Psychometrie: LES 1  
Inleiding

* Het is een vak zonder oefeningen, een ander aspect is de timing ervan!
  + Hij zou het willen in het 2de semester van de 1ste master!
    - Want vrij veel dingen die ze doen in de psychometrie zijn vrij geavanceerd!
    - Maar hier is het vroeg in de opleiding, dit is een hele uitdaging voor de lesgever, want statistiek 2 komt pas in het tweede semester 🡪 Opbouw is gebaseerd op Stat II wat niet zo veel is
  + Dit vak moet eigenlijk genoemd worden “introductie tot” psychometrie, er is veel die nog niet aan bod komt!
    - Zowel hijzelf als wij moeten ons bescheiden houden wat dit vak betreft!
      * Op het einde van de rit krijg je een goed voorsmaakje ervan!
      * We zullen essentiële concepten te horen en te zien en dat is het topje van de ijserg!
  + Je zult moeten wachten tot de eerste master 🡪 Toegepaste data-analyse!

# Wat is psychometrie

* Wat is psychometrie? We vroegen het aan studenten, collega’s, en psychometrici.
  + Iedereen kan er wel iets bij voorstellen!

### Studenten

* Slide

### Collega’s

* Slide

### Handboeken

* Slide

### Psychometrika 🡪 Dat clubje, merendeel van de medici zitten erin!

* + Grafiekje 🡪 De impact van psychometrie
  + Officieel vakblad van de Psychometric Society
    - Het staat dus voor een methode, een manier van theorievormen en onderzoek doen in het vakgebied (een specifieke manier) om bepaalde zaken te onderzoeken, deze tak (kwantitatieve) probeert veel van de betere ideeën een beetje te importeren in het vakgebied van de psychometrie/psychologie
  + Op de cover
    - A Journal of Quantitative Psychology
  + Meer dan enkel ‘measurement’ & ‘psychological tests’

### Gerelateerde disciplines

* Mathematische psychologie
  + Ze proberen het neer te schrijven in een taal die zeer streng en eenduidig is, de taal van wiskunde
    - Het voordeel is dat het communiceren makkelijker is tussen de wetenschappers en er is minder ruimte voor interpretatie en nuance
      * Een enorm strenge manier!
* Data-analyse
  + Dat is enorm veel breder!
  + Is niet hetzelfde als statistiek
* Slides

# Belang van psychometrie

### Psychologie als wetenschap

* + Vele theorieën in de psychologie staan nog in hun kinderschoenen
  + Vele van deze theorieën postuleren het bestaan van psychologische constructen, zoals attitudes, motivatie, persoonlijkheidstrekken, gemoedstoestanden, intelligentie, emoties, depressie, racisme, vindingrijkheid, doorzettingsvermogen, weerbaarheid, empathie, bevlogenheid, ...
    - Deze onderwerpen/variabelen zijn redelijk latent! (Niet direct observeerbaar)
      * Dit compliceert de zaak aanzienlijk
  + De theorieën impliceren vaak relaties tussen meerdere psychologische constructen onderling, en/of relaties tussen psychologische constructen en niet- psychologische constructen (geslacht, leeftijd, context, land, ... )
  + Om deze theorieën wetenschappelijk te toetsen, willen we data verzamelen alwaar deze constructen op een of andere manier ‘gemeten’ en met elkaar in verband kunnen worden gebracht

### Psychologisch meten

* + Het meten van deze psychologische constructen is niet evident
    - Deze psychologische constructen zijn vaak niet direct te observeren maar latent en kunnen enkel indirect worden gemeten
    - De meetfout is vaak aanzienlijk
    - De metingen zelf worden vaak uitgevoerd door mensen
    - Hoe weten we nu zeker dat onze metingen het beoogde psychologische construct meten? (Construct Validiteit)
  + De constructie van ‘meetinstrumenten’ en psychologische testen (vragenlijsten, testboekjes, computergestuurde testen, …) is een vakgebied op zich
  + Meten is enkel nuttig (voor wetenschappelijke doeleinden) indien de meetinstrumenten aan bepaalde criteria voldoen: standaardisatie, objectiviteit, betrouwbaarheid, validiteit

### Psychologisch meten in de klinische psychologie

* + Het meten van psychologische constructen kadert vaak in het bepalen van psychologische problemen
    - Leerstoornissen
    - Depressie
    - Stress, burnout
  + Vaak een onderdeel van een breder assessment, waarbij verschillende aspecten gemeten worden, met verschillende bronnen (beoordelingen van ouders, beoordelingen van experten)
  + Meten is een gevaarlijk iets 🡪 Kwantitatief meten leidt tot een score/getal
    - * Het heeft heel vaak consequenties
    - Scores leiden vaak tot classificatie (normaal, klinisch) 🡪 Plots krijgt iemand een label! (KP zullen dat vaak moeten doen)
      * Het lijkt alsof dat het getal de waarheid is!
      * Er zijn heel wat regeltjes <7 🡪 Problematiek!
      * Stel dat je je vergist, dat is een grote verantwoordelijkheid, we hebben een score 🡪 Maar is dit wel belangrijk? Je moet de context weten, de voordleen, nadelen, variabiliteit, hoe te verantwoorden, … 🡪 De functie van psychometrie
    - Scores leiden vaak tot interventie (therapie, medicatie)
* Vanaf hier is het goed, hierboven opnieuw doen BBI

### Psychologisch meten in de bedrijfs-en organiatiepsychologie

* Mensen gaan vele verschillende dingen willen meten, dit is zelfs voordat je op gesprek komt!
  + De verantwoordelijke van deze lijsten heeft ook een grote voorwaarde
    - Want deze bepaalt wie er wel of niet in mag op basis van score
  + De opdrachtgever wilt mensen die minstens een dergelijke score haalt!
  + Het gebeurt constant 🡪 Niet enkel voor selectie maar eigenlijk constant, we worden constant gemeten (evaluatie)
    - Bv. Performance, functioneer ik goed als professor, … Dit wordt gemeten!
      * Als een van de cijfers niet goed is moet hij bij de decaan, blijft het negatief, buiten!! (2x na elkaar negatieve evaluatie!)

### Psychologisch meten in een onderwijscontext

* Wat meten wij 0-20, de bekwaamheid
  + We gebruiken een meetinstrument namelijk een examen = test (meetinstrument)
  + Er worden allerlei vragen gesteld, dit vak zal een MP zijn!
* In andere landen, net over de grens hebben ze veel meer gestandaarde testen
  + Deze test bepaalt of je door mag naar de universiteit
  + Amerika is het nog erger! Je score op het eind van de rit bepaalt welke universiteit je binnen mag en welke niet!
    - Zwarten scoorden lager dan blanken 🡪 Herzien van de test!! Discriminatie!

### De test-industrie

* In de test Myers-briggs Type Indicator zit een hele marketingmachine achter
  + Als je deze invult word je op het einde van de rit word je ingedeeld in een van de 16 types
    - Alle types hebben een type!! Poëtische termen en bedrijven houden hiervan en denken van ja een rustige reflector 🡪 zo iemand moet ik hebben
    - Mensen kopen dan de testen en gebruiken dan die test bij het zoeken van mensen!
  + De wetenschappelijkheid van deze test is quasi nihil!
    - Daar zijn alle wetenschappers het over eens, het is niet valide en het zorgt voor helemaal geen type
      * “Vandaag ben ik rustige reflector en morgen ben ik een agressieve aggregator”
  + Wetenschap zegt dat het niet werkt maar de bedrijven vinden het leuk en betalen veel geld ervoor!
    - De psychometrische kwaliteit wordt betwist, wij als KP kunnen ons dat niet veroorloven

### Het belang van psychometrie

* Iedereen wil meten, maar het probleem is dat meten niet gemakkelijk is
  + Psychologisch meten is moeilijk, aartsmoeilijk
  + Degene die meet en de score die eruit komt gebruikt om acties beslissingen te maken op basis van een meting 🡪 Deze moet geweldig bewust zijn van de problematiek van meten
    - Hiervoor krijgen we psychometrie want meten is zo fundamenteel in onze disciplines!

# Psychmotrie in een oogopslag

## Latente variabelen

* Een voorsmaakje van hoe we gaan redeneren is te zien op dit diagram, wat je moet weten is
  + Dit is de typsiche manier waarop we denken in de psychometrie
    - De n staat voor een latent variabele = Bv. Depressie
      * We gaan ervan uit dat het iets is dat echt bestaat en varieert van mens tot mens
    - De cirkels staan voor latente variabelen en de vierkantjes = geobserveerde variabele
      * In dit geval wordt er vaak gewerkt met zelfrapportage-vragenlijsten!
      * Je moet vaak een vragenlijst invullen; er staat een stelling en je moet zeggen of je het er mee eens bent of niet 🡪 Likert-schaal
        + We doen het met vele vragen!

Samen proberen ze te latente variabele te meten!

* + - In dit voorbeeld zijn er 5 vragen --> 5 scores! -2, -1, 0, 1, 2
      * Je zal 5 sores hebben, op basis van het patroon kan je zien waar de persoon zit een depressief persoon heeft een compleet ander patroon dan een “normaal” persoon
      * 1 item op zich zegt niet veel maar als we vele informatie samenleggen geeft het ons een idee om de mens op een as te zetten
        + Heel depressief – Helemaal niet
  + Factoranalyse
    - Is dergelijk soort van meting waarbij de latente variabele continu is/numeriek, alle mogelijk waarden zijn mogelijk!!
      * Als de geobserveerde ook numeriek zijn dan zitten we in de factoranalyse
    - Als het categorisch zou zijn 🡪 IRT (Het een of het ander, 1 of 0)
      * Wat restrictiever

Deze twee modellen samen vormen de kern van alle modellen! 90%

* + De categorische kolom daar gaan we niet op in gaan! = Categorische wereld waar de latente variabele discreet is 🡪 Gaan we nu niet zien!
    - Het zou zijn als “je hebt het of je hebt het niet, er is geen reiswaarde! Dat is een moeilijke discussie
      * Er is geen continuüm!!
    - De reden is 🡪 Als er iets kapot is dan heb je een ziekte, is het niet kapot dan heb je het niet!
  + In de psychologie hebben we de neiging te kijken in gradaties!!

## Structurele vergelijkingsmodellen

* Je hebt 4 latente variabelen
  + Bv. (random) Leesvaardigheid, Leesmotivatie (sommige zijn heel sterk gedreven)
    - Typisch is het zo dat als je sterk gemotiveerd bent blijkt het zo te zijn dat je ook goed kan lezen!
      * Daardoor wordt je ook beter want oefening baart kunst, want je zal sneller en meer lezen dan iemand die niet veel motivatie heeft en dus minder lees!
    - De andere kunnen variabelen zijn die motivatie stimuleert of juist verslecht!
  + De pijl = Het een veroorzaakt het andere (een causaal effect wordt gesuggereerd)
    - De hypothese zal ons toelaten om te toetsen of die klopt of niet
* Dit diagram reflecteert een theorie een hypothese
  + De psychometrie zal bestudeert één zo’n blokje en dan vooral de validiteit ervan!
    - We moeten alle subgroepen nagaan of ze valide zijn!

# Terminologie

## Meten

* Een getal heeft te maken met iets
  + Bv. Weegschaal dan heeft het te maken met ons gewicht
  + Men wil dit ook bekomen in de psychologie!
* 3 mogelijkheden en 3 categorieën met een ordening in 🡪 Kwalitatieve score en in sommige settingen is dat even nuttig als een kwantitatieve
* Om tot een score te komen hebben we een meetinstrument nodig!
  + Het kan een vragenlijst zijn, een procedure, een echt instrument dat je op je hoofd moet zetten, beeldmateriaal met de pc, …
* Meetfout 🡪 Afwijking tussen de gemeten waarde en de werkelijke waarde = discrepantie!
  + Stel je voor dat je een weegschaal hebt die zegt 10-20-30-40- … kg
    - Fout van nauwkeurigheid
    - Betrouwbaarheid 🡪 Wil zeggen dat je bv. Weegschaal en je staat er op en er staat 70, je gaat er af en diep ademhalen 🡪 60 kg

## Psychologisch meten

* Het onderscheid tussen kwalitatief en kwantitatief zijn heel belangrijk, maar de subverschillen zijn minder belangrijk!

## Soorten metingen

* + Vragenlijsten 🡪 Je wordt gevraagd jezelf te meten
* Fysiologische metingen zijn in zekere zin iet wat objectiever

## Psychologsiche test

* Een test kan op papier zijn, kan op de computer zijn maar het kan ook gedrag zijn!
* Het probeert een soort gedrag uit te lokken bij een individu!
  + Onder gestandaardiseerde omstandigheden zodat men het kan vergelijken onder individuen

## SChaaltechnieken

* Als er maar 1 schaal is 🡪 Is het uni dimensioneel
  + Maar er bestaan ook multi dimensioneel
* Schalen 🡪 Gemeten individu een plaats geven op de schaal

## Software voor psychometrie

* Moeilijk om er aan te geraken!
  + Vooral IRT, SEM
* Deze was niet altijd voor iedereen beschikbaar!
  + Het was ok niet zo makkelijk en transparant
* Maar heden ten dage is alles gratis en publiek beschikbaar woehoew!!

PM: LES2

# Geschiedenis van de PSYCHOMETRIE :’(

## Overzicht tijdslijn

* Bepaalde ontwikkelingen moet je kunnen plaatsen in de tijd
  + Bv. Binet, Simon de eerste intelligentietest 🡪 Feit dat het voor de eerste oorlog was en in een eeuw kunnen plaatsen

## Oorsprong van examens

### examens voor de staat!

* Hoe werd je kabinetsmedewerker? Omdat je de beste bent?
  + Niet noodzakelijk, het is altijd met lange armen en via via
  + Typisch voor de Westerse wereld!
    - Uiterlijk gaan we er van weg maar het speelt nog altijd mee!
* Maar in andere contexten hebben ze een andere kijk op 🡪 Oude china, ambtenaar van de staat dan hielp het niet als je vader of nonkel goed bevriend was met de keizer of zo
  + Je moest een examen doen!
  + Ze doen verschillende disciplines dat in die tijd beschouwd werd van dit moet je kunnen
    - Indien je dit niet kunt ben je geen intellectueel en kan je niet werken voor de staat!
* In pruisen hadden ze een zeer progressieve koning en die heeft dat dan ook voor zijn personeel gedaan!
  + Eerste staat!!
  + Napoleon heeft dat dan ook gedaan

### Examens voor studenten

* Oude Grieken deden geen examens!
  + Het examen was een wandeling en een goede babbel doen met juiste componenten, als dit goed was mocht je een nieuwe functie bekleden!

## De grondleggers

* De namen dat je ziet, zijn de typische namen die typisch geassocieerd worden met de individuele aspecten en dan heb je nog Fechner en Wundt die grondlegger zijn van de experimentele psychologie
  + Wundt had wel evenveel interesse in individuele verschillen als in het geheel, maar hij is gewoon gekender geworden met het geheel-aspect!
* Het is niet de bedoeling dat hij iedereen in kampen duwt!!

## Psychofysica en experimentele psychologie

### Gustav Fechner

* In deze periode bestond het fenomeen ‘psycholoog’ bestond niet, ze waren wetenschappers en fysici en dit en dat, …
* Op dat moment de 22ste oktober is het inzicht gekomen dat de psychische processen toch kon kwantificeren en bestuderen!
  + Dit was het moment dat psychologie ook een wetenschap kon zijn!
  + Dus niet louter een

### Hermann von helmoltz

SLIDE

### Wilhem Wundt

* Laboratorium in Leipzig!
  + Maar James heeft het ook opgericht in de VS!
  + Student van Wundt Spearman zou je kunnen zeggen dat hij de belangrijkste persoon is voor de psychometrie

## Intermezzo: meetfouten in de astronomie

### Galileo Galilei

* Hij bestudeerde sterren en hoopte dat er een patroon in zat
  + Maar hij had door dat observationele fouten onvermijdelijk zijn!
  + Maar de random fouten zijn meestal vrij klein!
    - Je hebt meer fouten in het midden
* Voor de rest hij heeft er niet veel mee gedaan, maar in de fouten zit er precies een systeem is!

### theoretische verdeling voor meetfouten

* Enkele personen zeiden hoe ze dachten dat het eruit zag!
* Gauss is niet de uitvinder van de curve!
  + Maar hij was degene die zei dat het patroon dat we zien in de fouten is deze curve!
  + = Normaalverdeling

## DE eerste studies naar individuele verschillen

### Bessel & Gauss

* Rond de 1800 in de UK

### Adolphe Quetelet

* Zijn interesse was sociale wetenschap!
  + Hij heeft er een wetenschap van gemaakt, een empirische wetenschap die data verzameld en vragen stelt!
    - Ervoor was het louter beschrijvend en filosofisch!!
* Hij is ook de eerste criminoloog

### Charles Darwin

* Elk ras, soort, dier, wezen komt voor in deze planeet met heel veel variëteit
  + Alle kenmerken van levende wezens 🡪 Heel veel variatie
  + Het is dankzij de variatie kan er evolutie zijn! Bepaalde zijn evolutionair in het voordeel!
* Hij was bezeten door de natuur “Fauna & Flora”
  + Calton had meer oog voor mensen!

### Galton (deze moet je googlen)

* Hij heeft heel veel gedaan!
* Dingen die van belang zijn 🡪 Bijdragen aan statistiek! Hij heeft de term regressie gelanceerd, …
* Zijn wetenschappelijke manier om individuele verschillen in kaart te brengen en te observeren en documenteren was van fundamenteel belang
  + Hij was de eerste die dit systematisch deed!
  + Zeer geïnteresseerd in vreemde dingen bv. Het effect van bidden
  + Hij was bezeten door gegevens van mensen en wou dit allemaal in kaart brengen!
* Eugenetica
  + Alle mensen die afwijken moeten eruit, enkel de intellectuelen mogen zich voortplanten!!

### Karl Pearson

* Was ook bezeten door eugenetica
  + Hij vond dat hij moest identificeren wat de goeie genen zijn en hoe die stimuleren en wat slechte zijn en deze mensen weg werken
* Vooral een statisticus
  + Hij heeft veel tools ontwikkeld die wij zullen gebruiken!!
* Het boek van Pearson lag op Einstein zijn kastje!

### Cattell

* De eerste echte grote Amerikaan!
* Eerste professor in de psychologie
  + Hj had heel veel moeite om andere collega’s van andere velden te overtuigen dat het effectief een wetenschap is!
* Grote meneer, heel veel gedaan voor de psychologie!

### Wissler

* In zekere zin was hij de eerste die zei als je intelligentie meet dat als iemand hoger scoort op een test van Cattell dat die betere prestaties heeft op een examen
  + Gemiddeld genomen zou je kunnen verwachten dat mensen die hoger scoren, een hoger IQ moeten hebben
    - Hij vond niks! De implicatie 🡪 (Iedereen was geweldig ontgoocheld)
    - Er was heel veel anticipatie, het trekt op niets die IQ-tests
      * Dat heeft een domper gezet op heel het intellegence
    - Had hij een nauwkeurigere studie gedaan, dan had hij het misschien wel beter kunnen doen!
      * Hadden ze toen beseft dat meetfout de grote problematiek was en achteraf kon je het corrigeren
        + Het hele vakgebied van intelligentie zou een totaal andere weg opgegaan
  + Maar hij heeft wel een opening gemaakt voor een andere benadering!

### Spearman

* Vele realisaties gedaan die van cruciaal belang zijn gebleken!
* Hij was bezeten van correlaties!
  + Hij had het alternatief voor ranggebaseerde correlaties
    - Dit is voor ordinale date en niet voor intervaldata (Pearson)
* Hij was een van de enige dat het probleem van Cattell, een probleem was van betrouwbaarheid!
  + De term is nog niet vaak specifiek aan bod gekomen!
* Een keer dat je weet hoe erg gesteld het is met de meetfout kan je het corrigeren!
  + Je kan dus een correlatie herstellen door rekening te houden met meetfouten!
  + Alle metingen in de psychologie bevatten geweldig veel meetfouten!
    - Een correlatie om het attenuatie effect te verwijderen!
* Hij heeft veel over intellegentie gedaan, het was in die tijd een mysterieus begrip en idereen wou er iets over zeggen!
* Spearman is de man die factoranalyse heeft bedacht!

## Meten van intelligentie

### Eerste STUDIES NAAR MENTALE ACHTERSTAND

### OORSPRONG IQ

* Hij heeft met zijn collega’s een test opgericht die de bedoeling had om kinderen met een achterstand te “vinden”
  + Zodat ze extra inspanning kunnen leveren!
  + Gestandaardiseerde testen zijn in het geheel meer betrouwbaar!
* Binet-Simontest 🡪 Vooral verbale aandacht, dit kan negatief zijn!
  + Hij wou gewoon de kinderen met een achterstand identificeren!
* Er waren zowel kinderen die voor op achterlopen op hun leeftijd!
* Om te weten te komen wat normaal is te kunnen binnen de tests
  + 🡪 Ze hebben een normeringstudie gedaan! (Een van de eerste!!)

### De erfenis van Binet

* Goddard
  + Gebruikte de test voor immigratie…
* Lewis 🡪 Hij was een bezige bij
  + Dit was de man die dacht dat zo’n test interessant was!
  + De test was ook in staat om de goeie er uit te halen!
  + Hij heeft de test wat herwerkt!
    - Het resultaat 🡪 Stanford-Binet IQ test
      * Hij had nog het fatsoen zijn bron te expliciteren
      * De bedoeling was om de slimme ervan tussen te halen

#### David Wechsler

* Hij had door dat de mentale leeftijd stopt en de chronologische leeftijd gaat door!
  + Dus moet je er verandering in brengen om het ook voor volwassene te gebruiken!
  + Kijken naar gemiddelden!

### De G-factor: Spearman & Burt

* Er is daal heel wat controverse over geweest!
* Burt heeft het bifactor model opgesteld
  + Burt was ook een fervent voorstander van de eugenetica
    - Na WOII niet, maar hij ging verder!
    - Burt Affair 🡪 Het schijnt niet te kloppen, de achterliggende complot theorie is dat hij op een zijspoor is gebracht en hij was een te fervent voorstander!

### Enkele tegenstanders

#### Thorndike

* Allemaal verschillende soorten intelligentie, het zijn afzonderlijke dingen, allemaal aangeboren, …

#### Thomson

* Het brein is eigenlijk een machine met allerlei eenheden die constant met elkaar communiceren
  + Voor sommige activiteiten heb bepaalde hersenstukken nodig die gelijk zijn als die dat je nodig hebt bij andere, maar voor sommige andere activiteiten heb je compleet andere delen van je hersenen nodig!

## Meten op grote schaal

### De eerst wereldoorlog: het “arma alpha/beta” project

* Oorlog is enorm destructief! Veel miserie en doden etc.
  + Wat er typisch is, omdat ze onder druk gezet worden gaat de wetenschap met rassenscheden vooruit!
  + Allerlei inzichten!!
  + Dit is geen uitzondering
* Ook psychologen hebben enorm ‘geprofiteerd’ van die oorlogen zeker aan de Amerikaanse kant!
* Ze hebben een test bedacht dat ze duizenden tegelijkertijd gedaan hebben! Ze hebben er bijna 2 miljoen gedaan!
  + Testen op pen en papier
  + DE bedoeling was om een soort IQ-achtig test te nemen om de allerzwakste er uit te halen!
* De eerste grootschalige test heeft heel veel problemen opgeleverd!
  + Er is ook achteraf niks meegedaan met de resultaten!
  + Het waren gewoon dozen met papieren!
  + Maar voor de psychologen die betrokken waren, was dit een enorm goeie leerschool!!

### Na WOI

* Er toen nog heel veel commentaar over de test dat het niet kan dat alle immigranten dom zijn

### DE tweede wereldoorlog

* De test richt zich op het feit dat je bepaalde skills moet hebben!

