



Share Knowledge

Choose right

Develop Talent

Cooperate

WWW.JOHO.NL

## Methoden & Technieken

## College- en werkgroepaantekeningen 2012-2013

50-10-117

\* 8101748\*

PSYCHOLOGIE, JAAR 1

## JoHo Samenvattingen

JoHo biedt een compleet pakket samenvattingen aan. In dé studentenwinkels van Nederland vind je bijvoorbeeld uitgebreide boekuittreksels, stampijsten, handige schema's, samenvattingen van arresten, collegeaantekeningen en oude tentamens met uitwerkingen. Deze producten zijn aanvullend op je lesstof en studieboeken, besparen je tijd, vergroten je slagingskans en worden niet voor niets door vrijwel iedere student in de grote studentensteden al jaren als onmisbaar ervaren!

Kijk voor het actuele aanbod op [JoHo.nl](http://JoHo.nl)

Bij de samenstelling van de samenvattingen proberen we zoveel mogelijk de kwaliteit te waarborgen. We kunnen echter geen verantwoordelijkheid aanvaarden voor het gebruik ervan. De materialen dienen als hulpmiddel en niet als vervanging van bijvoorbeeld het volgen van onderwijs of het bestuderen van boeken.

## Hoe te verkrijgen?

In het JoHo center  
*Het volledige aanbod is hardcopy verkrijgbaar in het JoHo center van jouw stad. Hoge kortingen voor JoHo leden.*

Online via [JoHo.nl](http://JoHo.nl)  
*Een groot gedeelte van het aanbod is ook online gratis te downloaden voor JoHo leden.*

## Werken voor JoHo

JoHo zoekt voortdurend afgestudeerden voor het maken en controleren van de samenvattingen.

## JoHo centers

Je vindt de JoHo centers in Amsterdam, Den Haag, Groningen, Leiden, Rotterdam en Utrecht. Voor adressen en openingstijden, kijk op [JoHo.nl](http://JoHo.nl)

## The World of JoHo



Waar je alles kunt boeken en zoeken voor je studie, stage, project, baan, cursus en reis in het buitenland

JoHo **helpt** wereldwijd talent te ontwikkelen, kennis over te dragen en samen te werken.

JoHo **werkt** wereldwijd voor studenten, reizigers, vrijwilligers, expats en ondernemers.

JoHo **geeft** wereldwijd kortingen, hulp, advies, bemiddeling en gratis gidsen.

# **Inhoudsopgave**

Hoorcolleges week 1 t/m 6\*

Werkgroepaantekeningen week 1 t/m 6\*

Extra oefenvragen

# Hoorcolleges

## Week 1 (05-09-2011)

In het jaar 1875 werd de wetenschappelijke psychologie geboren. De belangrijkste grondlegger hiervan was Wilhelm Wundt, een Duitse wetenschapper die onder andere het eerste psychologische laboratorium opzette en begon met het schrijven van theorieboeken en lesgeven aan studenten.

*Wanneer is een studie wetenschappelijk?*

1) Empirisme:

Je moet kijken naar de feitelijk zintuiglijk waarneembare zaken. Oorzaak-gevolg relaties moeten gecontroleerd worden. Een van de meest voor de hand liggende methodes om dit te doen is steeds van oorzaak veranderen, en de gevolgen met elkaar vergelijken. Belangrijk bij empirisch onderzoek is het opzetten van systematische onderzoeken, in overkoepelende theorieën zit vaak een algemene regelmaat.

2) Verificatie:

Andere onderzoekers moeten jouw onderzoek stap voor stap kunnen nadoen. Daarom is het belangrijk dat alle onderzoeksresultaten en methodes openlijk worden gepresenteerd. Het nadoen van een onderzoek kan een bepaalde theorie versterken (als er opnieuw dezelfde conclusie uitkomt), maar ook verzwakken (als er een hele andere conclusie uitkomt)

3) Onderzoekbare problemen onderzoeken:

Je moet wel problemen onderzoeken die je ook echt kunt toetsen. Een voorbeeld van een niet-toetsbare hypothese is: engelen bestaan. Dat we het niet kunnen onderzoeken wil niet zeggen dat ze daadwerkelijk niet bestaan, maar op dit moment hebben we geen instrumenten of methodes om een dergelijke theorie te onderzoeken.

Als wetenschapper heb je twee taken:

- ontdekken en beschrijven
- verklaringen opstellen en voorspellen

Deze twee taken hangen samen met de verschillende soorten onderzoek die er zijn, het ene onderzoek is geschikt voor beschrijvingen, het ander voor verklaringen.

*4 categorieën van gedragsonderzoek:*

1. Descriptief onderzoek: dit is beschrijvend onderzoek. Bij beschrijvend onderzoek ben je alleen maar bezig met het beschrijven van wat er gebeurt. Je gaat alle waarnemingen noteren en onderzoeken.
2. Correlationeel onderzoek: hierbij laat je de samenhang zien tussen variabelen. Het gaat bij correlationeel onderzoek om twee variabele en hun verband. Dit hoeft niet per se een oorzaak-gevolg relatie te zijn, dat twee variabele samenhangen wil nog niet zeggen dat de ene ook daadwerkelijk het gevolg is van de ander.
3. Experimenteel onderzoek: met experimenteel onderzoek kan je wel oorzaak-gevolg relaties aantonen. Het gaat hier dus om causale relaties.
4. Quasi-experimenteel onderzoek: dit is experimenteel onderzoek met al bestaande groepen. Je kunt de categorieën niet random toewijzen. Voorbeeld hiervan is onderzoek in een schoolklas, je kan de kinderen die in de klas zitten niet zelf uitkiezen.

*Empirische cyclus:*

Adriaan de Groot had een empirische cyclus als model voor het onderzoek in 5 fasen:

1. **Observatiefase:** in deze fase ontstaat het idee voor een onderzoek. Dit idee ontstaat vaak vanuit een observatie, en kan dus overal overgaan.
2. **Inductiefase:** In deze fase ga je van een specifieke observatie naar een algemene theorie. Dit wordt ook wel operationaliseren genoemd. Een *theorie* is een verzameling uitspraken die aan de hand van begrippen een relatie beschrijft.  
'leap of faith': je kan nooit zeker weten of je theorie, die dus vaak gebaseerd is op één observatie, klopt.
3. **Deductiefase:** uit de algemene theorie wordt een toetsbare werkhypothese (onderzoeksvraag) afgeleid. Hiervoor zijn een aantal criteria: de werkhypothese moet voldoende specifiek (afgebakend) zijn, het moet staan in de 'als... dan...' vorm, het moet falsifieerbaar (weerlegbaar) zijn en in de werkhypothese moet gebruik worden gemaakt van duidelijk omschreven begrippen. Voor het laatste criteria onderscheiden we 2 soorten definities. De eerste is de conceptuele definitie, hierbij gaat het om een woordenboekomschrijving, een abstracte omschrijving die je niet kan meten. De tweede is de operationele definitie, dit is de concrete definitie die precies aangeeft hoe een begrip wordt waargenomen, gemeten of gemanipuleerd.
4. **Toetsingsfase:** in deze fase wordt de werkhypothese getoetst door het onderzoek daadwerkelijk uit te voeren. In deze fase worden data verzameld, deze worden geanalyseerd en daaruit worden conclusies getrokken.
5. **Evaluatiefase:** wat zegt het resultaat over de algemene theorie? Het resultaat kan de theorie bevestigen of verwerpen. Daarna moet er geëvalueerd worden of de theorie wellicht moet worden aangepast, uitgebreid of verbeterd. In deze fase wordt ook gekeken naar eventuele tekortkomingen in het onderzoek.

Gedurende een onderzoek wordt de empirische cyclus (vanaf stap 2) vaak een aantal keer herhaald.

Kan een theorie bewezen worden?

- positief bewijs: de theorie is 100% waar. Dit is logisch gezien onmogelijk, omdat je nooit met zekerheid kunt zeggen of een theorie waar is. Hoe vaker de theorie bewezen wordt, hoe aannemelijker het is dat deze waar is.
- negatief bewijs: de theorie is niet waar. Dit kan logisch gezien wel (het resultaat bevestigt de theorie niet) maar is uit praktisch oogpunt onmogelijk. Dat een hypothese niet waar is betekent niet dat meteen de hele theorie verworpen is.

**Empirische generalisatie:** uit data (dit zijn bestaande gegevens) een samenhang vinden. Dit gaat via deductie volgens Leary.

*Variabiliteit:*

Variabiliteit: dat is datgene wat varieert. Het kan hierbij gaan om iets dat varieert tussen personen, tussen situaties of in de loop der tijd. Variantie is de index voor variabiliteit. Je bekijkt hierbij scores ten opzichte van een standaard (het gemiddelde).

De totale variantie = systematische variantie + foutenvariantie

De totale variantie is het te verklaren totaal, alle verschillen tussen individuen. Systematische variantie zijn de verklaarde verschillen tussen groepen. Foutenvariantie gaat over de onverklaarbare verschillen.

Effect size is de maat voor de sterkte van de relatie. Dit wordt berekend als de 'proportie verklaarde variantie', dit is de systematische variantie/totale variantie. Als deze 0 is, is er geen relatie (niets verklaard). Als deze 1 is, is er sprake van een perfecte relatie.

Vuistregel effect size:

0.01 → small effect

0.06 à medium effect

Boven 0.15 à large effect

**Week 2 (12-09-2011)**

Alfred Binet:

Het was de bedenker van de IQ test. Toen hij deze bedacht was deze nog niet geschikt voor iedere leeftijd. Later is de IQ test leeftijdsadequaat gemaakt. Dit betekent dat deze IQ test voor iedere leeftijd te testen was. Eerst deed Alfred Binet tranimetrie: dit is ook onderzoek met intelligentie. In zijn tijd klopte wat op het eerste gezicht klopt: face validiteit. Nu weten we beter en geloven we niet meer in face validiteit.

Binet staat ook aan de start van de psychometrie. Dit is een deel van de wetenschap dat zich bezighoudt met het meten van mensen of aspecten van hun gedrag, maar ook het meten van stimuli.

**Variabiliteit:**

Een variabele waarde moet voldoen aan 2 eisen. Het moet uitputtend zijn en wederzijds uitsluitend. Met behulp van de tabel hieronder wordt uitgelegd hoe dat in zijn werk gaat:

Categorie:	Leeftijd:
1.	0-18
2	19-35
3	36-70
4	>71

De tabel is uitputtend, want iedereen van iedere leeftijd kan zich in een categorie inschrijven. Wanneer bijvoorbeeld bij categorie 1 in plaats van 0-18, 1-18 stond, dan konden mensen die 9 maanden oud zijn zich nergens plaatsen.

De tabel is wederzijds uitsluitend, omdat je niet hoeft te twijfelen over de categorie waar je in hoort. Wanneer bijvoorbeeld bij categorie 2 in plaats van 19-35, 18-35 stond, dan konden mensen die 18 zijn gaan twijfelen of ze in categorie 1 of in categorie 2 moesten staan.

**Meetniveaus:**

1. Nominaal: dit zijn categorieën, of labels. Het gaat er hierbij alleen maar om of mensen gelijk zijn aan elkaar, of om te tellen hoe vaak iets voorkomt. Het nominale meetniveau is het laagste meetniveau. Voorbeelden: oogkleur, geslacht.
2. Ordinaal: het gaat hier om een bepaalde volgorde, er wordt iets 'geordend'. Het eerst gekozen doel heeft een hogere rangorde, maar de verschillen onderling zijn niet gelijk en doen er ook niet toe. Voorbeeld: op volgorde van lengte zetten.
3. Interval: dit is een volgorde met gelijke intervallen, er zijn dus gelijke afstanden tussen de getallen. Er is hierbij geen absoluut nulpunt. Een voorbeeld hiervan is IQ. A is niet 2 x zo groot als B, want er is geen nulpunt, maar het verschil tussen een IQ van 70 en 80 is evenveel als het verschil tussen een IQ van 130 en 140. Dit betekent dat IQ zich op het niveau van interval bevindt. Je kan dus optellen en aftrekken met scores op intervalniveau, maar niet delen of vermenigvuldigen.
4. Ratio: dit is een interval met een absoluut nulpunt. Een voorbeeld hiervan is de inhoud in centiliter. Hierbij is een absoluut nulpunt te vinden en de intervallen zijn gelijk. Dan spreek je dus van ratio. Bij dit meetniveau kan je naast optellen en aftrekken ook verhoudingen uitrekenen: dus ook vermenigvuldigen of delen.
5. Absoluut: Het nulpunt en de eenheid liggen vast, hier mag je zelf niks meer aan doen. Dit is het hoogste meetniveau. Voorbeeld hiervan zijn percentages.

**Betrouwbaarheid:**

Betrouwbaarheid: komt de variabiliteit in de getallen overeen met de werkelijke variabiliteit in het te meten gedrag? Het gaat hier om weinig toevalsfouten. Betrouwbaarheid komt er in het kort op neer dat er bij herhaald meten steeds hetzelfde wordt gemeten.

Methoden van herhaald meten:

1. Test- hertest betrouwbaarheid: het twee keer afnemen van dezelfde test en de overeenstemming in de resultaten bepalen. Voorbeeld: je neemt bij iemand een IQ-test af waaruit blijkt dat deze persoon een IQ heeft van 120. Twee maanden later neem je precies dezelfde test af, maar nu heeft diegene ineens een IQ van 80. Dan kan je ervan uitgaan dat er iets mis is met het meetinstrument.
2. Paralleltest betrouwbaarheid: in dit geval zijn de vragen hetzelfde ( ze meten hetzelfde), maar ze zijn niet identiek. Goed voorbeeld hiervan is het tentamen en de herkansing: ze testen dezelfde kennis/vaardigheden, maar zijn niet identiek.
3. Inter-item betrouwbaarheid: in hoeverre zijn items in een meetinstrument intern consistent? Dus meten de items in een meetinstrument wel wat ze moeten meten? Het gaat hier om een samenhang.
4. Item-totaal betrouwbaarheid: de score van 1 item moet samenhangen met het totale meetinstrument.
5. Replicatie: het onderzoek nogmaals uitvoeren en de overeenstemmingen bekijken.

Meten betrouwbaarheid: systematische (ware score) variantie/totale ( geobserveerde) variantie  
De betrouwbaarheidscoëfficiënt ligt tussen 0 en 1. De vuistregel is dat 0,70 of meer voldoende is.  
Geobserveerde variantie: systematische variantie + toevalsfouten

*Validiteit:*

Meet het instrument ook werkelijk datgene wat het claimt te meten? Het gaat hier om weinig systematische fouten.

