

Psychometrie

Hoofdstuk 1 Overzicht

1.1 Wat is psychometrie?

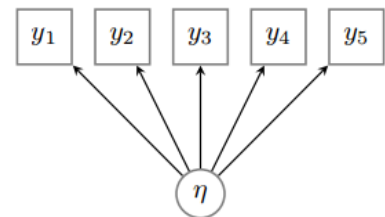
Psychometrika is het officiële vakblad van de Psychometric Society en stelt het volgende: “Psychometrika is devoted to the development of psychology as a quantitative rational science, including the advancement of theory and methodology for behavioral data analysis in psychology, education, and the social and behavioral sciences generally. Bijna elke discipline heeft een –metrics versie die handelt over het ontwikkelen en bestuderen van data-analytische technieken voor empirische data verzameld in de discipline.

1.2 Belang van psychometrie

Psychologische constructen zijn vaak niet direct observeerbaar (en dus latent), en kunnen enkel indirect worden gemeten. De meetfout is vaak aanzienlijk. Psychometrie biedt een referentiekader om meetproblematiek beter te begrijpen omdat verkeerd begrip hiervan dramatische gevolgen kan hebben. Het is van belang dat alle betrokkenen zich bewust zijn van deze problematiek.

1.3 Psychometrie in een oogopslag

In de afbeelding hiernaast is η een latente variabele en representeert een unidimensioneel psychologisch construct. y_1, y_2, \dots, y_5 zijn geobserveerde indicatoren van deze latente variabelen (vaak items in een vragenlijst). Er zijn 4 mogelijkheden welke staan beschreven in onderstaande tabel:



	η numeriek	η categorisch
y numeriek	Factor-analyse	Latente profiel modellen
y categorisch	Latente trek modellen (IRT)	Latente klasse modellen

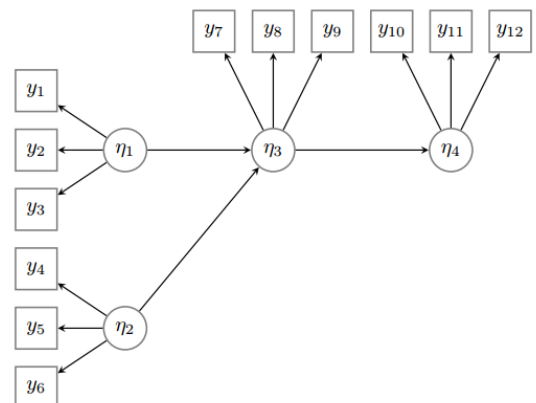
Hiernaast staat een voorbeeld van een structureel vergelijkingsmodel. Hier wordt later op ingegaan.

1.4 Terminologie

Metten: het omzetten van een waargenomen grootheid naar een getal (bij kwantitatief meten) of een categorie (kwalitatief meten).

Meetfout: afwijking tussen de gemeten waarde en de werkelijke waarde, dit hangt af van:

- Nauwkeurigheid van het meetinstrument
- Betrouwbaarheid (verkrijgen we hetzelfde getal als we twee keer meten?)
- Validiteit (meten we wat we willen meten? Is er sprake van een systematische over- of onderschatting?)



Het meetniveau bepaalt welke (wiskundige) bewerkingen zinvol zijn. Er zijn twee groepen meetniveau's van een meetvariabele, welke zijn onderverdeeld in elk twee soorten:

- Kwalitatief:
 - Nominaal (enkel categorieën)
 - Ordinaal (categorieën zijn geordend)
- Kwantitatief:
 - Interval (gelijke eenheid)
 - Ratio (vast nulpunt)

Een psychologische test probeert bij een individu een reeks van gedraging uit te lokken, dit gedrag wordt testgedrag genoemd. Het testgedrag dienen we te vertalen (te schalen) tot testcores. Schalen: Het vertalen van testgedrag tot testcores. Het geeft de gemeten individuen een plaats op een schaal. Er zijn unidimensionele schaaltechnieken (paarsgewijze vergelijking, cumulatieve schalen, succesieve intervallen) en multidimensionele schaaltechnieken.

Hoofdstuk 2 Geschiedenis van de Psychometrie

2.1 Overzicht tijdslijn

Een overzicht van de tijdslijn is te vinden op slide 4 t/m 6 van hoofdstuk 2. Dit overzicht hoeft verder niet gekend te zijn, maar is genoemd om overzicht en duidelijkheid te creëren.

2.2 De oorsprong van examens

De eerste gestandaardiseerde testen zijn gevonden in het oude China, waar men examens moest afleggen om te mogen werken voor de staat. Europese universiteiten zijn gestart met examens voor studenten in de 13^e eeuw.