### DE opkomst van educational testing

* Bv. Toegang tot universiteit

## Thurnstone en ‘multiple factor analysis’

### Thurnstone: factoranalyse met meerdere factoren

* Factoren zijn onderliggende dimensies die vaak geïnterpreteerd werden als fundamentele psychologische constructen
* Ook al was voor hem intelligentie zijn domein is hij voor ons van belang van factoranalyse met meerdere factoren!

### DE impact van ‘multiple factor analysis’

* SLIDES

## DE opkomst van de moderne psychometrie

* SLIDES

PM: LES3  
H3: Beschrijvende statistieken voor scores

# Beschrijvende statistieken voor scores

## inleiding

* Het is verstandig om thuis deze R-codes/oefeningetjes te doen!!
  + Vroeg of laat zal je het moeten kunnen!
* Je moet de R-codes begrijpend kunnen lezen!
  + Als je de R-syntax begrijpt en als je dat al gedaan hebt zal je dat gewoon kunnen lezen en vertellen wat de uitkomst is!

## Twee datasets

* Het grote verschil tussen de datasets
  + Het eerste zijn alle items gescoord op een continue/numerieke/interval niveauschaal
    - We kunnen er gemiddeldes, varianties, … mee berekenen
  + De tweede (vrijdag) dit is letterlijk een examen!
    - Elke examenvraag wordt gescoord in de zin dat het juist of fout beantwoord is
      * Eindscore = somsscore/testscore (kan eventueel getransformeerd worden)
        + 0-20
    - Elke vraag is een item en op elk item krijg je een score!
      * Kan ook een continue score hebben!
      * Likertschalen 🡺 Typisch van -2 🡪 +3
        + Meer range alleszins!
    - Met een beetje goeie wil zijn dit continue variabelen
* Waar staan scores voor
  + “p” subtest = “p” items

### Twee soorten scores

* Metrische scores hebben een grote range bv. Van 1-100
* Niet numeriek 🡪 Nominaal of ordinaal; speciaal geval is dichotome (0/1; J/F)
* Test score is in beide gevallen een somscore en zullen we beschouwen als numerieke variabele!

## DAtaset 1

* Het komt van een vragenlijst en wordt numeriek aanschouwd
  + Deze probeert levenstrevedenheid te meten!
    - Uw leven globaal!
  + “Satisfaction with life scale”
* Hoe?
  + 5 vraagjes
    - “Bijna continu” in een zekere zin
    - Likertschaal
  + De afstand tussen 1-2 is dezelfde afstand tussen 3-4
    - Of dit zo is, daarover is er wat discussie
    - Als je dit nu als een ordinale of als een continue schaal, alle eindresultaten zijn quasi identiek
* Testscore is de som
  + De 5 items = “p”
  + 25 = mijn testscore voor die schaal
* De interpretatie van die score 🡪 Is afhankelijk van de indeling!
  + Dit zijn momentopnames!!
    - De ene dag zie je het iets rooskleuriger dan de andere kant
    - Maar het toont wel aan welke kant van de schaal je staat!
* Het datasetje is artificieel, dus niet echt 100% zuiver

### Data

* N = 20 betekent hier dat je de eerste 20 personen ziet! (Van de 200)
  + De kolommen zijn de items/variabelen en de rijen zijn de individuen!
* Testscore 🡨 RowSums
  + = Het telt voor elke rij de som van alle kolommen!
    - Soms = testscore
  + Omgekeerd pijltje = testscore berekenen we als volgt rowSums
* Cbind 🡪 c= kolom & bind = binden
  + Als je stukken data hebt en ze moeten hetzelfde aantal rijen hebben
    - In dit stuk wil je een extra kolom toevoegen van testscores

### De meetschaal van een 7-punt Likertschaal

* Het is een ordinale schaal, maar het heeft intervaleigenschappen 🡪 De intervallen zouden doorgaans hetzelfde moeten zijn
  + Dus we beschouwen het toch als continu
  + Waardoor we berekeningen kunnen doen die nuttig/relevant zijn

### Beschrijvende statistiek: 1 variabele

* Bij de variantie hoe groter je sample hoe minder verschil tussen de uitkomsten van de twee soorten varianties!
* Varianties zijn gekwadrateerd en neem je er de wortel van 🡪 Standaarddeviatie!
  + Deviaties zijn iets makkelijker om te interpreteren
    - Want varianties zijn gelinkt nog aan covariantie!

### Mean en var in R

* DataSWL = dataset
  + In dat bestand staan er verschillende kolommen en je kiest er één uit
    - Sat1 met een dollar ervoor
  + Je moet zeggen dat Sat1 een kolom is in een bestaande dataframe
    - 🡺 Eerst naam databestand dan $ teken en dan de kolom
      * Alles aan elkaar
* Je kan het ook in 1 keer
  + Colmeans (dataset) 🡪 Alles in een keer!
* Welke items vertonen meer spreiding en welke minder!
  + Daar dienen varianties voor in dit voorbeeld!
    - Bv. Het 3de item 🡪 Daar zijn we het meer met elkaar eens dan met de andere
      * Hij heeft minder spreiding/minder verschil tussen de individuen!

### sd in R

Slides

### Rekenregels voor gemiddelden en variaties

* Als je een constante a toevoegt aan je scores 🡪 Gemiddelde verandert afhankelijk van de variabele
  + Als je alles met een constante vermenigvuldig met een gelijk getal
  + Het gemiddelde is een enorm getal en is het gemiddelde (oorspronkelijk) maal a
* Variantie verandert niet!
  + Als je er overal a bijtel, blijft de variantie identiek
    - Het vergelijkt met het gemiddelde, hoeveel afwijking is er ten opzichte van het gemiddelde
    - Als je iedereen omhoog schuift 🡪 Variatie blijft hetzelfde
  + Als je er een constante vermenigvuldigd met je scores
    - Als je ze vermenigvuldigd met een constante a wordt de variantie groter 🡪 Oorspronkelijke variantie moet je vermenigvuldigen met het kwadraat van de constante!

### Histogram

* Mfrow 🡪 Je zal 6plots maken in plaats van 6 slides ermee te vullen wou hij ze in 2 rijen zetten met 3 kolommen (dus het is niet zo belangrijk)
  + Dit hoeft niet, je moet maar 1 plot maken en that’s it
* Een histogram heeft een beeld waar het merendeel van de responses liggen!

### Boxplot

* P25 = score en 25% van de scores liggen eronder
  + P75 = score en 75% van de scores liggen eronder
* De staafjes geven aan waar het minimum/maximum is
  + Het is een alternatief voor het histogram!
  + Whiskers

### Beschrijvende statistiek: 2 variabelen

* Covariantie is een maat om te kijken over een een samenhang is!
* Dit geeft een getal
  + Bv. 13
* Pearson correlatiecoëfficiënt
  + Een herschaalde covariantie!
  + Dit heeft het voordeel dat het altijd tussen -1 en +1
    - +1🡪 perfecte correlatie 0 is niets!
      * Hoe meer naar 0 hoe zwakker!
    - -.9 is ook enorm sterk!
      * De zwakte wordt bepaald of je dicht bij nul zit of niet

### Covariantie en correlatie in R

* Cov(data$kolom1, data$kolom2)
* De variantie is een zekere speciale vorm van covariantie
  + Covariantie met zichzelf!
* Sommige items hangen sterk samen 1 en 3 en sommige totaal niet 1 en 4
  + Later zal blijken dat dit een zwakte is van een schaal
  + Want de somscore zal niet goed zijn, want de schaal zou moeten verwacht worden dat ze allemaal correleren want een schaal peilt normaal naar de gelijke richting!

### Rekenregels covariantie en correlatie

* Je kan bij beiden een constante a en b optellen/aftrekken
  + De covariantie verandert niets!
    - Ze gaan kijken of de ene evenveel stijgt als de ander
* Je kan bij beiden een constante a en b vermenigvuldigen
  + Je krijgt een heel ander getal!
    - Wat is de relatie 🡪 De originele maal a maal b
  + De correlatie heeft het geen effect op de absolute waarde van het getal!
    - Het voordeel is, het is vrij van schaal!
      * Schaalinvariant
    - Sign (a) 🡪 Teken van a = kan +1/-1 zijn

### Spreidingsdiagram

* Geeft informatie over de samenhang van in dit geval 2 numerieke scores!

### Beschrijvende statistiek: meerdere variabelen

* Variantie-covariantie matrix
  + De diagonaal = variantie
  + Andere getallen zijn de covariantie
  + Hij is ook symmetrisch! (Spiegelbeeld)
  + Round (……, 2) 🡪 Afronding op 2 cijfers na de komma
* Correlatie-matrix
  + De diagonaal = 1

### Relatie gemiddelde testscore en gemiddelde van de itemscores

* Als je de gemiddelden optelt van de itemscores dan bekom je de gemiddelde testscores

### relatie variantie testscores en (co)variantie(s) itemscoers

* De variantie van twee items (x&y) is de som van de individuele varianties en 2x hun covarianties
* Bij meerdere variabelen moet je veel meer covarianties hebben en meerdere varianties!
  + Je kan de variantie doen van de totale somscore als je de sum doet van cov(dataSWL) 🡪 Variantie van de testscore

### Correlatie tussen testscore en itemscores

* .79 is de correlatie tussen de eerste itemscore en de testscore
  + Die zijn relatief hoog (de 4de springt ertussenuit, niet zo goed)
* As.vector is niet van belang voor ons
  + Getallen die staan dan voor elkaar en niet naast elkaar
* .63 is .79 herberekend met een somscore van alle items zonder item 1
  + Dat is de itemrestcorrelatie, wat blijft erover als je het item 1 uit de testscore trekt
  + Dit kan men 5 keer doen en het zal 5x lager zijn

## De normaalverdeling

* Op de basis staan er mogelijke scores = Z-scores
* De formule 🡪 Voor dit vak staan er een aantal formules in, je moet ze niet vanbuiten leren, je moet ze proberen te lezen
  + Op het examen krijg je een formularium!!
* De standaardverdeling met z als variabele
  + Z-waarde
* Wat we vaak zullen moeten berekenen is wat de kans is dat een observatie kleiner is dan -1
  + Wat is die kans?
  + De kans < of gelijk aan 0 = .5
  + De totale oppervlakte = 1

### Cumulatieve distributiefunctie/curve

* Het zogenaamde inkleuren wordt hier anders ingegeven
  + Hoe meer je naar rechts gaat, hoe groter deze wordt!
* Het kan ook omgekeerd!
  + De curve werkt van beide kanten!
    - De ene keer wil je een kans berekenen maar de andere keer krijg je een kans en wil je weten met welke score ze correleert!

### De standaardnormaalverdeling in R

* D = densiteit
  + Dnorm 🡪 Zo bereken je wat op de Y-as zou komen! (gebruiken we niet zo vaak)
* P = probabiliteit
  + Pnorm 🡪 Je neem een z-score en er komt een probabiliteit uit!
    - .16 🡪 De kans dat je een score van -1 of kleiner 🡺 16%
* Q = omgekeerde beweging!
  + Je neemt een kans en je maakt er een z-score van!

### Optioneel) genereren van standaardnormaal verdeelde random scores

* Deze genereert random getallen
  + De r-norm berekent in zekere zin (pseudorandom)
    - Het enige dat je moet meegeven is hoeveel je er wilt
    - Ze volgen uiteindelijk de standaardnormaalverdeling

## Transformatie van testscores

### Transformeren naar z-scores

* We gaan de testscores her-schalen
* Alleen zodat de gemiddelde 0 wordt en de standaarddeviatie 1
  + Dit is enkel mogelijk als het mooi normaal verdeeld is
  + Dit heeft niets te maken met een normaalverdeling!
    - Je herstructureerd gewoon je scores!
      * Het is makkelijker want het blijft tussen -4 en 4
      * Je kan rap zien waar een persoon terecht komt!

### Z-scores in R

SLIDES

### Transformeren naar a priori gegeven gemiddelde en spreiding

* IQ-scores = arbitrair beslist dat het gemiddelde 100 is en de standaardafwijking is 15
* T-scores = variant van de Z-scores
  + Gemiddelde 50 en SD = 10
    - Je doet dit door je z-score maal 10 te doen en 50 erbij te tellen
* Wechscler IQ 🡪 Gem. 100 en SD = 15

### RAngscores

* Moeilijk als ze gelijk zijn, bevoorbeeld meerdere gelijke dingen
  + = Knopen
* Bv. Stel je hebt 5 unieke cijfers 🡪 unieke rank`
* Bij gelijke 🡪 Het gemiddelde 🡪 2.5 2.5 ze krijgen dezelfde rangnummers
  + = Gemiddelde van de rangnummers die ze hadden gehad mochten ze verschillend zijn
    - 3x gelijke 🡪 2,3,4 🡪 Gemiddelde = 3
  + Dit is de manier per defoo en deze volstaat!!
* Deze zijn zeer makkelijk te berekenen!
  + Nadeel is dat ze afhankelijk zijn aan het aantal van observatie
    - 17 op 200 is helemaal anders dan 17 op 20
    - Je zal ze delen door het aantal observaties en nog eens vermenigvuldigen met 100 🡺 Percentielen

### Rangnummers voor testscores

Slide

### Percentielscores en decielen

* Percentiel 80 🡪 Dan zit je hoog wan 80% zit lager dan jou! = P80
  + Het geeft aan op een schaal van 0-100 waar je ongeveer zit
  + 50 = mediaan
* Per 10 = decielen
  + Het eerste deciel is de eerste 10 procet

### Percentielscores en decielen in R

* Je moet ze eerst rangschikken
* 2de lijn is gevaarlijk
  + De rangnummers deel je door het aantal observaties (= length(testscore.rang) erna vermenigvuldig je met 100 🡪 Percentielen
    - Getallen tussen 0-100 met heel veel na de komma
    - Het stuk na de komma laten we vallen, dus we ronden af tot het dichtbijzijnde gehele getal!
      * BV. 53,8 🡪 54 etc.
* Floor 🡪 kapt alles dat na de komma is, kapt hij er af!
  + Maar eerst opgeteld met .5 🡪 Dat is om ervoor te zorgen dat de afrondingen kloppen!
    - Want als je dus eindigt op .7 dan zal je toch nog naar boven afronden!
* Cut 🡪 snijdt het in 10 stukjes
  + Soms willen we ruwer weten waar iemand is!

PM LES4

# Transformatie van testscores

## Genormaliseerde scores

* We doen achteraf een transformatie om ervoor te zorgen dat ze quasi normaal verdeeld zijn
  + 🡪 Genormaliseerde scores
* De eerste is percentielscores te berekenen
  + Deze moet je delen tot 100 🡺 Proporties tussen 0-1
    - 🡪 Op de Y-as bij de cumulatieve normaalfunctie (Z-score krijg je in de plaats van de probabiliteit)
  + Wat er in gaat is een getal tussen 0-1 🡪 Wat er uitkomt is een Z-score!
    - Andere manier om Z-scores te bekomen!

## Genormaliseerde T-scores in R

* 2de lijntje is niet altijd nodig!
  + Als je uitkomst 100 zou zijn 🡺 +∞
    - Dus je wilt stelt dat er een percentielscore is die exacte gelijk is aan 100, als dit waar is, vervang je het door 99.9 en anders blijft het originele!
* Q-norm is belangrijk 🡪 omzetting van proportie naar Z-scores!
* Je ziet nauwelijks het verschil tussen genormaliseerde T-scoresn en niet-genormaliseerde 🡪 Het was al vrij normaal verdeeld
  + Dit kan je zien op slide 52

## Stanines

* Standaardafwijking is niet exact 2, een beetje minder! (1,94)
  + Als je in een bepaalde zone valt krijg je je getal 1-9
  + 5 = Midden, als je 10 groepen hebt dan heb je 5 boven en 5 onder het gemiddelde
  + Maar soms is het handig om een middelpunt te hebben!
    - Het zijn vaak de extremen die belangrijk zijn om te detecteren
* Het zijn exact 9 groepen/intervallen en zo bekomen we 9 scores
* Er zijn nog vele maar deze zijn de belangrijkste

# Normen

* Dit is meer gericht op een grote doelpopulatie!
  + Bv. Mannen tussen 20 en 30 met een bepaald profiel!

## Voor dit soort mensen, wat is het gemiddeldeConversie

* Getransformeerde scores voor een grotere populatie
  + Veel data verzamelen!

### Voorbeeld 1

* Slide 57 🡪 Allemaal studies telkens over een welbepaalde doelpopulatie!
  + Bv. Nonnen in een klooster (zijn ze tevreden met hun leven)
    - Gemiddeld genomen is hun score 25/35 🡪 De hoogste score
  + Bv. Veteranen in een hospitaal
    - Gemiddeld genomen is 11/35 🡪 Helemaal niet tevreden
  + …
  + Zo kan je het zien per doelpopulatie!
    - Je hebt ook standaardafwijkingen, deze zijn ook relevant!
    - Lage SD 🡪 Homogene groep!
  + Laat toe individu beter te bestuderen!

### Voorbeeld 2

* Zeer competitief!
  + De Universiteiten eisen dat ze eerst een gestandaardiseerde test maken!
  + GRE is er een voorbeeld van!
* Dit is een zeer lange test!!
  + Het ingangsexamen arts en tandarts valt er mee te vergelijken!
* SLIDE 59 🡪 Rechterkant zijn percentielen
  + Hier vergelijken ze met andere leeftijdsgenoten
  + Jouw score = rangnummer

## Conversie-tabel (optioneel)

Side 61

# Dataset 2: 10 itemscores examen statistiek II

* Binaire data 🡪 1tjes en 0tjes

## Data-EX in R

Slides

## Testscore

Sldies

## GEmiddelde scores en varianties

* Een item dat veel mensen juist hadden krijgt een hoger gemiddelde
  + Hoe moeilijker het item, hoe lager de proportie correct is
  + Gemakkelijke item hebben een iets hogere proportie!!
    - 4 was een weggever en 2 was niet zo evident
* P1 = proportiecorrect
  + De proportie correct voor item 1 en P7 voor item7
  + Deze proportie = p1 en de variantie is
* N = aantal observaties = 556
  + De variantie is op een nadere wijze berekent
  + De eerste formule deel je door n-1
    - De tweede deel je door n!
  + Piepklein verschil maar moet wel aangegeven worden!
* Het blijkt zo te zijn bij binaire items, als je de proportie kent
  + De proportie correct = 1 – proportiecorrect
    - Daarom moet je delen – 1 en niet n-1!

## Bemerkingen

* 1- porportiecorrect = maat van moeilijkheid!
* Variantie = p(1-p)
  + Hoogste variantie dat je kan bekomen is .25!

## DE relatie tussen 2 dichotome items

* Pjk = de proportie die rechtsonderaan staat = proportie juist,juist
  + De P en de 1 en 2 staan voor items 1 en 2
    - Item j en Item k
      * Proportie = Pjk = .29
* De covariantie tussen twee binaire items
  + Strikt genomen mag/kan dit niet!
* Als je binaire/categorische data hebt dan zijn al die functies en bewerkingen voor continue data, mag je deze niet zomaar toepassen!
  + Vaak alternatieve!
  + Typisch voor proportie is dit de maat van samenhang tussen item j en k
    - Maar de formule levert exact hetzelfde op tussen de 2
    - Dus in de praktijk gaan we toch de covariantie gebruiken

## De phi-coëfficiënt

* Je mag in principe mag je de ordinaire correlatiecoëfficiënt niet gebruiken op categorische data!
  + Maar eigenlijk lijkt dit twee druppels water op de correlatiecoëfficiënt!
  + Maar las je gewone correlatie berekent, krijg je toch identiek hetzelfde
    - Er is een conceptueel verschil!
    - Maar er is nog iets als een computationeel verschil
      * Je komt hetzelfde uit!
* Correlatiecoëfficiënt

## tetrachorisch correlatiecoëficiënt

* Alternatief van de phi-coëfficiënt
  + Dit blijkt numeriek gezien exact hetzelfde te zijn voor de correlatiecoëfficiënt
* Er is ergens een continue score, maar die is ergens in 2 stukjes gekapt en die grenswaarde (treshold)
  + Zit hij er erboven krijg je 1 eronder 0
    - Soms is deze zichtbaar en soms is die niet zichtbaar (latent) maar wel continue score
    - Wat we wel observeren is de dichotome versie!
    - We kunnen er soms van uitgaan dat er een continue meting aan de grondslag was en dat dat de reden waarom we 1 of 0 zien
* Probleem 🡪 Je weet niet op voorhand waar de grenswaarde ligt!
  + Je probeert dus een samenhang te vinden tussen de twee items!
* Het idee 🡪 Die continue maar niet observeerbare scores gaan we observeren met een sterretje!
  + Y-star
  + Stel dat we deze continue scores hadden geobserveerd dan konden we gewoon de klassieke correlatiecoëfficiënt meten!
* Dus we veronderstellen dat de y\* samen een bivariaten standaardnormaalverdeling is
  + Hoe je die moet berekenen is iets anders, maar dit is het concept
* De tetrachorische is de correlatie die je zou bekomen moest het een continue meting geweest zijn!
  + Maar je hebt enkel binaire data, kan je dit berekenen??
    - Ja dat kan, mits een kleine assumptie (de latente score, deze is standaardnormaal verdeeld!)
      * Hij is latent dus het maakt niet uit, dus dat kan arbitrair bepaald worden
  + Elk item apart is standaardnormaal verdeeld!
  + Maar er is nog een bijkomende assumptie (slide 74)
    - Ze volgen een bivariatenverdeling!
      * Samen een tweedimensionale verdeling en daar staan een paar voorbeelden!
    - De kegels verschillen van elkaar afhankelijk van hoe sterk de items met elkaar gecorreleerd zijn!
      * We gaan ergens halverwege snijden!
        + Want dat is meer praktisch!

Dan zien we de contouren en dan zie je beter wat de structuur is!

* + Cirkel 🡪 Dikke cirkel, wel gecorreleerd 🡪 Ellips
    - Hoe sterker de correlatie 🡺 Streep of/smalle ellips!!
    - = Bivariatennormaalverdeling
      * Het enige dat kan variëren is van 0-1 of negatief
        + Dit is allemaal met continue data, maar we zitten met binaire data!

### Bivariate standaardnormaalverdeling + grenswaarde

* De verticale stippellijn is de grenswaarde voor het ene item en dan het andere item heeft ook een grenswaarde een horizontale stippellijn!
  + Je hebt dus voor je data (slide 68)
    - Je kan voor 2 items zien hoeveel observaties er in elk van die kwadranten zitten!
    - Het is hem vooral te doen om de proporties!
* Hoe bereken je de tetrachorische correlatie
  + Je hebt dergelijk figuur (Slide75)
    - Je weet hoeveel procent en welke “vakjes” behoort!
      * 1. Je kan de correlatie aanpassen (kan zwak, stevig, heel sterk, tussenin)
      * 2. Je kan ook met de lijnen spelen (verticale lijn verplaatsen 🡺 De proporties (oppervlakte in het vierkant) vergroten/verkleinen als ze verschuift)
        + Dit kan je doen op zo’n manier dat de oppervlaktes van de kwadranten min of meer gelijk zijn aan de proporties die daar staan!
        + Eens de lijnen vastzitten 🡪 Met je correlatie spelen! (draaien met de hoek)

Dan kan je zien wat het percentage is van kwadrant dat overlapt met elk van die cirkels!!

* + - * + Het is dus een beetje T&E

Een PC krijgt een tabel met frequenties, omzetten naar proporties, maakt hij een dergelijke figuur maken, spelen met de lijnen, dan begint hij te draaien met de hoek en deze hoek bepaald de correlatie!

Tussen -1 en +1

* + - * + Het resultaat hiervan is tetrachorische correlatie!
        + Heel moeilijk gebleven, pas in de jaren 60 is er een betere manier gekomen!

Het is een lange tijd een groot mysterie gebleven!