*Soorten validiteit:*

1. Face validiteit: het onderzoek is valide naar het oordeel van de onderzoeker of participanten. Dit is dus niet heel betrouwbaar.
  2. Inhoudsvaliditeit: wordt het gehele domein gedekt door het instrument? Dit vereist onafhankelijke beoordelaars
  3. Construct validiteit: is er samenhang met andere metingen?
    - Convergente validiteit: er is een sterke samenhang met instrumenten die hetzelfde of een vergelijkbaar construct meten
    - Discriminante validiteit: er is een zwakke of geen samenhang met instrumenten die een ander construct meten
  4. Criterium validiteit: stelt een criterium, doet een voorspelling en kijkt vervolgens of het klopt. Deze vorm van validiteit is onder te verdelen in 2 componenten.
    - Concurrente validiteit: deze validiteit wordt tegelijkertijd gemeten met de uitvoering van het onderzoek.
    - Predictieve validiteit: deze validiteit wordt voorspeld aan de hand van eerder onderzoek
- De voorspelbaarheid is echter laag. Voorbeeld hiervan is selectie bij de poort voor studenten  $\hat{a}$  is dit valide?

Opletten: validiteit vereist betrouwbaarheid, maar dat een test betrouwbaar is wil nog niet zeggen dat deze ook valide is!

Benaderingen psychologisch meten:

- observatie
- fysiologisch
- zelfrapportage
- archiefdata
- inhoudsanalyse



Interne consistentie is een vorm van betrouwbaarheid (meten we nauwkeurig).

Correlatiecoëfficiënt: mate van samenhang tussen variabelen. Het gaat hierbij om een lineaire samenhang en ligt altijd tussen de -1 en +1. Is de uitkomst 0, dan is er geen samenhang.

We meten de interne consistentie met Cronbach's alpha. Hoe groter Cronbach's alpha, hoe groter de interne consistentie. Cronbach's alpha geeft het gemiddelde aan van alle mogelijke split-half opties. Bij split-half deel je de vragenlijst in tweeën en bereken je van elk deel de variantie. Een instrument bestaat uit items die allemaal hetzelfde proberen te meten. Ieder item is op zich een klein meetinstrumentje. Met de interne consistentie kijk je of alle items wel goed in de vragenlijst passen. Visueel voorbeeld: je stelt je eigen voetbalteam samen. Dit bestaat uit allerlei bekende voetballers, maar vervolgens staat premier Rutte in het doel. Het is dan meteen duidelijk dat Rutte er niet thuis hoort, dus niet consistent is met de rest.

Formule van Chronbach's Alfa

$$\alpha = \frac{k}{k-1} * \left[ 1 - \frac{\sum S_i^2}{S^2 x} \right]$$

Van elk afzonderlijk item moet de variantie berekend worden, en deze moeten uiteindelijk bij elkaar worden opgeteld. Chronbach's Alfa ligt tussen de 0 en de 1. Als het 0 is, is er geen samenhang tussen de items. Hoe dichterbij de 1, hoe beter.

Als iemand hoog scoort op een concept, moet deze persoon ook hoog scoren op alle afzonderlijke items.

*SPSS programma:*

Met SPSS is Cronbach's alpha ook te berekenen en via de verschillende opties kun je ook nog andere dingen meten:

1. Alpha if item deleted: Dit is de optie waarbij je kunt kijken of je een item kunt verwijderen waardoor Cronbach's alpha groter wordt. Als de alpha die bij een item staat onder dit kopje groter is dan de alpha die je nu hebt, dan kun je dat item het beste verwijderen.
2. Scale variance: deze wordt kleiner als je er 1 verandert. Dit is de variantie die de hele schaal heeft.
3. Corrected item Total correlation: Hierbij kan je zien hoe goed de items correleren met het geheel.

*Verschillende manieren van onderzoek:*

1. Cross-sectioneel: hierbij neem je 1 keer een steekproef in een populatie af.
2. Opeenvolging onafhankelijke steekproeven: hierbij neem je meerdere keren in dezelfde populatie een andere steekproef af.
3. Panel/ longitudinaal: hierbij heb je 1 steekproef die je meerdere keren over een lange tijd ondervraagt.

Een van de methode van onderzoek is het beschrijvende onderzoek. In een beschrijvend onderzoek bekijk je de eigenschappen en het gedrag van een populatie. Een populatie kan in grootte erg verschillen. Bij hele grote populaties is het niet mogelijk om iedereen te ondervragen, en zal je een steekproef moeten trekken.

Een steekproef zal nooit helemaal overeenkomen met de gehele populatie.

*Manieren van afnemen van steekproeven:*

Kanssteekproef (probability samples): Hoe groot is de kans dat een willekeurig individu in de steekproef komt. Dit hangt af van het 'sampling frame': de populatie.

Er zijn verschillende probability samples:

- 1) Simple random steekproef (SRS): dit is het beste te vergelijken met het trekken van namen uit een hoge hoed, iedereen heeft dus gelijke kans om in de steekproef te vallen. Dit is een praktisch probleem bij grote populaties ( hoe krijg je de namenlijst?)
- 2) Systematische steekproef: hierbij wordt een vaste regelmaat aangehouden. Bijvoorbeeld na elke tien voorbijgangers weer iemand ondervragen.
- 3) Gestratificeerde random steekproef: hierbij verdeel je de populatie eerst onder in bijvoorbeeld mannen-vrouwen of verschillende leeftijden (stratum). Vervolgens haal je uit beide delen mensen voor het onderzoek.
- 4) Clustersteekproef: hierbij gaat het niet om een inhoudelijk stratum, maar wordt de steekproef ingedeeld op geografische locatie.
- 5) Multistage Clustersteekproef: hetzelfde als de clustersteekproef, alleen behelst het meerdere stappen. Voorbeeld: lijst met gemeentes opvragen: daar er 10 uittrekken en vervolgens uit alle 10 één basisschool trekken. Deze steekproef moet wel representatief zijn ( zorgen dat niet alle gemeenten op het platteland liggen)

Niet-kanssteekproeven (non-probability samples)

- 1) gelegenheidssteekproef: afgaan op wat er voorbij komt. Voorbeeld: voor het FSW mensen ondervragen.
- 2) Quota steekproef: hierbij moet de steekproef gelijk zijn aan de percentages binnen de populatie. Voorbeeld: onder psychologiestudenten is 20% man en 80% vrouw. Een steekproef moet dan uit 20% mannen en 80% vrouwen bestaan.
- 3) Doelgerichte steekproef

*Florence Nightingale:*

Florence Nightingale is de uitvindster van het Pooldiagram (polar area chart). Tijdens de Krimoorlog zocht ze een overzichtelijke manier om de doden onder te verdelen in diegene die gesneuveld waren tijdens de strijd, aan ziekten enz. Een pooldiagram bestaat uit verschillende taartpunten, waarvan de hoeken allemaal gelijk zijn. De lengte van de punten zegt iets over de hoeveelheid.

Haar pooldiagram werd al snel vervangen door een taartdiagram, hierin verschillen de taartpunten zowel in hoek als in vorm. Een pooldiagram wordt bijna niet meer gebruikt, het taartdiagram is een manier om de verdeling van variabelen weer te geven.

Uit een onderzoek komen vaak een heleboel cijfers (ruwe data). De ruwe data zijn lastig te interpreteren. Daarom vatten we ze samen. Dit kan numeriek en/of grafisch. Een grafische weergave van een verdeling heet een plot. Dit verschilt van een numerieke verdeling (bijvoorbeeld een tabel). Voor een goede beschrijving van de data moet de data accuraat, beknopt en begrijpelijk zijn.

*Frequenties en frequentieverdelingen:*

$f$  = absolute frequentie

$p$  = relatieve frequentie  $\hat{=}$  hoeveel procent van het geheel

Bij de absolute frequentie moet je het getal steeds relateren aan de populatiegrootte, daarom gaat de voorkeur uit naar de relatieve frequentie. Die bereken je door de absolute frequentie te delen door het aantal proefpersonen (dus  $p=f/n$ )

Frequentieverdelingen kun je in een gegroepeerde tabel plaatsen. Je moet dan eerst intervallen maken. De vuistregel voor de hoeveelheid intervallen is de wortel van  $N$ . Bij 49 groepen heb je dus 7 intervallen. Daarna moet je het bereik (range) bepalen. Range = hoogste waarde – laagste waarde. Als laatste moet je ervoor zorgen dat je gelijke intervalbreedtes hebt.

De intervalbreedte is:  $\frac{\text{Range}}{\text{Aantal intervallen}}$ . Een interval moet uitputtend & wederzijds exclusief zijn. (zie week 2)

Als je de intervallen bij elkaar optelt, dan moet je ook de frequenties van de bij elkaar behorende intervallen bij elkaar optellen.

*Diagrammen en meetniveaus:*

Taart/staaf- diagram is van een nominaal en ordinaal meetniveau (categorische meetniveaus). Voor een intervalniveau gebruiken we de frequentiepolygoon, het histogram of de stem&leaf plot. Intervalniveau wordt ook wel numeriek meetniveau genoemd.

Een staafdiagram onderscheidt zich van een histogram door de ruimtes tussen de staven. Dit geeft de scheiding aan tussen verschillende waarden. Bij een histogram gaat het om een continue variabele.

Cumulatieve frequentieverdeling: wat is de kans dat iemand die score heeft of lager? Je telt dan dus de huidige klasse op bij de voorgaande klassen.

In sheet 17 is te zien dat er verschil wordt gemaakt tussen de schijnbare grens (vb.: 2-3) en de ware grens (vb.: 1.5-3.5). Het midden blijft 2.5 en is dus gelijk. Het verschil tussen de schijnbare grens en de ware grens is dus dat het lijkt dat de grens van 2 tot 3 loopt, maar dat deze werkelijk van 1.5 tot 3.5 loopt als je mensen tussen deze waarden in een categorie moet plaatsen. Het opvallende hieraan is dat 2.5 altijd het middelste getal blijft.

*Formule voor de variantie:*

Waarom wordt bij de variantie altijd de 'sum of squares' gekwadraterd?

Je moet ervoor zorgen dat alle afwijkingen positief zijn, anders tel je altijd op tot 0.

Waarom delen we door  $n-1$ ?

Het gemiddelde van de steekproef zit er altijd een klein beetje naast. De spreiding/ verdeling zegt iets over de steekproef/ fluctuatie. Het gemiddelde van het gemiddelde is de goede waarde. Daarom deel je door  $n-1$ . Zo wordt de schatting gecorrigeerd.

#### *De boxplot:*

De mediaan is de middelste waarneming. Om de mediaan te kunnen bepalen moet je eerst ordenen (dit is vanaf een ordinaal meetniveau). De modus is de meest voorkomende waarde en is alleen een kwestie van tellen (dit is dus vanaf een nominaal meetniveau).

De interkwartiele range (IQR) is het verschil tussen het 3<sup>e</sup> kwartiel en het 1<sup>e</sup> kwartiel. Het 3<sup>e</sup> kwartiel is de mediaan rechts van de gewone mediaan (dus het getal op 75% van de hele score). Het 1<sup>e</sup> kwartiel is de mediaan links van de gewone mediaan (dus het getal op 25% van de hele score).

De boxplot kun je tekenen na het maken van een five number summary. De five number summary bestaat uit: het minimum, het 1<sup>e</sup> kwartiel (Q1), de mediaan (Q2), het 3<sup>e</sup> kwartiel (Q3) en het maximum. Je hebt ook een modified boxplot. Hierbij worden uitbijters (waarnemingen die ver weg liggen van de meerderheid van de waarneming) in de boxplots als puntjes weergegeven. De vuistregel is dat een meting een uitbijter is als die  $1,5 \cdot \text{IQR}$  boven de Q3 of onder de Q1 ligt.

Voor een ordinaal of interval meetniveau is een boxplot handig.

#### *Stem&leaf plot:*

Het maken van een stem&leaf plot gaat als volgt. Eerst noteer je de takken. Dit zijn de allergrootste cijfers van je getal. Dus als je het getal 20 hebt, is 2 je tak en als je het getal 34 hebt, is 3 je tak. Dan maak je de bladeren. Dit zijn de getallen die na de tak komen. Bij 20 is dat een 0 en bij 34 is dat een 4. Je ordent deze getallen van klein naar groot en maakt zo een stem&leaf plot voor ieder getal. In een stem&leaf plot kan je dus nog de originele data aflezen, dit in tegenstelling tot een histogram.

*Adolphe Quetelet:*

Hij was degene die erachter kwam dat mensen eigenschappen hebben die, als je die registreert, een normale verdeling vormen. De Queteletindex is een andere naam voor de body-mass index. Hij had uitgevonden dat er een samenhang is tussen lengte en gewicht. Hij had hierdoor invloed op de statistiek met het concept van de (normale) natuurlijke variantie.

*Dichtheidcurves:*

Een histogram geeft exact de verdeling (in intervallen) weer, maar men kijkt liever naar een versimpelde verdeling. Daarom moet er gekeken worden of de verdeling is weer te geven als een gladde curve, ook wel dichtheidscurve of 'density curve' genoemd. Het is een vloeiende lijn door de toppen van het histogram. Curves hebben een algemene vorm.

De dichtheidscurve is een benadering van een empirische verdeling. De lijn geeft een goede benadering van het histogram weer.

Bij een normaal verdeling zijn de gemiddelde verdeling en de standaard deviatie genoeg om een model te kunnen tekenen. Een normaal verdeling is altijd symetrisch en heeft één top. De normale verdeling is een model maar ook een norm.

Cumulatieve frequentieverdeling: hoe een verdeling aangroeit

Relatieve frequentieverdeling: proporties

Het totale oppervlak onder de curve is altijd 1. Als staafjes een relatieve frequentie voorstellen, kan je ze optellen en er iets over zeggen. Dit betekent dus ook dat ze in een histogram staan (week 4). Je kunt via 2 benaderingen een berekening doen naar cumulatieve staafjes: er zit wel een kleine onnauwkeurigheid in, maar deze is niet groot genoeg om je er zorgen over te maken.

Manier 1: de staafjes optellen, dus het oppervlakte van de staafjes bij elkaar optellen. Manier 2: het oppervlak onder de curve berekenen. De staafjes zitten soms wat boven de curve en soms wat eronder en dit reken je weg door alleen te kijken naar het oppervlak onder de curve.

Er bestaat wat abstractie boven wat je observeert. Daarom gebruiken we nu Griekse symbolen voor de standaardafwijking en het gemiddelde.

*Eigenschappen normaalverdeling:*

- Ze zijn klokvormig, ze hebben 1 top, ze zijn symmetrisch
- $\mu$  (het gemiddelde) is de locatie van de top, maar zegt verder niks over de vorm.
- $\sigma$  is zichtbaar. Het is de afstand van de top (het gemiddelde) tot het buigpunt.
- Hoe wijder de standaardafwijking, hoe kleiner het gemiddelde. Dit komt omdat het oppervlak van de grafiek altijd 1 is en dus constant. Bij een grotere spreiding, krijg je dus een lagere top en andersom.
- $\mu$  doet verder niks, heeft geen invloed op het maxima, dat doet alleen  $\sigma$ . Dit geeft de vorm van de grafiek aan.