2.3 De grondleggers

Er zijn twee scholen van denkers die een impact hebben gehad op het ontstaan van de psychometrie:

- Het bestuderen van individuele verschillen
 - Quetelet
 - Darwin
 - Galton
 - Cattell
- Wetmatigheden die gelden voor alle mensen (kenmerken van de gemiddelde mens)
 - Fechner
 - Wundt

Fechner en Wundt waren de grondleggers van wat later experimentele psychologie zou worden.

2.4 Psychofysica en experimentele psychologie

Gustav Fechner (Duitsland, 1801-1877)

Fechner wordt beschouwd als de grondlegger van de psychologie als wetenschap. Hij kwam tot het inzicht dat je de menselijke geest kan bestuderen door menselijke reacties op fysieke stimuli te kwantificeren. Hij publiceerde de wet van Fechner ($S = k \log R$), waarbij R de maat is van een fysieke stimulus, S een maat is van de perceptie van deze stimulus en k een constante is die experimenteel moet worden bepaald op basis van data. Ook introduceerde Fechner het concept mediaan als een maat van centrale tendentie.

Hermann von Helmholtz (Duitsland, 1821 – 1894)

Helmholtz hanteerde de methode van Fechner voor de studie van de perceptie van beeld en geluid.

Wilhelm Wundt (Duitsland, 1832 – 1920)

Wundt was de eerste academicus die zichzelf 'psycholoog' noemde en richtte in 1879 het 'eerste' laboratorium voor psychologie op aan de universiteit van Leipzig. Hierdoor wordt hij vaak de vader van de (experimentele) psychologie genoemd, ondanks dat Willem James al in 1875 een psychologisch laboratorium oprichtte aan Harvard University.

2.5 Intermezzo: meetfouten in de astronomie

Galileo Galilei (Italië, 1564 – 1642)

Galilei is de vader van de moderne astronomie en kwam tot de conclusie dat observationele fouten onvermijdelijk zijn bij het uitvoeren van metingen en experimenteel onderzoek. Hij maakte een onderscheid tussen systematische fouten, welke afhankelijk zijn van het meetinstrument of de meetmethode, en random fouten. Deze random fouten hebben de volgende eigenschappen:

- Kleine fouten komen meer voor dan grove fouten
- Negatieve waarden voor fouten komen evenveel voor als positieve waarden
- Indien men meerdere metingen uitvoert dan is gemiddeld genomen de fout (bijna) nul

Carl Friedrich Gauss (Duitsland, 1777 – 1855)

Gauss stelde de normaalverdeling voor, welke hiernaast is te zien.



2.6 De eerste studies naar individuele verschillen

Friedrich Bessel (Pruisen, 1784 – 1846)

Bessel realiseerde als eerste dat er inderdaad consistente en onvrijwillige discrepanties waren tussen observators. Hij ontwikkelde met behulp van de normaalverdeling van Gauss "The personal equation", welke een correctie kon maken voor elke observator.

Adolphe Quetelet (België, 1796 – 1874)

Quetelet is de uitvinder van de body mass index. Hij paste ook als eerste de statistische methode toe op sociale data. Hij beschreef ook het concept van 'de gemiddelde mens' ('l'homme moyen), welke op elk karakteristiek een gemiddelde waarde heeft en waarbij individuen hiervan afwijken volgens de normaalverdeling. Hij noemde deze wetenschap 'sociale fysica'.

Charles Darwin (Engeland, 1809 – 1882)

Darwin publiceerde "On the Origin of Species", een boek over de evolutietheorie. Het werk van Darwin inspireerde Galton, die dezelfde 'variatie' ook bij mensen zag.

Francis Galton (Engeland, 1822 – 1911)

Galton heeft veel bijgedragen tot de statistiek (standaard deviatie, bivariate normaalverdeling en de connectie met regressie, regressie naar het gemiddelde en het symbool r voor de hellingsgraad van een eenvoudige regressie). Hij bestuurde ook op een systematische manier individuele verschillen en toonde aan dat deze verschillen op een objectieve en repliceerbare wijze konden worden gemeten. Hij dacht dat hij zowat alles kon meten. Galton was vooral gemotiveerd door zijn onderzoek naar menselijke eugenetica (rasverbetering). In de Angelsaksische wereld wordt Galton beschouwd als de grondlegger van de differentiële psychologie (wij beschouwen Quetelet als grondlegger).

Karl Pearson (Engeland, 1857 – 1936)

Pearson is de vader van de mathematische statistiek. Hij heet veel bijgedragen tot de statistiek (formele definitie van de correlatie coëfficiënt, schattingsmethode 'method of moments', p-waarde, hypothese toetsen, beslissingstheorie, chi-kwadraat toets, principale componenten analyse, histogram, etc.)

