

STATISTIEK VOOR DE PSYCHOLOGIE

**DEEL 1:** VERDELING VAN EEN VARIABLE,  
SAMENHANG TUSSEN TWEE VARIABLEN

# Statistiek voor de psychologie

*Deel 1: Verdeling van een variabele, samenhang tussen twee variabelen*

Jules L. Ellis

Tweede druk

Boom Lemma uitgevers  
Den Haag  
2014

# Inhoud

	Voorwoord voor docenten	11
	Voorwoord voor studenten: relevantie van het vak	18
Deel A	Verdeling van een variabele	23
HOOFDSTUK 1	Observatie, vragenlijst, test	24
1.1	Inleiding	24
1.2	Methoden om gegevens te verzamelen	24
1.3	Observatie	25
1.3.1	Datamatrix	25
1.4	Vragenlijsten	26
1.4.1	Datamatrix	27
1.5	Psychologische tests	28
1.5.1	Zijn tests wel te vertrouwen?	29
1.5.2	Datamatrix	30
1.6	Het meetniveau van een variabele	34
HOOFDSTUK 2	Elementair rapport van een variabele	35
2.1	Inleiding	35
2.2	Het elementair rapport	35
2.3	De frequentieverdeling, het histogram, en 'N'	36
2.4	Het vijfgetallenresumé, uitschieters en het gemodificeerde boxdiagram	40
2.4.1	Het vijfgetallenresumé	40
2.4.2	Uitschieters	42
2.4.3	Het gemodificeerde boxdiagram	42
2.5	Gemiddelde en standaardafwijking	45
2.5.1	Berekening	45
2.6	Indicatie van normaliteit	47
2.7	Samenvatting: het elementair rapport van Bdv	50

---

HOOFDSTUK 3	Normeren	53
3.1	Inleiding	53
3.2	Percentielscores	54
3.2.1	Berekening van percentielscores	54
3.3	Standaardscores	57
3.3.1	Berekening van standaardscores	57
3.4	Normaalscores	59
3.4.1	Berekening van normaalscores	59
3.5	Vergelijking van ruwe, standaard-, percentiel- en normaalscores	60
HOOFDSTUK 4	Normaalverdelingen	64
4.1	Inleiding	64
4.2	Notatie	65
4.3	Eigenschappen van normaalverdelingen	66
4.4	De standaard-normaalverdeling	69
4.5	Rekenen met normaalverdelingen	69
4.6	Enkele belangrijke getallen van een normaalverdeling	74
HOOFDSTUK 5	De intuïtieve psycholoog en steekproeven	75
5.1	Inleiding	75
5.2	Ongevoeligheid voor steekproefgrootte	77
5.2.1	De statistische norm over steekproefgrootte	78
5.2.2	Beperking van het intuïtief denken met steekproefgrootte	80
5.2.3	Onderzoeken over ongevoeligheid voor steekproefgrootte	82
5.2.4	Het dagelijks leven: voorbeelden van ongevoeligheid voor steekproefgrootte	82
5.3	Ongevoeligheid voor steekproefbias	83
5.3.1	De statistische norm over steekproefbias	84
5.3.2	Beperking van het intuïtief denken bij steekproefbias	87
5.3.3	Onderzoeken over ongevoeligheid voor steekproefbias	89
5.3.4	Het dagelijks leven: voorbeelden van ongevoeligheid voor steekproefbias	92
HOOFDSTUK 6	Visualiseren	94
6.1	Inleiding	94
6.2	Het lezen van een histogram	94
HOOFDSTUK 7	Opgaven deel A	98

HOOFDSTUK 8	Leerdoelen en zelftoetsen deel A	116
8.1	Leerdoelen	116
8.2	Zelftoets A1	119
8.3	Uitwerkingen van zelftoets A1	123
8.4	Zelftoets A2	126
8.5	Uitwerkingen van zelftoets A2	131
Deel B	Samenhang tussen twee variabelen	135
HOOFDSTUK 9	Samenhang tussen twee kwantitatieve variabelen	136
9.1	Inleiding	136
9.2	Het elementaire rapport	137
9.3	Het design	137
9.4	Spreidingsdiagram	139
9.5	Correlatie, hellingscoëfficiënt en intercept	140
9.5.1	Berekening	141
9.6	Tekenen van de regressielijn	142
9.7	Indicatie van lineariteit	144
9.8	Invloedrijke waarnemingen	148
9.9	Samenvatting: elementair rapport van samenhang	150
HOOFDSTUK 10	Voorspelde scores en residuen	151
10.1	Inleiding	151
10.2	Voorspelde scores en residuen	151
HOOFDSTUK 11	De correlatiecoëfficiënt	153
11.1	Inleiding	153
11.2	Correlatie en spreidingsdiagram	153
11.3	Correlatie en gemiddelden en variantie	156
11.4	Het verband tussen correlatie en regressie	157
11.5	Het verschil tussen correlatie en regressie	158
11.6	Variantie van voorspelde scores en residuen	158
11.7	Correlatie en proportie verklaarde variantie	161
11.8	Correlatie en causaliteit	163
HOOFDSTUK 12	Visualiseren	165
12.1	Inleiding	165
12.2	Schatten van regressie en correlatie	165
12.3	Het regressie-effect	166
12.4	Bepalen van lineariteit	168
12.5	Geschatte correlatie versus echte correlatie	170

HOOFDSTUK 13	Samenhang tussen twee kwalitatieve variabelen	172
13.1	Inleiding	172
13.2	Elementair rapport	172
13.3	Voorbeeld van data en vraagstelling	172
13.4	Design	173
13.5	De kruistabel	173
13.6	De voorwaardelijke verdelingen	174
13.7	Het gesegmenteerde staafdiagram	175
13.8	Beoordeling onafhankelijkheid	175
13.9	Samenvatting: elementair rapport van samenhang	176
HOOFDSTUK 14	Suppressie en de paradox van Simpson	178
14.1	Inleiding	178
14.2	De paradox van Simpson	178
14.3	De paradox van Simpson bij kwantitatieve variabelen	179
14.4	De paradox van Simpson bij kwalitatieve variabelen	180
14.5	Processen die de paradox van Simpson veroorzaken	183
HOOFDSTUK 15	De intuïtieve psycholoog en samenhang	185
15.1	Inleiding	185
15.2	Aandacht voor de present-presentcel	185
15.2.1	De statistische norm over samenhang in kruistabellen	186
15.2.2	Beperking van het intuïtief denken over samenhang in kruistabellen	186
15.2.3	Het dagelijks leven: Voorbeelden van aandacht voor de present-presentcel	187
15.3	Illusoire correlatie	190
15.3.1	De statistische norm over het bepalen van een correlatie	190
15.3.2	Beperking van het intuïtief denken bij het bepalen van een correlatie	190
15.3.3	Onderzoeken over illusoire correlaties	192
15.3.4	Het dagelijks leven: Voorbeelden van illusoire correlaties	194
HOOFDSTUK 16	Opgaven deel B	197
HOOFDSTUK 17	Leerdoelen en zelftoetsen deel B	212
17.1	Leerdoelen	212
17.2	Zelftoets B1	215
17.3	Uitwerkingen van zelftoets B1	221
17.4	Zelftoets B2	226
17.5	Uitwerkingen van zelftoets B2	231

Appendix	235
A.1 De BDI en de ATQ	235
A.2 De formule voor de standaardafwijking	241
A.2.1 De standaardafwijking van een populatie	241
A.2.2 De standaardafwijking van een steekproef	242
A.3 De formule voor de correlatie	243
A.4 Tabel A: de standaard-normaalverdeling	244
Referenties	246
Register	249

# Voorwoord voor docenten

Er zijn tientallen inleidende statistiekboeken voor het Nederlandse taalgebied. Moet er echt nog een bijkomen? Wat denk ik nog toe te voegen? Wel, diverse unieke kenmerken, en die zet ik hieronder uiteen.

*Structuur.* Alle analyses worden behandeld met globaal dezelfde structuur, die hier het ‘elementair rapport’ wordt genoemd. Hiermee onderscheidt de huidige serie zich van alle boeken die ik ken. Het zoeken van de structuur in de stof is iets waar de student normaal gesproken een groot deel van de tijd mee kwijt is. Dat wordt in deze serie aanzienlijk eenvoudiger gemaakt door steeds ongeveer dezelfde structuur te gebruiken.

