

Design: within-subject
afhankelijke variabele = Hoeveelheid pret = {Wpret,
Zpret} (kwantitatief)
onafhankelijke variabele = Dag (kwalitatief: woensdag,
zaterdag)