## Tetrachorische correlaties in R

* Vierkante haakjes worden gebruikt om rijen en kolommen te selecteren in een tabel!
  + Het getal voor de komma selecteert de rij en erna is de kolom!
    - Je wilt niet een bepaalde rij selecteren, dus je laat het weg!
      * “Niks”, kolom
        + Dus alle rijen en enkel kolom v1 en v2
  + C = functie die twee dingen aan elkaar plakt!
    - Het eindresultaat is een tabel met slechts 2 van de 10 kolommen
      * Voor die twee bereken je de tetrachorische correlatiecoëfficiënt
      * Maar daarvoor moet je nog “ordered” toevoegen!
        + Je moet ze niet beschouwen als binair, maar als ordinaal!

Het is speciaal geval van ordinaal!!

* + Dan krijg je een correlatietabel met 1 op de diagonaal!
    - .257 is de fameuze correlatie!
      * Je hoeft dat niet constant voor 2 apart te doen, je kan dat direct met allemaal doen! (Slide 77)
* Tetrachorische waarde is altijd groter (verder af van 0) in absolute waarde!!
  + Het zal altijd iets straffer zijn, iets sensitiever 🡪 Een volledige range afdekken!!
  + Terwijl de phi-coëfficiënt is de range vrij beperkt!! (Niet ideaal!)
* Tetrachorisch wordt als beter beschouwd maar de traditie heeft de andere gebruikt omdat het te moeilijk was, maar nu is dat geen excuus meer!

## Intermezzo Pearson vs. Yule

* Yule had een pamflet, zijn argument was 🡪 Als iemand dood is of niet, ja dat is 1 of 0
  + Maar er is niet iets zoals een continue dood-achtig iets 🡺 Je bent dood of je bent het niet!
  + Hij vond het onderliggende Y\* maar niets!!
  + Er was in die tijd ook een mazelen plaag en ze hadden ook zoiets 🡪 Ofwel heb je ze ofwel heb je ze niet!!
    - JE HEBT HET OF JE HEBT HET NIET!!
      * Er zit daar niets tussen!
* Pearson kon niet zo goed tegen kritiek en zeker niet tegen een jonge snotaap!
  + 200 pagina wederwoord! “In yo face bish”
* 2 kampen ontstaan
  + Is het zinvol om een tetrachorische te brekenen?

## Relatie testscore en dichotome items

* Correlatie tussen testscore (continu) en binair item?
* De onderste rij en de laatste kolom, zijn identiek, deze bevat de biseriële correlatie tussen de 10 items en de testscores
  + De volledige matrix zit in out
  + Hij heeft enkel de eerste kolom afgedrukt die correspondeert met de laatste lijn!
  + Deze items zeggen iets over de itemdiscriminatie
    - Iemand die hoog scoort op totaalscore, heeft die dan ook meer kans dat hij een individueel item juist heeft?
      * De grootste is .71 🡪 Dit is een item dat goed discrimineert in die zin dat een item wanneer je hem juist hebt, dan heb je veel kans dat je totaalscore ook juist zal hebben!
        + Dat is een item dat de goede studenten van de minder goeie studenten kan onderscheiden!
      * .47 🡪 Minder goed
  + Als er nauwelijks resultaten zijn daarvoor 🡪 Slecht item!!
    - Validiteit van de eindscore staat of valt met de kwaliteit van de individuele items!
      * Je wilt graag vragen stellen die er goed discrimineren!

## Overzicht maten van correlatie

SLIDE

Het 1-factor model

# HEt lineaure 1-factor model

## Inleiding

* Sprong van beschrijvende statistiek naar een model!
  + We gaan spreken over de populatie! (Eigenlijk “ongekend”)
* We praten over gemiddelden/varianties/covarianties van “de populatie”/doelpopulatie

## HEt grieks alfabet

* Elke avond voor het slapengaan, probeer je het op te dreunen

## Kansvariabelen

### Discrete en continue kansvariabelen

* Het onderscheid numeriek vs. Categorisch komt hier weer terug!
  + Dus discrete kansvariabelen (discreet
  + Continue kansvariabelen
    - Andere beestjes, andere wetmatigheden
* Bij categorische 🡪 Uitkomst vrij beperkt! De een zal meer of minder zijn
  + Bv. Dobbelsteen!
* Continue 🡪 Oneindig veel uitkomsten!!
* Cumulatieve verdelingsfunctie geeft de kans aan dat de waarde van een variabele X kleiner of gelijk aan x kan zijn!

### Het gemiddelde in de populatie: de verwachting

* Gemiddelde = E(X) bij discrete
* Voor continue = mu x

### Variantie van een kansvariabele

* V(X) of Var(X)

### Covariantie & Correlatie tussen twee kansvariabelen

* Slide

### Rekenregels verwachtingsoperator (gemiddelde)

* Slide

### Rekenregels variantie

* Slides

### Rekenregels covariantie

* Slide

PM: LES5

# H4: 1-factor model

## Het lineaire 1-factor model

* 1904 🡪 Meer dan een eeuwigheid geleden!
  + Het is onwaarschijnlijk dat die man dat heeft bedacht, want het is niet zo makkelijk!
* 1 factor = 1 éta 🡪 Zijn tijd was het 2-factor
  + Het kwam omdat hij zei, elk item is afhankelijk van 2 factoren, een gemeenschappelijke en unieke!
  + Wij nemen gewoon 1, maar bedoelen hetzelfde
* We gaan er van uit dat het continu is
  + In dit model stelt hij vooraf dat er 1 factor zal zijn
    - Geen discussie! Op voorhand beslist (kan foute keuze zijn)
  + We willen bevestiging krijgen 🡪 Confirmatorische factoranalyse
    - Later komt de exploratieve 🡪 We weten dan nog niet hoeveel factoren er zullen zijn!

### Indicatoren

* Op het einde van de rit resulteert dit in een score
  + Alle metingen moeten bedoeld zijn om dat construct te meten!
  + Niet-lineair bestaat ook, maar dat is eentje dat we zullen bespreken later!

### Grafische voorstelling van het 1-factor model

* Éta = fameuze factor
  + = Psychologisch construct die we wensen te meten!
    - Depressie, motivatie, attitudes, …
    - Latent natuurlijk!
      * Niet direct observeerbaar, maar we veronderstellen dat iedereen die heeft en dat iedereen een andere score/niveau heeft daarop!
* Bv. Depressie/bevlogenheid
  + Bevlogen, deze zijn meestal extravert, zijn passievol bezig, …
  + Maar te veel is niet goed van deze bevlogenheid!
    - Mochten we een meetinstrument hebben, direct, zou dat wel lukken!
    - Die scores die we niet kunnen zien, mochten we hem kunnen meten, verschilt die van individu tot individu
    - Maar dit kunnen we niet meten direct
  + We doen het indirect 🡪 Y1-Y6 op deze figuur
    - Dit zijn allemaal scores die we bekomen via een ander instrument die weldegelijk een getal produceert en dit kunnen we zien (observerend)
    - Dit kan van alles zijn afhankelijk van wat je meet
      * Bv. Vragenlijsten, …
  + Het zijn allemaal pogingen dit te meten
    - Het typische van factoranalyse zijn dat er verschillende metingen zijn!
    - Ander alternatief, anders geformuleerd, …
    - Maar wel allemaal met de bedoeling je éta te meten!
      * Hopelijk is er veel overlap tussen die scores
      * Het is wat gemeenschappelijk is, die halen we eruit, wat hebben ze gemeenschappelijk? Misschien is het datgene wat ze zoeken! (kan ook fout zijn)
  + Elk item bevat een stuk gemeenschappelijk en een stukje uniek!
    - Factoranalyse probeert dit te splitsen!
      * = Meetfouten = unieke 🡪 die “e”-tjes
      * Lamda = gemeenschappelijke!
  + Dit diagram is statistisch 🡪 We veronderstellen iets
    - Het is de factor die de items beïnvloedt, hij is de oorzaak!
    - Het idee is, als je een vraag hebt bij een item/meting zal het een invloed hebben!
      * Als de variabele stijgt, dan zouden de items ook moeten stijgen
  + De relatie tussen het item en de factor is niet perfect want deze is verstoort door enkele factoren van ruis/meetinvloeden 🡺 Distortie (cirkeltjes/e-tjes)

### Lineaire relatie tussen de latente variabele en een item

* Je pikt 1 van de items op de Y-as 🡪 Dat is een meting van een item
  + Dat impliceert als de scores stijgen op de latente variabele dan mag je ook hogere scores verwachten op dat item!
  + Kan ook omgekeerd zijn hé!
  + Als de score 0 is op de latente variabele dan is de corresponderende waarde = Mu!
    - = Intercept (van de regressielijn)
    - De waarde voor Y die je bekomt als je x = 0
    - Als je stijgt van 0 🡪 1 = 1 eenheid bij, wat gebeurt met de Y-waarde
      * Hij stijgt met de griekse letter lamda! = Regressieslope/helling
      * Vakjargon 🡪 Factorlading! = Hellingsgraad van een regressierechte
* X-as = continue score (hier Z-scores, maar dat maakt niet echt uit)

### Het statistisch model

* Yj = 1 van de items
  + Zijn score volgens dit factormodel (hoeft niet juist te zijn = assumptie)
  + Het stelt dat deze score geschreven kan worden als een intercept hij stijgt of daalt afhankelijk van de éta score (latente variabele zijn score)
    - Zal een impact hebben op Y en hoe groot deze zal zijn is afhankelijk van de hellingsgraad! (dit moet worden geschat)
    - Deze relatie is niet perfect en wordt vervuild
      * Epsilon = ruis!
        + Deze is een normaalverdeling! (gemiddeld genomen is de meting oké)
  + Dit geldt voor alle indicatoren/items
  + Elke indicator heeft zijn eigen aspecten!!
    - Foutterm, intercept, …
* De variantie van errorterm 🡪 Theta kwadraat!
  + Deze is enorm belangrijk!
  + Elk item heeft een andere variantie, er zijn items die meer ruisgevoelig zijn dan andere
* De variantie van eta is ook een symbooltje xi kwadraat 🡪 Maar deze is volledig arbitrair want de latente variabele bestaat niet
  + In deze cursus gaan we er altijd voor zorgen dat de variantie van de latente variabele gelijk is aan 1 en een gemiddelde van 0 en hij is normaal verdeeld
    - 🡪 Eta = standaardnormaal!
    - Je kan ze niet zien, maar we veronderstellen dit gewoon!
* Foutterm is gemiddeld genomen 0 en er is geen enkele relatie tussen de foutterm en éta
  + Het kan niet zo zijn als eta stijgt dat de fout ook stijgt! Dat kan niet de bedoeling zijn!
* Als we data hebben dan gaan we model fitten!
  + Het komt er op neer dat we in dit model de vrije parameters zullen schatten op basis van data
    - Intercept, lamda, thetakwadraat zijn ongekend
      * Deze moeten worden geschat op basis van dit model
      * Deze 3 🡪 Voor elk item samen = modelparameters!
  + Data zijn metingen van de geobserveerde variabelen!
    - Steekproeven van 500 studenten (rijen) en 10 kolommen (items)
    - Deze gaan in de data-analyse gezet worden
      * Wat er uit komt zijn 10 schattingen voor die trio van de modelparameters
* We willen wel eerst weten of het model adequaat is!
  + We gaan dus letterlijk wat het model beschrijft, vergelijken met de echte data en kijken of ze dicht bij elkaar liggen!
    - Als dit zo is, een goeie fit 🡪 Niet goed passen 🡺 Parameters die we zijn bekomen zijn waardeloos!!
* Het diagram dat we hebben gezien = eenvoudig model 🡪 Congeneriek model

### Het congeneriek factor model

* De fouttermen moeten uniek zijn voor elk item!
  + Mocht dit niet zo zijn 🡪 Zie niet-congeneriek model
    - Hij is deels te verklaren door onze factor en ernaast ook nog iets anders die ervoor zorgt dat ze gecorreleerd zijn
    - Bij 5 en 6 zijn er zelfs 2 factoren in het spel!
* De enige reden waarom indicatoren samenhangen is onze factor en niet omdat er nog andere factoren van belang rijk!
  + Omdat ze allemaal hetzelfde moeten meten zijn ze van dezelfde soort!
* Het mooie en misschien ook het lastige, een keer dat we zo’n model postuleren en we snappen het, willen we nagaan of dat model in staat is belangrijke karakterstieken van de data te beschrijven!
  + Er zijn er minstens 2 🡪 Vector van de gemiddelden
  + En ten tweede 🡪 De samenhang tussen de items
    - Deze is van belang en hoe gaan we deze typisch kwantificeren = variantie/covariantiematrix!
      * Op de diagonaal = varianties en onder de diagonaal de covarianties en deze geven aan hoe de relatie is tussen alle mogelijke paren!
* Je vraagt aan het model, je hebt nu de parameters, vertel mij nu eens, wat zou volgens jou de variantiematrix, hoe zou je er moeten uitzien
  + Op het einde van de rit, zal je een modelversie creëren en we hebben ook een echte datavariantie 🡪 S (sigma)
  + Hoe krijgen we de modelversie?
    - Je moet twee dingen te weten komen (de varianties voor elk item en alle mogelijke paren voor de covarianties)

### Voorbeeld van een niet-congeneriek model

* Dit is wat we niet zullen beschouwen
* Van de fouttermen zijn er die gerelateerd zijn!
  + We hebben er die een andere oorzaak hebben dan de latente construct die wij hebben opgesteld!
    - Maar dit willen we niet!
    - We willen geen resterende correlaties tussen de fouttermen!
  + Die dubbele pijl of iets anders dat mag niet in een congeneriek model!

### Implicaties van het model: de variantie van Yj

* Als we geloven dat dit model klopt, wat is de variantie van Yj
  + Dit moet uitgewerkt worden aan de hand van rekenregels!
    - Zie slides

### Toepassing van de rekenregels (optioneel)

* De parameters mu en lamda dat zijn constante!!
  + Doordat het een constantie is mogen we die er uitsmijten
  + 🡺 We krijgen een andere rekenregels! Soms van de varianties
    - Door nog een rekenregel in de formule (assumpties)
      * Deze covariantie valt weg 🡪 0
        + Variantie van éta is 1
        + Dus kortom de variantie is zie formule!
* Zie slide 23 voor bewijs van de formule!
  + De constante die er nog uit moet, wordt gekwadrateerd
  + Dus je variantie is lamda kwadraat + variantie van je errorterm epsilon
    - Dan heb je volgens het model de variantie van eta, dat kan er goed op lijken maar er ook ver van af liggen!
  + Variantie wordt gesplitst in 2 stukjes
    - Deze twee corresponderen met het gemeenschappelijk deel en het unieke deel
      * De lamda = gemeenschappelijk met het factor
        + = Communaliteit
      * De epsilon = unieke variantie

### Implicatie van het model: covariantie Yj en Yk

* Het is gewoon rekenregels!
  + Je moet dit niet doen op het examen de rekenregels afleiden!
    - Zie slide 25
* We kunnen nu de hele matrix vullen, we weten alle formules!
* Als je S (covariantiematrix) kan vergelijken met het model geïmpliceerde versie!
  + Als ze goed op elkaar gelijken 🡪 goeie fit anders ga je op zoek naar een beter model!

## Generatie van data volgens 1-factor model in R

* We gaan spelen met de parameters om voeling te krijgen waar ze voor staan en wat er gebeurt als je deze verhoogd/verlaagd
  + Zelf data genereren volgens het model, we volgen het principe van het model!
  + We gaan ideale data genereren, niet noodzakelijk realistisch!
    - = Artificiële data volgens hun principes
* Dit doen we aan de hand van een virtuele steekproef die telkens 100 scores genereert volgens de standaardnormaal verdeling
  + Ze staan voor 100 individuen
* We draaien het om
  + Normaal geven mensen allerlei instrumenten om ze te bekomen, maar nu draaien we ze om!
    - We gaan ervan uit dat we scores kennen! Als je ze kent, wat zouden dan de observaties moeten zijn?
  + Als je ze kent en je hebt het model, wat zouden dan de observaties moeten zijn!
* Ene keer hij de scores heeft van de 5 factoren op basis van 3 termen
  + Je zal er dan mee spelen, kijken naar het effect op verschillende beschrijvende statistieken!
  + f.scores, daar zal hij het in steken!
  + Hij zal een aantal items, een intercept waarde, factorlading en een thetakwadraat!
    - Er mee spelen, verhogen, … Wat is het effect op de statistieken!

### Genereren data in R

* Rnorm 🡪 Laat hem toe om 100 scores uit een standaardnormale verdeling te halen deze stopt hij in f.scores
  + Het gemiddelde van de 100 getalletjes zullen niet exact 0 en de standaardafwijking zal niet 1 zijn, heel dicht maar niet perfect, om het wat net te doen zal hij de scores transformeren zodanig dat het gemiddelde exact 0 is en standaardafwijking exact 1 is
    - Hij zal de F-scores het gemiddelde ervan afnemen, centreren, en delen door de standaardafwijking 🡪 Betere f-scores, nu ben je zeker dat het exact is!
      * Z-scores!
* Dit zijn de eerste 6 van 100 scores!
  + Ze zitten allemaal op de Z-scores range (-2 en +2)
    - Uitschieters zijn zeer uitzonderlijk
    - Latente scores die je normaal nooit ziet!! (maar voor 1x wel)
* De 6 indicatorn, scores op Y1-6 is letterlijk door de formule van de factormodel toe te passen!
  + De intersect is random 3,4,5
    - = Arbitrair en ervoor gekozen dat het niet hetzelfde is
  + Factor lading = .9
    - Je doet .9x de factor score + 5
      * Dit zou ideaal zijn, waar het niet dat er nog wat ruis zou zijn!
        + Random ruis!!

Gemiddelde is 0, maar de standaardafwijking zal wat verschillen

* + - * + Variantie = kwadraat van die getallen
  + Zo zie je hoe de data tot stand komt!
    - Je zal de 6 kolommen verzamelen in een data frame! deze heeft honderd rijen, de eerste 3 rijen staan er op!
    - We hebben nu data en we gaan kijken naar de karakteristieke ervan!
* De twee matrixen, zijn van de S!
  + Ze correleren vrij sterk in de correlatiematrix
    - Dat is wat we graag hebben, we weten ook waarom ze zo goed correleren
    - Ze zijn allemaal vertrokken van de factorscores!

### Gemiddelde + covariantie/correlatie geobserveerde scores

* De factorscores is eigenlijk de oorzaak waardoor ze allemaal vrij sterk correleren!
  + Ze verschillen wel onder elkaar want de factorwaardes zijn niet altijd hetzelfde!
* Ongeveer 5, 5, 4, 4, 3, 3 🡪 Lijkt op de intercepten!
  + De gemiddelde correspondeert gewoon met de intercepten!
* De twee matrixen, zijn van de S!
  + Ze correleren vrij sterk in de correlatiematrix
    - Dat is wat we graag hebben, we weten ook waarom ze zo goed correleren
    - Ze zijn allemaal vertrokken van de factorscores!
      * Iets gemeenschappelijk!

### Effect intercepten

* Genereren van nieuwe data!
* Ze hebben een direct effect op de gemiddelden!
  + Groot effect!
    - Dat is duidelijk
* Maar dit is ook het enige!
  + Dit heeft geen effect op de covariantie en de samenhang/correlatiestructuur
    - Het is er ongevoelig aan!
    - Schaal/nulpunt doet er niet veel toe
    - De intercepten, het gemiddelde is anders, maar de samenhang (hetgeen ons interesseert blijft hetzelfde)
  + De samenhang blijft onbewogen zeg maar als je deze doet variëren!
    - Dit is ook de reden waarom ze later zeggen dat we ze niet nodig hebben en zullen ze vaak op 0 gezet worden omdat men het meeste naar de correlatie/covariantie structuur kijken!

### effect factorlading!

* Alle factorladingen gedeeld door 2!
* Effect op gemiddelden heeft het niet!
  + Maar de correlaties zijn allemaal kleiner!!
  + De correlaties zijn gezakt!
  + Hoe sterker de factorlading, hoe sterker de correlaties!
* De rest is terug normaal van parameters

### Variaite errortermen (meer ruis)

* Rest is terug origineel!
  + De gemiddelden zijn quasi gelijk
  + Maar net zoals voordien zijn de correlaties een stuk zwakker!
* Hoe meer ruis, hoe zwakker de correlatie!

### Observaties

SAMENVATTING (zie slide 33)

## FActoranalyse

* Analyse van data + laat het op een model los om zo goed mogelijk de waardes te schatten van bepaalde parameters!
  + We gaan proberen parameters (intercepten, factorlading, variantie errortermen) te schatten voor elk item!!
  + Dat is wat factor analyse doet!
* Praktijk 🡪 R = Factanal
  + En een andere = CFA (confirmatorische factoranalyse)
    - Deze vind je in de Lavaan bibliotheek!

### Stappen in factoranalyse

* Je moet op voorhand een goeie selectie maken van welke items in je model gaan
  + Bepalen van de indicatoren/items!!
    - Desnoods in overleg met collega’s etc.
  + Dus je moet redelijk zeker zijn dat die items bedoeld waren om de constructor te meten!
* Eens je dit weet kan je al die scores bekomen in een steekproef
  + Na het bepalen van de indicatoren
* Dan heb je data!
  + De ruwe data volstaat! (evenveel rijen/kolommen als je data)
    - CFA 🡪 Ze stoppen de ruwe data in het model en dat volstaat
* Het eerste dat achter de schermen gebeurt is de S (variantie/covariantiematrix) berekenen
  + Een van de twee programma’s zullen ook r berekenen!
    - Correlatiematrix
  + De ruwe data gooien we weg en de gemiddelden ook
    - We kunnen met het dataoverschot de parameters schatten!

### FActoranalyse in R

* 1 lijntje en dat is het!
  + Al de gegevens zitten in data!
  + Je verlangt 1 factor 🡪 Moet er bij anders zit je in een exploratieve modus!
    - Al wat eronder staat is de output
    - Loadings = belangrijkste 🡪 Factorladingen!
    - SS loadings 🡪 Variantie van de error term! (5.44)
    - Proportion var kan je voorlopig negeren!
* Het getal onderaan
  + P-waarde
    - Deze krijg je altijd als je een statistische toets uitvoert, we starten met een hypothese 🡪 nulhypothese = model is een goed model
      * Je wilt dit weten
    - De mate van evidentie dat mijn nulhypothese waar is!
      * Vrij groot want het is een getal tussen 0-1
      * Hoe groter hoe liever!
        + Dit is te mooi om waar te zijn want hij heeft zelf zijn data gegenereerd!
    - Als het te klein wordt 🡪 (Te klein daar is discussie rond!)
      * Hangt van de complexiteit van je model af, het is ook kwestie van interpretatie en overleg met collega’s!
      * Hier is de boodschap dat het goed past!
* Vrijheidsgraden geeft aan wat het getal erboven zou moeten zijn!
  + Zeker niet veel groter!
    - Bv. Stel dat het 100/200 was 🡪 JE zal zien dat de P-waarde dan echt enorm zou dalen!