*Design en analyse.* Er wordt systematisch en doorlopend aandacht besteed aan het kiezen van de juiste analyse bij een vraagstelling. Daartoe moet men in de eerste stap van het elementaire rapport het ‘design’ beschrijven: afhankelijke variabelen, onafhankelijke variabelen, hun meetniveaus, en de within- en between-subject-structuur. In een latere stap van het elementaire rapport moet dan de analyse worden gekozen. Ook zijn er opgaven die speciaal zijn gericht op het oefenen in het specificeren van het design en het kiezen van de analyse.

Vaak ziet men gevorderde studenten die geen flauw idee hebben welke analyses zij moeten doen met hun data. Het probleem zit vaak al een stap eerder, namelijk dat zij niet op een rijtje hebben welke variabelen er zijn en wat hun rol is. Door altijd eerst het design te specificeren wordt het kiezen van de analyse een stuk gemakkelijker. Deze stap wordt in traditionele statistiekboeken echter verwaarloosd, omdat hij zo gemakkelijk lijkt. Maar deze stap is, zelfs als hij niet moeilijk is, wel cruciaal om de volgende stap te kunnen maken: het kiezen van de analyse. Daarnaast blijkt dat studenten met deze stap toch vaker fouten maken dan men in eerste instantie misschien zou denken. De concepten zijn weliswaar eenvoudig, maar de toepassing daarvan blijkt voor een ongetrainde persoon toch enige haken en ogen te hebben.

*Formuleren van conclusies.* Er wordt systematisch en doorlopend aandacht besteed aan nauwkeurige formuleren van conclusies. De meeste andere boeken behandelen dit slechts terloops of in een aparte sectie die in nogal vrijblijvende toon is geschreven, zodat geen enkele student het serieus neemt. Dat merkt men dan aan de ver-



ontrustende ongenueanceerdheid waarmee studenten op tentamens hun conclusies formuleren. Dat verandert niet als de docent zich aan de koffietafel beklagt over hoe dom de studenten dit jaar weer zijn. Het enige dat helpt is: die studenten leren hoe ze een conclusie moeten opschrijven. In deze serie wordt de student bij elke analyse weer opnieuw verplicht om de conclusies volgens een vaststaand format op te schrijven. Daartoe moet men in de laatste twee stappen van het elementaire rapport de ‘beslissingen’ en de ‘causale interpretatie’ formuleren. Ook zijn er opgaven die speciaal zijn gericht op het oefenen in het formuleren van de conclusies.

Het *belang* van nauwkeurig formuleren wordt vaak onderschat door studenten. Zij zijn vaak al blij als de berekeningen goed zijn. Maar het trekken van een conclusie is, zelfs als het niet moeilijk is, wel waar het om gaat in de analyse. Daarnaast leidt het verwaarlozen hiervan ertoe dat studenten het gevoel krijgen niet te weten waar ze mee bezig zijn. Dat is funest voor hun motivatie.

De *moeilijkheid* van het formuleren wordt volgens mij vaak onderschat door zowel studenten als docenten. Het wordt als een vanzelfsprekendheid gezien. Maar het blijkt dat studenten hier in eerste instantie veel fouten in maken. Zelfs studenten die zeggen te snappen waar ze mee bezig zijn en die alle berekeningen goed hebben. Zij weten wel *ongeveer* wat de uitkomsten betekenen, maar kunnen het vaak nog niet *precies* opschrijven. Omdat ze denken dat het formuleren heel eenvoudig is en dat het eigenlijk geen statistiek is, gaan ze ervan uit dat het toch wel lukt, en doen geen poging de conclusies daadwerkelijk op te schrijven. Of ze schrijven de conclusies wel op maar gaan ervan uit dat hun antwoord goed is en vragen daarom geen feedback. Daardoor komen deze studenten er niet achter dat hun formulering gebrekkig is en onaanvaardbaar is volgens de wetenschappelijke normen.

*Causaliteit.* Er wordt systematisch en doorlopend aandacht besteed aan causaliteit. De meeste andere boeken behandelen ook dit onderwerp slechts terloops of in een aparte sectie die in nogal vrijblijvende toon is geschreven. Het gevolg daarvan is dat studenten geneigd zijn de meest voor de hand liggende causale verklaring aan te zien voor een feitelijke constatering, zonder de vraag te stellen of er nog alternatieve verklaringen zijn. In de meeste boeken wordt daar weliswaar tegen gewaarschuwd, maar vervolgens doen studenten dat massaal toch, zonder dat zij daarin gecorrigeerd worden. In deze serie wordt dit ondervangen door de student bij elke analyse weer opnieuw te dwingen een causale interpretatie te geven en daarbij de vraag te stellen of er alternatieve verklaringen zijn. Daartoe moet men in het begin van het elementaire rapport bepalen wat de controlemaatregelen zijn, en in de laatste stap een causale interpretatie geven. Ook zijn er opgaven die speciaal zijn gericht op het oefenen in causale interpretaties.

De moeilijkheid van het geven van causale interpretaties wordt volgens mij eveneens vaak onderschat. De meeste studenten zijn zonder veel moeite in staat om bij een samenhang tussen twee abstracte variabelen X en Y drie mogelijke verklaringen te geven: X beïnvloedt Y, of Y beïnvloedt X, of er is een variabele Z die zowel X als Y beïnvloedt. Dat is waarschijnlijk de reden dat zowel studenten als

docenten niet geneigd zijn er veel aandacht aan te geven. Het probleem is dat zodra het over een concreet onderzoek gaat, bijvoorbeeld neuroticisme en geluk, dat veel studenten dan het algemene principe lijken te vergeten. Zij geven dan nog maar één van de drie verklaringen en stoppen daarna. Bovendien, als er op gewezen wordt dat er nog een storende variabele ( $Z$ ) kan zijn, blijken zij vaak niet in staat daar een plausibele kandidaat voor te bedenken. Nu is dat soms ook moeilijk, maar veel studenten kunnen het zelfs niet in onderzoeken waar de storende variabele heel voor de hand ligt.

*Visualiseren.* Er wordt bij de meeste analyses aandacht besteed aan het trekken van conclusies op grond van een weergave van de data in een figuur. De meeste andere statistiekboeken *gebruiken* figuren om bepaalde principes uit te leggen. Veel auteurs en docenten zijn visueel ingesteld en voor hun maakt een figuur de stof gemakkelijker. Wat zij zich meestal niet realiseren is dat er een grote groep studenten bestaat die juist niet visueel zijn ingesteld. Deze studenten hebben moeite met het 'lezen' van figuren. Voor die studenten maakt het gebruik van een figuur het niet gemakkelijker, maar juist moeilijker. In deze serie wordt bij veel analyses in een aparte sectie uitgelegd hoe de figuur moet worden gelezen en *hoe* men er conclusies uit kan trekken. Ook zijn er visualiseringsopgaven die speciaal zijn gericht op trekken van conclusies uit figuren.

Bij het bespreken van de visualiseringsopgaven met studenten blijkt vaak *waarom* studenten moeite hebben met sommige van die opgaven. De docent ziet de uitkomst vaak in een oogopslag. Maar wanneer de docent gedwongen wordt de redenering expliciet te maken, blijkt deze vaak nog behoorlijk lang te zijn. Geen wonder dat een student dat niet meteen kan! De snelheid waarmee de docent het antwoord 'ziet' is vaak ontstaan door jarenlange training die grote delen van het redeneerproces heeft geautomatiseerd. Daardoor is de docent zich niet meer bewust van de vele stapjes in de redenering en heeft moeite ze uit te leggen. Inderdaad zijn de visualiseringsopgaven bij uitstek ook te zien als inzichtsvragen. Het bijzondere van deze serie boeken is dat niet wordt verondersteld dat zulk inzicht vanzelf komt, maar dat stapje voor stapje wordt geprobeerd zulk 'inzicht' ook bij studenten te kweken.