*Hernstein & Murray:*

Deze 2 onderzoekers deden samen bepaalde aannames:

Een hiervan is: IQ is normaal verdeeld en genetisch bepaald. De vraag is of dat zo is. We ontkennen veel dat diefstal enzovoorts genetisch bepaald is, omdat dit voor ons een angstaanjagende gedachte is. Deze aanname is aan te vallen met tweelingonderzoek. Hierbij kun je namelijk stellen dat tweelingen identiek zijn qua erfelijkheid en dus zouden ze hetzelfde moeten doen op het gebied van criminaliteit. Maar dit blijkt in de werkelijkheid niet het geval te zijn.

*Toepassingen normaalverdeling:*

Steekproefvariabiliteit: dit zijn de verschillen in steekproeven. Het zegt iets over de theoretische variabiliteit van de steekproef. De normaalverdeling is hier het model voor.

### Standaardregel:

Dankzij de symmetrie en de standaardafwijking kan je in iedere normaalverdeling kijken naar bepaalde gebieden in de verdeling. Voor de vraag tussen welke x-waardes liggen hoeveel mensen kun je de 68-95-99.7% regel gebruiken.

Tussen 1 standaardafwijking links en 1 standaardafwijking rechts van het gemiddelde ligt 68% van de individuen onder de normaalverdeling; dus 68% van de individuen ligt tussen deze grenzen. Bij 2 standaardafwijkingen links en rechts is dat 95% en bij 3 standaardafwijkingen links en rechts is het 99.7%

Alle normaalverdelingen zijn in principe hetzelfde, omdat we er hetzelfde mee omgaan. Daarom kan er een standaard normaalverdeling worden gemaakt, waarmee je in iedere verdeling kan uitrekenen wat een bepaald stukje in de verdeling is.

### Z-scores:

Z-scores zijn de standaard waarden die bij de standaard normaalverdeling horen. De z-score geeft aan hoeveel standaarddeviatie een score afwijkt van het gemiddelde. Z-scores heb je nodig bij het standaardiseren van data. Bij deze waarde op de x-as hoort een cumulatieve waarde. Deze scores zijn te vinden in de tabel in je werkboek.

Ruwe scores omzetten in z-scores:

$$z = \frac{x - \mu}{\sigma}$$

Van z-scores terug naar ruwe data (ontstandaardiseren):

$$x = z * \sigma + \mu .$$

Bij een standaard normaalverdeling is het gemiddelde altijd 0 en de standaarddeviatie altijd 1. Dus  $N(0,1)$ . Is de z-score negatief, dan ligt deze links van het gemiddelde (gem. is 0). Positieve z-scores liggen rechts van het gemiddelde.

Belangrijk om te onthouden is dat standaardiseren alleen kan als de oorspronkelijke verdeling al een normaalverdeling was. Dit kan je snel zien door de ruwe data te plotten in een histogram en globaal te checken of het histogram ruwweg symmetrisch is en één top heeft (kan je er een density curve door tekenen?). Er verandert dus niets aan de vorm, standaardiseren is alleen om scores met elkaar te kunnen vergelijken ten opzichte van het gemiddelde. Met standaardiseren kun je scores omzetten in kansen en andersom. Met standaardiseren bewerkstellig je dus geen normaalverdeling.

### Tabel A:

Op sheet 17 is een stukje van Tabel A te zien. Als je aan de linkerkant naar beneden gaat in deze tabel zie je de hele cijfers van het getal en wanneer je bovenaan in de tabel naar rechts gaat, kun je het getal aanvullen met eventuele decimalen.

Dus als je bijvoorbeeld bij een z-score van -0.22 een proportie wilt weten, ga je eerst naar -0.2 (dat doe je door naar beneden te gaan in de tabel). Daarna zoek je de decimalen hierbij, dus ga je naar 0.02 en kom je op een proportie van 0.4129.

### Normal Quantile plot:

Deze plot is bedoeld om na te gaan of je data normaal verdeelt zijn. Je kunt dit nagaan door je data en de bijbehorende z-scores te plotten in een normal quantile plot. Wanneer deze plot een rechte lijn weergeeft, heb je te maken met een normaalverdeling. Je berekent de relatieve frequenties.

*Karl Pearson:*

Hij ontdekte de product- moment correlatie. Hij is de vader van de correlatie coëfficiënt. Dit is de mate waarin wij samenhang tussen variabelen kunnen aangeven.

Hij was ook erg geïnteresseerd naar regressie (week 7)

*Samenhang:*

Als er sprake is van samenhang wil dat niet meteen zeggen dat er ook sprake is van een oorzaak – gevolg relatie. De twee variabelen die samenhangen kunnen ook nog beïnvloed worden door een onbekende derde factor.

*Afhankelijkheid:*

Bij afhankelijkheid is er wel sprake van een oorzaak – gevolg relatie. De ene onafhankelijke variabele (de x-variabele) bepaalt dan de andere afhankelijke uitkomstvariabele (de y-variabele).

*Meetniveaus:*

Kwantitatief meetniveau is te vergelijken met het interval-meetniveau en hoger.

Kwalitatief meetniveau is te vergelijken met een nominaal meetniveau. Het is ook een categorisch meetniveau. Je classificeert scores.

*Scatterplots:*

Als je veel variabelen hebt bij dezelfde cases is het handig om een scatterplot (puntenwolk) te maken.

Bij een scatterplot zijn er drie dingen belangrijk:

- 1) Richting: de richting is positief als lage waarden op de x-as samenhangen met lage waarden op de y-as en hoge waarden met hoge. De richting is negatief als hoge scores op de x-as samenhangen met lage scores op de y-as en andersom. Dat de richting negatief is wil niet zeggen dat er geen verband is.
- 2) Sterkte: hoe meer de punten op één lijn liggen, hoe sterker het verband tussen de variabelen.
- 3) Vorm: lineair of niet-lineair. Zijn er clusters te ontdekken (heterogeen) of liggen alle punten bij elkaar (homogeen)?

Als er geen verband is tussen de variabelen kan je uit een scatterplot niks aflezen. Zie voor voorbeelden sheet 12.

Uitbijters liggen ook bij scatterplots buiten de puntenwolk, het zijn punten die duidelijke afwijkingen van het algemene patroon. Een uitbijter kan een grote invloed hebben op de correlatie, maar je kan ze niet zomaar uit je data set halen!

Bij het observeren van een scatterplot moet je altijd in je achterhoofd houden welk meetniveau je aan het bekijken bent en wat je daaruit kan concluderen.

*Covariantie:*

De covariantie is een maat voor sterke samenhang: als we de ene variabele weten, kunnen we dan de andere variabele voorspellen? 'educative guess'

Herhaling: bij variantie kijken we naar hoeveel een score afwijkt van het gemiddelde. Het gaat dan om de spreiding van alle individuen. Dit is echter niet geschikt voor 2 variabelen, dus als je werkt met 2 gemiddelden en 2 standaardafwijkingen.

De formule voor covariantie zie je op sheet 21. Het gaat bij covariantie om een kruisproduct, daarom hoef je de getallen niet te kwadrateren. Covariantie kan ook negatief zijn.

Met de covariantie kan je alle mogelijke eenheden gebruiken. Dit maakt het lastig om covarianties onderling met elkaar te vergelijken. Correlatie is het gestandaardiseerde model van de covariantie.

Pearson correlatie: de maat voor lineaire samenhang tussen 2 variabelen op interval niveau. De correlatiecoëfficiënt wordt aangegeven met 'r'. De correlatiecoëfficiënt is dus afhankelijk van de mate waarin elk individu afwijkt van het gemiddelde.

Je deelt de deviatie door de standaarddeviatie. Dit proces heet standaardiseren. Dit standaardiseren kwamen we ook tegen bij het berekenen van de z-scores, waardoor je die nu weer in de formule kunt zetten. De formule is:  $R_{xy} = \frac{\sum z_x z_y}{n-1}$ . De correlatie ligt altijd tussen de -1 en 1. Bij 0 is er geen sprake van correlatie.

Als je in een scatterplot kijkt en een lineaire verdeling ziet, doe je een aanname van een bepaalde meetnauwkeurigheid. Bij een getal is het veel betrouwbaarder weergegeven. R is niet robuust (bestand) tegen uitbijters, waardoor je er ook voor moet zorgen dat deze uitgebannen worden.

$$R_{xy} = \frac{\text{covariantie}}{S_x S_y}$$

Criteria voor causaliteit:

- variabelen moeten covariëren of correleren
- discriptionality: de oorzaak gaat in de tijd aan het gevolg vooraf
- alternatieven verklaringen zijn uitgesloten (dit is altijd lastig om te bepalen)

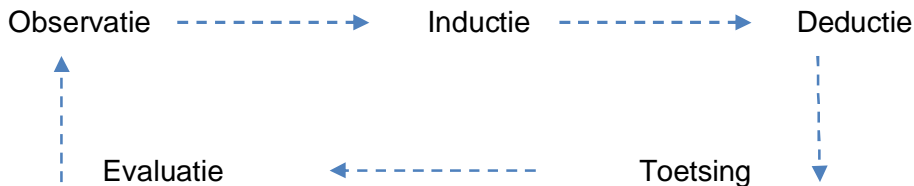
Als er sprake is van correlatie is er nog niet automatisch sprake van causatie. Correlatie bestaat alleen bij lineair verband!!



# Werkgroepen

## Week 1

Hieronder is de empirische cyclus te vinden. Het begint bij het waarnemen van verschijnselen, waarna je je hypothese gaat vaststellen:



Een voorbeeld van *inductie* is het volgende: *je loopt over straat en ziet een witte zwaan. Daarna zie je er weer een. Wanneer je een paar witte zwanen hebt gezien zeg je tegen jezelf: alle zwanen zijn wit. Dit is inductie. Je gaat bij inductie van verschijnselen ( de witte zwanen) naar een hypothese (alle zwanen zijn wit).* De hypothese die je bij inductie ontwikkeld moet falsificeerbaar zijn. Dit betekent dat het mogelijk moet zijn om de stelling aan te vechten. In het bovenstaande voorbeeld doe je dat door bijvoorbeeld op zoek te gaan naar een zwarte zwaan.

Bij deductie gaat het om concrete begrippen. Conceptuele begrippen zijn te abstract ( woordenboekbegrippen) en zeggen dus niks over hoe we ze kunnen meten. Operationele begrippen kunnen we wel meten, en zijn ook navolgbaar. Als iemand anders de hypothese wil testen, moet deze persoon dus precies weten wat er gedaan is.

Variante:

Variabiliteit is de mate waarin een variabele varieert (schommelt) om een bepaalde waarde. Variantie is de meetkundige maat voor variabiliteit.

De totale variantie = systematische variantie + foutenvariantie

Dit wordt ook wel als volgt aangeduid:  $SS_{\text{totaal}} = SS_{\text{tussen}} + SS_{\text{binnen}}$

*Variante en de bijbehorende formules:*

De formule voor variantie is:  $\frac{SS_{\text{totaal}}}{n-1}$

$SS_{\text{totaal}} = \sum (y_{ij} - \bar{y})^2$

$n$  = het aantal observaties of het aantal personen.

$SS$  = de som van de gekwadeerde afwijkingen

$\bar{y}$  = groot gemiddelde. Dit bereken je door alle scores te delen door het aantal scores. Het gemiddelde wordt aangegeven met het streepje boven de  $y$ .

$y_{ij}$  = de individuele score. Hierbij is  $U_i$  de groep en  $U_j$  het individu

$\sum$  = het Griekse symbool dat we gebruiken als somteken.

Om de variantie te berekenen moet je dus eerst van alle individuele observaties het groot gemiddelde afhalen. Deze kwadrateer je, omdat je anders bij het optellen altijd uitkomt op 0. Na het kwadrateren tel je ze bij elkaar op en deel je ze door  $(n-1)$  Bij 10 individuen worden de scores dus gedeeld door 9.

$SS_{\text{totaal}}$ : individuele scores verschillen rondom het gemiddelde

$SS_{\text{binnen}}$ : individuele scores verschillen binnen elke groep

$SS_{\text{tussen}}$ : groepsgemiddelden verschillen tussen groepen

$VAF$  ( de proportie verklaarde variantie) =  $SS_{\text{tussen}} / SS_{\text{totaal}}$

**Antwoorden opdrachten:**

Opdracht 1.1.a:

In de onderstaande tabel is te zien in welke paragrafen van het werkboek de volgende componenten uit de empirische cyclus te vinden zijn.

1	Observatie	9	Toetsing+evaluatie
2	Inductie	10	Inductie
3	Toetsing	11	Toetsing
4	Inductie	12	Inductie
5	Inductie	13	Deductie
6	Deductie	14	Toetsing
7	Inductie	15	Evaluatie
8	Deductie+toetsing	16	Evaluatie

- c) in alinea 4 en 7 worden conceptuele definities gebruikt
- d) In alinea 6 + 13 worden operationele definities gebruikt

Opdracht 1.2.

b) Voor  $SS_{\text{totaal}}$  moet je eerst het groot gemiddelde van de individuele score aftrekken en deze uitkomst kwadrateren.  $SS_{\text{totaal}}$  is het totaal van al deze kwadrateringen en is in dit geval 0.34

c)  $0.34 / (9-1) = 0.0425$

g) De proportie verklaarde variantie is het verklaarde/ te verklaren, dat is  $SS(\text{tussen})/SS(\text{totaal})$ ,  $0.24/0.34 = 0.706$

0.706 is erg veel, want 0.15 wordt al als veel gezien.

**Week 2**

Herhaling week 1: operationaliseren betekent dat we een begrip zo definiëren dat we het kunnen meten.

Verschillende soorten onderzoek:

- experimenteel: hierbij gaat het om het onderzoeken van een causaal verband in random groepen.
- correlatieel: verband tussen twee variabelen ( niet per se oorzaak – gevolg)
- quasi-experimenteel: hetzelfde als experimenteel, met als enige verschil dat je hierbij niet zelf je onderzoeksgroep kan samenstellen.

Variabele: datgene dat verschillende waarden aan kan nemen. Deze moet aan twee eisen voldoen: moet uitputtend zijn ( alle waarden die mogelijk zijn moeten voorkomen) en het moet wederzijds uitsluitend zijn ( iemand kan niet in twee categorieën vallen).

Uitputtend		Niet uitputtend	
1	<6	1	1-6
2	7 t/m 22	2	6 t/m 22
3	23t/m 50	3	22 t/m50
4	>50	4	50t/m70

*Meetniveaus:*

1. Nominaal: dit zijn categorieën. Een voorbeeld van een nominaal meetniveau is de oogkleur, maar ook rugnummers van voetballers. Er zit geen volgorde in.
2. Ordinaal: hierbij vindt men wel een volgorde. Een voorbeeld hiervan is het maken van een rijtje met je gymklas van klein naar groot. Je ziet dan wel dat er een volgorde is, maar de verschillen in grote zijn niet gelijk tussen alle leerlingen. De intervallen zijn dus niet gelijk. Een ander voorbeeld is een kledingmaat; het verschil tussen xs en s is niet hetzelfde als dat tussen m en l.
3. Interval: dit is een volgorde met gelijke tussenruimtes. Er is hierbij geen absoluut nulpunt. Een voorbeeld hiervan is IQ. A is niet 2 x zo groot als B, want er is geen nulpunt, maar het verschil tussen een IQ van 70 en 80 is evenveel als het verschil tussen een IQ van 130 en 140
4. Ratio: dit is een interval met een absoluut nulpunt. Een voorbeeld hiervan is de inhoud in centiliter. Hierbij is een absoluut nulpunt te vinden en de intervallen zijn gelijk. Dan spreek je dus van ratio. Een absoluut nulpunt houdt in dat het nulpunt bij 0 ligt en er geen negatieve getallen mogelijk zijn.
5. Absoluut: Het nulpunt en de eenheid liggen vast, hier mag je zelf niks meer aan doen. Dit is het hoogste meetniveau. Voorbeeld hiervan zijn percentages.