### Interpretatie

* De eerste 4 items hebben een hogere factorlading dan de laatste 2
  + De interpretatie hiervan is letterlijk dat de eerste 4 items blijkbaar betere indicatoren zijn van je latent construct dan de andere!
  + Zo’n analyse laat toe je kwaliteit van je items meteen in te schatten!
  + Je ziet het ook bij de varianties, de laatste twee hebben duidelijk meer ruis dan de eerste!!
    - Ze lopen hand in hand dus ze zijn symmetrisch gelinkt aan elkaar!
      * 🡪 Hoge factorlading 🡪 Weinig ruis

### FActoranalyse in R (2)

* DE manier om dit te lezen 🡪 f = arbitraire naam!
  + Kon bevlogen zijn
  + De naam die je geeft aan je latente variabele
* Opeenvolging van 2 symbooltjes, gelijkheidsteken en tilde 🡪 2 toetsen die je moet indrukken
  + 🡪 1 symbool!
    - Je moet dat lezen als: “De factor f wordt gemeten door de indicatoren y1 - … y6
    - ‘Wordt gemeten door’
* Een lijntje 🡪 Later gaan we meerdere factoren hebben en gaan we meerdere factoren hebben!
  + Nu we onze items hebben gaan we cfa toepassen!
    - Std.lv = true 🡪 Latente variabele daarvan wordt de standaarddeviatie 1 wordt en gemiddelde = 0
    - Meanstructure : true 🡪 Niet echt nodig
      * Maar is voor de intercepten te zien!
* Alles komt in een fit!
  + Met summary fit krijg je een soort overzicht en extra standardized is true (zien we straks)
* Je krijgt een wat langere output!
  + Het eerste wat je ziet is de modeltoets 🡪 Dat i opnieuw dezelfde toets die nagaat of je model past!
    - Deze p-waarde die niet identiek is aan degene van de vorige slides
    - Deze ziet er goed uit, vrij groot!
    - Conclusie is dat het model past!
* Wat van belang is voor ons is de eerste kolom!
  + 3 blokjes factorladingen, intercepten en variantie! 🡪 ESTIMATE
  + Factor moest variantie 1 ebben
  + Nu zie je wel dat ze heel dicht bij de originele waarden liggen!
    - Zowel de factorladingen bijvoorbeeld
* Mochten we 1000 observaties hebben zouden we er nog dichter op zitten
  + Hoe meer info, hoe preciezer we de waardes kunnen bepalen!
* Z-value, als we met een andere steekproef dezelfde parameter gaan schatten dan zullen we
  + Hoe groter hoe meer ruis, hoe kleiner hoe preciezer!
  + Z-scores 🡪 Deze moeten rond de 0 liggen als mijn stistiek dicht bij 0 ligt, maar deze zijn heel extreem!
    - Ze liggen ver buiten de -4/+4
      * Omdat je aan het toetsen bent of deze parameter gelijk is aan 0 en nee, ze zijn redelijk ver van 0
        + Dus in alle gevallen moet je de nulhypothese verwerpen!
        + Wat in deze kwestie goed is!
  + Als mijn intercept 0 zou zijn zou dat vreemd zijn want ze waren 5/4/3!!
    - De laatste is van belang degene ervoor niet!
    - De eerste is een herschaling van de eerste
      * De eerste is de ruwe!
    - De laatste 🡪 Gestandaardiseerde parameters!
      * We hebben de schaal een beetje aangepast!
      * We doen dit omdat we het nu kunnen vergelijken onderling!
      * Dus ruwe, niet gestandaardiseerde kan je niet vergelijken!
        + Maar dit willen we wel zien, welk is beter dan de ander!
        + Je kunt ze niet rangorden omdat de schaal anders is!
  + Bij all 🡪 Zie je dat de eerste beter samengebracht zijn en de andere zijn minder, je ziet het hier veel beter!
* Het eerste tabelletje de laatste kolom, die getallen zie je in factanal!
  + Daar zijn ze direct gestandaardiseerd en zie je de originele niet!
* De 9 staat voor een getal dat zou moeten rond liggen!
  + Stel dat het 100x groter zou zijn, dan zou dat enorm slecht zijn!

### Interpretatie

Zie bovenstaande en slides!

### Kwalitatieve interpretatie

* We gaan vooral kijken naar de factorladingen!
  + .7/.8/.8 vinden we vrij tot zeer goed
  + .5/.4 🡪 Vrij tot zeer zwak!
  + We kunnen eventueel kijken wat zijn de beste items, …
  + Je zou ze kunnen rangschikken!
* Errorvarianties op zich zijn niet zo gek veel van belang, want ze volgen het patroon van de factorlading (eentje van de twee volstaat)
  + Als je de factorlading bekijkt zie je de spiegelbeeld!

## FActorscores

* Deze kan je berekenen, maar blijft altijd een soort schatting
  + Je zal per individu een score krijgen, deze zullen standaardnormaal verdeeld zijn!
    - Lijkt op z-scores
  + Je kunt zien wie er rond het gemiddelde ligt, hoger, lager etc.
  + Deze zijn enkel zinvol als het model goed past bij de data!
    - =Minimumvoorwaarde! (Maar nog niet voldoende)
  + Er is nog een voorwaarde
    - De factorladingen, deze moeten sterk/hoog genoeg zijn!
    - > .7 🡪 Relatief sterk; .8 heel sterk; .9 is uitstekend,
  + Factor die lager zijn 🡪 Ook al heb je een goed passend model dan heb je meer en meer ruis!
    - Dus als je factorlading zakt 🡪 Stijgen de unieke variantie en deze zorgen voor ruis en daardoor zullen ze minder betrouwbaar zijn!
      * Hoe meer ruis, ruis, ruis hoe minder betrouwbaar en minder zinvol!
* Eens je ze hebt kan je ze wel gebruiken om mensen te vergelijken met het gemiddelde!
  + “Wie zijn de mensen die hoog score, laag, gemiddeld, …

### FActorscores in r

* 1 lijntje predicted 🡪 Factorscores geschat zoals mijn model
  + Ernaast staan de originele factorscores die je zelf hebt gecreëerd!
  + Je zult zien dat ze vrij goed op elkaar gelijk
* F = factorscores
* F.scores = de scores die je zelf hebt gecreëerd
  + Ze lijken goed op elkaar en correlatie tussen die twee kolommen is maar liefst .99755
    - Beter dan perfect, (praktijk zal dit niet zo goed zijn)
    - Dit is het ideaalbeeld!!
  + Als het echt goed past en je hebt goeie factorladingen 🡪 (Bovendien weinig ruis) dan krijg je goeie factorscores!!
    - Slechte modellen 🡪 Geen zin
    - Zwakke factorladingen 🡪 Ook geen zin!

## Factormodellen met minder parameters

* Bruggetje naar vrijdag 🡪 Klassieke testtheorie = factoranalyse maar een speciaal geval!
* Het heeft te maken met de vrijheid dat je geeft aan al die parameters!
  + Intercepten kunnen voor elk item anders zijn, voor de factorlading kan voor elk item anders zijn, elk item heeft ook zijn eigen errorvariantie!
    - Dus elk item had hier zijn eigen waarden!
* Wat kan je doen?
  + Je kan eisen dat het model zo te werk gaat dat alle factorladingen hetzelfde moeten zijn 🡺 Minder vrijheid! = Restrictie 🡪 Minder goed passen dan het origineel!
    - Maar als het verschil nauwelijks merkbaar is dan prefereren we het eenvoudig model
    - Hoe eenvoudiger hoe liever want het heeft weinig parameters!
* De vraag zal zijn willen we intercepten? Ja of nee
  + 🡺 NEE

### 6 verschillende modellen

* Het essentieel congeneriek 1facor model
  + Wat is er verandert 🡪 De intercepten zijn weg!
* 3 zijn essentieel en de 3 andere niet, essentieel is met intercepten en anders is het zonder!
  + Totaal 6 modellen!
  + Meestal zonder intercepten want ze hebben toch weinig “nut”
* Tau equivalent 🡪 Factor lading is hetzelfde als iedere item!
  + Er is maar 1 getal voor een gemiddelde factorlading, deze gebruiken we voor alle items 🡪 Geen indexering meer!
    - Kan dit? Is er evidentie voor? In de data dat we dit kunnen permitteren 🡪 Dat moeten we testen!
* Parallel model 🡪 Het gaat over de factorladingen zijn gelijk en de errorvarianties zijn gelijk aan elkaar
  + Als de j index wegvalt en ook de lamda, alles is hetzelfde voor elk item 🡪 Alle items gedragen zich identiek

### Congeneriek, tau-equivalent en parallel model

* Congeneriek 🡪 Alles is verschillend
* Tau 🡪 Factorlading hetzelfde
* Parallel 🡪 Meest restrectief, items zijn gelijk voor iedereen!

### Fitten congeneriek, tau-equivalent en paralelle modellen in R

* Hoe kunnen we dit nagaan?
  + We kunnen dit empirisch nagaan!
  + We gaan eerst een model fitten van een congeneriek model waar we alles loslaten en de intercepten spelen geen rol!
    - Erna vinden we een tweede alternatief model
    - Hierin worden de factorladingen verplicht gelijkgesteld, deze zal per definitie minder goed passen dan het eerste
      * Maar dan zullen ze een modelvergelijkingstoets toepassen!
    - Nulhypothese is 🡪 Twee modellen zijn gelijkaardig
      * P-waarde groot 🡪 Behouden van 0-hypothese => Gerestricteerde versie is even goed!
      * Maar bij kleine waarde (.05) = slecht nieuws want dat model is te restrictief! 🡺 Terug naar het origineel!
  + Hoe doen we dit?
    - Bovenaan het originele model
      * F wordt gemeten door 5 indicatoren en we fitten dit
    - 2de blokje 🡪 Tau-equivalent model, hier zal hij zijn factorlading gelijk zetten
      * De letter l 🡪 Lx y1 + lx y2 + …
      * Dit noemen we labels!
        + We gaan de factorlading gewoon een naam geven!
        + Deze komen in de tabel 🡪 Zoals je ziet zijn de factorladingen gelijk aan elkaar!
    - Dit komt door hetzelfde label te gebruiken voor als je factors!
    - Het past nog steeds goed, we zien een vrij hoge p-waarde!
    - Fouttermen mogen nog altijd uniek zijn
  + Parallel model!
    - Twee keer een tilde 🡪 Is een teken voor variantie/covariantie!
      * Als de naam van de variabele links en rechts hetzelfde is, dan is het een variantie!
      * E 🡪 Error en we geven 5 keer hetzelfde labeltje en dat zal als gevolg hebben dat het pakket weet dat de variantie allemaal hetzelfde moeten geschat worden
        + 🡺 Gelijkheidsrestrictie
    - Oei oei, P-waarde is zeer klein!
      * 🡪 Dit model, past niet in de data!
        + Dat is iets dat je kan leren
* Dus je hebt 3 modellen!
  + De factorladingen leken goed op elkaar 🡪 Daardoor kan je ze gelijkzetten!

### Modelvergelijkingstoetsen

* WE kunnen formeel toetsen of het ene model slechter is dan het eerste, is het derde echt slechter dan het eerste, …
* Anova 🡪 Zeer ongelukkig gekozen, dit kan geassocieerd worden met iets van experimentele designs
  + Deze functie dient voor modelvergelijkingstoetsen, wat er in gaat zijn modellen en die gaan vergeleken worden met elkaar en je krijgt drielijntjes
  + Tau-equivalent past goed p-waarde, parallel niet! Klein getal!!
    - Parallel past niet!!
      * Er zijn enkele formules die enkel werken bij parallelle modellen!
* We gaan van flexibel naar restrictief!
  + We hebben 4 vrijheidsgraden want we hebben 4 restricties, de eerste moet gelijk zijn aan de 2de, moet ook gelijk zijn aan de 3de, …

## FActoranalyse van de “satisfactoin with life” schaal

* 5 items
  + 200 observaties geedaan!
* Hij fit een congeneriek model
  + Dit is meteen balen, maar dit is de realiteit, meestal van de eerste keer krijg je geen goed nieuws
  + Het goede nieuws is, je kan het nagaan of het goed werkt of niet!
  + WE hebben een dataset, we hebben een schaal en ze wrden gemeten op hoe tevreden ben je
    - Je stop deze in een factor analyse, je wilt nagaan of ze hetzelfde meten
    - Het antwoord is neen 🡪 Want het 1-factor model lijkt niet te passen!
      * 🡪 Slecht nieuws voor de schaal!
        + Later zien we een oplossing, het zou kunnen dat er twee factoren nodig zijn!
  + Je gaat na of items die bedoeld zijn hetzelfde te meten, of ze wel echt dit meten!
    - Vanaf dat je iets meet en je bent van plan hiermee aan de slag te gaan om beslissingen te nemen (slagen of niet, ziek of niet, …)
      * Controleer dit!!
      * Als het model niet past bij de data ben je onzin aan het verkopen

### Interpretatie

* Als je zo’n model ziet dat nietpast bij de data, hoef je niet kijken naar de parameters
  + Deze zijn wel meer informatief!
    - De eerste is goed, derde ook, maar de 4de is bedroevend… (.345 is te zwak voor een ernstig instrument)
    - Er moet iets mee zijn waardoor het niet past!
      * Dit item wordt door de mensen die het lezen anders geïnterpreteerd dan wat er bedoelt was!
        + Het meet iets en een deeltje van wat het bevraagt is inderdaad gemeenschappelijk, maar er is teveel uniekheid!
  + Je wilt boven de .7 geraken!
* Latere meerdere factoren proberen!

## somscores

* Met meetinstrumenten is dit hetzelfde
  + We hebben items en wat velen doen is de scores optellen! (eventueel delen 🡺 Gemiddelde)
  + Wel nu 🡪 Somscore
* Eigenlijk zet je daar een sterke assumptie, je kunt je dit enkel permuteren als je een factormodel hebt met slechts 1 factor!!
  + Dus je moet een factormodel hebben die goed past bij de data, waarbij de factorladingen gelijk zijn aan elkaar en de ruis hetzelfde is voor iedereen
    - 🡺 Parallel model
* GEEN PARALLEL MODEL 🡪 Mag je in se geen somscores berekenen, je kunt het doen, maar zo veel ruis dat het onzin wordt!
  + Waarschuwing!! Enkel verantwoord als we hebben nagegaan of het congeneriek past, dan tau en dan parallel!
    - Dan enkel dan mag je somscores toepassen want anders is niet betrouwbaar!

# Voorbeeldvragen

* 1ste 🡪 Inleiding!
  + DE eerste vragen zijn vrij logisch en letterlijk
* Vraag 4
  + 🡪 Sws variantien hieromtrend!
  + Originele x gedeeld door de standaardafwijking 🡪 Z-scores dus de Y zijn z-scores
  + Tweede lijn 🡪 Transformeer je de z-scores \*7 + 20
    - Makkelijkste dat hij kon bedankeen!
    - Gemiddelde wordt 0 🡪 \*7 blijft 0
      * Zal niet zo evident zijn!
* Vraag 5
  + Rangnummers worden omgezet naar proporties
  + Pnorm
  + Daarna nog eens iets
  + Er staat een fout in zijn syntax
    - Pnorm moet Qnorm zijn…….
* Vraag 8!
  + Bij tau-equivalent moet je kijken naar de estimate
  + D klopt!

# Klassieke testtheorie

## INleiding

* De term is zeer slecht gekozen!
  + Het suggereert dat er zoiets is als een moderne (mensen denken dat het de IRT is)
  + Klassieke oudere literatuur, dat wel, van rond 1900
* Eigenlijk zou dit hoofdstuk “betrouwbaarheid” moeten genoemd worden!
* Het concept van betrouwbaarheid van een score
  + Bv. 17/20 hoe betrouwbaar is deze, we gaan proberen dit te formaliseren, hoe wordt dit gedefinieerd, hoe kunnen we dit te weten komen?

## KTT voor 1 subject

### Een gedachtenexperiment

* Een experiment die we niet echt kunnen uitvoeren, dus we doen het virtueel!
  + Een populatie en je neemt kinderen van 9 jaar = doelpopulatie
  + Je zal iets meten
  + Het is een gedachtenexperiment 🡪 Want je kan alle herinneringen wissen vanuit het geheugen tot net het moment vooraleer het werd afgenomen en hij vult het nog eens in de test
    - Verbeter het opnieuw dit keer is de score anders 🡪 76/100
      * De ene keer had hij toevallig iets juist en heeft hij gegokt, de eerste plafond en toen kwam het op het plafond
      * Alle scores zullen dicht bij elkaar liggen 🡪 rond de 80!
    - Bv. 10 000 scores bij dezelfde scores 🡪 Wat is nu zijn ware/echte score van alles 🡪 Het gemiddelde!
      * Door het gemiddelde te nemen zal de ruis verdwijnen van de ruis is soms meer of soms minder!
* De notatie 🡪 X is een score en de index i staat voor het subject ‘i’
  + Eens je een score ziet x, is een reële score een getal!
  + Terwijl X = net voor de test, dat is de score dat je nog niet gezien hebt!
    - Je moet hem nog afnemen = kansvariabele, deze is nog niet ingevuld!
    - We weten nog niet wat het zal zijn, dat hangt ervan af!

### Basisassumpties KTT voor 1 subject

* De echte waarde is de Griekse letter tau
  + Het verschil tussen echte waarde en de nog te observeren test score
    - 🡪 Is de meetfout
  + Bijgevolg bekomen we het klassiek moeetmodel voor 1 subject (testscore die we niet hebben) die stelt dat de testscore kan worden opgedeeld in twee delen 🡪 Ware score en de meetfout!
    - Deze zal een functie zijn van de echte score en de foutterm!

### KTT voor 1 subject in R

* Stel je voor dat de echte score 15 is
  + Je doet het nog vele keren (100 000X)
    - De eerste 6 staan er op!
    - Dat schommelt rond de 15
      * Er is redelijk wat variatie want de standaardafwijking van 2 is best wel veel
  + De verwachte waarde (mean(rep.scores)) = E
    - Zo goed als 15, als je het meer en meer doet is het nog dichter en dichter naar 15!
  + De variantie is 3.996 🡪 Te verwachten want de sd = 2 🡪 4
  + Sd, getal zeeer dicht, het is niet exact omdat je maar 100 000 replicaties gedaan hebt

### Implicaties van het KKT model voor 1 subject

* Zie A1 = Zie appendix
* Aan de hand van de assumpties van KTTA volgens deze implicaties
  + De verwachte waarde van de foutterm = 0
    - Iedere keer als je meet zit je er een beetje naast, onder/boven 🡪 Varieert, wat is het residu krijg je weer een distributie met een gemiddelde 0
      * Dit kan worden bewezen!!
  + Variantie foutterm = zelfde als variantie van Xi
    - Als je de variantie weet, weet je ook de standaardafwijking = wortel!
      * = Standaardmeetfout voor subject “i”
  + De informatie over de ware score is het omgekeerde van de variantie!
    - Hoe meer spreiding, dus hoe groter de fout kan zijn hoe minder informatie we zullen weten!

## KTT voor een populatie van subjecten

* De ware score van een nog niet geselecteerd subject noteren we als een T (hoofdletter tau)
* Bij X en T staan er geen indexen!!
  + Waar staan ze voor, dat is een potentiële score die we niet hebben geobserveerd van een kind die we nog niet hebben gekozen
    - Dan moeten we dit toepassen op een kind om een x te krijgen
    - Heel veel onzekerheid 🡪 Kans variabele!
  + Maar we weten niet wie dat kind zal zijn en of wat zijn of haar score zal zijn!

### Het Klassieke meetmodel voor een testscore X

* Het is dezelfde formule maar de index is verdwenen omdat je niet weet welk kind het is!
  + Alle 3 zijn kans variabelen
  + Dit is wederom een assumptie en we moeten het maar aanvaarden! (Valt niet echt te controleren)
    - Kan een beetje maar dat is voor later!

### Implicaties van het KKT model voor een populatie van subjecten

* Eerste implicatie is gelijk aan de vorige!
* 2de 🡪 Variantie errorterm = gemiddelde van de variantie van verschillende varianties van de errortermen van de individuen in je populatie!
* Standaardmeetfout van een test!
  + Als je weet dat er veel meetfout op zit dan moet je de cijfers achteraf met een korrel zout moet nemen!
    - Standaardmeetfout moet bijgevolg zo laag mogelijk zijn!
* De T is een latente variabele, de X kunnen we potentieel observeren, maar de ware score, wat we pogen te meten gaan we nooit observeren maar de gemiddelde scores zijn gelijk!
* Variantie van onze testscore die we zien kan gesplitst worden in twee stukjes
  + De variantie van de ware score en de variantie van de meetfout

### Meerdere testscores

* Check slide

### KTT voor N subjecten in R

* Ware score die je in t stopt zijn de echte scores!
  + Deze zou je graag willen zien met je test
    - Maar zal niet perfect zijn en omwille van de meetfout zal je iets anders zien!
* E 🡪 Je maakt er Z-scores van, de gemiddelde zal een exacte 0 hebben! En Exact 5 als variantie!
  + Gemiddelde is zo goed als nul!
* Bij de geobserveerde score 🡪 Gemiddelde is nog altijd zo goed als 50
  + Dus de verwachting van 50 en de verwachting van de populatie is ook 50
* Het zijn illustraties om implicaties te testen!

## Betrouwbaarheid: definities

* De waarde wordt gekwadrateerd zodat het altijd een positief getal tussen 0 en 1 is!
  + Het kwadraat van de correlatie tussen de gemeten scores en de scores die we zouden willen zien
    - 🡪 Maar in de praktijk kunnen we er niets mee doen, want de T is latent en kunnen we nooit zien
  + Variantie T/Variantie X
    - Hoe groter de meetfout hoe zwakker de betrouwbaarheid!
    - Elke definitie waar de T in staat, daar kunnen we niets mee doen!
* De variantie van doelpopulaties kunnen verschillen van populatie tot populatie
  + Dus de variantie van echte/ware scores kunnen wijzigen!
* Betrouwbaarheid van een test 🡪 Het kan dus zijn dat de test, betrouwbaarder is bij 9-jarigen dan bij 11-jarigen of omgekeerd
  + Het is dus niet per se een eigenschap van de test, het is een eigenschap van een test op hoe hij is afgenomen van een bepaalde doelpopulatie!
  + De betrouwbaarheid is niet louter een eigenschap van de test zel maar ook van de doelpopulatie!

### parallelle testen

* Er is een tweede definitie, de T is verdwenen 🡪 alternatieve X (X-prime/accent)
  + Stel dat je een alternatieve test hebt, ook voor leesvaardigheid
  + We gaan ervan uit dat ze hetzelfde meten, laten we er even van uit gaan dat ze beiden equivalent zijn
    - Stel dat je de ene neemt of de andere, maakt niet uit want je komt hetzelfde uit!
    - Perfect inwisselbaar!
      * De correlatie tussen de X en de X-prime is een maat van betrouwbaarheid
        + Het enige verschil dat er zou kunnen zijn is door de meetfouten!
* Dit is al iets praktischer, want het is zeer moeilijk in de praktijk, maar we kunnen iets verzinnen dat erop gelijkt!
  + De definitie is hetzelfde mits er voldaan is aan de “hieruit volgt”
  + Maar dit opent de deur naar praktische manieren om dit te doen!