*De intuïtieve psycholoog.* In de eerste twee delen van deze serie wordt uitgebreid aandacht besteed aan de beperkingen van intuïtieve menselijke redeneringen in vergelijking met formele statistische procedures, zoals onder andere is onderzocht door Kahneman en Tversky (Kahneman & Tversky, 1973; Tversky & Kahneman, 1974). Het is ongebruikelijk om deze stof in een statistiekboek op te nemen, maar er zijn goede redenen voor. Naar mijn mening is het voor een goede academische vorming van de student wenselijk dat docenten van psychologievakken regelmatig statistiek gebruiken in hun cursus, en evenzogoed ook omgekeerd. Immers, juist deze resultaten van de sociale psychologie maken de student duidelijk waarom statistiek zo belangrijk is. Niet omdat een paar ouderwetse strenge mannen het ver-

plicht stellen, of omdat de resultaten dan wat nauwkeuriger zijn; maar omdat men anders systematisch  *totaal*  foute conclusies kan trekken. Daarnaast maken de onderzoeken van Kahneman en Tversky duidelijk dat het voor een goede toepassing van statistiek in alledaagse situaties onvoldoende is om de formele statistische regels te beheersen; ook statistisch goed geschoolde personen maakte deze denkfouten. Het is dus nodig om studenten in het statistiekonderwijs te wijzen op deze valkuilen. De onderzoeken van Kahneman en Tversky waren onder andere geïnspireerd door de fouten die zij studenten elke keer weer zagen maken in het statistiekonderwijs. Waarom zouden we dat niet gebruiken om het onderwijs te verbeteren?

*Formules.* In deze serie worden heel weinig formules gebruikt, en de formules die er zijn worden vaak verbaal geformuleerd. Er zijn meer statistiekboeken met deze eigenschap (bijvoorbeeld Aron & Aron, 1994), maar de meeste boeken gebruiken meer formules. Voor de meeste docenten maken formules het gemakkelijker, en zij zijn dan geneigd de formules uit te leggen met andere formules. Voor veel psychologiestudenten maken formules het daarentegen juist moeilijker, en zo'n uitleg van de docent maakt het dan alleen maar nog moeilijker. In deze serie worden alleen de belangrijkste, centrale formules gegeven. De formules die er zijn worden veelal uitgelegd door te laten zien dat de belangrijkste aspecten van de formule een kwestie van gezond verstand zijn. De betekenis van de uitkomst wordt zo mogelijk in woorden omschreven. Sommige formules worden als rekenprocedure beschreven. Het nut van de formule wordt besproken.

Bijna net zo belangrijk is wat er *niet* wordt uitgelegd. Afleidingen worden vrijwel overal vermeden, omdat zij het voor de meeste studenten, anders dan bij de docenten, obscuur en onbegrijpelijk maken. Daarnaast wordt niet geschroomd om toe te geven dat de specifieke vorm van de formule deels een kwestie van afspraak is (bijvoorbeeld het gebruik van F-verdelingen in plaats van beta-verdelingen), deels redenen heeft die nog te moeilijk zijn en die men maar moet geloven (bijvoorbeeld de optimale power van een t-toets), en deels betrekkelijk willekeurig in de historie is gegroeid (bijvoorbeeld om bij de berekening van een steekproefvariantie te delen door  $N - 1$  in plaats van  $N$ ). Veel statistiekboeken maken in mijn ogen de fout dit soort dingen uit te willen leggen terwijl ze eigenlijk niet uit te leggen zijn. Dan komt er bijvoorbeeld een vaag verhaal over 'vrijheidsgraden' waarbij de student terecht een mystiek gevoel krijgt.

*Keuze van voorbeelden.* Elke analyse wordt uitgelegd met een 'doorlopend voorbeeld' dat het hele hoofdstuk of zelfs meerdere hoofdstukken wordt gebruikt. Dat heeft het evidente voordeel dat de student zich niet steeds opnieuw hoeft in te lezen in het voorbeeld. Deze doorlopende voorbeelden betreffen meestal een werkelijk onderzoek uit de psychologie. Daarbij is geprobeerd bij elke analyse een voorbeeld te kiezen dat goed kan dienen als een 'stereotype' voorbeeld. Ten eerste is dat gedaan door te kiezen voor onderzoeken met variabelen die gemakkelijk te

begrijpen zijn zonder al te veel uitleg. Ten tweede is dat gedaan door steeds een voorbeeld te kiezen uit een deelgebied van de psychologie waarin de betreffende analyse vaak wordt gebruikt: bij 2-factor ANOVA is gekozen voor een experiment uit de sociale psychologie; bij regressie-analyse is gekozen voor een onderzoek uit arbeids- en organisatiepsychologie; bij quasi-experimenten is gekozen voor een voorbeeld uit klinische psychologie.

Naast de doorlopende voorbeelden zijn er ook veel voorbeelden die zijdelings worden gegeven. Die betreffen vaak alledaagse situaties en doen een beroep op het gezonde verstand van de student.

*Praktische regels.* Veel statistiekboeken vertonen koud-watervrees als het erom gaat praktische richtlijnen te geven; dat wordt als onacademisch ervaren. Maar een student moet wel weten wat er op het tentamen ‘goed’ of ‘fout’ zal worden gerekend, en wat in een artikel wel of niet acceptabel is. Neem bijvoorbeeld het beoordelen of een variabele bij benadering normaal verdeeld is. De meeste boeken schrijven alleen dat dit door grafische inspectie moet worden gedaan. Maar waar ligt dan de grens tussen ‘wel’ en ‘niet’ normaal verdeeld? Dat staat er nooit bij. In dit boek wordt geprobeerd om daar eenvoudige richtlijnen voor te geven. Men kan kritiek hebben op de door mij geformuleerde richtlijnen, maar ze zijn tenminste expliciet. Een ander voorbeeld is de keuze van het significantieniveau. Veel boeken doen alsof dit vrijelijk te kiezen is. Inderdaad is er vanuit de theoretische statistiek geen dwingende reden om 5% te kiezen. Maar in psychologisch onderzoek is het de conventie om 5% te kiezen, dus in de praktijk moet een student ook altijd voor 5% kiezen. Dus dat schrijf ik dan op.

*Toepassing voorafgaande aan de theorie.* In de meeste boeken wordt de steekproeven-theorie behandeld voorafgaande aan de t-toetsen. Dat is wiskundig gezien de logische volgorde, maar is het ook didactisch optimaal? Steekproeven-theorie is een abstract onderwerp. Hoe kan men zich hiervoor interesseren en de hoofdzaken van de bijzaken onderscheiden, als men nog geen zicht heeft op de toepassingen ervan? Dat vraagt wel erg veel geduld van de student. Daarom is de volgorde van deze onderwerpen in deel 2 van deze serie omgedraaid. Eerst wordt in deel 2A behandeld hoe men t-toetsen moet uitvoeren. Daarna wordt in deel 2B de steekproeven-theorie behandeld. En dan blijkt dat studenten wel degelijk in staat zijn de toetsen goed uit te voeren en de strekking ervan te begrijpen zonder de finesses van de abstracte theorie te kennen.

Deze zelfde strategie is gevolgd bij veel andere onderwerpen, bijvoorbeeld de behandeling van het begrip ‘interactie’ in 2-factor ANOVA. Veel boeken proberen min of meer gelijktijdig uit te leggen wat interactie is en hoe je een ANOVA-tabel uitrekt. Maar dat is te veel tegelijk. Daarom wordt in deel 3 eerst behandeld hoe de ANOVA wordt uitgevoerd, en wordt pas daarna uitvoerig ingegaan op het begrip interactie.

*Opbouw van eenvoudig naar moeilijk.* De serie begint extreem eenvoudig. Er wordt nog uitgelegd hoe je een gemiddelde uitrekent met een rekenmachine. Veel auteurs gaan ervan uit dat de student dat al kan. Maar de werkelijkheid is dat er studenten zijn die het niet kunnen, of het weer vergeten zijn. Waarom zou je het dan niet uitleggen?