*Centrale tendenties:*

De modus is de meest voorkomende waarneming.

Kijk bijvoorbeeld naar de volgende getallenreeks: **2.8.8.8.6.5.4.3.6.6.8.8**

8 is hier de modus.

De mediaan is de middelste waarneming. Hiervoor moet je alles eerst op volgorde zetten.

Kijk bijvoorbeeld naar de volgende getallenreeks: **2.4.6.9.12**

6 is hier de mediaan

Bij de volgende getallenreeks zie je 2 middelste getallen: **2.4.6.7.9.12**

6.5 is dan de mediaan (  $6+7= 13$ ,  $13/2= 6,5$ )

Bij een nominaal meetniveau hoort de modus

Bij een ordinaal meetniveau hoort de mediaan (+modus)

Bij een interval hoort het gemiddelde (+mediaan +modus)

Bij ratio hoort het gemiddelde( +mediaan+modus)

Met interval en ratio kun je gaan rekenen, omdat je hierbij het gemiddelde kunt berekenen.

*Betrouwbaarheid en validiteit:*

Betrouwbaarheid geeft dezelfde uitkomst bij het herhalen van dezelfde meting. Een voorbeeld hiervan is de weegschaal: je stapt erop en weegt 60 kg. Als je er een minuut later weer op gaat staan en hij geeft nu ineens 70 kg aan, dan weet je dat er iets mis is met de weegschaal.

Validiteit is het meten van een meetinstrument wat het klinkt te meten. Het voorbeeld hierbij is weer de weegschaal: Meet de weegschaal het goede gewicht? Meet het instrument wat hij behoort te meten?

#### *Betrouwbaarheid:*

Parallelbetrouwbaarheid is de betrouwbaarheid waarbij je 2 parallel testen afneemt om de meetresultaten te controleren. De testen lijken op elkaar. Een probleem wat hierbij kan optreden is dat ze misschien niet parallel zijn. Een ander probleem kan zijn dat 1 van de 2 testen onbetrouwbaar is.

Inter-item betrouwbaarheid is de betrouwbaarheid waarbij je nagaat of de items in je vragenlijst wel allemaal hetzelfde meten. Je gaat bijvoorbeeld kijken of de vragen die je in een vragenlijst hebt geplaatst wel allemaal onzekerheid meten, in plaats van dat bijvoorbeeld een vraag daarentegen depressiviteit meet.

Splithalf betrouwbaarheid is het vergelijken van 2 delen van een onderzoek. Je kunt hierbij bijvoorbeeld de oneven en even vragen met elkaar vergelijken. Je verdeelt het onderzoek random in tweeën. Belangrijkste voorbeeld is Chronbach's Alpha ( week 3).

#### *Validiteit:*

Face validiteit: men heeft hierbij de indruk dat de test valide is. Dit zegt weinig.

Inhoudsvaliditeit: het hele domein moet gemeten worden. Is het een goede afmeting? Er wordt gekeken naar de inhoud van het onderzoek.

Een voorbeeld hiervan is de geschiedenis toets: je moet leren van de jaartallen 1000 tot 2000 en alleen het tijdsinterval 1900 tot 2000 wordt gevraagd. Wanneer dit het geval is, is het onderzoek niet valide.

Criteriumvaliditeit: meet het instrument wat het wil meten? Criteriumvaliditeit is onder te verdelen in 2 componenten:

- Concurrente validiteit: dit wordt tegelijk gemeten met de afname van het onderzoek. 'concurrerende onderzoeksmethoden met elkaar vergelijken'.
- Predictieve validiteit: hierbij wordt geprobeerd een voorspelling te maken aan de hand van het onderzoek.

Constructvaliditeit: deze vergelijkt 2 testen in plaats van een criterium. (het vergelijkt het instrument met de test).

- Discriminante validiteit: testen die iets anders meten moeten laag correleren.
- Convergente validiteit: testen die hetzelfde meten correleren hoog

#### **Antwoorden opdrachten:**

##### Opdracht 2.1.b:

In de onderstaande tabel staat van iedere variabele het meetniveau vermeld.

1	Nominaal
2	Ratio
3	Nominaal
4	Ratio / Ordinaal
5	Nominaal
6,7	Ordinaal
8	Ratio
9 t/m 19	Ordinaal

##### Opdracht 2.1.c: waarom is vraag 5 nominaal?

Bij vraag 5 is de volgorde onduidelijk. Je kunt beter de vraag omzetten tot een ordinale variabele.

Opdracht 2.1.d: wat is het probleem van vraag 4 t/m 8?

Het zijn teveel variabelen. Je kunt het beste alles 1 eenheid geven. Het resulterende niveau wordt het rationiveau.

Opdracht 2.1 e: hercoderen naar 1=romantisch, 2=twijfelachtig, 3=niet-romantisch. Optie 4 is de enige neutrale. Let op: niet alle vragen wijzen dezelfde kant op ( ompolen).

Opdracht 2.1 g: er zijn altijd mensen die een vragenlijst klakkeloos invullen. Om dit tegen te gaan zijn niet alle waardes aan elkaar gelijk, daardoor moeten mensen blijven lezen. Dit trucje heet ompolen.

Opdracht 2.2:

In de onderstaande tabellen is terug te vinden in welke alinea welke vorm van betrouwbaarheid/validiteit staat.

Test- hertest betrouwbaarheid	Alinea 5/6
Paralleltest betrouwbaarheid	Alinea 4/6
Inter-item betrouwbaarheid	Alinea 7/10
Repliceerbaarheid hele onderzoek	Alinea 10/11

Face validiteit	Alinea 1
Inhoudsvaliditeit	Alinea 2
Criterium validiteit	Alinea 14/18
Construct validiteit	Alinea 5,6, 13
Statistische validiteit	Alinea 6
Interne validiteit	Alinea 8/9
Externe validiteit	Alinea 4,10,12

**Week 3**

Interne consistentie: alle items moeten onderling correleren. Als er een item tussen zit dat niet hetzelfde meet als de rest, moet deze eruit om de betrouwbaarheid en validiteit van de test te verhogen.

Cronbach's alpha is een methode om de interne consistentie te bepalen. Dit is het gemiddelde van de scores op alle items.

Als je de somscore wilt berekenen, dan ga je na wat iedereen heeft gescoord op een schaal. Hierbij moet je wel eerst de vragen die andersom zijn gesteld ompolen.

$$\text{Variantie} = \frac{\sum(u_{ij} - \bar{u})^2}{n-1}$$

$$Sx^2 \hat{=} \frac{(\text{gem.vraag 1} - \text{afzonderlijk antwoord})^2}{n-1}$$

Correlaties: ligt altijd tussen de -1 tot 1. Is de score 0, dan is er geen samenhang. Als de score -1 of 1 bedraagt is er een sterke samenhang. -1 betekent: hoe hoger op het ene, hoe lager op het andere. 1 is dezelfde richting. De correlatie is een rekenkundige maat van lineaire samenhang.

Gedragsonderzoek is descriptief, het is beschrijvend onderzoek. Je verzamelt feiten. Een voorbeeld van een populatie is alle psychologistudenten (uit het hele land). Vaak is een populatie erg groot.

*Onderzoeken:*

Surveys: vragenlijst, interview

Demografisch: statistisch onderzoek

Epidemiologisch onderzoek: voorkomen (plaatsvinden) van ziekten.

*Manieren van onderzoeken:*

Cross-sectioneel is eenmalig onderzoek.

Opeenvolgende onafhankelijke steekproeven: dit zijn 2 of meer steekproeven. Je gaat deze steekproeven met elkaar vergelijken. Het gaat namelijk om verschillende mensen. De onderzoeken moeten wel op elkaar lijken.

Longitudinaal: je onderzoekt hierbij 1 groep op meerdere tijden. Dit kan voor problemen zorgen als er mensen uitvallen. Ze kunnen bijvoorbeeld geen zin meer hebben in het beantwoorden van al die onderzoeken.

Internet surveys: dit is makkelijk en goedkoop. Je weet hierbij nooit wie er in je steekproef zit.

**Antwoorden opdrachten:**

Opdracht 3.1.a:

In de onderstaande tabel zijn de itemvarianties te vinden.

	Gemiddelde	SS(totaal)	variantie
Item 1	$(3+6+3)/3=4$	$(3-4)^2 + (6-4)^2 + (3-4)^2$ = 6	$6/(3-1) = 3$
Item 2	$(4+6+5)/3=5$	$(4-5)^2 + (6-5)^2 + (5-5)^2$ = 2	$2/(3-1)= 1$
Schaal 7,12,8	$(7+12+8)/3= 9$	$(7-9)^2 + (12-9)^2 + (8-9)^2$ = 14	$14/(3-1) = 7$

Opdracht 3.1.c:

Hieronder is de berekening van Cronbach's alpha te vinden.

$$\alpha = \frac{k}{k-1} * \left[ 1 - \frac{\sum S_i^2}{S^2_x} \right]$$

$$\alpha = 2/(2-1) * (1 - (3+1)/7) = 0.86$$

Opdracht 3.2.b:

Hoe is het probleem van de samengevoegde gegevens over de 10 vragen op te lossen?

Voor dit probleem zullen de items die de verkeerde kant op staan eerst moeten worden omgepold.

Dit zijn de items 1,2,10. Hierbij is het eens zijn met de vraagstelling van de tegenovergestelde betekenis als bij de andere items.

Opdracht 3.2.f:

Hieronder is de berekening van alpha te vinden.

$$\alpha = \frac{k}{k-1} * \left[ 1 - \frac{\sum S_i^2}{S^2 x} \right]$$

$s^2 x = 54.480$ , want dit is gegeven.

$$\sum S_i^2 = 21.430$$

Cronbach's alpha is dan gemakkelijk uit te rekenen door de gegevens in te vullen. Bij k vul je 10 in, want er zijn 10 items. Dan kom je op een Cronbach's alpha van 0.674.

Opdracht 3.2.g

Wat is de definitieve conclusie over het item dat verwijderd kan worden?

De Cronbach's alpha is 0.674. In kolom 5 in de tabel in het werkboek staat namelijk dat als je item 10 eruit haalt deze dan 0.747 wordt. Dit is hoger dan de oorspronkelijke alpha dus kun je item 10 er definitief uithalen.

Belangrijk is wel om te onthouden dat naarmate een testschaal meer items heeft, deze betrouwbaarder wordt.

**Week 4**

Opdracht 4.1.a:

Wat kun je allemaal zeggen over de verdeling?

De modus is 23, want dit getal komt het vaakst voor

De mediaan is 22. Dit is nummer 38 van de 75 waarnemingen als je alles op volgorde zet.

De range is de hoogste waarneming – de laagste waarneming, dat is  $41-16= 25$

41 is misschien een uitbijter, maar dat weet je niet zeker.

Opdracht 4.1.b:

Bij welke aspecten kun je er minder goed iets over zeggen en bij welke aspecten beter?

Als de personen op volgorde staan is het veel lastiger om er iets over te zeggen, omdat je dan de hele volgorde van hun scores nog moet bepalen. Met de scores op volgorde gaat dat dus beter.

$f$  = absolute frequentie. Dit is het aantal personen/geheel van personen

$P$  = relatieve frequentie. Dit is de percentages/ proporties

$F$  = cumulatieve frequentie. Hierbij moet je de waarnemingen optellen. Dit loopt op tot maximaal 1, als je alles bij elkaar hebt opgeteld. Je kijkt naar de scores en die daarboven tel je erbij op. Dus alles wat voorafging aan de score waarvan je de cumulatieve frequentie wilt weten, tel je erbij op. De cumulatieve absolute frequentie houdt in dat je de absolute frequenties bij elkaar optelt.

Opdracht 4.2:

Wat kun je zeggen over de gegevens wanneer je een frequentietabel hebt gemaakt voor de gegevens?

In de frequentietabel geef je ook de 0 weer. Dat is voor het overzicht. Het plaatje wordt zo compleet gemaakt. De modus is door de tabel goed te zien: 23

Scheefheid kun je zo ook zien. Eerst heb je hoge scores, dan lagere scores en dan waarschijnlijk nog een uitbijter.

De gegroepeerde frequentie tabel bevat intervallen die exclusief zijn en even groot.

Aantal klassen ( $k$ ) =  $\sqrt{n}$  ( $n$  = aantal mensen)

Intervalbreedte= (i) =  $\frac{R(\text{range})}{k(\text{aantal klassen})}$ . Je moet alles afronden naar boven op een geheel getal

Dan kun je met die intervalbreedte de gegroepeerde frequentietabel maken.

Opdracht 4.3.a:

Hoe groot worden de intervallen als je een gegroepeerde tabel gaat maken?

$k = \sqrt{75}$ , dat is ongeveer 9

De range is  $41-16=25$

Het klasseninterval =  $\frac{r}{k}$ , dat is  $\frac{25}{9}$  en dat is ongeveer 3. De intervallen krijgen dus een intervalbreedte van 3.

Opdracht 4.3.b:

Wat zijn  $K$  en  $I$  bij 320 mensen tussen 20 en 65 jaar?

$K$  is  $\sqrt{320}$  en is dus ongeveer 18.  $I = \frac{r}{k} = \frac{45}{18}$  en dat is ongeveer 3

Opdracht 4.3.c:

Maak de gegroepeerde tabel.



Je doet dit door de intervalbreedte 3 te maken. Dus het eerste interval is 16 t/m 18. De frequentie hiervan is 14. Je komt aan de intervalbreedte via opdracht 4.3.a. De frequentie vind je door alle scores van 16 t/m 18 bij elkaar op te tellen.

Het tweede interval is 19 t/m 21. De frequentie is 18. Zo kun je doorgaan totdat je iedere score in een interval hebt geplaatst. Dat doe je steeds in groepjes van 3 (dat is de intervalbreedte).

Opdracht 4.3.d:

Probeer iets te zeggen over de gegroepeerde tabel.

De modus vindt je dan bij 22 t/m 24. In deze groep zit het meest voorkomende getal.

De symmetrie/scheefheid is naar rechts. En waarschijnlijk is er ook nog een uitbijter(41).

De spreiding is hier iets groter omdat het laatste interval 40 t/m 42 is. 42 komt eigenlijk niet voor in deze data maar omdat het een gegroepeerde tabel is lijkt het alsof deze wel voorkomt.