### praktisch belang betrouwbaarheid

* Cijfers komen van meetinstrumenten, meetinstrumenten die zijn ofwel heel betrouwbaar/minder betrouwbaar
  + Je moet daar een idee van hebben om het cijfer te plaatsen/bepalen
* Natuurlijke variabiliteit = standaardmeetfout 🡪 Dit zegt het cijfer kan naar boven of beneden gaan met ongeveer zoveel
  + Schatten van de ware scores en vooral hun precisie (optioneel)
    - Zie later
      * Het laat ons toe om meetlijsten aan te passen??
    - Er is een gulden regel 🡪 Hoe meer vragen op het examen, hoe betrouwbaarder het eindexamen!
      * Bv. Maar 4 vragen 🡪 Is niet echt nauwkeurig!
        + De lengte van het examen is te kort
        + Meer 🡪 Betrouwbaarder
        + In praktijk is dit een probleem mag niet langer dan 3uur duren! (Logistiek, aandachtsspanne, fysiologische verschijnselen)

## Toepassingen van betrouwbaarheid

### Standaardmeetfout van een test

* Slide

### testlengte en de spearman-Brown formule

* Stel X is de soms van P-items en de betrouwbaarheid van X = Rel(X)
  + Algemene formule van een langere test 🡪 Slide
    - Hoe groter hoe betrouwbaarder
* Stel: betrouwbaarheid is .60
  + Je zal een test 4x langer maken!
  + Of dat hij nu is! 🡪 Het maakt niet uit hoeveel hij nu bevat hij zal 4x langer worden gemaakt!
  + Voor een examen .90 is streefdoel, al maak je hem 4x langer zit je nog maar aan .86
  + Dus S-B-formule laat toe de lengte van de test te relateren aan de betrouwbaarheid!
* Je kan hem ook korter maken!
  + Zeer hoge betrouwbaarheid maar je hebt geen tijd
    - K kan dus ook kleiner zijn dan 1
* Stel 🡪 Rel(X) = .60 men wenst .80 🡪 Wat is de factor van k
  + Dit komt vaker voor in de praktijk
  + Rel(X(k)) 🡪 = 80

### Correctie voor attenuatie

* 2 testen! Wiskunde en leesvaardigheid!
  + Je wilt de correlatie weten tussen de wiskundevaardigheid en de leesvaardigheid
    - Dus de correlaties tussen die hun true scores, maar deze hebben we niet!
  + Je gebruikt de normale correlatieformule 🡪 Correlatie van .5
    - Altijd een onderschatting door de ruis!
      * Het fenomeen dat je correlatie zwakker is dan hij zou moeten zijn (onderdrukt) met als oorzaak meetfouten! = Attenuatie
* Wel opletten met deze formule!
  + Als we de betrouwbaarheid Xa Xb onderschatten 🡪 Dan krijgen we een overcorrectie!
  + Hoe meer ruis, hoe zwakker de geobserveerde correlatie is!
  + Dus hoe zwakker de betrouwbaarheid, hoe groter het effect van de attenuatie

### (Optioneel) schatten van ware scores: 1ste benadering & 2de benadering

* Lezen

## Het schatten van de betrouwbaarheid

### Methode 1: 2 parallelle testen!

* De constructie is niet echt evident!
  + Dit is enorm moeilijk!
* In de feiten werkt dit niet, niemand doet dit nog, het is al zo veel werk om 1 op te stellen die betrouwbaar zijn laat staan meerdere die parallel zijn!
  + Bovendien als je één hebt dat betrouwbaar is en een parallelle zou je ze beter samenhangen zodat je een nog betrouwbaarder examen hebt

### Methode 2: test-hertest methode

* Idee is eenvoudig, maar je zou om goed te zijn moet je het geheugen wissen!
  + Voor een test is dit moeilijker dan voor een weegschaal!
* Bv. Weegschaal 🡪 1ste keer 70kg 2de keer 🡪 65 🡺 Niet betrouwbaar
  + Betrouwbaar = altijd hetzelfde!
  + Zelfde test, zelfde persoon, twee keer meten zelfde uitkomst, tenzij er sprake is van geheugeneffecten etc.

### Methode 3: ‘split-half’ betrouwbaarheid

* Dat is een optie
  + Zeker als je veel items hebt, kan je dit makkelijk doen
  + Maar een probleem is, je kan even en oneven nemen, de eerste 50 en laatste 50
  + De keuze is arbitrair!
    - Er zijn oneindig veel manieren om ze in 2 stukjes te snijden
    - Elke keuze leidt tot een ietwat andere correlatie en een andere betrouwbaarheid en het wordt afhankelijk van de arbitraire keuze!
* Stel je voor dat je het niet 1x doet, maar vele keren
  + Allemaal op andere random manieren!
  + En je berekent weer de betrouwbaarheid en je doet dit zoals het ware oneindig keer veel
    - 🡪 Het lijkt goed om het gemiddelde te nemen ervan….
      * Maar hoe doen we dit? Goeie paper

### Coëfficiënt Alpha

* = Split half vele keren herhalen (alle mogelijke manieren)
  + Deze formule bestond al, maar er werd nooit een link gelegd met de split-half!!
  + P = aantal items = lengte van je tests
  + Var(X) 🡪 Variantie van je testscore die te observeren valt
  + Var(Yj) 🡪 Variantie van het individuele item
  + Dit wordt een beetje herschaald en je krijg een getal tussen 0-1!!
    - Dit is een elegante en praktische manier om de betrouwbaarheid te meten met 1 test op 1 tijdstip!!
* Dit is de meest gebruikte maat!

### Praktische berekening Cronbachs alpha

* S = Covariantie matrix!
* O is alles wat onder de diagonaal staat maar niet wat er op staat!
  + Enkel ONDER!
* D is wat opde diagonaal staat 🡪 4 varianties
* S = volledige som = 2x O + D
* Dus ze kunnen de formule vereenvoudigen!

### Cronbachs alpha in R

* Fluitje van een cent!
  + P = aantal kolommen
  + De som nemen van al die elementen
  + Diag = diagonaal van s en die kan je ook sommeren!
* EZ

### Interne consistentie!

* Over de tijd heen is er een andere term op de proppen gekomen en hij heeft dit zelf een beetje in de hand gebracht
  + 🡪 Als alle items een beetje hetzelfde meten 🡪 Consistent
  + Hoe meer consistent, hoe betrouwbaarder de scores zullen zijn!
* Interne consistentie = meer dan betrouwbaarheid
  + Is er 1 dimensies, 2 dimensies; 1 of 2 factoren
* Alpha is geen goede maat van interne consistentie!
  + Mocht het zo zijn dat we meer dan 1 factor nodig hebben, dan is de cronbachs alpha helemaal niet van toepassing
  + Dit is enkel iets dat je kan gebruiken voor een zuiver één-dimensioneel construct
  + De betrouwbaarheid is wel groot, maar dat getal zegt niks, want het slaat op data die uiteindelijk niet kunnen worden gecapteerd door een 1-factormodel!

### Historiek van cronbachs alpha

* Nu heeft hij een slechte reputatie, wat wel jammer is want hij is een van de belangrijkste geweest in dit vakgebied
  + Hij zou zelf zeggen dat het eigenlijk een maat is die betrouwbaarheid te schatten mits je weet dat het een-dimensioneel is
  + Je moet ook kijken naar de context, …
* Zijn paper is een van de meest geciteerde papers!
  + Toch moeten we erop wijzen dat er al voorlopers waren!
  + Ze hadden een formule nummer 20 🡪 KR20
    - Niet voor continue maar voor binaire items, de continue versie zijn zeker niet door cronbach bedacht!
* Hij heeft het ingebed met de split-half!
* Het is vaak een onderschatting
  + Dus je berekent cronbach en je bekomt .6 uit, het zal zeker niet minder zijn!
    - Het gebeurt niet veel dat er een pessimistische maat is!
  + Onder sommige voorwaarden zitten ze er op (straks)

PM LES7

### Historiek van Cronbachs alpha

* Het meest gebruikte maat is zeker een understatement 🡪 99.9% gebruikt deze test!
  + Dat .1 procentje is wel aan het evolueren!

### Kritiek op cronbachs alpha en alternatieven

* Zijn Achillespees is dat het maar om een unidimensionaliteit van de test te bepalen!
  + De formule van cronbach zal je niet tegenhouden om de alfa te berekenen ook al heb je meerdere factoren!!
    - Maar dat is niet de betrouwbaarheid, enkel als je test uni dimensioneel is
    - Dit kan je wel nagaan! (Exploratieve factoranalyse)
* Het is dus enkel voor unidimensionele testen!
  + Het is dan ook geen maat van hoeveel factoren je nodig hebt!
  + Dus als je een slechte score krijgt hiervoor, wordt er geredeneerd dat je eventueel twee factoren nodig hebt
    - Daarvoor dient het niet!
    - Het is dus niet geschikt om te evalueren of het aantal factoren die opstelt oké is
      * 🡪 Het is dus een primitief is want je moet het al een uni-dimensionele test geven en enkel in deze setting als het aan de test voldoet, enkel dan kan je het gebruiken!
* De score die je krijgt Bv. .70 = geschatte betrouwbaarheid
  + De werkelijke betrouwbaarheid zal beetje groter zijn!
    - Doorgaans is deze te laag!!
* Er zijn verschillende alternatieven die deze vervangt!
  + Deze zijn typisch een beetje groter!
    - Van deze alternatieven geven ook een score, deze worden nu meer en meer gebruikt en ze zullen stijgen!
  + De scores liggen allemaal in de buurt
    - .7;.75; … Maar Cronbachs Alpha blijft het kleinste!!
      * Maar idealiter schakelen we allemaal over naar Omega! (Zal niet voor meteen zijn)

### Optioneel: berekening coëfficiënt omega in R

* Als je het ooit van je leven nodig zou hebben, kan je deze gebruiken die toont hoe het moet!
  + Vroeger was dit veel moeilijker maar nu is dat “water uit de kraan”

## Relatie met het 1-factor model

* Ze hebben dezelfde true-score en dezelfde error je ziet dat de “j”-index verdwenen is!
* De true score waar het we de hele tijd over hadden is eigenlijk de factor weliswaar een beetje omgeschaald de lamda maakt niet echt uit, dat is toch latent!
  + De true-score en de latente 1-factor, is dus exact hetzelfde ding
* Klassieke testtheorie is dus niets anders dan gewoon het 1-factormodel maar plots met een andere terminologie
  + Anders spreken we met True-score = T
    - Hier spreken we over latente variabele lamda 🡪 T = herschaalde versie van eta (Als lamda = 1 🡪 Hetzelfde)
      * Maar eta en T zijn hetzelfde!
      * Dat is wat we pogen te meten!
* We kunnen eventueel in het factormodel het een of ander loslaten, de errorterm kan anders zijn van item tot item en in het congeneriek model kunnen we terug de index toevoegen 🡪 We hebben dus iets meer vrijheid in het factormodel!
  + Dus we kunnen bepalen voor onze test of we een congeneriek model nodig hebben of hebben we de chance of een tau-equivalent model past of kunnen we naar een parallel model gaan!
    - We kunnen het toetsen (adhv 1-factormodel!)
* Stel voor 🡪 1-factormodel dat het congenerieke niet past bij de data!
  + De fit is niet goed, dit betekent dat een 1-factormodel voor die test in onze data niet past en dit impliceert dat heel dit hoofdstuk, klassieke testtheorie niet van toepassing is want deze theorie veronderstelt dat het 1-factormodel wel past bij je data en als dat niet het geval is dan zullen alle formules niet geldig!

### Implicaties congeneriek, tau-equivalent, parallel

* Als tau-equivalent niet gelijk is 🡪 Zal je cronbachs alfa te klein zijn!
* Dus je moet minstens een tau-equivalent hebben!
  + Dan zijn de omega en de cronbach goed voor elkaar!
* Spearman-Brown fomule is enkel toepasbaar voor parallelle testen!
* 1-factor model = breder kader om te testen wanneer welke theorieën/rekenregels van toepassing zijn of niet!
  + Ideaal is dat we een parallel model hebben en dan kunnen we alles doen!
  + Maar dit komt amper voor…
    - De praktische haalbaarheid slaat er op dat de “mooie” formules niet veel kunnen gebruikt worden
* Als je al een 1-factormodel aanvangen is het meestal congeneriek, tau-equivalent komt ook nog vaak voor!

## Conclusie

* Het heeft het voordeel van eenvoud, maar eigenlijk is het zijn taak om ook uit te leggen wat de hedendaagse voorbeelden zijn!
  + Dat is best wat geëvolueerd! (Namen zijn van belang!!)
    - Hij heeft een andere theorie bedacht
      * Generaliseerbaarheidstheorie
        + Hij zei dat het ook zeer context gevoelig was, afhankelijk daarvan is de betrouwbaarheid anders! (Dat is geen constante maar afhankelijk van vele andere factoren)
        + Zeer mooi kader! Maar het vereist een goed begrip van ANOVA etc. (nog niet voor meteen)
    - Ander kader 🡪 Latent State-Trait Theory
      * Is nog eleganter, de meeste psychometrici werken hier mee, een beetje te moeilijk, maar weet dat het bestaat!

## Appendix

* Geen examenstof! Bewijsjes 🡪 Geniet van de eenvoud

# factoranalyse met meerdere factoren

## Inleiding

* Slide

### De lokale onafhankelijkheidshypothese

* Het idee van heel de manier van denken met latente variabele
  + Het idee is dat deze een verklaring biedt, waarom wij patronen zien in de geobserveerde data
  + De scores die we meten en zien, daar zit een patroon in!
    - Bij de ene groot en bij de andere ook, vice versa ook geldig!
    - Latente variabelen proberen uit te leggen waar de correlatiestructuur vandaan komt, het is blijft een hypothese
      * 🡪 We kunnen er zeker naast zitten en het kan een andere reden zijn!
* Mocht je het effect van de variabele er helemaal uit halen 🡪 Het enige dat nog overblijft is ruis
  + De Y’s worden beïnvloed door de latente variabelen, als de invloed wegneemt is er ruis en deze mag niet meer gecorreleerd zijn!
    - Want als er daar correlatie is, dan moet er nog iets zijn!
* Je moet voor ogen houden dat het een hypothese blijft, maar deze kunnen wel empirisch geverifieerd worden, we kunnen toetsen of deze klopt met de data die we verzamelen of niet!

## Voorbeelden van factormodellen

### 1-factor congeneriek model

* H4 🡪 Het congeneriek 1-factormodel
  + Er is geen enkele correlatie tussen de errorvarianties!
    - = CONGENERIEK
* Als dit model klopt, de enige reden waarom we correlatie zien tussen de Y’s dan is de enige reden waarom dit is 🡪 Die ene factor!
  + Als we rekening houden met de gemeenschappelijke oorzaak dan schiet er niets meer van structuur over! (Lokale onafhankelijkheidshypothese
* Deze matrix van 1 kolom = Lamda-matrix
  + Deze is een matrix die het model weergeeft!
    - 6 rijen = de geobserveerde variabelen
    - Kolommen = factoren (1 factor)
  + Matrix van de factorladingen!
    - Deze zes getallen zetten we op een rijtje en dat zijn 6 factorladingen, er staan geen nulletjes
    - Stel dat je enkel deze matrix ziet dan kun je veronderstellen dat het model er uitziet als een 1-factor congeneriek model
    - = Korte manier om het model neer te schrijven

### 1-factor niet-congeneriek model

* Er is correlatie tussen de errortermen
  + Het is vaak problematisch 🡪 Het situeert zich hier tussen 2 en 4
  + Na de factor schiet er nog resterende correlaties over tussen twee reeds uit-gepersonaliseerde items, er zit nog extra structuur in
    - Dus er moet nog een andere factor zijn?
      * Die ook een effect heeft op beide items
      * Want de invloed van eta is er al uitgehaald!
  + Dit hebben we niet graag…
    - Als je dit merkt in de data, dan gaan we vaak kijken of we kunnen te weten komen waarom dit is, is er iets met deze items, waarom hebben ze iets gemeenschappelijk dat niets te maken heeft met het latent construct
    - Vaak weten we niet waarom 🡪 Weggooien en vervangen!
      * Soms gaan we op zoek naar een ander model!

### Factormodel met twee gecorreleerde factoren!

* De hypothese is dat de eerste drie enkel iets te doen hebben met de eerste factor en de laatste 3 enkel met de 3de factor!

### We kunnen dit toetsen of dit letterlijk zo is

* + = Confirmatorische factoranalyse
    - 🡺 Het klopt of niet, niet 🡪 Herzien van model
* Factorieel eenduidig 🡪 ER is een nette scheiding tussen de factoren!
* Lamda-matrix
  + Heeft 2 kolommen 🡪 2 factoren
  + En enkel waar lamda staat, is er ook een pijltje
  + De 0 = geen pijl van de tweede factor naar het eerste item
* De dubbele pijl is ook ee assumptie
  + Je gaat er van uit dat deze ook gecorreleerd zijn je latente variabelen
    - Bij de ene is dit vrij logisch, maar het kan ook zijn dat het niet zo is!
    - Maar het kan ook zijn dat de pijl weg is omdat men denkt dat ze niet correleren, je kan dit testen!

### Factormodel met twee gecorrleerde factoren + kruisladingen

* Y4 wordt beïnvloed door 2 factoren!
  + Ze kruisen elkaar 🡺 Kruisladingen
    - Dit hebben we ook niet graag
* = Factorieel complex 🡺 Items eten van twee walletjes!
  + Het is een indicator van de ene variabele en ook van de andere
  + Dus idealiter geen kruisladingen (vorig model)
* Lamda-matrix
  + Hier zie je dat er rijen zijn waar meerdere lamdafactoren aanwezig zijn!

### Mimic model

* Slechte naam
* Je hebt X’n erbij, je hebt X 1 en X2 🡪 Geslacht en leeftijd bijvoorbeeld
  + Het zijn observeerbare variabelen en de individuen die we gemeten hebben, daarnaast hebben we geslacht en leeftijd
  + Het meetmodel is soms afhankelijk (of sommige aspecten van het meetmodel) van de wat we covariaten noemen
    - Deze variëren mee met de twee Y’s
      * Je hebt degene die rechtstreeks effect hebben op de indicator!,
        + Ondanks gelijk score van latente waarde toch een grotere score op Y1
      * Ander type is X2 🡪 Effect op de latente variabele
        + Deze verklaart waarom we verschillen zien op de latente variabelen
        + Het heeft misschien te maken met een verklarende predictor

Bv. Een hormoon, geslacht, …

* + - * Er is maar 1 getekend maar er kunnen er velen zijn!
        + Dit is uiterst interessant om te proberen achterhalen waarom mensen differentiëren van elkaar!
* Gegeven een grote steekproef kunnen we alle effecten testen!
  + Het is dus mogelijk om deze na te gaan!
* X1 🡪 Deze hebben we niet graag, hij interfereert met je meetmodel!
  + Als de latente stijgt dan stijgt Y1 ook en hoe snel dat hangt af van de lamda
  + Maar wat gebeurt nu, hij stijgt hier bv. Sneller als iemand een man is dan wel een vrouw
  + Er is dus iets anders die de relatie beïnvloedt/verstoort
    - Idem voor examenvragen!
      * Stel dat deze beter zou beantwoord worden door mannen of vrouwen of vice versa!
  + Er zijn veel testen die afhankelijk zijn van de cultuur, sommige items zijn blijkbaar gevoeliger aan externe invloeden (Item-bias)
* X2 🡪 Deze vinden we uiterst boeiend en interessant, om te weten te komen welke variabele die we kunnen observeren verklaren waarom mensen andere scores hebben op de latente variabele die ons interesseert!
* Combinatie van twee types van modellen 🡪 zuiver meetmodel en een structureel model
  + Het structureel gedeelde zijn de zaken die neits te maken hebben van een meetmodel 🡪
  + 🡺 Het is eigenlijk een Structure Equasion Model (SEM)

### Multiple group factormodel

* Meerdere groepen testen om te zien of ze hetzelfde gedragen binnen twee groepen!
  + Het zou kunnen dat de relatie tussen de latente variabelen en de indicatoren dat die anders is bij groep 1 en groep 2
    - Dit is noodzakelijk om te weten want het kan bv. Een klinische groep zijn vs. Normale groep!
* Je moet ervoor zorgen dat wat je meet (latente variabele) hetzelfde is bij beide groepen!
  + Als het meetmodel hetzelfde is in groep 1 en groep 2 dan is het invariant en dit is een goeie zaak want we willen hetzelfde meten over de groepen heen en dan kunnen we ze vergelijken!
    - Als dit niet zo is, zouden we appels en peren vergelijken… We komen iets uit maar het is niet zo nuttig…
* Dingen worden vergeleken met elkaar maar dit is enkel nuttig als het meetinstrument zich hetzelfde gedraag in de ene groep dan wel de andere groep
  + Een minimumvoorwaarde is dat de factorladingen gelijk zijn in zowel groep 1 als groep 2

### Bifactor model

* Elk item wordt beïnvloed door de algemeen en 1 en slechts 1 specifieke!
  + Op elk items komen twee pijltjes toe!
  + Een ander iets dat van belang is, is dat er geen dubbele pijlen zijn tussen de latente variabelen!
    - De correlatie tussen de latente variabelen zijn nul dus de factoren zijn orthogonaal! Ze zijn niet gecorreleerd!!
    - Typisch in een standaard bi-factor model
* Wanneer komt dit voor?
  + Dit is wanneer er zoiets is dat geld voor een algemeen iets en dan nog iets dat is voor specifieke zaken
    - Bv. G-factor is algemeen iets en dan heb je ook nog verbale intelligentie en spatiale intelligentie en die uiten dan invloed op specifieke items!
* Andere setting 🡪 Er is maar 1-factor
  + De specifieke hebben niets inhoudelijk te zeggen 🡪 Bv. De eerste 3 items zijn afgenomen via een vragenlijst en de laatste 3 zijn via de pc gedaan
    - Heeft te maken met de methode!
      * Als je dezelfde methode gebruikt over metingen heen krijg je onherroepelijk samenhang tussen de meting = Common factor effect
    - Dit zorgt er voor dat alles via vragenlijsten op elkaar lijkt en andere methode deze lijken dan ook op elkaar (Bv. Pc-experiment)

### Tweede-orde factormodel

* 1ste orde latente factoren zijn meteen gelinkt aan de indicatoren!
* 2de orde latente factoren worden gemeten door andere latente factoren
  + Dit zie je wel nog vaker!

### Structureel vergelijkingsmodel

* Alle variabelen zijn latent, dit hoeft niet, dit kan vervangen worden door observeerbare
* Je hebt 4 meetmodellen, maar je hebt ook structurele delen, dat zijn de enkele pijlen
  + En vaak zit de interesse in die pijlen!
    - Wat is het effect van het ene op het andere?
    - Hoe sterk is dit verband? Misschien is dit op zijn beurt dat het voorspelt wordt door twee andere latente variabelen!
* 12 indicatoren
  + Een SEM-analyse wil nagaan of deze hypothese correct is of niet
    - Dit model zegt dat eta2 indirect een effect heeft op eta 4
  + We denken dat de relaties tussen de constructen er zo uit zien!

### Exploratief factormodel

* Je weet nog niet hoeveel andere latente factoren er zullen zijn!
  + Je weet ook niet welke indicatoren samenhangen met welke factor!
    - Alle pijltjes staan er nog, elke factor is nog gerelateerd aan alle items omdat je nog niet weet hoe de structuur er zal uit zien
    - Het is de bedoeling om te zien hoeveel factoren zijn er nodig!
* Je ziet 2 kolommen maar dit kunnen er nog twee of drie worden of zelfs meer!

## Exploratief vs Confirmatorisch

* Exploratief komt meestal eerst!
  + EFA 🡪 Is iets wat men toepast indien men wilt weten hoeveel factoren men hier kan bekomen!
  + Er moet vaak aan gesleuteld worden vooraleer het doet wat men moet doen!
* De twee methodes hebben allebei hun rol, maar idealiter eindigen we met confirmatorisch 🡪 We hebben dan al een beetje zicht op de hoeveelheid van factoren en hoe zijn de indicatoren gelinkt aan de factoren!

## Historiek van factoranalyse

* Vroeger was het 1-factor, meer dan 1 dat ging gewoon niet, we hebben moeten wachten tot het interbellum en vooral Thurnstone
  + Hij was degene die aangegeven heeft hoe je het model met meerdere factoren kon doen!
* Hoe komen we van die data tot die factoren, dat waren allemaal trucjes en regeltjes
  + Het heeft een paar decennia nodig gehad en een paar goeie statistici die dat een grondige statistische onderbouw hebben gegeven
  + Ervoor was het iets als voodoo een soort van black-box
  + Hoe mooi en theoretisch het model werd geschreven, het bleef/bleek onhaalbaar!
    - In de jaren 40 was het onmogelijk uit te rekenen want het moest manueel zijn!
* We moeten wachten tot de computers 🡪 Jaren 50-60 = Mainframes!!
  + Toen zijn er duizenden exploratieve factoranalyses uitgevoerd, dat was plots iets heel mysterieus!
    - Het was fenomenaal, je steekt data in de machine en factoren kwamen eruit, hoeveel er uit kwamen was een mysterie!!
      * Draaien 🡪 Padum 2 factoren
      * Lang draaien padum 4 factoren!
  + Iedereen deed het in de tijd, want het kon!
    - Dit heeft geleid tot enkele theorieën die nog altijd bestaan die wel niet altijd zo solide waren!
    - We zitten dus met een erfenis van die periode, theorieën die toen ontstaan zijn op basis van factoranalyse die gedraaid werden (letterlijk)
* Deze hebben wel de weg geëffend naar confirmatorische factoranalyse, maar we hebben moeten wachten tot Karl
  + Hij heeft een methode bedacht in 1967, om confirmatorische factoranalyse te doen een soortgelijke er op!
    - Dus van 1904 van Spearman tot 1967 waar er nog een decennium nodig was om te verspreiden!! 🡺 Vrij tot zeer laat!
* 2 jaar later ’69 dan heeft hij het effectief uitgebracht!!
* LISREL daar moest je in de tijd voor betalen!
  + Je kon met de gratis versie SPSS geen CFA doen!
  + Want merendeel van de psychologen hadden CFA nodig en ze hadden de software niet!
    - Dus deden ze dan maar een exploratieve factoranalyse en ze geraakten er mee weg ook want iedereen deed het!
    - Slechts enkele konden met dat programma werken!
  + SPSS is niet bepaald geïnteresseerd in wetenschap, dus je kunt het bedrijf het niet kwalijk nemen… Ze dachten vooral aan winst!
* Vanaf heden kunnen we gratis dergelijke software krijgen!
  + Lavaan = voor latente variabele analyse!
  + Vanaf 2000 is de aandacht ook verschoven!!