Ondanks dit eenvoudige begin, eindigt de serie op hoog niveau, met MANOVA-varianten zoals dubbel multivariate repeated-measures ANOVA in deel 4. Er zijn maar weinig inleidende statistiekboeken die dat laatste aandurven. Vaak stopt men bij eenvoudigere analyses, of behandelt men MANOVA op zeer technische wijze in een aparte cursus. Dat het in deze serie wel lukt, komt onder andere door de gedoseerde opbouw van de moeilijkheid. Door de aanvankelijke, vrij eenvoudig behaalde succeservaringen bij deel 1 is de student bij deel 2 bereid wat langer te blijven proberen en niet zo snel op te geven. Dat leidt dan weer tot een nieuwe succeservaring, waardoor de frustratiedrempel bij deel 3 nog hoger ligt. Enzovoort.

*Integratie met SPSS.* Er is een aantal boeken die specifiek gericht zijn op het gebruik van SPSS. De meeste van die boeken behandelen alleen de eenvoudige procedures, zoals Select cases en t-toetsen. Verder blijft de statistische theorie in die boeken meestal onderbelicht. Daarnaast zijn er veel statistiekboeken die juist de theorie behandelen, zonder uit te leggen hoe de analyse met SPSS kan worden gedaan. Dit wordt ondervangen in deel 4 van deze serie. Voor praktische oefeningen met geavanceerde analyses (zoals MANOVA) kan men daar terecht. Maar ook in de delen 1 - 3 wordt daar al op ingespeeld door steeds SPSS-termen te gebruiken (bijvoorbeeld 'repeated-measures ANOVA' in plaats van '2-factoriële variantieanalyse zonder herhalingen').

*Behandeling van testtheorie.* In deel 2 is een hoofdstuk opgenomen waarin de relatie wordt gelegd tussen klassieke testtheorie en ANOVA. In de meeste boeken wordt testtheorie niet behandeld, of de evidente relatie met ANOVA wordt niet vermeld. Ook worden in dit hoofdstuk enige begrippen uit de generaliseerbaarheidstheorie behandeld.

*Gebruik van Engels termen.* In sommige statistiekboeken worden krampachtig altijd de Nederlands termen gebruikt (bijvoorbeeld 'kwadratenom' in plaats van 'sum of squares'). Toch zijn de Engelse termen gebruikelijker, ook als Nederlandse onderzoekers onder elkaar praten. Het is nu eenmaal de taal van de wetenschap. Verder zal de student uiteindelijk toch de Engelse termen moeten kennen, bij het lezen van output en internationale literatuur. Daarom is niet geschroomd om Engelse termen te gebruiken. In het bijzonder is ook gekozen voor de gebruikelijke afkorting van sum of squares tot SS, ook al zou dat bij sommige lezers onaangename associaties kunnen oproepen.

*Beperkingen.* Uiteraard is het nodig geweest de onderwerpen in deze serie enigszins te beperken. Hier volgt een lijstje met mogelijke aandachtspunten.

- Bij de ‘mate van controle’ wordt alleen een onderscheid gemaakt tussen ‘experiment’ of ‘geen experiment’. Bij counterbalancen wordt slechts 1 vorm (intra-group counterbalancen) behandeld.
- De student wordt in eerste instantie geleerd bij het formuleren van conclusies van statistische toetsen de term ‘in de populatie’ te gebruiken. In artikelen is dat ongebruikelijk.
- Bij causale interpretaties wordt het gebruik van inhoudelijke theorieën vermeden, hoewel men die in een echt onderzoek natuurlijk ook zou moeten bespreken.
- Het controleren van assumpties van een analyse wordt steeds pas achteraf behandeld. Dit, om niet af te leiden van het hoofddoel van de analyse.
- Om duidelijkheid te scheppen bij de student worden vaak regels geformuleerd, maar sommige studenten raken te veel gefocussed op de regeltjes en verwaarlozen de grote lijn.
- Het tekenen van histogrammen wordt erg simpel gehouden. Veel boeken besteden uitvoerige aandacht aan allerlei regels over klassegrenzen en interpolaties, waar ik het nut niet van inzie.
- Het handmatig berekenen van correlatie wordt niet behandeld. Dit, omdat het veel tijd kost en alleen maar afleidt van het concept.
- Kanstheorie en binomiaalverdelingen worden niet behandeld. Zij zijn op zichzelf niet van belang in de praktijk. Wat een kans is, snapt iemand ook wel zonder de ‘en-of’-regels. En-of-regels komen wel ter sprake bij MANOVA, waar ze veel relevanter zijn. De binomiaalverdeling dient in de meeste boeken als inleiding tot de Centrale Limiet Stelling, maar is op zichzelf voor de meeste studenten zo moeilijk dat het zijn doel voorbij schiet.
- Er wordt geen inleiding in SPSS gegeven. Er zijn vele andere boeken die dat behandelen. In de cursus die ik geef volstaat het om bij het begin van deel 4 een inleidend praktikum te geven van ongeveer twee uur (een praktikumhandleiding met opgaven kan worden gedownload van de internetsite van deze serie boeken). Ook staan er tussen de opgaven van deel 4 steeds aanwijzingen, bijvoorbeeld hoe men Select Cases moet gebruiken.

Op de productpagina van dit boek op [www.boomlemma.nl](http://www.boomlemma.nl) is extra materiaal te vinden.

Nijmegen, december 2003

# Voorwoord voor studenten: relevantie van het vak

Psychologie baseert zich op het gedrag van mensen. Het is dus van belang om met onze eigen ogen na te gaan hoe die mensen zich gedragen. Het probleem waar je dan op stuit is dat mensen een grote verscheidenheid en variatie in gedrag tonen. Meer dan een mens zomaar kan bevatten. Daarom maken onderzoekers in de psychologie gebruik van statistiek. Statistiek helpt bij:

- het systematisch verzamelen en coderen van observaties
- het ordenen en samenvatten van gegevens
- het interpreteren van de resultaten
- communicatie over dit alles met andere onderzoekers.

Kortom, zonder statistiek geen psychologie.

## *Misverstanden over de inhoud van statistiek*

Statistiek heeft tot doel om vragen te beantwoorden met behulp van data. Veel mensen zijn daarvan te weinig doordrongen, en dat doet een aantal misverstanden ontstaan. Voor ik in ga op de aard van die misverstanden, eerst iets over de oorzaak er van. De oorzaak is denk ik dat veel mensen denken dat statistiek een vorm van wiskunde is. Dat is het niet. Statistiek gaat over data. Wiskunde is nou juist de enige wetenschap die geen data gebruikt. Wiskundigen zullen nooit in hun leven data verzamelen. Weliswaar wordt er in de statistiek veel wiskunde gebruikt, maar het is zelf geen wiskunde. Weliswaar wordt statistiek op middelbare scholen door wiskundeleraren gegeven, maar eerlijk gezegd betwijfel ik of dat wel zo verstandig is. Want dat zijn niet de mensen die zelf statistiek gebruiken.

Als gevolg van dit misverstand wordt er soms geklaagd dat sommige opgaven onduidelijk zijn. Dan wordt er bijvoorbeeld gevraagd een numerieke samenvatting te maken, terwijl in dit boek niet staat gedefinieerd wat een numerieke samenvatting is. Dat zou een terechte kritiek zijn als statistiek een vorm van wiskunde was. Maar dat is het niet. Statistiek heeft tot doel om vragen te beantwoorden met behulp van data. Die vragen zijn meestal onduidelijk. Dat is de praktijk in vrijwel elk onderzoek. Daar moet je dus mee om leren gaan. In het bovenstaande voorbeeld moet je dus zelf bedenken wat een numerieke samenvatting is. Zo moeilijk is dat niet: er moeten nummers in staan en het moet een samenvatting zijn. Die samenvatting moet een goede samenvatting zijn, dus er moet antwoord worden gegeven op de belangrijkste vragen die men zou kunnen stellen. Jij moet dus

bedenken welke vragen belangrijk zijn. Bij wiskunde hoefde je dat misschien nooit te doen, *maar hier wel*.