James McKeen Cattell (Verenigde Staten, 1860 – 1944)

Cattell was de eerste professor in psychologie in de Verenigde Staten. Hij bestudeerde samen met Wundt intelligentie, maar ook had hij regelmatig contact met Galton, waardoor hij zich het potentieel realiseerde van de combinatie van Wundt's psychofysica en Galton's mathematische bandering van de studie van individuele verschillen. Cattell was de oprichter van het vakblad Psychological Review en heeft de rest van de academische wereld ervan kunnen overtuigen dat psychologie een echte wetenschap was. Net zoals Galton was Cattell geïnteresseerd in individuele verschillen. Hij waagde zich niet aan de meting van 'hogere' mentale processen met het argument dat deze het resultaat worden van 'lagere' mentale processen die gemakkelijker te meten waren. Hij startte systematische metingen bij studenten en introduceerde de term 'mental tests'.

Clark Wissler (Verenigde staten, 1870 – 1947)

Wissler deed een vroeg voorbeeld van een valideitsstudie toen hij de relatie bestudeerde tussen de mentale testen van Cattell en de scores op examens van 300 studenten van Columbia University.

Charles Spearman (Engeland, 1863 – 1945)

Spearman bedacht de Spearman rangcorrelatiecoëfficiënt. Hij oordeelde dat het gebrek van correlaties tussen Cattell's testen en scores op examens het gevolg was van een probleem van de betrouwbaarheid. Elke test bevat een meetfout en dus zal de correlatie tussen twee scores die elk verstoord zijn een onderschatting zijn. Een manier om de mate van betrouwbaarheid te schatten is het meervoudig meten (zelfde test, zelfde persoon) en vervolgens de correlatie te berekenen tussen deze metingen. Ook bedacht Spearman een correctie om het attenuatie effect te verwijderen om zo een betere schatting te krijgen van de correlatie tussen twee scores.

Attenuatie effect: het te klein uitvallen van de correlatie door een toevallige meetfout.

Spearman bedacht de eerste versie van factor analyse. De reden waarom verschillende (test)scores (positief) met elkaar correleren is omdat ze allen het gevolg zijn van een gemeenschappelijke oorzaak (de factor). Hij geloofde sterk in het bestaan van een onderliggende intelligentiefactor 'g', ter verklaring van de correlaties die hij observeerde tussen afzonderlijke intelligentie domeinen.

2.7 Meten van intelligentie

In 1904 vroeg het Franse ministerie van onderwijs aan Alfred Binet (Frankrijk, 1857 – 1911) om een methode te ontwikkelen om kinderen met achterstand te detecteren. Samen met Théodore Simon (Frankrijk, 1872 – 1961) publiceerde hij de Binet-Simon test: een reeks van dertig vragen om de geestelijke ontwikkeling te kunnen meten. In 1908 wordt de schaal herzien waarbij het begrip 'mental level' wordt geïntroduceerd: van elk item werd het niveau vastgesteld, waardoor de test een mentale leeftijd kon meten. Wilhelm Stern stelde in 1912 voor om de score 'mentale leeftijd' te delen door de werkelijke chronologische leeftijd. Lewis Madison Terman reviseerde de Binet-Simon test en publiceerde in 1906 de Stanford-Binet IQ test. Hij introduceerde de term 'intelligence quotiënt' (IQ): $Terman\ IQ = \frac{mental\ age}{chronological\ age} \times 100$. Terman wou IQ hanteren om de beter begaafde kinderen te identificeren om hen te stimuleren verder te studeren. Problemen met

het Terman IQ is dat het niet geschikt is voor volwassenen en dat bij kinderen het IQ niet constant bleef over de leeftijd. Om deze redenen bedacht David Wechsler (Verenigde Staten, 1896 – 1981) een nieuwe definitie voor IQ: $\text{Wechsler IQ} = \frac{\text{individuele score}}{\text{gemiddelde score zelfde leeftijd}} \times 100$. Deze definitie is geschikt voor zowel volwassenen als kinderen. Het gemiddelde is 100, de standaardafwijking is 15. Dit is arbitrair vastgesteld. Wechsler ontwikkelde twee testen, de WAIS en de WISC.

Spearman ging er vanuit dat scores op testen het resultaat waren van een algemene g-factor enerzijds en een taakspecifieke s-factor anderzijds. 'g' was aangeboren, s-factoren konden worden aangeleerd. Cyril Burt (Engeland, 1883 – 1971) werkte nauw samen met Spearman en dacht dat er naast een algemene g-factor er ook andere group-factors waren: factoren die een subset (maar niet alle) testcores beïnvloeden (zoals verbale en spatiale intelligentie).