Opdracht 4.4.a:

Wat is het verschil tussen een histogram en een staafdiagram? Bij een histogram zit er geen ruimte tussen de staven en is het meetniveau interval/ratio. Het maakt hierbij uit wat de volgorde is. Bij een staafdiagram zit er ruimte tussen de staven. Het is een categorische variabele. Dit is ook wel nominaal/ ordinaal en het maakt niet uit waar de staven staan. Een histogram is voor deze scores het beste, omdat de volgorde betekenisvol is.

Opdracht 4.4.d:

Op welke manieren komen in een histogram de frequenties tot uitdrukking?

De oppervlakte hangt samen met de frequentie: hoe groter het oppervlak, hoe groter de frequentie.

Opdracht 4.5.a:

Arceer het oppervlak dat hoort bij de cumulatieve verdeling van klassengrens 3.

De eerste 3 intervallen is 17 t/m 23. Je arceert dus t/m 23. Je leest in de vorige tabel af wat de hierbij horende cumulatieve relatieve frequentie is en arceert dit.

Opdracht 4.5.b:

Controleer het door een berekening.

De relatieve frequentie is  $14+18+20 / 75 = 0.693$ . Dit geldt bij klassengrens 3. Dus is dit ongeveer 70 % van het geheel. Als je het arceert zie je dat dit klopt.

Opdracht 4.5.d:

Lees Q1, Q2 en Q3 af.

Q1 zit op 25%. Dit is ongeveer bij 19 (crf)

De mediaan zit op 50%. Dit is ongeveer bij 22 (crf)

Q3 zit op 75%. Dit is ongeveer bij 25 (crf)

Opdracht 4.5.e:

Wat zet je op de x-as en wat op de y-as? Kun je een frequentieverdeling van de variabele religie cumulatief maken?

Op de x-as zet je de klassengrens, op de y-as zet je de cumulatieve relatieve frequentie.

Religie is op een nominaal meetniveau, dus niet op te tellen. Pas vanaf ordinaal kan je het optellen.

*Boxplot:*

Een boxplot maak je met de five number summary: min, eerste kwartiel (Q1), mediaan(Q2), derde kwartiel(Q3), max.

$IQR = 3e \text{ kwartiel} - 1e \text{ kwartiel}$

$Q1 - 1,5 * IQR < \text{Uitbijter} > Q3 + 1,5 * IQR$ . Als de scores voldoen aan de volgende eisen, dan zijn het uitbijters. Dus als de score lager is dan  $Q1 - 1,5 * IQR$  of als de score hoger is dan  $Q3 + 1,5 * IQR$ .

Opdracht 4.6.

Bepaal five number summary.

1e kwartiel =  $(1+37)/2 = 19$ ste = 19

Mediaan =  $(75+1)/2 = 38$ ste = 22

3e kwartiel = 25

Minimum = 16

Maximum = 41

Range:  $25 - 19 = 6$ .  $1,5 * IQR: 1,5 * 6 = 9$

$19 - 9 = 10 < \text{Uitbijter} > 34 = 25 + 9$ .

Dus 41 is werkelijk een uitbijter

Je maakt dus een boxplot waarbij je 41 er buiten plaatst. Je doet dit door deze als een sterretje of een stipje buiten de boxplot te plaatsen.

**Week 5***Normaalverdelingen:*

De kant van de staart is de kant waar de grafiek scheef naartoe gaat. Zit de top dus meer naar links en loopt de staart rechts uit, dan is de grafiek scheef naar rechts.

Het oppervlak onder de curve van een normaalverdeling is 1. Als de grafiek symmetrisch is, zijn de modus, de mediaan en het gemiddelde hetzelfde. De normale verdeling =  $N$  (gemiddelde, standaarddeviatie).

Het gemiddelde is de plek waarop de x as het midden is.

Een curve berekent nooit een 0-punt.

*Z-scores:*

Z-scores: hierbij is het gemiddelde 0 en de standaarddeviatie is 1.

Z geeft het aantal standaarddeviaties boven of onder het gemiddelde aan. Een positieve z-score is een score boven het gemiddelde en een negatieve z-score is een score onder het gemiddelde.

Normale verdeling:  $N(\mu, \sigma)$  standaardnormaal:  $N(0,1)$

Een standaardnormale verdeling: hierbij kan je de cumulatieve proporties zien. Proporties die beneden of boven een bepaalde waarde liggen. Proporties liggen altijd links van de grafiek. Tabel A: z-scores staan langs de lijn. Binnenin de tabel vindt je de cumulatieve proportie.

## Opdracht 5.1.a.b:

Schrijf in de tabel de F- bovengrens erbij en de klassenproporties.

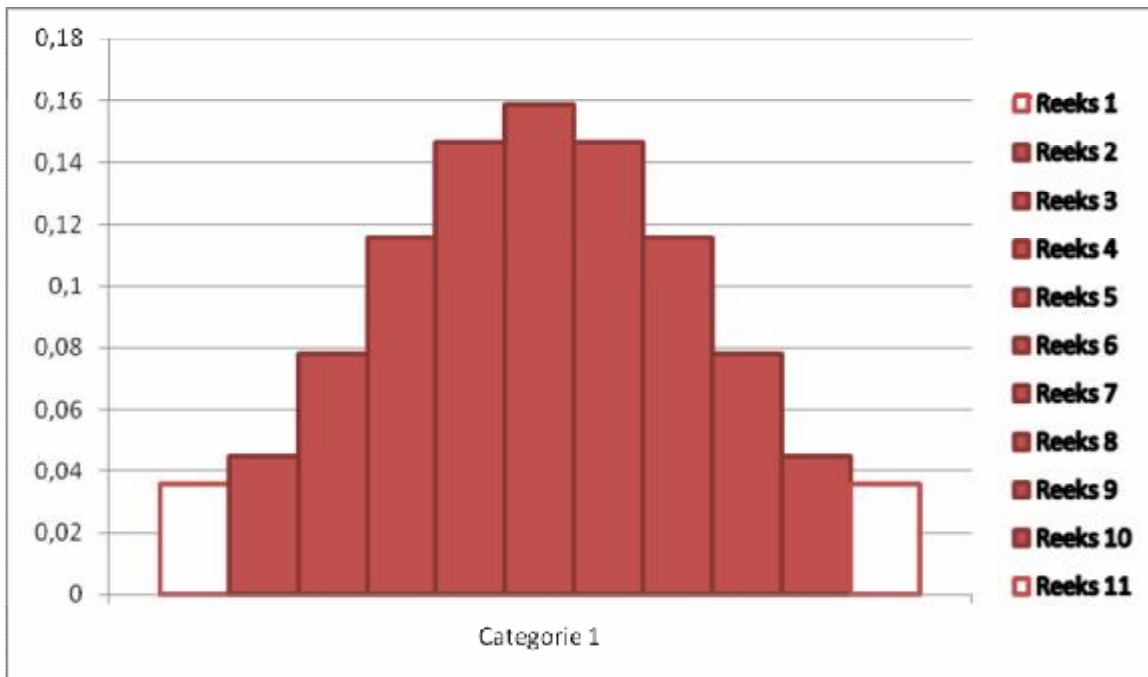
F- bovengrens: dit is de proportie van het meest rechtse getal in de eerste kolom. Je vind deze door in tabel A te kijken. De klassenproportie is de bovengrens- ondergrens. Je moet dan uiteindelijk op 1 uitkomen.

Z-intervallen	F(bovengrens)	Klasseproporties
$z \leq -1,80$	0.0359	0.0359
$-1.80 < z \leq -1.40$	0.0808	0.0449
$-1.40 < z \leq -1.00$	0.1587	0.0779
$-1.00 < z \leq -0.60$	0.2743	0.1156
$-0.60 < z \leq -0.20$	0.4207	0.1464
$-0.20 < z \leq 0.20$	0.5793	0.1586
$0.20 < z \leq 0.60$	0.7257	0.1464
$0.60 < z \leq 1.00$	0.8413	0.1156
$1.00 < z \leq 1.40$	0.9192	0.0779
$1.40 < z \leq 1.80$	0.9641	0.0449
$1.80 < z$	1	0.0359
Totaal		1.0000

## Opdracht 5.1.c:

Teken een histogram van de klassenproporties.

Als je goed kijkt naar de tabel met de klassenproporties, moet iets opvallen. Het 6<sup>e</sup> getal (het middelste getal) is het hoogste getal van allemaal. En daaromheen zie je dat er steeds op evenveel afstand van dit hoogste getal het zelfde getal staat. Dus 1 stap van 0.1586 staat zowel omhoog als omlaag het getal 0.1464, enzovoorts. Als je dit gaat tekenen, krijg je dan ook een parabool met 1 top. Het is een normaalverdeling. Je kunt voor het eerste en voor het laatste getal niet aangeven waar dit stopt, en daarom hou je deze staven open. Je kunt dit als volgt doen: je geeft met de lege rode rechthoeken aan dat reeks 1 en reeks 11 anders zijn dan de rest. Hiervan weet je niet waar ze stoppen.



Het is een symmetrische verdeling, dus de modus, de mediaan en het gemiddelde zijn 0.

Opdracht 5.2.a:

Hoeveel personen zullen een score X hebben?  $75 \leq X \leq 120$

Je hebt totaal 800 eindexamenkandidaten. Hun gemiddelde is 100( $\mu$ ) en hun standaardafwijking is 15( $\sigma$ ).

De formule die je hierbij gebruikt:  $z = \frac{x - \mu}{\sigma}$ .  $z(75) : \frac{75 - 100}{15} = -1.67$

$z(120) : \frac{120 - 100}{15} = 1.33$

Als je deze z-scores met tabel A vergelijkt, dan krijg je de bijbehorende proporties. Bij  $z = -1.67$  hoort een proportie van: 0.0475. bij  $z = 1.33$  hoort een proportie van: 0.9082.

$0.9082 - 0.0475 = 0.8607$ , dus 86% van de leerlingen.  $800 * 0.86 = 688$  leerlingen. Je moet naar beneden afronden.

Opdracht 5.2.b:

Tussen welke grenzen liggen de 80% minst extreme personen?

Het gaat om een normaal verdeling, dus de 20% extreemste personen liggen aan de zijkant, waarvan 0.1 aan de linkerkant extreem en aan de rechterkant ook 0.1.

Bij een proportie van 0,1 hoort een z van -1.28, bij een proportie van 0,9 hoort een z van 1.28

-1.28:  $100 + (-1.28 * 15) = 80,8$ ,

1,28:  $100 + (1,28 * 15) = 119,2$

Dus de 80% minst extreme personen liggen tussen 80,8 en 119,2

Opdracht 5.2.c:

Wat is de laagste score van de 200 hoogst scorende personen.

200 hoogst scorende personen.  $200/800 = 0.25$  (dit is de proportie). Rechts van 0.75 heb je dus de 200 hoogst scorende personen, want de cumulatieve proportie gaat van links naar rechts. Z bij 0.75 is 0.67

$X = 100 + 0,67 * 15 = 110,05$ . De laagste score is 111. Je moet dit naar boven afronden.

Opdracht 5.2.d:

Met welk percentiel komt de scorewaarde  $X = 80$  overeen?

$z = (80 - 100) / 15 = -1.33$ . Bij een z van -1.33 hoort een proportie(F) van 0.0918

$0.0918 * 100 = 9.18\%$ , dus het 9,18<sup>e</sup> percentiel. Dat is ongeveer het 9<sup>e</sup> percentiel.

Opdracht 5.2.e:

Wat is de scorewaarde van het 80<sup>e</sup> percentiel?

80<sup>e</sup> percentiel, dus de proportie(F) =0.8. Hierbij kun je in tabel A een z vinden van 0.84.

$$x = 100 + 0.84 \cdot 15 = 112.6$$

112,6 is dus de scorewaarde van het 80<sup>e</sup> percentiel.

Opdracht 5.2.f:

Welke scorewaarde heeft het 80<sup>e</sup> percentiel als bij iedere score 10 wordt opgeteld?

Als bij iedere score 10 wordt opgeteld, schuift de hele verdeling 10 punten omhoog.  $112,6 + 10 = 122,6$

122,6 is dan de nieuwe scorewaarde van het 80<sup>e</sup> percentiel.

Opdracht 5.3.b:

Met hoeveel moet het gemiddelde worden opgehoogd om de norm te bereiken?

500 of hoger moet de bovenste 40% hebben. 0.6 zit er nog onder (linkerkant van de grafiek). De

proportie is 0.6. Dit geeft een z-score van 0.25.  $z = 500 - \mu / 14 = 0.25$ .  $\mu = 496,5$ . Het gemiddelde is nu

dus 496,5 en moet:  $496,5 - 480 = 16,5$  punten omhoog.

Opdracht 5.4.a:

Wat is het gemiddelde?

Het gemiddelde is 17

Opdracht 5.4.b:

Wat is de variantie en wat is de standaardafwijking?

De variantie is 49. Dit kun je berekenen met de formule voor de variantie uit de voorafgaande weken.

De wortel trekken uit de variantie geeft de standaardafwijking: 7

Opdracht 5.4d/e:

Wat zijn: het gemiddelde en de standaardafwijking van de z-scores?

Het gemiddelde van de z-scores is 0 (dit betekent dat hij normaal verdeelt is). De

standaarddeviatie=1.

Opdracht 5.4.g:

Hangt af van de oorspronkelijke verdeling, als deze een normaalverdeling was zijn de z-scores ook normaal verdeeld. Het maken van z-scores bewerkstelligt echter geen normaalverdeling als de ruwe data dit niet was.

Stappenplan voor het rekenen met z-scores:

- 1) Teken en  $\hat{a}$  wel percentage wil je weten?
- 2) Standaardiseer
- 3) Fwaarden (proporties) berekenen
- 4) Bereken oppervlakte
- 5) Conclusie  $\hat{a}$  x aantal personen

**Week 6***Meetniveaus:*

Het categorisch meetniveau is hetzelfde als het nominale en het ordinale meetniveau.

Het kwantitatieve meetniveau is hetzelfde als het interval en ratio.

## Opdracht 6.1.1.a en b:

Voorzie alle scores van een + (hoger dan gemiddeld) of een – (lager dan gemiddeld). Als het gelijk aan het gemiddelde is, noteer je een 0. Noteer ook de som van de plussen en minnen.

In de onderstaande tabel vindt je de gegevens. De minnetjes en plusjes tussen haakjes bij de getallen, geven aan of het hoger, gelijk aan, of lager dan het gemiddelde is. De +/- verbanden geven aan over beide scores een – hebben, beide een + ( dat zijn positieve verbanden). Als de een een – heeft en de ander een + of een 0 is het een negatief verband.