Een ander misverstand is dat het bij statistiek om de getallen en de berekeningen gaat. Dan je zie dat iemand een hoop berekeningen heeft gemaakt en de uitkomst is bijvoorbeeld 42. En dan houdt die persoon op. Maar statistiek heeft tot doel om vragen te beantwoorden met behulp van data. De vraag is bijvoorbeeld: ‘Wat zeggen deze gegevens over de verschillen tussen mannen en vrouwen.’ En dan geeft zo iemand als antwoord: ‘42.’ Dat is toch ronduit belachelijk, zeg nou zelf. Je moet conclusies trekken en die conclusies nauwkeurig formuleren en opschrijven. Dus een goed antwoord is niet ‘42’ en ook niet ‘42 → verschil’ maar wel: ‘Er is een verschil van 42 punten tussen Nederlandse mannen en vrouwen van middelbare leeftijd met betrekking tot hun gemiddelde stress-tolerantie gemeten met de SQV-5.’ Misschien denk je dat zo’n zinnetje op het tentamen toch niet belangrijk is, maar dan zou je wel eens lelijk op je neus kunnen kijken. Veel statistiekdocenten vinden dat juist het belangrijkste. Bij tentamens die door mij worden afgenomen geldt in ieder geval: de meeste mensen zakken op het weglaten of verkeerd formuleren van de interpretaties. Want wat heb je eraan om alles perfect te berekenen als je er vervolgens de verkeerde conclusie uit trekt? Of zelfs helemaal geen conclusie; dat is nog stommer, lijkt me.

Verder moet je bij het maken van opgaven in gedachten houden dat communicatie en gezond verstand in de statistiek nog belangrijker zijn dan in de rest van het leven. Dus als er in dit boek nergens staat dat je de meeteenheid moet vermelden, dan kun je misschien zelf bedenken of dat een handige manier van communiceren is.

### *Mythen over het leren van statistiek*

Ik heb een aantal jaren statistische analyses gedaan op de cijfers die studenten haalden voor hun statistiektentamens in de cursussen die ik geef. De conclusies daarvan waren steeds hetzelfde. De belangrijkste reden waarom studenten het tentamen niet haalden, was dat ze überhaupt niet aan het tentamen meededen. Het statistiekcijfer correleerde nauwelijks met het wiskundecijfer dat men op de middelbare school had, maar sterk met de hoeveelheid gemaakt huiswerk. Globaal gesproken was de slaagkans ongeveer gelijk aan het percentage gemaakt huiswerk. Wie altijd alles had gemaakt, had 100% kans het tentamen direct of bij de herkansing te halen. Wie 50% had gemaakt, had ongeveer 50% kans. Mijn boodschap is dus simpel: maak al je huiswerk en doe mee aan het tentamen.

Mijn boodschap is simpel en mijn argument is helder, maar toch wordt ik vaak niet geloofd. Dat komt doordat mijn argument statistisch van aard is. Daarom zal ik hieronder wat uitleggen over de psychologie van het leren voor statistiek. Daar zijn natuurlijk veel meer woorden voor nodig.

Een van de belangrijkste redenen waarom mensen zakken voor statistiektentamens is misschien moeilijk voor je te geloven. Dat is namelijk omdat ze een foute



psychologische theorie hebben. Het is een *psychologische* theorie, geen statistische, want hij zegt iets over menselijke denkprestaties. De inhoud van die theorie is: *statistiek is een kwestie van begaafdheid en dat heb je of je hebt het niet*. Alvorens op die theorie in te gaan, moet je je eerst eens afvragen in hoeverre jijzelf in die theorie gelooft en waarom.

Uit verscheidene onderzoeken komt naar voren dat juist de mensen die in deze theorie geloven slecht presteren in statistiek. Daarentegen zijn degenen die goed presteren meestal mensen die geloven dat hun prestatie vooral een kwestie is van hard werken en veel geduld. In dat opzicht is het toch een beetje vergelijkbaar met wiskunde. Onder wiskundigen bestaat de uitdrukking dat je ‘zitvles’ moet hebben. Wie dat niet heeft, kan het wel vergeten.

Je hoeft geen groot psycholoog te zijn om in te zien dat de begaafdheidstheorie desastreus is voor iemands prestaties in statistiek. Want als hij waar is, dan heeft het geen zin om te leren. Dus dat doen die mensen ook niet. En dan zakken ze. En dan zeggen ze ‘Zie je wel; ik kan het niet’. Tja, zo kennen we er nog wel paar!

Als je in de begaafdheidstheorie geloofde, ben je nu waarschijnlijk nog niet overtuigd van het tegendeel. Maar waarom geloof je er eigenlijk in? Misschien zaten er in jouw klas jongetjes die het allemaal kwam aanwaaien en meisjes die alsmaar moesten ploeteren? Leraren of ouders die tegen je zeiden ‘dat kun jij niet’? Dan moet je bedenken dat het voor veel mensen psychologisch voordelig is om deze mythe te laten voortbestaan. Degenen die het zogenaamd allemaal komt aanwaaien zijn daar trots op. Daardoor hebben ze er belang bij om de hoeveelheid tijd die ze er in staken veel kleiner voor te stellen dan hij in werkelijkheid is. Verassend genoeg zijn ook degenen die ervoor moesten ploeteren daar vaak trots op. Die hoor ik bij het maken van een opgave dan uitspraken doen zoals: ‘Ik ben een alfa - ik kan dat niet.’ Dat wordt dan met een zekere blijdschap uitgesproken. Voor je doorleest in de volgende alinea moet je eens een minuut proberen te bedenken hoe zo iemand een statistiekopgave maakt. Een doorsnee statistiekopgave bevat vragen waarvoor een normaal mens enige minuten moet nadenken eer hij of zij aan het beantwoorden kan beginnen. Stel je nu iemand voor die er heimelijk trots op is geen statistiek te kunnen. Hoe zal die de statistiekopgave maken?

Dat gaat zo iemand dus niet lukken. Want die denkt na vijf seconden al: ‘Ik weet het antwoord niet. Zie je wel, ik ben een alfa!’ Dat is namelijk precies wat diegene wil denken over zichzelf. Dat zou nog niet zo erg zijn als de persoon daarna gewoon aan de vraag begon. Maar het verraderlijke is dat zo’n persoon gelijk stopt met de opgave en vindt dat een ander hem of haar moet helpen. Of misschien wel probeert door te gaan maar minutenlang deze en soortgelijke gedachten blijft herhalen. Dus die opgave komt nooit af. Vervolgens wordt dat door iedereen gezien als een bevestiging van de theorie. Terwijl het gewoon een self-fulfilling prophecy is.

Over het alfaschap bestaat nog een andere mythe, namelijk deze: *een alfa is geen beta*. Er zijn tientallen onderzoeken die hebben gevonden dat alle cognitieve prestaties positief met elkaar correleren. Mensen die goed zijn alfavakken, zijn meestal

ook goed in betavakken, en andersom. Toch wil de mythe dat het andersom is. En ook hier ontstaat een self-fulfilling prophecy: iemand die eenmaal van zichzelf heeft geconcludeerd dat hij of zij een alfa is, en gelooft dat zulks niet is te verenigen met betacapaciteiten, is al gauw geneigd de betavakken te verwaarlozen omdat daar toch geen eer is te behalen.

Nauw verwant aan de begaafdheidsmythe is deze: *statistiek is een kwestie van inzicht*. Mensen die dat zeggen, denken er vaak stiekem bij: ‘en dat heb je of je hebt het niet’. Dat is weer de begaafdheidsmythe. Maar ook zonder die aanvulling is het een mythe. Statistiek, althans zoals het in dit boek wordt onderwezen, is 70% oefenen, 30% uit je hoofd leren, en 10% inzicht. Ja, je leest het goed, en je kan ook goed rekenen. Want inzicht is niet iets waar je mee begint, of wat een voorwaarde is; het is iets waar je mee eindigt en wat een gevolg is van oefenen en uit je hoofd leren. Wat is ‘inzicht’ eigenlijk? Daar heb ik nog nooit een duidelijke definitie van gehoord. Volgens mij is inzicht weinig meer dan goed georganiseerde kennis en vaardigheid plus een goed gevoel daarover. Daarmee komen we op een verwante mythe: *uit je hoofd leren is voor de dommen*. Tot mijn grote spijt is deze mythe jarenlang als vanzelfsprekendheid uitgedragen op middelbare scholen. Een behoorlijk aantal studenten, met name degenen die voor slim willen worden aangezien, willen dan ook geen dingen uit hun hoofd leren omdat ze zich daar te goed voor voelen. Alsjeblieft, geloof me: je kan dingen pas begrijpen als je ze grotendeels uit je hoofd kent.