## exploratieve factoranalyse: oude stijl

### STappen om een EFA uit te voeren

* Zo gaat het er aan toe in de ideale wereld!
  + 1. Wat gaat er allemaal in de machine, de geobserveerde waarden
    - Allemaal zaken waarvan je vermoed dat ze de latente variabele waar je op zoek naar bent zullen meten
  + 2. Een zo groot mogelijke steekproef nemen
  + 3. Ruwe data 🡪 Deze kunnen we in de machine steken, maar we kunnen het samenvatten in een correlatiematrix!
    - Dit gaat in de machine
  + 4. “Hoe berekenen we de factoren”
  + 5. Dit is de achillespees 🡪 Hoeveel factoren
  + 6. Als er meer dan 1 is, bijkomend probleem!
    - We gaan moeten roteren! (Zie Straks)
  + 7. Scores bereken op basis van de latente variabelen
  + 8. Idealiter wil je zien of dit valt te herhalen met een andere steekproef!
    - Misschien zijn ze zo specifiek voor deze steekproef dat ze niet stabiel is, andere steekproef misschien andere resultaten

### Optioneel: factor extractie

Zie slide

### Voorbeeld: 8 subschalen omtrent persoonlijkheid

* Deze items meten allemaal dingen in de sfeer van persoonlijkheidspsychologie
  + Uiteindelijk zal je terecht komen bij 2 factoren (maar dit weet je nog niet)
    - Je zal proberen 2 dingen te meten
* 8-tal scores
  + Je hebt gewoon 8 items, voorlopig moet je de labels negeren
    - Je hebt correlaties tussen de items en er zit precies structuur
    - En je probeert te verklaren waarom je deze observeert!

### Invoeren correlatiematrix

* Slide

### Model 1: factor-factor extractie: maximum likelihood

* Je begint met 1 factor! 🡪 Correlatiematrix
* Laatste lijn 🡪 P-value
  + De e-notatie is met een - 🡺 Zeeeer klein getal!!
* 45% van de variantie wordt verklaard door de ene factor
  + Wat best al veel is
    - Maar door de P-waarde zie je dat niet voldoende is!

#### Interpretatie

* Slides (voor meer informatie) 🡪 Zie je hoe je aan de variantie komt!
* De modeltoets is niet goed doordat je een lage p-waarde hebt

### model 2a: 2 factoren, zonder rotatie

* P-waarde 🡪 Het model past goed bij de data, onze nulhypothese ziet er goed uit, dus we hoeven het niet te verwerpen
  + Het lijkt dus oké
* Kijken naar onze proporties
  + Het eerste is een beetje verandert
    - .484
  + Maar er is een 2de bijgekomen 🡪 .20
    - Maar samen is het in dit geval bijna .70 % (68% van de variantie, idealiter moet dit naar 100 maar dat zal het nooit gebeuren, zeker niet in de psychologie, want dan zou je meten zonder meetfouten!)
      * Maar dit is al vrij hoog!

#### INterpretatie

* Slide voor meer info

### Model 2B: 2 factoren varimax rotatie

* Het aantal factoren staat min of meer vast!
  + Hoe minder factoren hoe liever, je kan eventueel 3 doen, maar je hebt niet de behoefte om veel factoren te bekomen
    - Idealiter een klein aantal
      * Maar te klein is niet goed, want dan capteer je niet de belangrijkste factoren uit je data!
* De ene factor daar? (Zonder rotatie slide 28)
  + Wat meet die eigenlijk?
  + Manier om dit te doen dit kan je zien de eerste 4 🡪 .8 en dan heb je 4 negatieve
  + De tweede factor is al wat moeilijker…
  + Maar het is typisch de intiiële factoranalyse is moeilijk te interpreteren
    - Waarom? Je moet nog roteren!!
* Model 2B 🡪 Resultaat en dan uitleg
  + Deze oplossing, nog steeds 2 factoren behalve op het einde roteer maar en gebruik deze techniek (Varimax)
    - Wat is het effect van de rotatie?
      * Het enige wat anders is, is het patroon van de getallen van de onder je factoren is een beetje anders en je proportie van je factoren ook!
        + Het is meer even verdeeld!
        + Het is elke ongeveer 32-36%

Het totaal is hetzelfde

De toets is hetzelfde

De unieke varianties zijn hetzelfde maar de factorladingen zijn anders!

* + - * Hopelijk, die rotaties hebben we gedaan met 1 bedoeling, we willen beter in staat zijn te begrijpen waar ze voor staan
  + Met een beetje goeie wil de eerste vier staan goed bij N en dan die andere beter bij E
    - Stel dat ze allemaal 0 waren eronder 🡪 De eerste vier enkel bij N
  + 2de kolom de laatste 4 horen samen en definiëren in zekere zin wat ze meten maar er zit nog wat ruis op bovenaan!
    - DE eerste 4 meten NEURO HOERA!
    - De 2de vooral deze EXTRAV HOERA!
      * Ik heb twee factoren gevonden, een patroon in de items!

### Orthogonale rotatie (elke as roteert evenveel)

* Zwarte 🡪 Model 2A
* Groene 🡪 Model 2B
* Dus je draait rond z’n as en je kan helemaal draaien en op een bepaald moment zal ze stoppen
  + Wat ze graag hebben is dat het groene clusters zo dicht mogelijk bij de stippellijn liggen
  + Je moet even ver draaien met de ene as als de andere as
    - Ze liggen beiden iets dichter bij de stippellijnen!

#### Interpretatie

* Slide

### Model 2C: 2 factoren, oblique rotatie

* De getallen zijn zo klein geworden, kleiner dan .1 in absolute worden, dus ze worden niet meer afgedrukt (ze zijn niet exact nul) maar < .1
  + Daarom kiest het programma om ze niet eens te printen!
    - Dus valt het mooi uiteen in twee blokken!
    - Ze zijn ongeveer nog altijd hetzelfde getal en de fitmaat is ook nog altijd hetzelfde
    - Het enige extra dat je hier krijgt is de correlatie tussen de factoren die hier maar liefst .434 is wel een grote correlatie is!
* Realistischer maar wel moeilijker uit te rekenen

#### Interpretatie

* Slide

### Oblique rotatie (de ene as roteert meer dan de andere)

* Je hebt veel meer vrijheid!!

PM LES 8

### Bepalen van het aantal factoren

* Goodness-of-fit is eigenlijk de P-waarde van daarnet
  + Deze zijn niet zaligmakend
  + Dus het is goed dat we weten wat de andere alternatieven zijn (deze hebben wel andere nadelen!)
* Soms zullen de verschillende manieren niet altijd sporen met elkaar!
  + De ene zegt 1, 2, 3 of 4
  + Op het einde van de rit, welke je ook hebt gehanteerd hebt, je gaat meerdere hanteren!
  + Er is geen enkele perfect (ook niet de modeltoets)
    - Een groot nadeel aan statistische toetsen met een P-waarde
      * 🡪 Sensitief aan de steekproefgrootte
      * Als je grootte stijgt wordt je P-waarde sowieso klein
    - Hoe meer info, hoe groter steekproef, hoe sensitiever/gevoeliger je proef wordt op kleine triviale afwijkingen
* Op het einde van de rit zal de oplossing die ons het nuttigst lijkt, conceptueel gezien
  + Deze gaan we kiezen!!
  + We gaan straks criteria zien!
    - Geen rekening met de inhoud
* Er was een enorm sterk geloof in het kunnen van de software
  + Het gevolg ervan, tot op vandaag zitten we opgescheept met theorieën die steunen op bevindingen!
    - Het aantal factoren zijn op een bijzonder naïeve bepaald, maar als deze aanslaan en wijd verspreidt gaat er niemand terug om die analyse opnieuw te doen
      * Het wordt op een bepaald moment de waarheid
  + Bv. Big Five
    - 🡪 Je kijkt naar de criteria die ze gebruikt hebben
    - De manier waarop ze er gekomen zijn 🡺 Minstens disputabel!
* Kaiser criterium (BF houdt er rekening mee, dit is slecht)
  + Dit krijgen we niet uitgeroeid
  + We moeten iets vertellen over eigenwaarden

### Eigenwaarden (optioneel)

* Dit zijn cijfers die te maken hebben met matrixalgebra
  + Een symmetrische 8x8 matrix
  + Deze matrix, hierin kan je een eigendecompositie uitvoeren
    - Deze wordt opgesplitst in 3 stukjes
    - Het product ervan is weer de oorspronkelijke
    - 🡪 Herschreven tot eenvoudige matrices!
* De eigenwaardecompositie 🡪 3 stukjes
  + 1ste en 3de zijn gelijk, maar het is de middelste die ons interesseert 🡪 Diagonaal matrix (allemaal nullen behalve de diagonaalà
    - Deze diagonaal = eigenwaarde
    - Ze worden meestal zo georganiseerd dat ze gerangschikt worden van groot naar klein
* In R kan je dat doen met commando “eigen”
  + Slide 35 🡪 Eigenwaardes van je matrix!
  + We delen ook nog eens de waarde door het aantal!
    - Eerste = .53 🡪 53% wordt gecapteerd door de eerste component
      * 1 loopt al weg met 53%
    - De tweede nog eens goed voor een extra 23%
    - 3de is maar 6 procent…
  + De eerste twee 🡪 Leeuwendeel!!

### Kaiser criterium

* Stelt hoeveel factoren we moeten nemen, enkel die factoren die een eigenwaarde hebben groter dan 1!
  + (De rode lijn)
* Dit is het criterium die automatisch wordt toegepast door SPSS!
  + Je zegt niet op voorhand hoeveel factoren er worden gegenereerd! (Dus niet over het aantal)
  + Hij zal doodeenvoudig een eigenwaarde compositie doen en kijken hoeveel er een Y-waarde hoger heeft dan 1
    - 🡪 2 factoren!
  + Soms hebben ze het bij het juiste eind zoals op slide 37, maar het ziet er niet altijd zo netjes uit!
* Probleempje met deze criterium, waar komt de 1 vandaan
  + Stel dat ze niets met elkaar te maken hebt!
    - 1 op de diagonaal en eronder minimale resultaten rond de 0 (door de ruis)
      * Quasi geen correlatie!
        + Als je dit in een decompositie stopt krijg je overal 1 1 1 1 1
    - Als er structuur zit 🡪 Kantelen en je krijgt meer informatie in de eerste componenten en weinig in de laatste!
  + Als je en oneindig grote steekproef hebt, en je hebt een matrix niet gecorreleerd is, dan is dit exact 1
    - In deze setting 🡪 Is het kaiser criterium nog zo gek niet
      * De grootte van de steekproef is ook de achillespees
  + Rond de 10000 zit je al bij kleiner
    - Dir werkt niet meer dit idee!
      * Gemiddeld genomen is dit niet meer 1 maar er zit gemiddeld genomen een stijging in!
* Het is dus onbetrouwbaar!!
  + Te veel factoren 🡪 Over factoring; Te weinig 🡪 Underfactoring

### SCree test

* Deze zegt zoek de knik in de curve
  + Waar zit deze? Zoek de elleboog
    - Het gaat naar beneden en dan is het plat!
* Deze suggereert 3 factoren
  + Niet twee maar 3, want daar pas vlakt het af!!
    - Als mensen dit hanteren, zouden ze suggereren van kijk 3 factoren is wat je eruit afleest
    - Hier is het nog relatief duidelijk!
      * Maar soms zit er geen in, wat dan?
      * Het is dus redelijk subjectief en open voor discussie!!

### Parallel analysis 95%

* De moderne versie van het Kaiser criterium, deze houdt rekening met de steekproefgrootte
  + Lauter op basis van kans, hoe zou het kaiser criterium zijn die niet correleren met elkaar
    - Je doet dit vele malen en je houdt wel rekening met de grootte van de steekproef
    - Deze houdt er rekening mee!
  + Vroeger kon dit niet, duurde te lang!
* Eens je de groene lijn hebt is de criterium hetzelfde als bij het Kaisercriterium!
* Deze methode zou ook eindigen met de conclusie dat er tee factoren zijn!

### Rotatie

* Varimax en oblique rotatie 🡪 Groot probleem geweest
* Belangrijk 🡪 Rotatie heeft niets te maken met het aantal factoren, je moet deze eerst bepalen en daarna pas roteren we!
  + Deze verandert niets aan de kwaliteit van het aantal factoren
  + Deze is enkel bedoeld voor de interpretatie!
* Als we ze op 1 van de assen willen leggen dan noemen we dit simple structure
  + Zorgt ervoor dat je een makkelijke factoroplossing hebt!
  + Als er een item is dat nergens dicht bij een as ligt 🡪 Het item hoort nergens thuis omdat het op alle dimensies een lage factorlading heeft
    - Wat is laag?
      * 🡪 Meestal groter dan .50 soms .40
      * Er is een tendens om dat nog op te krikken!
  + Items die te laag laden 🡪 Verwijdert en vervangen door andere items die beter interpreteerbaar zijn!
* Vroeger orthogonaal 🡪 EZ, zat in de software
  + Oblique is moeilijker, maar we opteren hier voor!
  + Het is bijna altijd zo dat de factoren conceptueel gecorreleerd zijn!

### Orthogonale rotatie

* Meest gebruikt!
  + Er zijn bepaalde schalen overboord gegooit
  + Terwijl als men ze oblique had geroteerd zouden we ze wel kunnen gehouden hebben!
* In SPSS is het de eerste methode, bijzonder jammer

### Oblique rotatie

* slide

### Interpretatie van de factoroplossing

* Op het einde van de rit hebben we onze factoren aantal, wat hebben we, is dat nu zinvol of niet?
* De ene factor welke heeft een hooglading
  + Je kijkt op basis van de inhoud van de items en kijkt wat ze meten, Bv. Neuroticisme, Extraversie, …

## Exploratieve factoranalyse: via CFA

* In plaats van iedere keer totaal exploratief te werk gaan en dat je niet weet wat te verwachten
* Nee eigenlijk gewoon op verwachting testen!! (confirmatorisch, adhv CFA)
  + Het is iets dat langzaam schuift, maar binnen een jaar of 10-20 zou dit wel meer gedaan worden!

### Uitvoering exploratieve factoranalyse via CFA (optioneel)

Slide

### Echelon patroon (optioneel

Slide

### Exploratieve CFA in R (optioneel)

* E4 doen we maal 0 dat is de rechterbovenhoek
  + Dit is confirmatorisch testen
  + Je krijgt heel veel info!!
  + Slide
* “Het kan”

### Rotatie in R (optioneel)

* “Dit kan ook”

## Factoranalyse van de SWF schaal

* Steekproefgrootte kan geen probleem zijn voor je P-waarde
* Het is al enorm veel gestegen maar net niet boven de .05
* Sat1-2-3 (Na protax) (inhoudelijk bekeken)
  + Deze hebben de hoogste waarden, ze zijn wel goed
  + Waar staan die en waarom zijn die anders? (Blijkt dat er een inhoudelijk reden is!)
* 2de factor is relatief laag, dicht bij nul (in vergelijking met de rest)
  + opletten met laag en hoog 🡪 Afstand ten opzichte van nul!
  + Sat 4 is goed, maar het mag wat hoger!
* Wat je graag had gezien is dat ze netjes uiteen vallen!
  + De eerste 3 enkel bij de ene factoren en De laatste bij de ander factor
  + Bij de laatste 2 lijkt het alsof ze nergens thuishoren?
* 3 factoren zou de logische volgende stap zijn, maar je hebt te weinig items 🡪 Het zou een overkill zijn
* slide 52
  + Slechte geval is er geen variantie 🡪 0
    - Wat je niet kan hebben in principe is negatieve variantie
    - Maar toch heb je hier negatief getal 🡪 De software laat dit toe
    - Dit is om aan te geven van kijk als je er iets negatief krijgt, dan is er eits aan de hand
  + 5 items worden in 2 factormodel gestoken
    - Ze mogen kiezen tot welke factor ze behoren, het wringt aan alle kanten en dat gaat gewoo n niet
      * Je ziet het aan de toets, maar dit ‘negatieve variantie’ is een echte killer
* De 2 factor deugd niet en de 1factor ook niet
  + 🡪 De schaal lijkt niet de deugen…
* Slide 53
  + Echt confirmatorisch
    - Je gaat ervan uit 2 zaken, postuleer 2 factoren
      * Als er al zo zijn, moet het wel zo zijn dat eerste factor samenhangt met de eerste 3 en tweede met de eerste 2
        + Komt omdat je ze goed gelezen hebt en het viel op dat de eerste 3 naar de huidige situatie pijlen (Present) en de laatste 2 die pijlen naar een beetje het verleden (Past)
    - Je hebt nog steeds je negatieve variantie!
      * Er is iets dat wringt
      * 🡪 Je bent genoodzaakt te concluderen dat deze schaal niet deugt…

### Modificaties indices

* Je hoeft niet te weten wat ze zijn en hoe ze berekent worden
  + Heel kort 🡪 Het probeert vanalles aan te passen in het model om te zien of er iets is dat we kunnen toevoegen dat de fit heel sterk beter wordt
    - Dit kan informatief zijn!!
      * Daar is het probleem en dit kan beter worden
* Hier zie je een top 3 die mogelijkheden zijn om het model beter te krikken!
  + Sat4 komt er elke keer in voor
    - Hier schort iets mee, dit item past niet in het rijtje!
* Je kunt/moet nagaan of de schaal deugt of niet, je moet het nagaan want als het niet zo is
  + Dan is er een fout gemaakt!
  + Want als je er beslissingen mee maakt dan is dat niet verantwoord!

## Confirmatorische factoranalyse

* We willen bevestigen of het zo is?
  + We moeten op voorhand iets zeggen over de factoren
    - Ze zijn meestal gecorreleerd, maar er zijn soms settingen waar we eisen dat ze expliciet niet gecorreleerd mogen zijn!
* Bij grotere steekproeven is de modeltoets minder interessant
  + Hoe groter de steekproef hoe kleiner je p-waarde
* CFI is een getal tussen 0-1
  + Hoe hoger hoe beter!
    - Modellen die bij perfect passen krijgen ze een waarde van rond de 1
      * Hoe slechter hoe kleiner…
    - Als het kleiner is dan .95 is het niet goed genoeg (deze wordt daar vaak opgelet!)
      * Er zit snel wat subjectiviteit op
* RMSEA
  + Als het model perfect past 🡪 0
  + Onder .05 zijn de tevreden!
* Er is een grijze ruimte soms
  + Hier is er wel wat ruimte voor discussie, idealiter willen we een heldere binaire uitspraak maken, maar in de praktijk zitten we vaak in de grens 🡪 Het vergt ervaring, context, input van meerdere mensen, … Om het te kunnen inschatten
* Als het model niet goed past hebben we een hele lading middelen die kan aantonen waar het aan ligt
  + Hieruit kunnen we veel leren$

### Voorbeeld 1: Neuroticisme & Extraversie

* Je ziet dat N wordt gemeten door de vier items van N1-4
  + De andere hoorden enkel bij E
    - 🡺 De hypothese
    - Je gaat er ook van uit dat ze (de twee factoren) gecorreleerd zijn!
* De CFA = commando
  + Correlatiematrix 🡪 antwoord is ja!
    - Dit is hier ingevoerd, maar dit moet een covariantie zijn!
  + Ik stop er een correlatiematrix in maar strikt genomen mag dat niet
    - Het zou moeten een covariantiematrix zijn
    - Geen informatie over de variantie!
    - In principe moet brown een covariantiematrix zijn!
  + De sample size = 250
* Het start met een model toets die nagaat of het past
  + Je kijkt naar de P-waarde! Woehoew het past bij de data
    - .824 (Zo is dit met schoolvoorbeeldjes)
    - Dit is een beetje te mooi om waar te zijn
* Eronder staat een andere modeltoets!
  + Die p-waarde is super klein
    - We hebben een ander model gefit dan dat wij hadden aangegeven!
      * Zogezegd achter de schermen heeft het ongevraagd een ander model proberen te fitten
    - Dit model postuleert dat ze allemaal onafhankelijk zijn van elkaar!
    - Hier is dit zeker niet zo 🡪 Barslechte fit!
  + Hoeveel beter doe ik het nu?
    - Hoe groter de vrijheidsgraden, hoe minder vrije parameters!
* Deze twee modellen worden vergeleken en ze herschalen het getal tussen 0-1
  + CFI 🡪 Comparative Fit Index
    - Hij vergelijkt jouw model met het meest slechte denkbare model
      * Als het niet veel beter is, dan heb je een slecht model.
      * Deze index resulteert in een getal 0-1
      * Zeer rooskleurig!
  + TLI: broer ervan, maakt niet uit welke je rapporteert!
* RMSEA 🡪 Root Mean Square Error App…
  + Het is zo geformuleerd, hoe kleiner hoe beter, hier perfect!
  + Je krijgt er ook een confidentie interval voor!
    - P-value die er bij staat moet ook groot zijn
* Standardizes root mean square residual SRMR
  + Dit model genereert een modelgeïmpliceerde variantie/covariantiematrix
  + Dit voorspelt hoe die er zou moeten uitzien als het model past (de sigma)
  + Ernaast heb je de data gebaseerde/geobserveerde data covariantie/variantiematrix
    - Je neemt het verschil en je hebt positeve residuen en negatieve residuen, kwadrateert ze je gaat ze optellen en daar een gemiddelde van nemen en daar nog eens een wortel van nemen!
    - 🡪 SRMR
      * Als ze identiek zijn is dat 0
      * Als ze niet identiek zijn is er wat afstand tussen, als het afwijkt wordt de afstand groter en groter!
* Het idee van al die fit indexen
  + Er is niet 1 die je er kan uithalen
  + Je mag niet enkel diegene rapporteren!
    - Je moet in de praktijk allemaal toepassen!
    - Ze moeten tezelfdertijd worden gerapporteerd
      * = Minimumvoorwaarde!
* Hoogste ruis zit bij item E4 en daardoor is de factorlading ook het laagste!
  + Het enige waarom we kijken naar de varianties is om zeker te zijn dat er niets negatief is!
  + Niet meteen concluderen dat dit item altijd slecht is!
    - Het is in deze steekproef!!
    - De conclusies die je trekt slaat op deze steekproef!!
      * Hoe kleiner, hoe onzekerder!