Kind	Mentale leeftijd	Leesscore	+/- verband
1	6 (-)	5.5(-)	+
2	8.5(+)	7.0 (0)	0
3	7.0(-)	7.5(+)	-
4	5.5(-)	6.0(-)	+
5	10.0(+)	8.0(+)	+
6	8.0(+)	7.5(+)	+
7	7.0(-)	6.5(-)	+
8	7.5(0)	8.0(+)	0
9	8.0(+)	7.0(0)	0
10	7.5(0)	7.0(0)	0

## Opdracht 6.1.1.c:

Bedenk een maat voor samenhang op basis van het aantal positieve en negatieve paren.

Je kunt hier als maat verzinnen dat je in de tabel maar 1 minnetje ziet, en je ziet 5 plusjes. Dit betekent dat er 5/6<sup>e</sup> aan plusjes is en je dit kunt zien als genoeg bewijs voor een positieve correlatie.

*Scatterplot:*

Een scatterplot is een diagram waarin je 2 variabelen met elkaar vergelijkt. Vaak heb je een afhankelijke en een onafhankelijke variabele. Op de y-as zet je vaak de afhankelijke variabele. Dit is de variabele die beïnvloed wordt door de andere variabele. Op de x-as zet je de onafhankelijke variabele. Dit is de variabele die de andere variabele beïnvloed.

Een voorbeeld om dit proces te verduidelijken: je hebt twee variabelen, dat zijn in dit geval lengte en gewicht. Je gewicht is afhankelijk van je lengte( hoe langer je wordt, hoe zwaarder je wordt). Daarom zet je het gewicht op de y-as en de lengte op de x-as.

Je kunt een scatterplot beoordelen op 3 aspecten:

1. richting: je kijkt dan of er sprake is van een positief of een negatief verband.
- 2.vorm: je kijkt dan of het een rechte lijn is. Je kijkt dus naar de mate van het lineair zijn van de lijn.
- 3.sterkte: hierbij kijk je hoeveel punten er op 1 lijn liggen. Hoe meer punten er op 1 lijn liggen, hoe sterker de correlatie is.

Je hebt niet met een verband te maken als er alleen een horizontale lijn doorheen kan in plaats van een lijn met een helling.

## Opdracht 6.1.3.abc:

Teken een puntenwolk voor de gegevens en vertel wat over de vorm van het scatterplot:

De leesscore is de afhankelijke variabele. Dus deze zet je op de y-as. En de mentale leeftijd is de onafhankelijke variabele. Deze zet je op de x-as. Het is een positief verband. Er zijn geen uitbijters. Het is een lineair sterk verband.

*Covariantie:*

De covariantie is de samenhang tussen 2 variabelen. Je kunt deze berekenen met de volgende formule :  $S_{xy} = \frac{1}{n-1} \sum (X_i - X_{gem})(Y_i - Y_{gem})$

Het getal zegt weinig over de correlatie, maar als het een positief getal is, heb je wel sowieso een positieve samenhang en als het een negatief getal is, heb je dus negatieve samenhang.

Correlatie is een getal dat loopt van -1 tot 1. Een negatieve score betekent dat je te maken hebt met een negatief verband. Als je een positieve score hebt, heb je te maken met een positief verband.

De correlatie is r. R is te berekenen met de volgende formule:  $r = \frac{1}{n-1} * \sum (zxzy)$

Opdracht 6.2.a:

Wat zijn de voorwaarden voor een zinvolle correlatie?

1: je moet te maken hebben met twee kwantitatieve variabelen

2. er moet een ongeveer lineair verband zijn.

Geen correlatie als er geen sprake is van een lineair verband!

Opdracht 6.2.f:

Bereken de covariantie.

0.7778 is de covariantie. Je antwoord bij 6.2.d was namelijk 7. Als je deze kruisproducten bij elkaar optelt. En daarna deel je dit door n-1. Dat is  $7/(10-1) = 7/9 = 0.7778$

Opdracht 6.2.hij:

Is de covariantie afhankelijk van de meeteenheid?

Ja, als je de leeftijd in maanden gaat weergeven, wordt alles 12 x zo groot. Dus is de covariantie afhankelijk van de meeteenheid. De oplossing daarvoor is het berekenen van de correlatie:

$0.7778/(1,2693*0,8165) = 0.75$ . 0.7778 was je antwoord bij opdracht 6.2.f en tussen de haken staan de gegevens die je onder de tabel op bladzijde 52 terugvindt.

Opdracht 6.3.c:

Wat is het meetniveau van deze variabele?

Het gaat hierbij om een ordinaal meetniveau

Opdracht 6.3.e:

Wat is het beste voor het overzicht het beste om hier de gegevens te ordenen?

Het maken van een kruistabel is hierbij veel handiger, dan het maken van een scatterplot.

# Oefenvragen I

## Opgave 1

Welke van de volgende variabele(n) zijn nominaal?

- A. Lengte
- B. Temperatuur
- C. Sekse en de rugnummers van een korfbal elftal.
- D. Gewicht

\*Het goede antwoord is: C, omdat een nominale schaal getallen/waarden toekent aan bepaalde labels. Bij sekse kan dat zijn: 1=man 2=vrouw. En bij de rugnummers van een korfbal elftal is het zo dat elk rugnummer voor een speler codeert. Bijv. 1=Jantje 2=Klaas 3=Peter etc. Antwoord is A is in dit geval ratio, omdat er een absoluut nulpunt is (bij 0 cm is er afwezigheid van lengte). Dit geldt ook voor gewicht (bij 0 kg is er afwezigheid van gewicht).

## Opgave 2

Wat is een voordeel van longitudinaal onderzoek ten opzichte van een cross-sectioneel ontwerp?

- A. Bij een longitudinaal onderzoek kan je meer zeggen over de stabiliteit van een fenomeen. Bij een cross-sectioneel ontwerp kan dit niet.
- B. Bij een cross sectioneel ontwerp is het zo, dat je proefpersonen in de loop der tijd kwijt kunt raken. Dit is bij een longitudinaal onderzoek niet het geval.
- C. Bij een longitudinaal onderzoek kun je iets zeggen over het moment waarop het onderzoek wordt gedaan. Dit kan bij een cross-sectioneel ontwerp niet.
- D. Longitudinaal onderzoek heeft ten opzichte van een cross-sectioneel ontwerp geen voordelen.

\*Het goede antwoord is: A, omdat bij longitudinaal onderzoek er op meerdere momenten wordt getest over langere tijd. Dit is bij een cross-sectioneel ontwerp echter niet mogelijk, omdat dit ontwerp een momentopname is (eenmalig).

## Opgave 3

In welke fase van de empirische cyclus gaat men van een idee naar een algemene theorie?

- A. Observatie fase
- B. Inductie fase
- C. Deductie fase
- D. Toetsfase

\*Het goede antwoord is: B, omdat tijdens de inductie fase men een idee heeft uitgewerkt tot een algemeen principe. In de inductie fase is het zo dat men van een specifieke observatie naar een algemene theorie gaat.

## Opgave 4

Welk van de volgende onderzoeken vindt vaak plaats in een natuurlijke omgeving?

- A. Range onderzoek
- B. Correlationeel onderzoek
- C. Experimenteel onderzoek
- D. Quasi-experimenteel onderzoek



\*Het goede antwoord is: D, omdat quasi-experimenteel onderzoek niet plaatsvindt in een laboratorium. Men wil de variabele meten zoals die natuurlijkerwijs voorkomt. Bij experimenteel en correlatieel onderzoek wordt vaak geprobeerd om alle omstandigheden te manipuleren in een laboratorium.

#### *Opgave 5*

Bekijk de volgende stellingen en geef aan of deze (on)juist zijn.

1. Operationele definities zijn definities die aangeven hoe je een fenomeen kunt meten. Bijv.: het aantal keren dat een kind per dag huult (meer dan vijf) geeft aan of een kind een huilbaby is.
2. Conceptuele definities zijn omschrijvingen van een fenomeen. Bijv.: een huilbaby is een kind dat excessief huult en graag aandacht krijgt.

- A. Stelling 1 is juist
- B. Stelling 2 is juist.
- C. Stelling 1 en 2 zijn beide juist.
- D. Stelling 1 en 2 zijn beide onjuist.

\*Het goede antwoord is: C, omdat operationele definities gekenmerkt worden door het feit dat zij een concept nauwkeurig definiëren. Hierdoor kan men het concept meten en/of manipuleren. Een conceptuele definitie is een woordenboek definitie. Het gaat hierbij om de omschrijving van een concept en niet om de meetbaarheid ervan.

#### *Opgave 6*

Maak de volgende zin af.

De systematische variantie tussen twee variabelen kunnen we ook wel....

- A. De onverklaarde variantie noemen.
- B. De verklaarde variantie noemen.
- C. De proportie verklaarde variantie noemen.
- D. De totale variantie noemen.

\*Het goede antwoord is: B, omdat de systematische variantie het verband tussen twee variabelen is, dat we kunnen verklaren. Dit kunnen we toeschrijven aan de onafhankelijke variabele. De variantie komt systematisch voor, wat betekent dat de variantie met de twee variabelen te maken heeft.

Antwoord A is niet goed, omdat de onverklaarde variantie (error) niet systematisch voorkomt. Het zou kunnen zijn dat deze error ontstaan is door fouten tijdens het meten of bijvoorbeeld door situationele verschillen tijdens het onderzoek. De error verklaart dus niets.

Antwoord C is niet goed, omdat de proportie verklaarde variantie aangeeft hoe groot het deel is dat we kunnen verklaren ten opzichte van het totaal. Oftewel: hoe groot is de systematische variantie ten opzichte van de totale variantie.

Antwoord D is niet goed, omdat de totale variantie niets te maken heeft met de verklaring van variantie. In de totale variantie zit zowel de verklaarde als de onverklaarde variantie.

#### *Opgave 7*

Een onderzoeker is nieuwsgierig naar de voorspellende waarde van de entree-toets die kinderen in groep 7 maken. Hij gebruikt de entree-toets om het niveau van kinderen in de brugklas te voorspellen. Welke vorm van validiteit is dit?

- A. Content Validity
- B. Face Validity
- C. Test bias
- D. Predictieve validiteit

\*Het goede antwoord is: D, omdat de onderzoeker in dit geval de testresultaten uit groep 7 (de entree-toets) gebruikt om het toekomstige gedrag (niveau) van kinderen in de brugklas te voorspellen. Antwoord A is hier niet goed, omdat inhoudsvaliditeit te maken heeft met het feit of het domein wordt gedekt door het meetinstrument. Bijv.: is een tentamen representatief voor alle leerstof? Antwoord B is niet goed, omdat Face validity aangeeft dat een meetinstrument in eerste instantie valide lijkt, maar het niet is. Antwoord C is geen vorm van validiteit.

### Opgave 8

'Mensen vullen in wat er van hen verwacht wordt of wat anderen in zouden vullen.'

Waar is de bovenstaande zin een voorbeeld van?

- A. Schattingsfout
- B. Sociale wenselijkheid
- C. Ja knickers of nee schudders
- D. Sociaal angstig personen.

\* Het goede antwoord is B, omdat mensen invullen wat zij denken dat anderen in zouden vullen of wat zij denken dat de onderzoeker van hen verwacht dat zij in gaan vullen. Dit kan zorgen voor een bias in een test. Onderzoekers proberen dit te minimaliseren. Antwoord C is hier niet goed, omdat ja knickers en nee schudders mensen zijn die overal ja of nee invullen. Dit wordt tegengegaan door vragen om te polen oftewel de andere kant op te stellen, zodat dit het onderzoek niet schaadt. Antwoord A en D zijn beide fout, omdat ze hier niets mee te maken hebben.

### Opgave 9

Een onderzoeker wil iets te weten komen over het bestuur van gemeenten in verschillende Europese landen. Hij wil Spanje en Nederland vergelijken.

Hij gaat als volgt te werk bij het nemen van een steekproef:

Europa  Nederland  Zuid-Holland  Wassenaar  
Europa  Spanje  Catalonië  Barcelona

Dit is een voorbeeld van:

- A. Een gestratificeerde steekproef
- B. Een Quota steekproef
- C. Een niet kanssteekproef
- D. Een multisage clustersteekproef

\*Het goede antwoord is: D, omdat hier meerdere malen wordt geclusterd. De totale populatie is te groot en daarom worden er eerst groepen geselecteerd. Het clusteren gebeurt meerdere keren (vandaar 'multisage').

### Opgave 10

Bekijk de volgende stellingen en geef aan of deze (on)juist zijn.

De mediaan wordt beïnvloed door uitbijters.

De modus wordt berekend door het totaal van alle scores te delen door het aantal scores.

- A. Stelling 1 is juist
- B. Stelling 2 is juist
- C. Stelling 1 en 2 zijn beide juist
- D. Stelling 1 en 2 zijn beide onjuist.

\*Het goede antwoord is: D, omdat de mediaan juist niet tot weinig wordt beïnvloed door uitbijters. De mediaan is het middelste getal, waardoor 50% boven het getal ligt en 50% eronder. Door deze verdeling zijn de uitbuiters weinig van invloed op de mediaan. De modus wordt niet berekend door het totaal van alle scores te delen door het aantal scores. Dat is het gemiddelde. De modus is het getal dat het meest voorkomt.

### Opgave 11

Wat is geen kenmerk van een correlatie ( $r$ )?

- A. Geeft de lineaire samenhang tussen twee variabelen weer.
- B. Ligt altijd tussen 1 en -1.
- C. Is als je hem kwadrateert gelijk aan de proportie verklaarde variantie.
- D. Is een andere maat voor een z-score.

\*Het goede antwoord is: D, omdat de correlatie niets te maken heeft met de z-score (waar iets in de verdeling ligt). De andere antwoorden zijn wel kenmerken van een correlatie.

### Opgave 12

De volgende getallen zijn gegeven:

11    15    16    16    17    20    22    24    28

Wat is hier de mediaan en de modus?

- A. Mediaan: 17 Modus: 16
- B. Mediaan: 11 Modus: 28
- C. Mediaan: 17 Modus: 17
- D. Mediaan: 16 Modus: 24

\*Het goede antwoord is: A, omdat de mediaan het middelste getal is (17). Hier staan er zowel 4 getallen boven 17 als eronder. 17 is dus het middelste getal. De modus is het getal dat het meest voorkomt. Dat is hier 16, want die komt twee keer voor. De rest maar één keer.

### Opgave 13

Een normaal verdeling heeft het volgende kenmerk:

- A. Een normaal verdeling is symmetrisch.
- B. Een normaal verdeling heeft maar één top.
- C. Bij een normaal verdeling ligt het gemiddelde, de modus en de mediaan in het midden.
- D. Een normaal verdeling bezit alle bovengenoemde kenmerken.

Het goede antwoord is: D, omdat een normaal verdeling symmetrisch is, een top heeft en het gemiddelde, de modus en de mediaan altijd in het midden liggen.

### Opgave 14

Welke informatie kun je uit een scatterplot (puntenwolk) halen?

- A. De vorm, de richting, de sterkte en uitbijters.
- B. De vorm, de lengte, de sterkte en uitbijters.
- C. De vorm, de richting, het gemiddelde, de sterkte en uitbijters.
- D. De mediaan, de vorm, de richting, de sterkte en de uitbijters.