Een andere mythe is: *meisjes zijn minder goed in statistiek dan jongens*. Er zijn talloze onderzoeken naar gedaan, en die hebben bijna allemaal dezelfde uitkomst: meisjes scoren gemiddeld niet significant lager dan jongens, eerder omgekeerd. Ik vraag me af of deze mythe misschien iets te maken heeft met de alfa-betamythe en de heimelijke trots geen statistiek te kunnen. Is het misschien zo dat degene die verklaart geen statistiek te kunnen zichzelf als extra vrouwelijk profileert? Want de ideale vrouw is knap – en een beetje dom?

De hier besproken mythen hebben met elkaar gemeen dat zij prestaties in statistiek koppelen aan onveranderlijke personeigenschappen. Deze mythen zijn vaak heimelijke *keuzen* die mensen maken over wat ze willen geloven over zichzelf. Daarmee negeren zij het feit dat statistiek leerbaar is, net zoals je kan leren piano spelen. Maar dan moet je wel oefenen. Niemand, ook niet wonderkinderen zoals Mozart, kan bij geboorte al piano spelen. Er moet eerst worden geoefend, met bloed, zweet en tranen. Datzelfde geldt voor statistiek. Daarmee wil ik niet zeggen dat iedereen een Rachmaninoff voor de statistiek kan worden. Maar het leren van de standaarddeuntjes moet te halen zijn voor iemand die niet geestelijk gehandicapt is. Grote delen van statistiek zijn weinig meer dan veredeld gezond verstand. Wie zegt: ‘Ik kan geen statistiek,’ zegt eigenlijk: ‘Ik ben geestelijk gehandicapt, ik heb geen gezond verstand.’ Wie dat echt van zichzelf gelooft, moet de consequentie trekken: sla dit boek dicht en beëindig je studie.

# Deel A Verdeling van een variabele

## Leerdoelen deel A

- Systematisch kunnen rapporteren over de verdeling van een variabele.
- Genormeerde scores kunnen berekenen en interpreteren.
- Op grond van een grafische weergave (histogram) conclusies kunnen trekken over diverse statistische maten.
- Kunnen rekenen met normaalverdelingen.
- De bekende beperkingen van het intuïtief-menselijk denken bij het trekken en beoordelen van steekproeven kunnen onderbouwen en toepassen.

Aan het eind van deel A wordt een volledige specificatie van de leerdoelen gegeven, alsmede twee zelftoetsen.

# 1 Observatie, vragenlijst, test

## 1.1 Inleiding

### *Doel*

Waar het in dit hoofdstuk om gaat, is dat je weet hoe we in de psychologie aan gegevens komen. Deze gegevens worden dan met behulp van de statistiek geanalyseerd. Ook moet je kunnen bepalen wat het meetniveau van variabelen is.

## 1.2 Methoden om gegevens te verzamelen

Bij onderzoek in de psychologie baseert men zich op het gedrag van de proefpersonen. Dat gedrag heeft natuurlijk altijd erg veel aspecten. Te veel om voor de onderzoeker te bevatten. Daarom beperkt men zich maar tot enkele aspecten van het gedrag van de proefpersonen. Dit wordt dan systematisch geregistreerd, dat wil zeggen: voor elke proefpersoon. Daar zijn verschillende methoden voor:

- 1 observatie
- 2 vragenlijsten
- 3 tests

Deze methoden zullen we nu bespreken. Het gaat er daarbij nu nog niet om dat je precies alle voor- en nadelen van deze methoden kent – alhoewel het natuurlijk goed is om daar zelf over na te denken. Waar het om gaat, is dat je zicht krijgt op hetgene dat vooraf gaat aan alle statistische analyses: het verzamelen van gegevens.

De gegevens die we verzamelen zijn normaal gesproken afkomstig uit de ‘werkelijkheid’ buiten ons. Die werkelijkheid noemen we de **empirie**. Onderzoek hier naar heet empirisch onderzoek. Er zijn een aantal regels die door vrijwel alle onderzoekers in de psychologie worden geaccepteerd:

- De werkelijkheid waar je het over hebt moet in principe voor iedereen waarneembaar zijn. Je mag dus geen conclusies trekken op basis van introspectie van jezelf. Je mag daar overigens wel vermoedens uit putten. Ook uitspraken over je vorige levens, of over een transcendente wereld die alleen jij kan ervaren, zullen niet serieus worden genomen.
- De uitspraken die je doet, moeten duidelijk zijn. Dat wil zeggen: slechts voor één interpretatie vatbaar. Iedereen moet zonder nadere discussie kunnen con-

troleren of het klopt wat je zegt. De uitspraak: ‘Die jongen is lief,’ is niet duidelijk.

- We beperken ons tot feitelijk beschrijvingen en geven geen morele oordelen.

### 1.3 Observatie

Je kunt bijvoorbeeld in een café gaan zitten om te observeren hoe mensen elkaar ontmoeten. Er spelen hierbij natuurlijk heel veel zaken een rol die je zouden kunnen interesseren. Om het allemaal bij te houden moet je dus een selectie maken: welke aspecten interesseren je genoeg om te observeren? Dat moeten natuurlijk wel aspecten zijn die inderdaad ‘observeerbaar’ zijn. Zoiets als ‘of er een relatie is tussen non-verbaal gedrag en geslacht’ is veel te vaag. Wat is ‘non-verbaal gedrag’? Wat bedoel je met ‘relatie’? Er zijn wel duizend ‘relaties’ te bedenken.

Je zou bijvoorbeeld het volgende onderzoek kunnen doen. Ik ga er voor het gemak vanuit dat je steeds een persoon uitkiest die alleen in het café zit of staat, en die persoon observeert tot hij of zij een ontmoeting heeft met iemand anders, dan wel het café verlaat. Dan schrijf je op wat je hebt gezien. Vervolgens kies je een nieuwe persoon uit om te observeren.

Bij de observatie kun je dan steeds de volgende vragen beantwoorden. Tussen haakjes staan de antwoordmogelijkheden.

- 1 Wat is het geslacht van de eerste persoon? (man/vrouw)
- 2 Wat is het geslacht van de tweede persoon? (man/vrouw)
- 3 Hoe lang duurt de ontmoeting? (1 minuut, 2 minuten, ...)
- 4 Raken de personen elkaar aan? (ja/nee)

Deze vragen leveren je een selectie van het gedrag van de personen. Er zijn immers heel veel dingen die je niet registreert. Je schrijft bijvoorbeeld niet op waar de personen elkaar aanraken, of ze lachen, of ze veel praten, hoe hard ze praten, enzovoort.

#### 1.3.1 Datamatrix

Als je deze observatie voor 20 paren uitvoert, zou je bijvoorbeeld de volgende gegevens kunnen krijgen. Voor de overzichtelijkheid is het gebruikelijk de gegevens in een tabel te zetten, zoals in tabel 1.1. Deze tabel noemt men de **datamatrix**.

Tabel 1.1

Paar	Vraag 1	Vraag 2	Vraag 3	Vraag 4
1	man	man	1	nee
2	man	man	28	nee
3	vrouw	man	15	ja
4	man	vrouw	2	nee
5	vrouw	man	1	nee
6	vrouw	vrouw	44	ja
7	vrouw	vrouw	1	nee
8	man	vrouw	10	nee
9	man	vrouw	23	ja
10	man	man	14	nee
11	vrouw	man	69	nee
12	man	vrouw	21	ja
13	vrouw	vrouw	12	ja
14	vrouw	vrouw	9	ja
15	vrouw	man	9	nee
16	man	man	3	ja
17	man	vrouw	16	nee
18	vrouw	vrouw	23	ja
19	vrouw	man	33	ja
20	man	vrouw	21	nee

Op grond van de bovenstaande gegevens kun je al een aantal simpele, concrete vragen beantwoorden. Maak nu opgave 1 in hoofdstuk 7.

## 1.4 Vragenlijsten

Als je wilt weten hoe mensen hun vrije tijd besteden, zou je dat als volgt kunnen onderzoeken. Je zoekt een groep personen die je wilt onderzoeken. Je maakt een lijst met vragen die je aan deze personen wilt stellen. Je legt een kopie van de vragenlijst voor aan elk van die personen, met de instructie om de vragen met betrekking tot zichzelf te beantwoorden. Je zou bijvoorbeeld een groep psychologiestudenten de volgende vragen kunnen voorleggen.