### Interpretatie

Slides

### DE covariantiematrix zoals voorspeld door het model

Slide

### optioneel: residuen in correlatiematrix

Slide

### Optioneel: modeltoets

Slide

### Voorbeeld 2: DE Holzinger & Swineford data

* Grafische voorstelling zie slide

#### CFA in R

* P-value 🡪 modeltoets is .000 🡺 kleiner dan .05
  + Ligt niet aan sample size!
* Mocht dit het enige criterium zijn de CFI 🡪 Zou je geneigd zijn te aanvaarden dat het toch past
* De andere 🡪 RMSEA
  + Minder rooskleurig 🡪 Vrij hoog getal!!
    - Net boven de nul, maar hier hebben we .09!
      * Flink stuk boven de grens
* SRMR 🡪 Weer een stuk boven de grenswaarde
  + Een stuk beter, maar nog altijd niet fantastisch
    - Dit is vaak een matige fit, het is een slechte fit maar we willen hem niet weggooien!
* Dit model past niet voldoende bij de data, maar wat blijkt!
  + Van de 100 modellen, die je zult tegenkomen
* Deze modellen zijn twijfelachtig van kwaliteit!
  + We moeten ervan af
  + Eigenlijk moeten we het bord uitwissen en opnieuw starten
    - We moeten ze tegen het licht houden, wat is nu de empirische evidentie dat deze modellen echt goed zijn!
* Ze blijven bestaan door inertie!
  + Vernieuwing is moeilijk en lastig!

### Interpretatie

* Slide

### FActorscores CFA model

* We zien de data van latente variabelen die we normaal niet zien!
* Enkel maar van toepassing als het model goed-zeer goed past bij de data!
* We zien allemaal Z-scores!!
  + Je kan zien of ze boven onder, …

LES9:

Item respons theorie IRT

# Item Respons Theorie (IRT)

## Inleiding

* Overzicht
  + We maken een onderscheid tussen geobserveerde variabelen of latente
    - Ze kunnen categorisch zijn of numeriek
  + We hebben enkel de eerste kolom!
    - Factoranalyse komt vaak voor in deze cursus!
    - Het was zo dat beiden variabelen continu waren
* In dit verhaal 🡪 Hetgeen we observeren zal categorisch zijn!
  + (Latente blijven continu)
  + Oude latente trek modellen! Is eigenlijk een betere benaming dan IRT
  + Het enige dat anders is, de metingen die we observeren (de Y’s zijn categorisch)
    - Wij gaan vooral spreken over binaire items 1-0/J-F
      * Kon ook ordinaal zijn!
      * Het kan zelfs met nominale schalen (geen orde, gewoon categorie 🡪 Zien we niet woehoew)

### FActoranalyse met dichotome items

* Lange tijd, leek het alsof er twee soorten theorieën zijn (twee scholen/kampen)
  + Factoranalyse en IRT
    - 🡪 Elk traditie en geen gelijkpunten!
* In factoranalyse 🡪 Constant continue items
  + Maar je kan dit ook doen met binaire items (dat gaat!)
  + Dit is ontwikkeld geworden!
    - Het eerste was dat ze een correlatiematrix berekende voor binaire items!
  + Later staken ze de tetrachorische correlatie erin en dat zette een stap voorwaarts!
* Een van de grondleggers van de IRT is eigenlijk vetrokken uit factoranalyse
  + Input 🡪 Binaire items!
    - Dit is niet voldoende, goeie stap maar een stukje te kort!!
* Jaren 70 🡪 Uppsala
  + Ene keer dat de gewichtsmatrix er was konden ze factoranalyse uitvoeren op binaire data en dat werkte prima!
* De ene school 🡪 De school van de factoranalyse!
  + Het is wel gelukt, ze hebben hun traditie, is niet perfect!

### Kenmerken van de factoranalyse benadering

* Voordelen
  + Het gaat heel snel is een groot voordeel!
  + We kunnen ook vele latente variabelen toepassen
* Nadelen
  + We vertrekken van een correlatiematrix 🡪 Veel info weggeworpen!
    - 🡪 Limited information approach!
      * Je kan zich wel voorstellen dat een full approach beter is!

### ONtwikkeling IRT (Andere school)

* Enorme invloed van M. Lord
  + ETS 🡪 Groot “bedrijf” die alle grote testen maken om te bepalen wie binnen mag in de universiteiten etc.
    - Ze hebben vele testen gecreëerd
* De latente variabele konden om het even wat zijn!
  + Persoonlijkheid, motivatie, om het even wat
* De IRT focust op het begin al op latente variabelen die te maken hebben met examens, proeven, testen
  + 🡪 Meer de focus op “ability”
  + Heel veel van het jargon valt te verklaren omwille van de focus op toetsen, examens en dat soort meer
    - Heel sterk binnen het educationele context!
* De focus wordt gelegd op de kwaliteit van elk van de individuele items
  + Goeie van de slechte onderscheiden en de slechte moeten eruit!
* Lord = pionier
  + Hij heeft ervoor gezorgd dat het de term IRT kreeg
* Deen 🡪 Georg Rash
  + Hij heeft een Rasch model gebracht!
* De Amerikaan is ok nog belangrijk (Allen)

### Kenmerken IRT benadering

* Voordeel
  + We gaan zowel probit als logit kunnen gebruiken
    - Factoranalyse is probit
  + GLMMs
  + Het is een full-information`
  + Een stuk eleganter, mooier theoretisch gezien!
* Nadelen
  + Om alles net te kunnen uitrekenen duurt het vrij tot zeer lang (zeker vroeger!)
    - 1 latente factor met een keer want meer ging niet!
    - (Scherp in contrast met meerdere factoren van vorige hoofdstukken)
    - Het is niet omdat ze het niet kunnen verdragen de hoeveelheid factoren
      * Maar het ging gewoon niet!
        + Vandaag meer met traditie dan limitatie!
      * De laatste jaren komen er MIRT’s aanbod
    - Het duurt lang!
      * Vandaag 1 factor 🡪 2 minuten
      * Factoranalyse 🡪 .2 seconden
  + Het meest jammere is
    - In de IRT-wereld hebben ze de focus gelegd op items analyseren
      * De effecten tussen de variabelen (als er 1 is, niet veel te vertellen, maar bij meer kan je je afvragen wat beïnvloedt wat?)
        + Nooit in de IRT voorgekomen!

### Recente ontwikkelingen

* Ze zijn hetzelfde maar andere jargon, andere schattingsmethoden, maar ze zijn eigenlijk gelijk!
  + Het onderscheid wordt heden ten dage niet meer gemaakt!
* Het is een beetje zoals 2 religies die plots beseffen dat de god die ze aanbidden hetzelfde is!
  + Velen zeiden “dat kan niet, dat kan niet”
* Samen gestoken in de moderne SEM
* IRT 🡪 1-factormodel met binaire items
  + Mooie opvolging van 1factormodel met continue items

## /Users/Sammy.Maipauw/Desktop/Schermafbeelding 2017-12-05 om 16.47.17.pngDe niet-lineaire relatie tussen 1 item en de latente variabele

### RElatie numerieke variabelen

* Bij continu item!!
* Hoe beter iemand de stof beheerst, hoe hoger en hoe beter zijn score
  + Dat is wat we hopen te zien!
  + Als dit 1 item zou zijn
    - Meerdere testjes, elke test is een item en heeft een score op 10
    - Je kan een factoranalyse doen met alle testscores als indicatoren en beheersing als latente variabele!!
* Dit is dus maar 1 van de vele items
  + De relatie is perfect lineair!
    - Hellingsgraad = factorlading
    - Intercept

### /Users/Sammy.Maipauw/Desktop/Schermafbeelding 2017-12-05 om 16.49.01.pngDistributie scores van 1 dichotoom item

* Y –as = de kans dat je dat item juist hebt
  + Hoe beter je de leerstof beheerst, hoe groter de kans dat je het item juist hebt!
* Wat ontbreekt is een soort functie (zoals de rechte)
  + Die voor elke mogelijke waarde op X zegt wat de kans is dat mijn item juist is
  + Maar je hebt niets in het midden
* Je moet iets hebben die ze allebei probeert te capteren!
  + Deze functie je moet zoeken, deze die de relatie legt tussen de latente variabele en de kans dat je item correct is
    - = Item responsfunctie

### de item responsfunctie

* Alle 3 de termen worden door elkaar gebruikt!
  + IRF, IRC of ICC
* Een lineaire relatie werkt alleszins niet!
  + Want op den duur als je maar extreem genoeg gaat met rechte kom je kansen uit die kleiner dan 0 zijn en groter dan 1
  + 🡺 Andere vorm!
* Alternatieven (3)

### /Users/Sammy.Maipauw/Desktop/Schermafbeelding 2017-12-05 om 16.52.50.pngDE stapfuncite van Guttman (1ste)

#### Itemmoeilijkheid = B-parameter

* Index j 🡪
  + Als je 40 vragen hebt dan krijg je 40x een moeilijkheidsparameter
  + We zullen deze moeten schatten voor een item
* Als B heel klein is 🡪 de -2 slaat op de schaal van latente variabele
  + Dat betekent dat de kans 1 wordt vanaf je aan -2 zit
  + De kans dat je het juist hebt is 100% een keer je voorbij de -2 zit!
* De andere is moeilijk bij B = 1 moet je minstens bij 1 zitten, dus dan is het een moeilijker item!
  + Hoe meer links, hoe makkelijker, hoe meer rechts, hoe moeilijker!
* Kleine reflectie
  + Heel dit hoofdstuk praat over items, examen, moeilijk, makkelijk 🡪 Heel het jargon valt onder die context, maar er is geen enkele reden waarom dit enkel daarin zou kunnen toegepast worden
    - Je kan ze allemaal toepassen overal op voorwaarde dat de items binair zijn
      * Vele vragenlijsten met JA/NEE!
        + Bv. Depressievragenlijst!

Sommige items zijn ernstiger dan andere!

Het ene zie je vaker dan het andere!

De ernst van het item = de moeilijkheidsgraad van de examenvragen

* + - * + “WAT IS DE KANS DAT IK IETS OBSERVEER”

### Normaalogief model (2de)

* Hoofdletter phi
  + Op de X-as staan Z-scores en op de Y-as probabiliteiten (de kans/proportie)
* Deze normaalverdeling bleek een functie te zijn die perfect deed wat het moest zijn
  + Hoe meer je naar rechts gaat, hoe groter de kans dat een item juist is
  + We kunnen deze curve een beetje aanpassen
    - Wat platter, scherper laten lijken! Opschuiven etc.
      * Meer flexiibiliteit = voordeel hiervan!
* Die Z-scores
  + Probability-units
    - Prob units 🡪 PROBITS = De getallen op de X-as (slaat op de metriek)
    - Pnorm en Qnorm
      * Pnorm is de functie dat je een probits in en je bekomt een probabiliteit en Qnorm deed het omgekeerde!

### Item moeilijkheid

Slide

#### Item discriminatie aj in het normaalogief model (bj = 0)

* Het eerste = slecht item!!
  + Je moet niets kennen, het is alsof je een muntje gooit!
    - We willen een item dat een grotere kans heeft indien je de stof beheerst
      * Eerder het 3de item
        + Heb je een bepaald niveau bereikt heb je een grotere kans het te kennen
        + Daartussen een grijze zone, kan heel snel gaan, maar ook veel trager!
      * Bij een scherpe stijging 🡪 Dit item discrimineert heel sterk (onderscheid wordt gemaakt tussen twee groepen)
        + Namelijk een goed discriminerend item is heel groot dat je het juist hebt als je in de goede groep zit en niet als je in de slechte groep zit!

A 🡪 kleine waarde is slecht discirminerend!

Typisch ergens tussenin de 4-10

* + - * Maar je weet dit niet op voorhand
        + Een van toepassingen is een examen afnemen en zo kijken welke slecht discrimineren

Deze zullen ze de jaren erop niet meer vragen!

Geneigd zijn ze niet meerekenen in het examen maar dat mag niet…

* + - * + De kwaliteit van je items, een voor een nagaan!

Dit staat los van moeilijkheid!!!

#### Combinatie item discriminatie en item moeilijkheid!

* De twee parameters fungeren los van elkaar!
  + Op deze plots teken je de respons curve en je weet al voor dergelijk items is dit de a en dat de b
    - De bedoeling is dat je net voor elk item de a en de b te weten komt
      * We willen ze schatten, we willen deze karakteristieke weten!

#### 3de item parameter 🡪 De giscoëfficiënt = c parameter

* C = 0 🡪 klassieke parameter
  + Puur op basis van toeval heb je nog steeds 40% kans (als c = .4 is)
    - Elk items heeft 4 antwoordalternatieven
      * Stel dat je er niks van kent en weet niet waar het over gaat
      * Blindelings invullen heb je nog steeds .25 kans! = c

#### Problemen met het normaalogief model!

* Er waren vroeger dikke boeken voor Pnorm uit te rekenen!!
  + Vele tabellen want er zijn vele curves!

### Het logisitsich (ogief) model (3De)

* Het is veel eenvoudiger 🡪 Het is een stuk makkelijker
  + Louter om praktische redenen gaven sommigen de voorkeur aan deze versie van de responscurve!
  + = Logit

#### Itemresponsfunctie: standaarnormaal versus logistisch

* Behalve dat de schaal en alle parameters een beetje anders, ze worden wel evenveel gebruikt

### Overzicht IRT modellen

* Om te beginnen enkel over de IRT-modellen met 1 factor en ook enkel modellen met maximaal 2 item parameters met de a en de b
  + De c van daarnet komt er niet meer tussen
* De laatste rij zal hij verder niet bespreken!
  + De keuze van responsfunctie is heel vaak standaardnormaal!
    - = 1Pnormaalogief model
  + Soms logistisch 🡪 1PL
* Deze cursus 🡪 1P en 2P ‘unidimensionele’ IRT modellen bespreken met 1 latente variabele

## Dataset 2: 10 itemscores examen statistiek II

* Persoon 1 heeft 1ste fout 3de mis
  + 0 = fout, 1 = juist
* Je kan hierop een IRT analyse uitvoeren
  + Het heeft 2 hoofdbedoelingen
    - 1ste 🡪 Kwaliteit van de individuele items bestuderen!
      * Wat was de moeilijkheidsgraad en wat is de discrimineerbaarheid van het item
        + Om slechte items weg te smijten!
        + Je krijgt de data te zien en zie je andere data, de profs weten niet meer wat makkelijk/moeilijk is, het is moeilijk in te leven in een student 🡪 Vergissen ze hun er vaak in

Soms is bepaald makkelijk item zeer moeilijk te zijn

* + - * Bedoeling is dat je kijkt naar de goeie en de slechte!
    - 2de 🡪 Een keer je al die itemkarakteristieke kent 🡪 Dan kan je ook net zoals in factoranalyse:
      * Je kan dan ook (scores op de latente schaal) die scores geven aan de studenten
      * Die scores, zijn niet noodzakelijk de scores die we krijgen, wat wij krijgen is de somscore
        + Som van de juiste antwoorden (los van Standard setting)
        + Je kan dus via IRT, eerst items analyseren en als dan blijkt dat het model goed past bij de data dan kan je factorscores berekenen op de latente variabelen en die kan je dan transformeren naar een score tussen 0-20

= Alternatieve score gebaseerd op IRT

≠ scores, correleren wel!!

* + In deze instelling gebruiken we enkel de somscores!
    - Hier mag je geen IRT-scores gebruiken!
    - Hij heeft het al proberen uit te leggen dat het geen goed idee is!
      * IRT er zit iets in (andere instellingen gebruiken ze wel!!)

## Essentie van IRT

* 2 doelstellingen: item analyse en een tweede is factorscores (ability scores)

## Het statistisch model voor IRT

* De meeste IRT-handboeken starten hiermee!
  + Een model moduleert de kans dat je een item ‘j’ dat je die juist hebt en deze kans gaan we ervan uit dat deze afhangt van eta (score op de latente variabele)
    - Beheers je de leerstof zeer goed dan is de kans groot!
    - Als je verschuift op de schaal, dan verschuift ook je kans
  + Pj = Kans op een juist antwoord voor item j en we houden in ons achterhoofd dat dit afhankelijk is van je plaats op de schaal!
  + Bij ons 🡪 een functie van enerzijds itemparameters (a & b) enerzijds en anderzijds een persoon specifieke parameters (score op latente variabele)

### RAsch model

* Meest benoemde model
  + Zeer elegant in zijn eenvoud!
    - Het heeft dus vele mooie eigenschappen (wiskundig)
  + Zeer geliefd, maar praktisch gezien niet zo goed en het wordt nauwelijks gebruikt!
* Er is maar 1 item parameter 🡪 Moeilijkheid (b) geen a!!
  + Je krijgt discrimineerbaarheid niet te zien!
  + Typisch
* Andere eigenschap 🡪 Gebruiken enkel de logistische versie!
  + Volgens het model is de kans die psi en wat er ingaat is enkel eta en b en geen a
    - Dit is eigenlijk een expliciete manier om de kans te berekenen!

### Item responsfunctie voor het rasch model

* S-vorm is hetzelfde en meer kan er niet gebeuren dan verplaatsen op de schaal!
* Die ziet er altijd hetzelfde uit kan enkel verschuiven volgens de moeilijkheidsgraad en dit kunnen we per item tekenen
  + Je kan dat ook samenvatten voor alle items samen, als je ze allemaal bijeen gooit en je hebt 100 items dan krijg je 1 vorm, een soort gemiddelde functie voor de hele test!
    - Het enige dat kan verschillen tussen de items is de moeilijkheidsgraad, ze schuiven
      * Als je ze allemaal optelt krijg je deze testresponsfnctie!
        + Op de Y-as, heb je de somscore/verwachting ervan over de 5 items

Geen probabiliteit van 0-1

Score 0-5

* + - Je ziet hoe de testscore zich verhoudt met je latente variabele

Pm LES10

## Het statistisch model voor IRT

### Rasch model

* Het heeft een speciaal statuut en het wordt aanbeden door sommigen
  + Dat het enige juiste model is!
    - Hardcore Rashianen
* Wat kunnen we doen als het model niet past
  + 2 filosofieën
    - De aanhangers van Rasch zeggen dat het aan de data ligt
    - Een andere manier van denken, we moeten op zoek naar een ander model dat wel in staat is om te karakteristieken van onze data te kapteren
      * Er is dus best wel veel discussie over!

### Item-responsfuncties voor het Rasch model

* De moeilijkheid is verandert!
  + Hier schuiven ze gewoon op van links naar rechts!
    - De vorm verandert niet 🡪 Enkele goede eigenschappen!
    - Als je alle items optelt en je kijkt naar de testresponsfunctie
      * Bv. 40 vragen en 40 van de curves en je telt ze als het ware allemaal op bekom 1 grote curve!!

### TEstresponsfunctie Rasch model

* Als de curve te veel naar links ligt 🡪 Te makkelijk
  + Te veel naar rechts 🡪 Te moeilijk!
    - Zo kunnen we zien over de gehele test of deze te makkelijk was of te moeilijk!

### Het 2-parameter IRT model

* Slide 31

### Item responsfunctie voor het 2-paramter model

* We hebben meer variaties
  + Ze kunnen meer vormen aannemen!
    - 2 parameters 🡪 Moeilijkheidsgraad
    - A parameter 🡪 die de “platheid” van de curve laat zien
      * We kunnen de hellingsgraad van de curve laten variëren voor elke curve apart!!
      * 🡪 Item karakteristieke curve noemen we deze lijnen soms

### Test responsfunctie 2P model

* Test responsfunctie is de som!!
  + Globaal gezien kunnen we zien of de test moeilijk was of niet!
  + Deze curve is hij min of meer plat of scherp 🡪 Zegt iets over het discriminerend vermogen van de gehele test
    - Soms wenst men een scherpere curve 🡪 Dus we moeten items zoeken die heel erg discrimineren!!
      * Soms willen we dat het vlakker is!

## Schatting van de item parameters (optioneel)

* Grosso modo twee methodes
  + Een deel van de data gebruik maken 🡪 Limited information
  + Of we kunnen het net doen! 🡪 Full information
    - Het volledig deel
* Een hele literatuur, als je je afvraagt wat die psychometrici doen in de kelder 🡪 Heel wat van dergelijke zaken!

## Schatting van subject parameters (optioneel)

* Conceptueel gezien is het hetzelfde als factorscores (wel wat moeilijker te berekenen door het binair zijn van de items)
* Het is altijd interessant om de IRT-gebaseerde scores, die niet observeerbaar is en een product is van het model
  + Dan voorspelt het model 🡪 Voor dat individu, met zijn profiel
    - Dit is zijn Eta waarden op de as, is het heel hoog/laag
      * Er zijn mensen die er schrik van hebben!
      * Want je kan het niet zien
* Een alternatief 🡪 Klassieke manier
  + We berekenen een somscore 🡪 We tellen het aantal juiste items en voilaa
    - Je kan ze wel wat transformeren etc.
      * Je hebt aan de ene kant de geobserveerde somscore
      * Maar bij 2factoren 🡪 Geen one to one mapping met de latente variabelen!

## IRT in de praktijk

* Elke keuze heeft implicaties, hoe veel parameters zal ik hebben? Dus je moet nagaan of het model past bij de data of niet!
  + Wat ook een belangrijke vraag is, is 1 wel voldoende? Misschien 2-3 …
    - We moeten dat nagaan (net zoals in de exploratieve factoranalyse)
    - We moeten het nagaan! Je kan niet zeggen het zal wel 1 zijn, dit mag niet
  + Wij zien enkel de modeltoets van lavaan!
* Als het past, enkel en alleen als het past, dan kunnen we kijken naar de a en de b scores van de items!
  + Dan kunnen in principe ook de factorscores berekenen
    - Als het niet past, dan heeft het niet zo veel zin, je krijgt uitkomsten, maar deze zijn vrij waardeloos!
* IRT-modellen zijn in zekere zin fantastisch, maar je mag dat niet zonder kennis van zaken zomaar toepassen!
  + Je moet weten wat je doet en de conclusies moet je goed kunnen afwegen ten opzichte van praktische bezwaren!
    - Zijn de assumpties wel geldig, …

### Rasch model in R

* Je moet een model specificeren (🡪 Hier is het 1 factor, het kon met 2/3 zijn, er is niets in de IRT die zegt dat het 1 moet zijn!)
  + Hij heeft hier 1 factor “f” en deze wordt gemeten door 10 items die binair gescoord zijn
    - Wat er speciaal aan is en het een rasch model maakt 🡪 “1\*items”
      * Deze 1\* = Factormodel, maar alle factorladingen worden op 1 gefixeerd en worden niet geschat
      * Want de factorlading communiceert met de a-parameter!
    - Het dogma hier is 🡪 alle factorladingen moeten 1 zijn zo voert men dit in!
  + Je kan een standaard CFA uitvoeren!
    - Je moet aangeven met ordered 🡪 Welke van de 10 items dat je moet beschouwen als ordinaal 🡪 Alle variabele namen (v1,v2, ..., V10)
      * Hij zegt ook al die 10 items beschouw ze allemaal als binair en niet als continu!
  + Dit is het dus
    - Het uitvoeren van een IRT-analyse 🡪 1 bibliotheek 2 lijntjes model 1 commando en dan kijken naar de summary
    - Daarna kunnen we de fit doen
      * 🡪 IEDEREEN KAN HET! (vandaag toch)
* SCHEMA
  + .002 = p-waarde 🡪 Vrij triestig
    - Samplesize kan een probleem zijn, maar hier kijkt men daar niet van op
      * 500-2000 is normaal/oké
      * Als de samplesize stijgt, daalt de p-waarde
    - Niettemin we zijn er niet gelukkig mee!
    - Skip de andere twee getallen 1.067/5.402

### Past het rasch model bij de data?

* Als we de summary zouden opvragen zouden we veel meer krijgen!
  + We spraken altijd over de a en de b parameters 🡪 Deze staan niet altijd zo in de output!
    - Je kan dit wel allemaal omzetten (maar hoe dit gebeurt, hoe je niet te weten!)

### Omzetting CFA paramters naar a en b (optioneel)

Slides

### De geschatte item parameters voor het rasch model (probit metriek)

* Hoe hoger hoe moeilijker!
  + Dit is de moeilijkheid geschat door het rasch model 🡪 Dit is niet altijd de waarheid
    - Makkelijkste = -1.62
      * Examinator wil weten wat het item was
        + Mss is hij stomverbaasd ofwel omgekeerd!
        + Discrepantie tussen verwachting en wat hij ziet!