Bifactor models: elke score wordt beïnvloed door een algemene factor, een domein-specifieke factor en een score-specifieke (error) term. Dit is de moderne terminologie.

Edward Thorndike (Verenigde Staten, 1883 – 1971) was pionier van de leerpsychologie, ontdekker van het operant conditioneren en uitvinder van de 'puzzlebox'. Hij maakte een onderscheid tussen drie soorten intelligente (sociale, abstracte en concrete intelligentie).

Godfrey Hilton Thomson (Engeland, 1881 – 1955) dacht dat het menselijk brein niet uit een 'algemeen' deel (dat overall betrokken was) en vele 'specifieke delen (voor elke activiteit een ander deel) bestond. Hij dacht dat het menselijk brein uit heel veel kleine delen bestond, waarbij telkens een groot aantal tegelijkertijd nodig zijn om een activiteit uit te voeren. Doordat er veel overlap is tussen de delen over activiteiten heen, krijgen we correlatiepatronen te zien, zoals Spearman die zag. Thorndike en Thomson waren beide tegenstanders van de andere bovenstaande wetenschappers wat betreft intelligentie.

2.8 Meten op grote schaal

Arthur Otis (Verenigde Staten, 1886 – 1964) heeft een methode ontwikkeld om multiple-choice items te scoren en zo groeppafnames van testen mogelijk te maken. Dit was de start van massale pen-en-papier testen. Na de oorlog ontwikkelde hij de "Group Intelligence Scale", wat de eerste commerciële mentale test geschikt voor groep-aafnames was.

Robert Yerkes was in 1917 president van de 'American Psychological Society' en startte een initiatief in de Eerste Wereldoorlog om bij alle nieuwe rekruten van het Amerikaans leger psychologische testen af te nemen. Een statistische werkgroep werd opgericht met onder meer Thorndike, Terman, Otis en Thurstone. Zij bedachte de Army Alpha: een test voor het meten van "verbal ability, numerical ability, ability to follow directions and knowledge of information". In de Tweede Wereldoorlog speelde John Flanagan de rol die Robert Yerkes had tijdens de Eerste Wereldoorlog. Flanagan organiseerde een ambitieus testprogramma voor de Army Air Force voor selectie en classificatie van rekruten. Er ging grote aandacht naar de validiteit van de testen.

Terman en collega's ontwikkelden in 1923 de "Stanford Achievement Test" om objectief te meten wat studenten hadden geleerd op school. Thurstone ontwikkelden in 1924 de "Psychological Examination for High School Graduation and College Freshmen", welke resulteerde in twee scores: een taal-score L en een wiskundige/kwantitatieve score Q. In 1947 werd de Educational Testing Service opgericht en die nam de Scholastic Aptitude Test (SAT) af, welke ook resulteerde in twee

scores: een 'verbal aptitude score' die hoog correleerde met Engels en geschiedenis, en een 'mathematical aptitude score' die hoog correleerde met wiskunde en wetenschappen.

2.9 Thurstone, en 'multiple factor analysis'

Louis Thurstone (Verenigde Staten, 1887 – 1955) ontwikkelde voor het eerst factor-analyse met meerdere factoren. Hij ging vooral op zoek naar factoren die (zowat) niet overlappen. Deze eigenschap van factoren noemde hij 'simple structure'. Hij schreef ook een boek waarin hij postuleerde dat intelligentie bestond uit een aantal onafhankelijke factoren. Daarnaast bedacht hij de wet van paarsgewijze vergelijking (law of comparative judgement), die aan de basis ligt van de Thurstone schaal. In 1936 startte hij het vaktijdschrift Psychometrika, samen met Thorndike en Guilford. Thurstone's multiple factor analysis lag aan de basis van vele nieuwe testen die uit meerdere dimensies bestonden of meer specifieke aspecten poogden te meten. De afzonderlijke factoren van intelligentietesten konden worden gecorreleerd door de multiple factor analysis, waardoor hiërarchische factor-modellen werden ontwikkeld, met een tweede-orde factor 'g' die deze correlaties veroorzaakte.

2.10 De opkomst van de moderne psychometrie

Betrouwbaarheid en validiteit werden precies gedefinieerd en dit had een grote impact op de kwaliteit van psychologische testen. Ook kwam er beter begrip van de gevaren en nadelen van een verkeerd gebruik van testen en psychologisch meten in het algemeen.

Hoofdstuk 3 Beschrijvende statistieken voor scores

3.1 Inleiding

/

3.2 Twee datasets

Beschrijvende statistieken:

- Gemiddelde
- Variantie
- Covariantie
- Correlatie

Totaalscore/testscore/somscore: De som van de p scores (voor elk subject)