Het goede antwoord is: A, omdat een scatterplot wat zegt over de vorm (lineair/ clusters/ curve), de richting (positief of negatief), de sterkte (hoe dichter de punten bij elkaar liggen, hoe sterker de samenhang) en over de uitbijters (punten die ver buiten de verdeling liggen).

### Opgave 15

Geef aan of de volgende stellingen (on)juist zijn.

Op de x-as plaatsen we altijd onafhankelijke variabelen en categorische variabelen.  
Op de y-as plaatsen we altijd afhankelijke variabelen en kwantitatieve variabelen.

- A. Stelling 1 is juist.
- B. Stelling 2 is juist.
- C. Stelling 1 en 2 zijn beide juist.
- D. Stelling 1 en 2 zijn beide onjuist.

\*Het goede antwoord is: C, omdat beide stellingen in dit geval juist zijn. Op de x-as plaatsen we namelijk altijd de onafhankelijke variabele tegen de afhankelijke variabele op de y-as. Dit is een regel die je uit je hoofd moet kennen. Als je een categorische variabele tegenover een kwantitatieve variabele in een grafiek uitzet, is het de regel dat de categorische variabele op de x-as staat en de kwantitatieve variabele op de y-as. Een categorische variabele is bijvoorbeeld een groep (groep 1, groep 2, groep 3 etc.). Een kwantitatieve variabele is numeriek. Hierbij kun je denken aan een score op een test (45, 67, 87).

### Opgave 16

Welk kenmerk kan een lurking variabele hebben in een onderzoek?

- A. Een lurking variabele kan een correlatie verbergen.
- B. Een lurking variabele kan zorgen voor een correlatie die er eigenlijk niet is.
- C. Een lurking variabele is vaak moeilijk te ontdekken.
- D. Alle bovenstaande kenmerken kunnen ontstaan door een lurking variabele in een onderzoek.

\*Het goede antwoord is: D, omdat zowel A, B als C correct is. Een lurking variabele wordt ook wel 'een verscholen variabele' genoemd. Deze verscholen variabele kan je onderzoek beïnvloeden door een correlatie in het onderzoek te verbergen of te suggereren dat er een correlatie is, die er eigenlijk niet is. Het is dan ook van groot belang dat de lurking variabele in een onderzoek wordt ontdekt of onder controle wordt gehouden, zodat de resultaten goed kunnen worden geïnterpreteerd.

### Opgave 17

Bezie de volgende twee stellingen met betrekking tot regressie en geef aan of de stellingen (on)juist zijn.

Interpolatie is betrouwbaarder dan extrapolatie.  
Hoe groter de slope, hoe steiler de lijn.

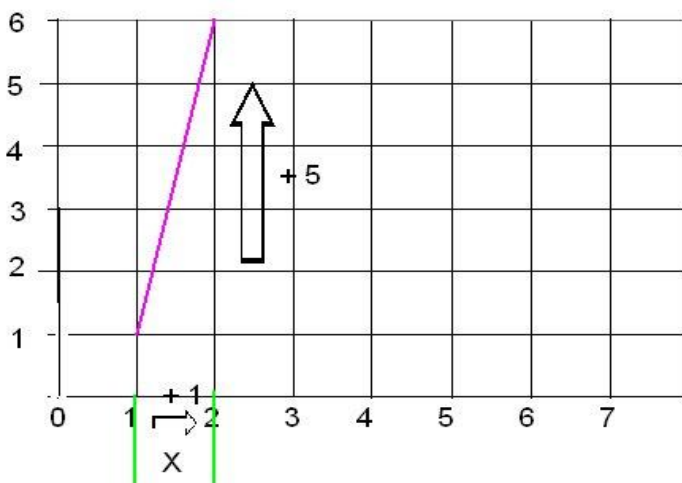
- A. Stelling 1 is juist.
- B. Stelling 2 is juist.
- C. Beide stellingen zijn juist.
- D. Beide stellingen zijn onjuist.

\*Het goede antwoord is: C, omdat beide stellingen juist zijn. Als je interpoleert voorspel je waarden die binnen de geobserveerde gegevens liggen. Deze waarden liggen dus binnen je bereik. Als je extrapoleert is het zo dat je een voorspelling doet over de gegevens die buiten de observaties liggen. Deze voorspelling is minder nauwkeurig en dus minder betrouwbaar. Dit omdat je waarden voorspelt die buiten je bereik liggen.

Stelling 2 is juist, omdat de slope de helling aangeeft. Bij de slope is het volgende principe van toepassing: als x 1 naar rechts verschuift, verschuift y ..... Hoe groter het getal is, hoe steiler de helling. VB 1: Als x 1 naar rechts verschuift (horizontaal), verschuift y 5 (verticaal). Dit is een zeer steile helling, omdat de y met 5 verschuift binnen een x van 1.

VB 2: Als je x 1 naar rechts verschuift, verschuift y 1. Deze helling is veel minder steil, omdat de y verschuift met 1 binnen een x van 1.

De volgende afbeelding is ter illustratie bij voorbeeld 1.



### Opgave 18

De gekwadrateerde correlatie ( $r^2$ ) wordt ook wel.....genoemd

- A. de lineaire samenhang tussen x en y.
- B. de partiële correlatie tussen x en y.
- C. de determinatiecoëfficiënt.
- D. de proportie verklaarde variantie.
- E. zowel C als D.

\*Het goede antwoord is: E, omdat de gekwadrateerde correlatie ( $r^2$ ) ook wel de determinatiecoëfficiënt of de proportie verklaarde variantie wordt genoemd. Nog een andere naam voor de gekwadrateerde correlatie is: de betrouwbaarheidscoëfficiënt. Al deze benamingen zijn gebaseerd op het feit dat als men een correlatie kwadrateert, men dan ziet hoeveel variantie er verklaard wordt (de proportie verklaarde variantie).

### Opgave 19

Op basisschool 'De Regenboog' wil men onderzoeken welke lesmethode het best werkt voor het vak aardrijkskunde. Er zijn twee lesmethoden: 1 en 2. De twee groepen 8 (groep 8a en groep 8b) krijgen respectievelijk de lesmethode. Na een tijdje evalueert de school welke methode het meest effectief is.

Welk soort onderzoek wordt hier gedaan?

- A. Beschrijvend onderzoek
- B. Correlationeel onderzoek
- C. Experimenteel onderzoek
- D. Quasi-experimenteel

\*Het goede antwoord is: D, omdat men hier bestaande groepen gebruikt die men gaat vergelijken. Antwoord B is niet goed, omdat er niet wordt gezocht naar een relatie tussen bijv. de lesmethoden. Antwoord A is niet goed, omdat het niet gaat om de beschrijving van bijv. de lesmethoden. Antwoord C is niet goed, omdat men niet in een laboratorium te werk gaat.

### Opgave 20

Kijk naar de volgende categorieën (leeftijd in jaren):

0-15, 15-30, 30-45, 45-60

De categorieën zijn:

- A. Uitputtend en wederzijds exclusief
- B. Niet uitputtend en niet wederzijds exclusief
- C. Niet uitputtend, maar wel wederzijds exclusief
- D. Uitputtend, maar niet wederzijds exclusief

\*Het goede antwoord is: B, omdat deze categorieën niet uitputtend en niet wederzijds exclusief zijn. *Uitputtend*: zijn alle waarden opgenomen in de categorieën? ► Nee, want iemand die 100 jaar is behoort tot geen van de categorieën. Waar moet deze persoon dan bij? *Wederzijds exclusief*: sluiten de categorieën elkaar uit? Oftewel overlappen de categorieën? ► De categorieën sluiten elkaar niet uit en overlappen elkaar. Dit zou betekenen dat iemand van 30 jaar oud in twee categorieën zou kunnen vallen (zowel 15-30 als 30-45).

### Opgave 21

Een bedrijfspsycholoog is geïnteresseerd in het aantal keer dat werknemers in hun leven geïntimideerd zijn op de werkvloer.

De volgende tabel was de uitkomst:

Aantal keren geïntimideerd	Frequentie
0	1
1	2
2	4
3	1
4	3

Wat is de modus, de mediaan, en het gemiddelde?

- A. Modus: 2 , Mediaan: 2, Gemiddelde: 2,27
- B. Modus: 2,27, Mediaan: 2, Gemiddelde: 2
- C. Modus: 2, Mediaan: 2,27, Gemiddelde: 2
- D. Modus: 1, Mediaan: 3, Gemiddelde: 2,99

\*Het goede antwoord is: A, omdat de Modus (het getal dat het meest voorkomt) 2 is, namelijk 4 keer. De mediaan (het middelste getal) ook 2 is ► 0 1 1 2 2 2 2 3 4 4 4. Hier is twee het middelste getal (zowel vijf getallen onder twee als boven twee). Het gemiddelde (alle getallen bij elkaar optellen en delen door het aantal getallen): 2,27 is. ►  $0+1+1+2+2+2+2+3+4+4+4: 25/11: 2,27$ .

De volgende vraag heeft betrekking op opgave 21.

Wat is het meetniveau van de variabele? En welke (minimale) centrale tendentie hoort bij dit meetniveau?

- A. Interval, gemiddelde
- B. Nominaal, Modus
- C. Ordinaal, Modus
- D. Ratio, gemiddelde

\*Het goede antwoord is: D, omdat ratio een absoluut nulpunt heeft ( je kan niet -1 of bijv. -3 scoren). En het is ratio, omdat de verschillen tussen de getallen even groot zijn. Dit zijn twee kenmerken van ratio. Interval kan niet, omdat interval geen absoluut nulpunt heeft. Interval heeft echter ook dat de verschillen tussen de getallen even groot zijn (net zoals bij ratio). De centrale tendentie maat die bij ratio hoort is het gemiddelde.

Ordinaal is hier niet van toepassing, omdat er geen sprake is van een ordening.  
Nominaal is hier niet van toepassing, omdat de cijfers niet coderen voor een label.

### Opgave 22

Welk van de volgende beschrijvingen geeft het best weer wat variantie inhoudt?

- A. Variantie is een spreidingsmaat. De variantie bekijkt alle scores ten op zichte van het gemiddelde.
- B. Variantie is een spreidingsmaat. De variantie bekijkt één score ten opzichte van het gemiddelde.
- C. Variantie is een spreidingsmaat. De variantie geeft aan hoeveel van de variantie verklaard wordt.

D. Variantie is een spreidingsmaat. De variantie geeft aan hoeveel van de variantie onverklaard wordt.

\*Het goede antwoord is: A, omdat dit de beste omschrijving is van variantie. Omschrijving B is de beschrijving van de standaarddeviatie. Omschrijving C beschrijft de systematische variantie (verklarend). Omschrijving D beschrijft de error (onverklaarde) variantie.

### Opgave 23

Een ontwikkelingspsycholoog wil weten wat het verband is tussen de hartslag (X) van kinderen en hun schoolprestaties (Y). Voor deze gegevens berekent ze het gemiddelde, de variantie en de covariantie. De gegevens zijn als volgt: gemiddelde x: 90 gemiddelde y: 100  $S_x^2$ : 25  $S_y^2$ : 36 en covariantie -20.

Bereken de correlatie.

- A. -0.67
- B. 0.67
- C. 0.20
- D. -0,20

Het goede antwoord is: A, omdat de volgende formule van toepassing is:

R: covariantie (x,y) /  $S_x \cdot S_y$ . In dit geval zijn  $S_x$  en  $S_y$  niet gegeven. Echter wel de  $S_x^2$  en de  $S_y^2$ . Door de wortel uit deze getallen te trekken, komen we uit op de  $S_x$  en de  $S_y$ . Dus  $S_x$ :  $\sqrt{25}$ : 5 en  $S_y$ :  $\sqrt{36}$ : 6. De covariantie is: -20

Dan vullen we de formule als volgt in:  $-20 / (5 \cdot 6) = -0.67$   
→ De andere antwoorden zijn verzonnen.

### Opgave 24

Een onderzoeker wil weten wat het verband is tussen de rijvaardigheid (X, op een schaal van 1 tot 10) en het aantal ongelukken (Y, per jaar) die vrachtwagenchauffeurs veroorzaken. De volgende vergelijking is opgesteld:

$$\hat{Y} = 3 + 0.66 \cdot x$$

Een vrachtwagenchauffeur is beoordeeld met het cijfer 8. Bereken hoeveel ongelukken hij per jaar veroorzaakt.

- A. 8
- B. -8
- C. 10
- D. 13

Het goede antwoord is A, omdat de berekening als volgt is.  $X=8$ , dus voer in  $3 + (0.66 \cdot 8) \square$   
 $3 + 5,28 = 8,28$ .  $\hat{Y}$  is dus 8,28. Dit zijn 8 ongelukken.

### Opgave 25

Een onderzoeker wil weten wat het drank gedrag is van voorbijgangers. Daarom wordt in 2008 op straat aan mensen gevraagd hoeveel alcohol ze de afgelopen maand gedronken hebben. Het



onderzoek wordt herhaald in 2012. Hierbij wordt opnieuw gevraagd hoeveel alcohol mensen de afgelopen maand gedronken hebben. Dit doen ze om te kijken of de hoeveelheid mensen die alcoholische drank drinkt toeneemt.

Welk soort onderzoek is dit?

- A. Longitudinaal
- B. Internet Survey
- C. Opeenvolgende onafhankelijke onderzoeken
- D. cross-sectioneel.

Het goede antwoord is: C, omdat hier gebruikt wordt gemaakt van opeenvolgende onafhankelijke steekproeven. Ten eerste is het een steekproef, omdat aan voorbijgangers (random geselecteerd) wordt gevraagd of ze alcohol drinken. Zij representeren de populatie. De steekproeven zijn onafhankelijk van elkaar, omdat de voorbijgangers één keer meedoen aan het onderzoek. Daarna ben je ze kwijt.

#### *Opgave 26*

'Hoe massiever de inhoud van je hoofd, hoe intelligenter je bent'

Dit is een voorbeeld van:

- A. Gebrek aan betrouwbaarheid
- B. Gebrek aan validiteit
- C. Gebrek aan meetinstrumenten
- D. Gebrek aan onderzoek

\*Het goede antwoord is: B, omdat hier gebrek is aan validiteit. (Validiteit: meet je daadwerkelijk wat je wilt meten). Hier gebruikt men de massiviteit van de inhoud van het hoofd om te bepalen of je intelligent bent. Dit is geen valide manier om intelligentie te meten. Dit doe je bijv. met bepaalde taakjes of een IQ test.

#### *Opgave 27*

Op een dag wil je weten hoe lang je nu eigenlijk bent. Je besluit om jezelf 3 keer per dag te meten met een speciale meetlat. Op moment 1 ben je volgens de meetlat 1,76 cm. Op moment 2 ben je volgens de meetlat ineens 1,70 cm. Op moment 3 ben je volgens de meetlat 1.85 cm. Dit is een voorbeeld van:

- A. Gebrek aan betrouwbaarheid
- B. Gebrek aan validiteit
- C. Gebrek aan meetinstrumenten
- D. Gebrek aan onderzoek

\*Het goede antwoord is: A, omdat de meetlat hier niet betrouwbaar is. De meetlat geeft namelijk steeds een andere lengte aan. De meetlat is hier dus geen *betrouwbaar* meetinstrument om te meten hoe lang je bent.