- 1 Ga je vaak naar de film? (nooit/soms/vaak)
- 2 Doe je veel aan sport? (nooit/soms/vaak)
- 3 Ga je vaak naar de kroeg? (nooit/soms/vaak)
- 4 Lees je veel romans? (nooit/soms/vaak)

En zo kun je nog wel een aantal vragen verzinnen.

### 1.4.1 Datamatrix

Als je deze vragenlijst aan 20 personen voorlegt, zou je bijvoorbeeld de volgende gegevens kunnen krijgen. Wederom worden de gegevens voor de overzichtelijkheid in een datamatrix weergegeven. Om het typewerk te besparen zullen we verder de antwoorden **coderen** met een getal:

- 1 = nooit
- 2 = soms
- 3 = vaak

Een mogelijk resultaat is de datamatrix in tabel 1.2. Iedere kolom (verticaal) scores in de matrix noemen we een **variabele**. De antwoorden op vraag 1 vormen dus tezamen een variabele. De antwoorden op vraag 2 vormen een andere variabele. Iedere rij noemen we een **subject**. Daar worden ook vaak andere namen voor gebruikt, zoals eenheid, unit of case. Meestal is dat een persoon of een dier.

Tabel 1.2

<i>Persoon</i>	<i>Vraag 1</i>	<i>Vraag 2</i>	<i>Vraag 3</i>	<i>Vraag 4</i>
1	1	1	1	2
2	2	3	3	2
3	3	3	1	1
4	1	3	2	1
5	1	2	1	1
6	1	2	2	1
7	3	2	2	1
8	3	1	3	3
9	2	2	1	2
10	1	1	2	2
11	2	1	2	1
12	3	3	1	1
13	1	2	2	1
14	2	1	1	2
15	1	1	3	3
16	1	2	3	2
17	3	3	1	2
18	2	2	1	2
19	2	1	3	3
20	2	3	1	2

Op grond van de bovenstaande gegevens kun je een aantal concrete vragen beantwoorden. Een aantal voorbeelden daarvan staan in opgave 2 in hoofdstuk 7. Maak nu die opgave.

## 1.5 Psychologische tests

In veel gebieden van de psychologie heeft men succesvol onderzoek kunnen doen door het gebruik van **psychologische tests**. Elke test heeft de volgende drie kenmerken volgens de *Standards* van de American Psychological Association (APA) (AERA, APA & NCME, 1985):

- De test bestaat uit een aantal vragen of opdrachten. Die vragen of opdrachten worden **items** genoemd. Een te onderzoeken persoon moet deze items maken. Dat wil zeggen: de vragen beantwoorden of de opdrachten uitvoeren. In dit opzicht is een test niet meer dan een vragenlijst.
- Voor elk item krijgt de persoon een score op basis van het gegeven antwoord. Deze scores noemen we **itemscores**. In het geval van een intelligentietest bijvoorbeeld, zal per item worden beoordeeld of het antwoord van de persoon ‘goed’ of ‘fout’ is. Voor een goed antwoord krijgt de persoon 1 punt, voor een fout antwoord 0 punt.
- De itemscores van de persoon worden samengevat met een enkel getal, de zogenaamde **testscore**. In vrijwel alle gevallen is dit gewoon de som van de itemscores van de betreffende persoon.

Een psychologische test heeft tot doel bij individuele personen te meten in hoeverre zij een bepaalde eigenschap bezitten. De mate waarin de persoon die eigenschap bezit, wordt met de testscore uitgedrukt (dat is tenminste de bedoeling). Een test is dus een **meetinstrument**.

Er zijn heel veel verschillende tests. Je kan bijna geen eigenschap verzinnen of er is al wel eens een test voor gemaakt.

### *Voorbeelden*

#### 1 Een multiple-choicetentamen

Een multiple-choicetentamen is een test. Voor elke vraag krijg je een aantal punten, afhankelijk van je antwoord, en die punten worden opgeteld om je cijfer te bepalen. Dat cijfer is dus een testscore. Hoe hoger je cijfer is, hoe meer je beheerst van het betreffende vak. Althans, volgens de examinator.

#### 2 Tijdschriften

In vrijwel elke *Viva, Elle*, enzovoort, staat wel een test om zaken te meten als: ‘Ben jij een goede vriendin?’, ‘Vertrouw jij je vriend?’, enzovoort.



### 3 BDI en ATQ

Beck's Depression Inventory (BDI) is een test waarmee men beoogd te meten hoe depressief (neerslachtig) de persoon in de voorafgaande week is geweest. De Automatic Thought Questionnaire (ATQ) is een test die beoogd te meten in welke mate de persoon werd gehinderd door 'automatische negatieve gedachten' in de voorafgaande week. Dit zijn negatieve gedachten die door het hoofd van de persoon malen en die in sommige situaties 'vanzelf' opkomen. ('De barkeeper ziet me over het hoofd. Hij wil me niet zien. Zie je wel, niemand wil me zien. Ik ben lelijk. Ik ben oninteressant. Alle andere mensen zijn leuker dan ik.') Automatische negatieve gedachten worden gezien als een oorzaak van depressiviteit.

In paragraaf A.1 van de appendix vind je de Nederlandse vertaling van deze twee tests. Merk op dat de antwoorden hier niet als 'goed' of 'fout' worden gescoord. Ze worden in feite gescoord als 'hoog-' of 'laagdepressief'. Merk verder op dat met deze tests niet wordt gepretendeerd een persoons eigenschap te meten die stabiel is gedurende het hele leven van de persoon. Het kan per week veranderen.

#### 1.5.1 Zijn tests wel te vertrouwen?

Met psychologische tests proberen we eigenschappen van personen uit te drukken in getallen. Een beetje raar. Veel eigenschappen zijn eigenlijk veel te complex om domweg in getallen uit te drukken. Mag dat zomaar, een stel vragen verzinnen en de antwoorden optellen? Daar valt heel veel over te zeggen.

Ten eerste kan deze vraag niet in zijn algemeenheid worden beantwoord. Het hangt er maar net van af welke eigenschap je wilt meten, en welke test je daarvoor gebruikt. Sommige eigenschappen, zoals intelligentie, zijn inderdaad te complex om met een enkel getal uit te drukken. Bij sommige andere eigenschappen lukt dat misschien weer beter. Je moet dus voor elke test opnieuw onderzoeken of hij geschikt is.

Ten tweede heeft het, bij een gegeven test, weinig zin om hierover theoretisch te speculeren. Je moet gegevens verzamelen waaruit blijkt dat de test goed is. Een veelgehoorde kritiek op tests is bijvoorbeeld dat een testscore slechts een momentopname is. Welnu, dat zou waar kunnen zijn. Dat kun je onderzoeken door de test tweemaal af te nemen bij dezelfde personen, met een tussen periode van (zeg) een jaar. Als iedereen bij de tweede afname ongeveer dezelfde score heeft als bij de eerste afname, dan is de testscore kennelijk geen momentopname.

Ten derde moet je oppassen niet in woordspelletjes te vervallen. Het is natuurlijk erg makkelijk om te roepen dat de BDI geen depressiviteit meet omdat jij iets heel anders onder depressiviteit verstaat dan meneer Beck. Maar daar schiet niemand iets mee op. Er bestaat geen algemeen geaccepteerde definitie van depressiviteit. Dus kunnen we gewoon afspreken dat we daarmee voortaan de score op de BDI bedoelen. Trouwens, zelfs als je je eigen depressiviteitstest zou creëren, zal die ongetwijfeld leiden tot een andere begrip 'depressiviteit' dan wat je oorspronkelijk

in gedachten had. Je uiteindelijke definitie dient namelijk precies te zijn, terwijl je oorspronkelijke begrip nogal vaag is. Dat kan dus nooit hetzelfde zijn.