Tussendeeltje even Teruggooi H3

* SLIDE 65
* De gemiddelde voor elke kolom berekent!
  + De proportie juist
  + .69 = makkelijker dan V2 want deze is .38 🡪 Deze is moeilijker
    - Je zal zien de volgorde van moeilijkheid is hier hetzelfde dan in het rasch model
* Waarom hebben we het Rasch model nodig als we het zo kunnen zien?
  + Als dit het enige dat je wilt weten is, dan hoef je het Rasch model niet gebruiken!
  + Je kan dan hier beschrijvende data gebruiken!
    - Hoe hoger het getal 🡪 Hoe makkelijker het item!
      * Maar de volgorde zal hetzelfde zijn!! Enkel zo bij het Rasch model is niet meer zo bij 2P model!

TERUG BIJ H7 IRT

### 2P normaalogief model in R

* Je hoeft de factorlading niet meer te fixeren!
  + Moet wel weer geordend worden 🡪 Binair
  + Latente variabelen zal gemiddelde nul hebben, maar we willen ze zo definiëren dat de variantie 1 is
    - Std. Lv
      * = Standaardisatie
  + 3de lijn 🡪 P-waarde = .87 🡪 Veel beter dan het Rasch model!
    - Maar het rasch heeft mooiere eigenschappen, maar de realiteit is hard, meestal past dit model beter dan het rasch!
* Je kan weer de parameters omzetten (hoe dit gebeurt hoef je niet te weten)

### GEschatte item paramters voor het 2P model (probit metriek)

* Je hebt nu 2 geschatte waardes!
  + Het makkelijkste en het moeilijkste item blijven gelijk!
    - Het midden (de grijze zone) daar kan de volgorde wel variëren!
      * Dus de volgorde zal niet noodzakelijk dezelfde zijn, maar dit komt omdat men rekening houdt met de discrimineerbaarheid
* V9 🡪 .88 = Heel goed gevolgd door V2/V10/V6 = .71
  + Maar V3 is heel slecht
    - Is het moeilijkste maar ook slechts determineerbaar!
      * .26
        + Dit is wat men nu noemt een SLECHT ITEM!!

Werkelijk “euwiweuw”

### Item responsfuncties 2P model

* Je ziet duidelijk welke je kan wegdoen en welke goed zijn!
  + Je krijgt feedback van je item
* 9-6-2 zijn best wel goed!
  + Item 4 en 3 zijn niet zo goed, 3 is de ergset!
    - Maar 4 is ook niet goed want bijna iedereen heeft hem juist!

### Subject paramters 2P omdel

* 1ste rij = somscore 🡪 rowsums 🡪 elke rij is een student!
  + Je telt ze allemaal op! = Totaalscore op 10
    - = Som van kolommen per rij!
    - = Testscore zoals deze is geobserveerd
* 2de lijn = Lavpredict 🡪 LAv = Lavaan en predict = ik voorspel de scores!
* Factorscore 🡪 Persoonscores wordt het ook genoemd! Subject parameters
  + = Scores op de latente trek
    - Dit zijn Z-scores, omdat deze zeer vervelend zijn, je wilt ze vergelijken met je somsscores
* Volgende lijntjes 🡪 Je zal er perfecte Z-scores van maken! (3de lijn)
  + Hierna kan je transformeren!
    - Gemiddelde op 6 in plaats van nul en sd = 2
      * Je zal afronden, getal na de komma afsmijten (kan beter met floor + .5)
        + Maar dit is om ruwweg te illustreren wat je krijgt
  + De laatste 3 lijntjes is om ze in een tabel te krijgen!
* 1ste kolom = Scores voor je factorscores ernaast staan de getransformeerde versies!
  + De laatste kolom = somscores!
    - Heel vaak kom je hetzelfde ooit, maar soms niet!
    - Soms is het in het voordeel van de student en soms niet!
      * Kleiner of groter!
    - Wat heb je liefst 🡪 De methode in mijn voordeel, maar dat is niet makkelijk om voor de hele groep te kiezen

### RElatie somscore en factorscore

* Gelukkig is de correlatie tussen beiden vrij hoog!!
  + Bijna hetzelfde, maar iedere keer rond de kritische waarden wordt er veel heen en weer gesprongen!
* Dus de ranking van de eerste 10 studenten!
  + Hier zie je de rangorde van de studenten, maar deze is verschillende bij de somscores en de IRT!

## Item en test informatie

* Informatie 🡪 Groter letter I
* Item informatie 🡪 met een index!!
* Je zal zien dat het geen vlakke lijn is!!

### Formules item informatie (optioneel)

* In het rasch model is dat nog zo gek niet, maar in het 2P model is dat moeilijker!

### Item informatie in R (Rasch)

* Voor elk item is het quasi identiek
  + De vorm is gelijk enkel verschuift het een beetje van links naar rechts
  + Dit is typisch voor Rasch
    - Bv. Item 4 🡪 Zie je dat het een beetje naar links schuift 🡪 Deze is meer informatief voor mensen die eerder aan de onderkant van de schaal zitten
      * Aan de bovenkant is dit niet informatief
      * Het leert ons niets als we aan de bovenkant zitten!
    - 🡺 EZ!!

### TEst informatie RAsch

* Erg informatief voor de middenstuk
  + Maar men wil een test voor een bepaalde zone!
    - Voor de hogere op de schaal is het minder informatief

### Item infromatie in R (2P)

* Item 3 🡪 Informatie is quasi nietes
  + Hoe platter de curve hoe minde informatief!
* Item 9 🡪 Heel informatief, voor een brede rang!
  + Het hoeft niet noodzakelijk breed te zijn, maar dan moet je wel meerdere items hebben die zich op verschillende plaatsen vinden!
  + Hoe hoger de piek, hoe lever!
    - Item 5 is ook een zwak beestje
  + Item 4 niet te vet
  + Item 5 ook niet
  + Item 6-7 kan er mee door
  + …

### TEstinformatie 2P

* Matige piek!
  + Het blijft informatief voor de mensen aan de lage kant van de schaal als voor de mensen aan de hoge kant!
  + Mensen aan de hoge kant daarvoor is het niet zo informatief want het maakt voor hen niet veel verschil tussen een 18-19/20
  + Het is belangrijk informatief te zijn voor de mensen die schommelen rond de grenswaardes!! (8-9-10)
    - Semi door de discrimineerbaarheid!
* Examinator 🡪 hoeft niet informatief te zijn om goed het onderscheid te maken aan de extreme lage kant of de extreme hoge kanten
  + Je moet geen examen opstellen dat zeel goed de mensen kan differentiëren van iemand die 18-19 zou verdienen!
  + Je hebt liever een examen die heel kritisch en nauwkeurig is rond de kritische grens die bij ons de 10/20 heeft
    - Liever hier onderscheid maken dat iemand eerder een 9 of een 10 verdienen!
  + Dus je kiest je items zo dat ze maximaal kiezen rond de kritische waarden van 10!!
    - Niet al je informatie moet daarrond gecentreerd zijn, maar daardoor zwakt het wel af naar de extremen toe!

## Toepassingen van IRT

### Item bank

* Enorm veel werk!
  + Voor een vak als psychometrie kan deze bank niet oneindig groot zijn
    - Maar bv. Pizza studies (huh? The fuq?) en andere die internationaal gebeuren
* Bv. Mastery test ontwerpen
  + Er is een kritische zone waar je super informatief over wil zijn!
  + Andere is adaptief

#### Ontwerpen van een mastery test

* Is een test waar eigenlijk het enige van belang is 🡪 Deze persoon wil weten of je een bepaald niveau hebt bereikt of niet?
  + Of je dit bereikt hebt 🡪 Dan is de persoon niet geïnteresseerd in de gradatie die erna komen
  + En is helemaal niet geïnteresseerd in de onderkant!
  + Het enige 🡪 “Kan de persoon het ja of nee?”
    - Kan hij over de grens springen of niet!
      * Examen zijn in zekere zin zo 🡪 Kritische grens 10
        + Als je een 10 hebt dan kan je het
        + Geen 10 🡪 Geen krediet
* Bv. Piloot 🡪 9/20 is niet zo fijn, een piloot die 10/20 haalt is dat ook niet zo waauw
  + Je zou graag hebben dat die 16-18/20 hebben
    - Een soort minimumniveau 🡪 dit willen we meten!
      * Eta’c’ van criterium = soort grens
      * Hoe goed kan de piloot zijn vliegtuig besturen?
    - Dus we meoten een test ontwerpen die rond de kritische test super inormatief is!
      * Rond de kritische waarde

#### TEst informatie mastery test voor (nc=1.2)

* Een selectie van items op zo’n manier dat de totale informatie rond de kritsche grens maximaal is
  + 🡺 Meestertesten worden op deze manier veel ontwikkelt!
  + Dit kan je enkel maar doen als je over een grote item bank beschikt met alle informatie over de informatie per item en op die manier kan je selecteren!

#### Ander toepassing = adaptief testen

* Examen is typisch 🡪 40 vragen en iedereen moet ze oplossen
  + Stel je voor 🡪 Moeilijkste vragen komen eerst (is niet zo)
* Je kan snel je score op een latente trek verkleinen en verkleinen
  + Je kan met een veel kortere test dezelfde score bekomen dan wel iedereen dezelfde items
    - Adaptief testen, na elke aanbieding ga je beslissen wat zal mijn volgend item zijn
      * Hier zit er een volledig algoritme achter 🡪 Vooral bij CAT
  + We kunnen dezelfde testen die 2uur duurden nemen in een kwartier
    - We weten dat we een item nodig hebben die wat moeilijker is!
      * Je biedt het aan 🡪 Heeft hij het juist of niet 🡪 Proberen een schatting te doen
        + Ik ga een item bieden rond dat typische kritsche waarde en zo item per item aanpassen!
* Dit heeft vele mogelijkheden, maar het is niet gemakkelijk om het goed te doen laten verlopen!
  + Het komt meer en meer op ons af

### ITem bias of DIF

* Stel 🡪 Je hebt een vraag en je hebt 2 mensen
  + Puur hypothetische je weet wat hun score is op de latente variabelen 🡪 Ze zijn gelijkwaardig
    - Idealiter is voor beide personenen is de kans dat ze item 13 juist hebben, dezelfde want ze hebben de gelijke eta waarden
    - Stel dat het niet zo is wegen geslacht 🡪 ITEM BIAS
      * Verschuiving van wat er zou moeten zijn veroorzaakt wat niets te maken heeft met item/latente variabele maar door een externe variabele = geslacht (Amerika = typisch “ras”)
        + Hypergevoelig!

Kan toch niet zijn dat het is dat iemand beter zou zijn met een andere ras!!

* Echte wereld komt dat toch voor!
  + Je hebt een bepaald woord gebruikt dat anders wordt geïnterpreteerd in een andere cultuur en is deze daardoor verwarrend voor andere culturen!
    - Het opsporen van dergelijke items is ene kunst op zich
* Het item gedraagt zich anders dan wat het zou moeten doen bij de ene groep dan wel de andere groep
  + Dit heeft te maken met het begrip fairness 🡪 het moet fair zijn en het mag niet dan andere mensen meer kans hebben ookal zitten ze op hetzelfde niveau!
    - Als testontwikkelaar moet je opletten!
* Bv. Het bleek zo te zijn dat Aziaten beter scoren op IQ-test dan andere rassen!
  + ITEM BIAS
  + Allerlei elementen waardoor de ene cultuur bevoordeeld werd ten opzichte van de andere

### en nog zo veel meer

* Detecteren van spiekgedrag 🡪 IRT kan van elk antwoordpatroon zeggen dat er enkele items zitten dat hij juist heeft terwijl dat hij dat niet juist zou mogen hebben
  + We kunnen vreemde antwoordpatronen eruit halen!
* Detecteren van verandering 🡪 Groei
  + We willen zien dat er evolutie in zit
    - Meten over de tijd is iets dat we kunnen doen
* We kunnen veel meten over de tijd, Bv. Motivatie
  + Deze moeten we dan ook meten
  + Dit moeten we er bij nemen en zo vorm je een structurle vergelijksmodel?

EXTRATJE PAPERS (niet te kennen sjanske)

* Extra informatie over de IRT 🡪 Er zijn twee sprekers gekomen!
  + Hij was in shock wawie
    - IRT is als rijden met een sportwagen, je moet het begrijpen hoe het werkt
    - Je kan je er pijnlijk mee vergissen

LES11 PM  
H8: Validiteit

# Validiteit

* Het meest mysterieuze en het moeilijkste te bevatten!

## Inleiding

* Het is een containerbegrip, het wordt ten pas en te onpas in allerlei contexten gebruikt!
  + We gaan voornamelijk over psychometrische validiteit praten
    - 🡪 Meetinstrument/test/examen kort door de bocht “meet het ding wat het meent te meten?”
* Bredere context
  + Is de studie in zijn geheel valide, zijn de conclusies die eruit komen te vertrouwen? Als dit niet zo is dan is de studie niet valide!
    - Hier helaas niet over hebben in het vak!
  + Hoe was de steeproef trekking?
    - Was het een makkelijke sample? Random samle? …
  + Wat is de onderzoeksopzet?
  + We zijn met zijn allen een beetje te weinig kritisch
  + Confounder
    - Uitkomstvariabele Y en X meet-item
    - Een andere variabele die een invleod heeft op beide variabelen (zowel op Y als op X)
      * Deze vervormt het verband en je meot het inrekening brengen
        + Ofwel ben je er vanop de hoogte en weet je wat ze zijn ofwel weet je het niet? Dus je moet je conclusies zo robuust mogelijk maken tegen de schending van confounders in het model!
  + Externe validiteit 🡪 Het opzetten van een wetenschappelijke studie is onzettend moeiijk!
* Ook in de praktijk volgt psychodiagnostiek!
  + Zowel voor klinische context, bedrijfscontext
  + Op het einde van diagnostiek komt er vaak een diagnose, maar de kwaliteit hiervan is deze correct of niet?

## Belang van validiteit

* Als het meetinstrument niet valide is dan is heel de toetsting van de theorie om zeep!
  + Het is cruciaal dat het meetinstrument doet wat het moet doen, het moet meten wat de bedoeling was om te meten
* In de prakijk, vanaf dat je een test hanteert in welke context dan ook, wil je zeker zijn dat deze test meet wat je wilt dat het meet!
  + Helaas in onze wereld, worden allerlei testen gebruikt die invalide zijn
  + “Meten is weten” 🡪 “Meten is zweten”
    - Het is moeilijk want het is delicaat en lastig!
    - Want er zijn vele gevolgen!

### Voorbeeld 1: Validiteit in wetenschappelijk onderzoek

* Vele studies die te maken hadden met videogames 🡪 Agressie met gewelddadige videogames
  + Een resem aan studies 🡪 Is het zo dat het spelen van dergelijke games, leidt het tot echt gewelddadig gedrag
  + Het is zo ver gegaan dat ze een hoorzitting gehouden!
* Er is onvoldoende evidentie!
  + De stelling is niet noodzakelijk fout!
  + Maar we hebben niet voldoende informatie om het met zekerheid te bevestigen of falsifiëren, we kunnen het niet zeggen of het juist of waar is
    - Dit is door meting 🡪 “Hoe meet je de neiging om zich agressief te gedragen?”
      * Dit gaat niet!
        + Ze hebben zelfrapportage gebruikt en ze hebben zelf gevraagd van kijk vind jij zelf dat je nadat je zo’n game hebt gespeeld dat je meer of minder/heb je dan te neiging je agressief te gedragen?
        + Goeie meting? 🡪 Heel veel discussie!!
* We zijn soms een beetje overmoedig
  + Vele dingen meten zijn moeilijk! Je moet aantonen dat het valide is
  + Dit is echter een lang proces

### Voorbeeld 2: Validiteit in een sollicitatiecontext

* Kleurenquiz 🡪 Hier wordt er een variant van dit testje gebruikt om personeel te seleceteren
  + Je krijgt een aantal kleurenvakjes te zien, wat is je favoriet kleur en zo moet je in volgorde de kleuren aanduiden degene die je het liefst ziet tot degene die het minst graag ziet
    - Ze willen zien of je consistent bent of niet
      * 🡪 Persoonlijkheidsprofiel: “Stevig in je schoenen, introvert, leiderschap”
        + ONZIN! Maar wordt gebruikt
* Ethisch is dit onverantwoord
  + Als je een instrument gebruikt en je weet dat er geen enkele empirische evidentie is voor de validiteit van dat meetinstrument dan vinden we dat ethisch niet kunnen!
* Een ziener!
  + We hebben allemaal een persoonlijke aura kan aflezen tijdens het interview
  + Je kan allerlei kristallen kopen!!
* Helaas in dit land zijn er dergelijke bureaus…
* Myers-Briggs 🡪 MBTI persoonlijkheidstest 🡺 Super populair!
  + 0 validiteit 🡪 Het is een schande, maar wordt wel als een wetenschappelijk iets gebruikt
    - Het punt is wij kunnen de literatuur raadplegen, dus als je geconfronteerd wordt met een test 🡪 vooraleer je die zelf gebruikt, ZOEK HET OP, het is je ethische plicht de literatuur te raadplegen of er voldoende evidentie is voor validiteit van de test!
      * Dat is de reden waarom we al die statistiek vakken krijgen, we moeten in staat zijn met een kritische blik de empirische evidentie te kunnen lezen om te zien of ze valide zijn of niet!

## Korte historiek van het concept validiteit

* Idealiter willen we een getal hebben net zoals bij betrouwbaarheid!
* We hebben dat decennialang geprobeerd, maar dit is te moeilijk/lastig!
* Clark Wissler 🡪 1ste validiteit studie
* Kelley 🡪 Hij heeft een definitie op papier gezet die heel populair is
  + Meet het instrument datgene wat het poogt te meten
* Cronbach heeft diep nagedacht over validiteit 🡪 Hij heeft een paper geschreven
  + Hij maakte een onderscheid van de 4 soorten! Deze worden nog toegelicht!
* Later
  + Messick 🡪 Wat je met de test doet is belangrijk!
  + Hij heeft het concepter vershoven en wat breder getrokken
    - Bepaalde test kan valide zijn in een bepaalde context maar niet in een andere context
      * Het is afhankelijk van de context/intentie, …
      * Dit is overgenomen in de APA 1999
* De officiële definitie is deze van Messick!
* Denny Borsboom 🡪 Collega 🡪 Deze man heeft veel geschreven over validiteit
  + Hij heeft veel invloed gehad 🡪 We gaan terug naar de oer-definitie
  + Het gaat terug over de test zelf!
* Validiteit heeft een heel parcours afgelegd!!
  + We hebben het einde waarschijnlijk nog niet gezien!

## Een veelvoud van validiteit definities!

* Er zijn heel veel definities 🡪 De belangrijkste eruit genomen die waarschijnlijk ook zullen terugkomen!!

### Construct validiteit/begripsvaliditeit

* Overkoepelende definitie
  + Een groot deel van de toegepaste psychologen denken aan dit concept wanneer ze over
* Gedraagt de meting zich zoals de theorie het voorschrijft
  + Het gaat dus eigenlijk over wat zegt de theorie mij over dat construct?
  + Wat hoort het te meten en hoe positioneert dat construct zich tenopzichte van andere constructen waarmee het gerelateerd is
    - Construct A moet sterk correleren met B
      * Je kan metingen verzamelen en kan je zien of ze hoog zijn
      * Als het hoog is 🡪 Evidentie voor construct validiteit = Convergerende validiteit!
  + Soms zijn er constructen waarvan we zeggen dat A niets te maken heeft met B
    - Dan verwacht geen correlaties tussen die meting, als je er toch vindt 🡪 Probleem = discriminerende/discriminante validiteit!
* Het nomologisch netwerk 🡪 alle constructen die gerelateerd of niet zijn, maar relevant zijn voor het construct

### COnvergerende en discriminante validiteit

* Doenbaar om nog na te gaan, maar het hangt heel sterk af van de kwaliteit van de theorie
  + Achillespees 🡪 Vaak is de theorie nog niet stevig genoeg!
  + Je poogt de meting net te verzamelen de theorie te vormen, maar als de theorie nog niet volledig/af is dan kan je de validiteit niet berekenen
    - We pogen dat de theorie af is
* Stel we hebben de theorie af
  + Zijn er verschillende manieren 1 = MTMM 🡪 Op basis van vele correlaties tussen al die concepten een uitspraak te kunnen doen
    - De moderne manier = CFA 🡪 Factor met zijn indicatoren en andere factor met zijn indicatoren en we kijken naar de correlatie

### FAce validity/validiteit op zicht

* De zwakste versie!
* Je kan mensen een vragenlijst laten invullen!
  + Iemand die dat item leest, wetende dat we “iets” meten
  + Als iemand vindt dat dat een goed item is 🡪 Ja dat meet dat, ik vind dat
    - In zekere zin is dat puur subjectief ‘ik vind het goed of niet”
      * Maar dit wordt wel heel vaak gebruikt!
  + De flauwste manier
* Zeer onwetenschappelijk

### INhoudsvaliditeit

* Wetenschappelijk variant van de “face validity”
  + WE krijgen scores bij experten en zo kan je zien waar de consensus ligt
    - Je krijgt een beetje zicht of de items goed zijn gekozen?
    - Ze hebben lang gezocht om het te formaliseren!
* Bekendste techniek 🡪 Lawsche
  + Per item kreeg je een score tussen -1 en +1
  + Hoe hoger de score, hoe meer valide het instrument was
  + Hoe meer inhoudsvaliditeit het instrument had
* Het wordt wel geaccepteerd als zijnde een zinvolle oefening, maar het volstaat niet!
  + Het is een belangrijk aspect, maar er moet meer zijn dan dit!
  + Het is logisch dat experts er moeten eens zijn!
    - Er moet een consensus zijn dat de inhoud van de test op zicht goed is door experts!
  + Deze kwam het dichts in de buurt van het ideale om een getal erop te plakken!
* Wel veel discussie!
  + Veel verschillende opinies 🡪 Zeer moeilijk om consensus te krijgen!
    - Een keer de metingen uitrolden, voor vele schalen, bleek dat de scores zeer laag waren!
      * Waren ze allemaal waardeloos 🡪 Mensen verschillen nogal van mening!
    - Hoe duidelijker de materie 🡪 Is het duidelijk wat gemeten moet worden
      * Voor dergelijke zaken werkt het goed!

### Criterium validiteit

* Dit was iets dat we graag deden 🡪 Het levert een cijfer op = “objectief”
* Hoe ga je na of iets kan voorspellen?
  + Kan je doen met correlatie, regressie,
  + Als beide criteria latent zijn moet dat gebeuren door structure equasion modeling
    - Als je het doet met correlaties en geobserveerde scores was er vaak ontgoocheling
      * Het was eerder .40 🡪 Het bleek zo te zijn dat door de band genomen is de validiteit heel laag!
* De slechte correlaties 🡪 Het heeft te maken dat mensen veranderen over de tijd, sommige zaken zijn moeilijk te meten
  + Maar ook een technische reden!
    - Range restriction
      * Je scores, als je 2 concepten hebt A en B 🡪 en je wilt weten of ze correleren
      * Stel dat de correlatie in de populatie .9 is!
    - In de steekproef (100 man)
      * Stel dat de rang een score is van 0-100
        + In de populatie heb je alles ertussenin!
* Er zijn bedrijven die niets doen dan beloven dat ze de beste werknemers kunnen vinden in de zin dat ze na 10 jaar het meeste profeit zullen hebben!

## Validiteit versus betrouwbaarheid

* Behoeden zoals in de media wordt er vaak in de validiteit gesproken, maar ze spreken over betrouwbaarheid en vice versa
* Betrouwbaarheid 🡪 Meetfouten!!
* Validiteit 🡪 Meet hij wat hij moet weten?
* Bv. Weegschaal 🡪 Weegschaal: 60 kg
  + Hij staat er 3keer op 🡪 60 kg 🡪 Heel betrouwbare weegschaal, maar hij is niet valide!!!
  + Want het meet niet wat het moet weten (hij weegt 70)
  + Het weet systematisch fout!

## Conclusie

* SLIDE