#### *Opgave 28*

Je wil weten hoe betrouwbaar een test is. Daarom leg je dezelfde test twee keer voor aan groep 3 van de basisschool.

Dit is een voorbeeld van:

- A. Paralleltest betrouwbaarheid
- B. Test-hertest betrouwbaarheid
- C. Inter-item betrouwbaarheid
- D. Algemene betrouwbaarheid

\*Het goede antwoord is: B, omdat hier twee testen worden afgelegd. Hiermee wil je de betrouwbaarheid meten. Je legt de eerste test voor en daarna de tweede om te kijken of de variaties met elkaar overeen komen. Een paralleltest zijn twee testen die hetzelfde proberen te meten. Dat is hier niet zo, omdat het om één dezelfde test gaat. Inter-item betrouwbaarheid is hier niet goed, omdat men bij inter-item wil meten wat de betrouwbaarheid is tussen de items van de test. De algemene betrouwbaarheid is hier veel te vaag.

### *Opgave 29*

Je wil weten wat het leesniveau is van kinderen in groep 3. Daarom besluit je twee verschillende testen aan de kinderen voor te leggen. Beide testen proberen het leesniveau van de kinderen zo nauwkeurig mogelijk te meten.

Dit is een voorbeeld van:

- A. Paralleltest betrouwbaarheid
- B. Test-hertest betrouwbaarheid
- C. Inter-item betrouwbaarheid
- D. Algemene betrouwbaarheid

\*Het goede antwoord is: A, omdat hier twee verschillende testen worden gebruikt die hetzelfde beogen te meten. De testen verschillen inhoudelijk, maar meten hetzelfde. Inter-item betrouwbaarheid is hier niet goed, omdat men bij inter-item wil meten wat de betrouwbaarheid is tussen de items van de test. De algemene betrouwbaarheid is hier veel te vaag. Voor uitleg Test-hertest zie vraag 26.

### *Opgave 30*

Je laat een vragenlijst invullen in een 2 VMBO klas waarmee je probeert te meten wat deze jongeren bezighoudt. Je bent benieuwd of je deze resultaten kunt generaliseren naar alle 2 VMBO klassen in Nederland.

Dit is een voorbeeld van:

- A. Externe validiteit
- B. Content validiteit
- C. Construct validiteit
- D. Predictieve/voorspellende validiteit

\*Het goede antwoord is: A, omdat men probeert de resultaten te generaliseren naar alle VMBO klassen in Nederland. Dit is externe validiteit, omdat het hoofddoel van externe validiteit het generaliseren van gegevens is naar de populatie. Content validiteit (ook wel inhoudsvaliditeit genoemd) is hier niet van toepassing, omdat content validiteit zich bezig houdt met het feit of het domein representatief is voor het te meten construct. Bijvoorbeeld: is een tentamen representatief voor alle leerstof?

Construct validiteit houdt zich bezig met het feit of twee constructen die hetzelfde proberen te meten, dat ook daadwerkelijk doen. Predictieve validiteit houdt zich bezig met het feit of we op basis van huidige meetresultaten toekomstige testresultaten kunnen voorspellen.

### *Opgave 31*

Je hebt een weegschaal gekocht. Je wil hiermee andere mensen gaan wegen. Op het eerste gezicht ziet de weegschaal er betrouwbaar uit.

Dit is een voorbeeld van:

- A. Face validiteit
- B. Content validiteit
- C. Face Validiteit
- D. Construct validiteit

\*Het goede antwoord is: C, omdat de weegschaal op het eerste gezicht een valide meetinstrument lijkt voor het meten van gewicht. Dit hoeft echter niet zo te zijn.

Antwoord A is niet goed, omdat content validiteit (ook wel inhoudsvaliditeit genoemd) hier niet van toepassing is. Dit komt, omdat content validiteit zich bezig houdt met het feit of het domein representatief is voor het te meten construct. Bijvoorbeeld: is een tentamen representatief voor alle leerstof?

Construct validiteit houdt zich bezig met het feit of twee constructen die hetzelfde proberen te meten, dat ook daadwerkelijk doen. Predictieve validiteit houdt zich bezig met het feit of we op basis van huidige meetresultaten toekomstige testresultaten kunnen voorspellen.

### *Opgave 32*

Bereken voor de volgende twee situaties de betrouwbaarheid.

Situatie 1: Een schooltest heeft een totale score van 80 en een ware score van 70.

Situatie 2: Een schooltest heeft een totale score van 80 en een toevalsfout van 60.

Welk antwoord is goed?

- A. Situatie 1: 0.50 Situatie 2: 0.22
- B. Situatie 1: 0.75 Situatie 2: 0.30
- C. Situatie 1: 0.88 Situatie 2: 0.25
- D. Situatie 1: 0.92 Situatie 2: 0.34

\*Het goede antwoord is: C, zie onderstaande uitleg.

Situatie 1. Gegevens → Totale score: 80 en ware score: 70. De betrouwbaarheid berekenen we door de ware score te delen door de totale score (ware score/totale score: betrouwbaarheid). In situatie 1 ziet dat er als volgt uit:  $70/80 = 0.88$

Situatie 2. Gegevens → Totale score: 80 en toevalsfout: 60. De betrouwbaarheid berekende we door de ware score te delen door de totale score. Hier moeten we echter eerst achter de ware score komen. Het model ziet er als volgt uit: Betrouwbaarheid → Totale score = ware score + toevalsfouten (error). Onze totale score is 80 en de toevalsfout is 60. Door de totale score van de toevalsfout af te trekken krijgen we de ware score. Dus  $80 - 60 = 20$ .

De betrouwbaarheid is in situatie 2 als volgt:  $20/80 = 0.25$ .

Opgave 33

Bekijk de volgende tabel.

Item total statistics				
	Scale mean if item Deleted	Scale Variance if item deleted	Corrected Item-Total Correlation	Cronbachs Alpha if item Deleted
Item1. Hoe vind je de begeleiding op school? (Goed/Slecht)	43,97	19,56	0.41	0.33
Item 2. Hoe vind je de docenten op school? (Goed/slecht)	45,55	20,01	0.34	0.29
Item 3. Hoe vind je de omgeving op school? (Goed/slecht)	45,23	18,78	0.38	0.31
Item 4. Hoe vind je het werkmateriaal? (Goed/slecht)	46,34	17,88	0.36	0.27
Item 5. Houd je van sporten? (Ja/Nee)	56,57	30,89	-.01	0.59
Item 6. Houd je van de zomer? (Ja/Nee)	54,99	32,65	-.0.09	0.61

Welk item kun je het best weglaten volgens deze tabel?

- A. Item 1
- B. Item 5
- C. Item 6
- D. Item 3

\*Het goede antwoord is: C , omdat als je hier Item 6 zou verwijderen, Cronbachs Alpha 0.61 zou worden (dit kun je zien door in de laatste kolom 'Cronbachs Alpha if item Deleted' te kijken). Dit is de moeite waard, want de Alpha wordt dan redelijk hoog (dit is gunstig voor de betrouwbaarheid).

De volgende vraag heeft betrekking op de tabel van opgave 33.

Is het verstandig om nog een item te verwijderen (als we kijken naar de bovenstaande tabel)?

- A. Ja, item 5 is ook niet geschikt.
- B. Ja, item 4 is ook niet geschikt.
- C. Nee, er is geen slecht item meer.
- D. Nee, dan heb je te weinig items.

\*Het goede antwoord is: A, omdat item 5 ook niet geschikt is. Als we item 5 zouden verwijderen dan wordt de Alpha: 0.59 (dit kun je zien door in de laatste kolom 'Cronbachs Alpha if item Deleted' te kijken). Dit is zeker de moeite waard (aangezien je altijd een hoge betrouwbaarheid wil). Antwoord D, is hier niet goed, omdat de Alpha zodanig omhoog gaat dat het de moeite waard is om item 5 toch te verwijderen.

### Opgave 34

Bereken met de volgende gegevens de Cronbachs Alpha.  $S_1^2=3$ ,  $S_2^2=4,5$ ,  $S_3^2=0,5$ ,  $S_x^2=13$ . Het aantal items is 3.

Wat is de Cronbachs alpha?

- A. 0.33
- B. 0.77
- C. 0.43
- d. 0.58

\*Het goede antwoord is: D (zie berekening).

$$\text{Cronbachsalpha: } \frac{k}{k-1} \left( 1 - \frac{\sum S_i^2}{S_x^2} \right)$$

**K: aantal items**

K is in dit geval 3 (aantal items). De som van de  $S_i^2$ 's =  $3+4,5+0,5=8$ .  $S_x^2=13$ .

We vullen de formule als volgt in:

$$\frac{3}{2} \left( 1 - \frac{8}{13} \right) = 0.567 \rightarrow 0.58: \text{Cronbachs Alpha.}$$

### Opgave 35

Vul de volgende zin in.

Voor een normaal verdeling geldt: hoe groter de standaarddeviatie is, hoe... de curve is.

Wat hoort op de puntjes te staan?

- A. Breder
- B. Horizontaler
- C. Verticaler
- D. Hoger

\*Het goede antwoord is: A, omdat de standaarddeviatie aangeeft hoe de verdeling is van de scores rond het gemiddelde. Als de standaarddeviatie groot is betekent dat, dat de scores over het algemeen veel afwijken van het gemiddelde. Dit zorgt voor een bredere normaal verdeling, omdat de curve door een grotere standaarddeviatie breder wordt.

### Opdracht 36

Geef bij de volgende situaties aan wat de *afhankelijke* en de *onafhankelijke* variabele is. Dit is *geen* meerkeuze vraag.

In dorpen waar meer koeien zijn, is meer melk.  
Hoe langer een relatie, hoe meer ruzies.  
Leerlingen die vaak ziek zijn, hebben minder sociale contacten.  
Iemand zonder fiets, is minder mobiel.

De stellingen worden één voor één toegelicht.

Wat is een onafhankelijke variabele?

→ De onafhankelijke variabele is de variabele die de afhankelijke variabele beïnvloedt.

Wat is een afhankelijke variabele?

→ De afhankelijke variabele is de variabele die beïnvloed wordt door de onafhankelijke variabele.

De onafhankelijke variabele is hier: het aantal koeien. De afhankelijke variabele is hier: de hoeveelheid melk. → Het aantal koeien bepaalt/beïnvloedt de hoeveelheid melk. De hoeveelheid melk is hier *afhankelijk* van het aantal koeien.

De onafhankelijke variabele is hier: de lengte van een relatie. De afhankelijke variabele is hier: het aantal ruzies. → De lengte van een relatie bepaalt/beïnvloedt de hoeveelheid ruzies in een relatie. Het aantal ruzies is hier *afhankelijk* van de lengte van de relatie.

De onafhankelijke variabele is hier: leerlingen die vaak ziek zijn. De afhankelijke variabele is hier: het hebben van minder sociale contacten. → Of een leerling vaak ziek is, bepaalt/beïnvloedt of een leerling sociale contacten heeft. De sociale contacten zijn *afhankelijk* de frequentie van het ziek zijn van een leerling.

De onafhankelijke variabele is hier: iemand zonder fiets. De afhankelijke variabele is hier: minder mobiel zijn. → Of iemand een fiets heeft, bepaalt/beïnvloedt zijn/haar mobiliteit. De mobiliteit is *afhankelijk* van het feit of iemand een fiets bezit.

# Werken bij JoHo - De ideale studentenbijbanen!

## ***Student-managers (vanaf 10 uur per week)***

Als student-manager ben je in één van de JoHo support centers samen met één of twee collega's verantwoordelijk voor het gehele traject rondom het uitgeven van samenvattingen. Een zeer veelzijdige functie waarbij je in 1,5 jaar met alle aspecten van de bedrijfsvoering te maken krijgt.

### Profiel:

- Enthousiaste student, binnen de relevante studies
- Zelfstandig en in teamverband kunnen werken
- Geen moeite hebben met zo nu en dan leiding geven en aansturen
- Organisatorische vaardigheden
- Commercieel inzicht

## ***Student-auteurs***

Voor het maken van de boekuittreksels en samenvattingen, maken wij gebruik van ervaren auteurs, voornamelijk Masterstudenten en/of (pas-)afgestudeerden. De hulp van studenten die het vak volgen is echter hard nodig om ons aanbod perfect te laten aansluiten op de wensen van de student!

### Dus:

1. Heb jij aanleg om netjes en overzichtelijk te schrijven en wil je deze vaardigheden verder ontwikkelen? Vind je deadlines geen probleem en vind je het prettig om in je eigen tempo daar naartoe te werken? Word dan **student-auteur!** Als student-auteur help je JoHo met het verbeteren van de samenvattingen, door bijvoorbeeld bestaande samenvattingen te controleren op inhoud en spelling, het schrijven van aanvullende teksten en het maken van collegeaantekeningen.
2. Naast de verdiensten voor de gemaakte opdracht verbeter je ook je schrijfvaardigheden en krijg je gratis hulpguides om effectiever te studeren en beter je tentamens voor te bereiden.

## ***Studie-coördinatoren (4 tot 8 uur per maand)***

- Sta jij graag veel in contact met je medestudenten en ben jij van alles op de hoogte rondom je studie? Zoek je een bijbaan voor maar enkele uren in de maand, die perfect aansluit bij je werkzaamheden voor je studie? Word dan **studie-coördinator!** Als studie-coördinator help je JoHo met het verzamelen van alle relevante info voor jouw studie en zorg je ervoor dat je medestudenten weten wanneer de samenvattingen beschikbaar zijn. Tevens help je JoHo met het vinden van nieuwe auteurs en je medestudenten met een passende bijbaan.
- Naast de vaste verdiensten per maand, maak je gratis gebruik van een groot deel van de samenvattingen voor jouw studie

Interesse in een van bovenstaande functies? Stuur je motivatiebrief en CV naar: [personeelszaken@joho.nl](mailto:personeelszaken@joho.nl) t.a.v de procescoördinator P&O (je hoeft de mail niet persoonlijk aan iemand te richten). Heb je nog vragen dan kan je iedere dinsdag tussen 12.00 en 17.00 uur via 088-3214567 contact opnemen met de Procescoördinator P&O.

# Europass: Essentieel voor je internationale carrière

Wil je werken, studeren of stage lopen in het buitenland? Dan is het belangrijk dat je in het buitenland duidelijk maakt wat je weet en wat je kunt. Europass is een set gratis documenten in alle Europese talen, waarmee werknemers hun vaardigheden en talenten kunnen laten zien. De basis van Europass is het Europass CV, de Europese standaard, dat je naar wens kunt aanvullen met documenten over je taalvaardigheid, je werkervaring en je diploma's.



[www.europass.nl](http://www.europass.nl)