Ten vierde is een test vooralsnog vaak het beste alternatief om op systematische wijze hanteerbare informatie over menselijk gedrag te verzamelen. In het geval van de BDI zal niemand ontkennen dat depressiviteit een complex verschijnsel is dat niet volledig wordt gevangen door de BDI. Niettemin is de BDI bruikbaar, omdat hij in ieder geval een deel van de relevante informatie wel geeft. Dat is nog altijd beter dan niets. En het is ook beter dan een chaotische stortvloed van alle mogelijke details, waar niemand iets mee kan.

Bij de bovengenoemde damesbladentests is vermoedelijk niet onderzocht of ze adequaat zijn. Zo'n test is dus wetenschappelijk gezien van nul en generlei waarde. Van de BDI en de ATQ is dat wel onderzocht. Hoe je de geschiktheid van een test kan onderzoeken, zal pas veel later in de studie worden behandeld. Op dit moment is dat nog te moeilijk. Voorlopig moet je dus maar aannemen dat de tests waar we het over hebben, redelijk goed zijn.

### 1.5.2 Datamatrix

In een onderzoek zijn de BDI en de ATQ bij een aantal jongeren tussen de 15 en de 19 afgenomen (samen met nog een aantal andere tests). Vervolgens deed een deel van deze personen mee aan de cursus 'Stemmingmakerij'. Dit noemen we groep 1. De andere personen deden niet mee aan de cursus. Dit noemen we groep 2. Een aantal maanden later, dus na de cursus, werden zowel de BDI als de ATQ nogmaals afgenomen bij alle personen. We geven hier de datamatrix met de testcores in tabel 1.3. Daarbij gebruiken we de volgende afkortingen:

Bdv	= depressiviteit voormeting	= score op de BDI bij de eerste afname
Bdn	= depressiviteit nameting	= score op de BDI bij de tweede afname
Atqv	= automatische gedachten voormeting	= score op ATQ bij de eerste afname
Atqn	= automatische gedachten nameting	= score op ATQ bij de tweede afname

Tabel 1.3

Groep	Bdv	Bdn	Atqv	Atqn
1	22	4	80	40
1	13	3	57	38
1	15	4	89	47
1	19		70	53
1	23	8	79	56
1	12	3	68	38
1	15	7	63	55
1	18	22	77	104
1	14	3	70	47
1	12	1	91	32
1	16	7	50	40
1	21	3	69	40
1	23	11	117	80
1	11	7	72	56
1	16	21	71	73
1	20	33	95	95
1	13	6	50	47
1	16	18	103	94
1	14	10	52	57
1	15	14	78	70
1	16	8	99	73
1	12	7	68	48
1	11	1	80	35
1	15	6	47	49
1	11	6	58	45
1	11	6	55	51
1	14	14	93	73
1	12	10	54	44
1	12	9	66	49
1	14	11	59	54
1	21	13	67	45
1	17	27	67	72
1	12	14	55	79
1	12	8	70	61
1	21	19	110	91
1	21	9	66	70
1	13	3	56	39
1	25	4	92	36
1	14	7	61	45
1	14	6	84	65

<i>Groep</i>	<i>Bdv</i>	<i>Bdn</i>	<i>Atqv</i>	<i>Atqn</i>
1	19	19	103	108
1	22	14	99	78
1	19	22	97	90
1	20	14	91	69
1	19	25	85	87
1	11	5	57	57
1	12	15	63	68
1	10	19	65	77
1	19	16	94	90
2	11	5	69	52
2	11	5	58	52
2	13	0	43	33
2	12	5	60	51
2	22	17	58	61
2	12	3	44	32
2	10	13	53	53
2	10	5	36	43
2	12	5	47	39
2	11	9	67	57
2	18	3	79	56
2	14	15	61	58
2	10	10	58	58
2	16	0	66	84
2	15	16	55	66
2	10	13	46	59
2	16	15	89	92
2	11	4	58	31
2	13	12	50	45
2	16	13	82	62
2	11	0	47	38
2	22	22	87	100
2	16	14	64	58
2	13	9	67	62
2	15	11	84	50
2	10	6	70	62
2	19	8	99	83
2	11	12	64	59
2	14	12	85	81
2	11	5	38	41
2	12	13	69	63

Groep	Bdv	Bdn	Atqv	Atqn
2	20	8	88	80
2	10	10	50	51
2	12	6	58	55
2	15	9	64	68
2	13	11	58	67
2	15	8	53	61
2	18	26	72	72
2	16	12	43	66
2	11	7	87	41
2	20	16	70	59
2	10	16	35	38

*Number of cases listed = 91*

Je ziet, we hebben nu wat meer proefpersonen, maar het weergegeven van de datamatrix is niet echt informatief. Een kind kan begrijpen dat het handiger is om de zaak samen te vatten. Een voor de hand liggende manier om dat te doen, is om in beide groepen voor beide variabelen het gemiddelde te berekenen. Dat zijn dus vier gemiddelden. In figuur 1.1 is een stukje computeroutput weergegeven waar die gemiddelden in staan. (Er staan ook nog andere dingen in, die nog niet behandeld zijn. Dat heb je nou altijd met statistische programma's: je vraagt iets simpels, en dan geven ze er iets bij wat je niet snapt. Daar wordt een mens zenuwachtig van.)

Op grond van de gegevens in de datamatrix en de output kun je een aantal vragen beantwoorden. Die vragen staan in opgave 3 in hoofdstuk 7. Maak die opgave nu.

GROEP		BDV	BDN
1,00	Mean	15,8571	10,8750
	Std. Deviation	4,02078	7,37051
	N	49	48
2,00	Mean	13,7381	9,7381
	Std. Deviation	3,44334	5,66142
	N	42	42
Total	Mean	14,8791	10,3444
	Std. Deviation	3,89254	6,61656
	N	91	90

*Figuur 1.1 Gemiddelden van Bdv en Bdn per groep*

### Jongen, 17 jaar, BDI-score 16

De laatste tijd ben ik nogal eens depressief. Meestal niet tijdens de les, maar thuis en vaak na een feest. Ik heb vaak het idee dat ik er maar een beetje 'bijhang'. Ik heb weinig zelfvertrouwen, ben erg verlegen, meng me nooit in gesprekken en ga na school ook haast niet met mensen om. Ik heb een paar goede vrienden, maar ik kan met hen ook niet goed over mijn problemen praten. Vaak denk ik dat zij best zonder mij kunnen maar ik niet zonder hen, ik maak namelijk moeilijk contact met mensen. Ik heb eens geprobeerd voor mezelf te bewijzen dat ik best zonder hen kon, maar toen voelde ik me toch heel erg alleen, kreeg nog meer het gevoel dat ik niets waard ben en baalde ik een paar dagen ontzettend. Af en toe denk ik ook aan zelfmoord en daar maken mijn vrienden zich best druk om. Ik weet wel zeker dat ik het (nu) niet zou doen (...) Meestal als ik depressief ben denken mijn ouders dat ik chagrijnig ben. Zo kom ik thuis ook over, maar thuis gedraag ik me heel anders dan op school. Thuis kan ik nooit echt zeggen wat me dwars zit. Altijd als ik baal ga ik over dingen nadenken. Hierdoor kan ik me moeilijk concentreren op mijn huiswerk en kom ik meestal slecht in slaap. Als ik kranten bezorg ga ik altijd piekeren over allerlei dingen en kan hier dan moeilijk mee stoppen

*Uit: Nine Veltman (1997), Stemming maken*

## 1.6 Het meetniveau van een variabele

Een variabele die een hoeveelheid uitdrukt, noemen we een **kwantitatieve** variabele. Andere variabelen noemen we **kwalitatieve** variabelen. Deze terminologie is enigszins ongebruikelijk. Vaker gebruikt men de termen 'continu' versus 'categoriaal', of 'interval' versus 'nominaal', in plaats van kwantitatief versus kwalitatief. Die begrippen hebben echter ook een veel sterkere betekenis, die meestal niet is gerechtvaardigd.

### *Voorbeeld*

BDI-scores zijn kwantitatief. Nationaliteit is kwalitatief, want Belgen hebben niet meer nationaliteit dan Duitsers. Geslacht is kwalitatief. Overigens kun je daar nog over discussiëren, en dat geldt voor alle variabelen met twee categorieën: weliswaar hebben jongens niet meer geslacht dan meisjes, maar als je de variabele 'mannelijkheid' noemt, dan is het wel een hoeveelheid.