

Oefenvragen bij Hoofdstuk 7

1. Wat is meten bij implicatie? Tegen welk type meten wordt 'meten bij implicatie' afgezet?
2. Wat is een populatie-onafhankelijke meting? Laat de klassieke testtheorie dit type meting toe? Waarom (niet)?
3. Waarom wordt gesproken van pseudokansniveauparameter en niet van gisparameter?
4. Bespreek de rol van de itemparameters δ , α en γ in het Rasch-model.
5. Van welk type testgegevens zou het Rasch-model een goede beschrijving kunnen geven?
6. Geef aan wat het belangrijkste verschil is tussen de modellen van Mokken (model van monotone homogeniteit en model van dubbele monotonie) enerzijds en het Rasch-model, het Birnbaum-model en het drie-parameter logistische model anderzijds.
7. Is de meting van personen in het model van monotome homogeniteit populatie-onafhankelijk? Leg uit hoe dat zit.
8. Op welk ander item-responsmodel lijkt het model van dubbele monotonie nog het meest? Licht dit nader toe.
9. Welke schaal wordt in de Mokken-modellen gebruikt om personen te meten?
10. Wat is het belangrijkste verschil tussen de modellen van monotone homogeniteit en dubbele monotonie?
11. Waarin verschilt in de klassieke testtheorie en de item-responstheorie de bepaling van de betrouwbaarheid of de nauwkeurigheid van de meting?
12. Wat is het meetniveau van de persoonsscores in
 - a. het Rasch-model?
 - b. het Birnbaum-model?
 - c. het Mokken-model van monotone homogeniteit?
 - d. de klassieke testtheorie?
13. De θ –schaal is volgens de auteurs van dit boek vooral handig voor allerlei toepassingen van de item-responstheorie, zoals adaptief testen. Leg uit waarom dit zo is.
14. De auteurs hebben daarentegen twijfels over de mogelijkheid om de meting op de θ –schaal te interpreteren in termen van hoeveelheden of afstanden. Licht deze twijfel nader toe.
15. Wel is een interpretatie van de θ –schaal mogelijk in termen van 'odds'.
 - a. Leg uit wat hieronder wordt verstaan.
 - b. Geef eveneens een voorbeeld van de vergelijking van twee personen in termen van de 'odds'.
16. Hoe wordt de informatiefunctie gebruikt bij adaptief testen?
17. Stel, we maken een test voor algemene intelligentie, die over de gehele schaal betrouwbaar moet meten. Hoe ziet voor deze test de doelinformatiefunctie eruit?
18. Noem voor- en nadelen van adaptief testen.
19. Wanneer is een item in twee doelgroepen zuiver? Dus wanneer is er geen sprake van item

bias of differential item functioning?

20. Stel, de items van een test worden in volgorde van oplopende moeilijkheden gepresenteerd. Licht toe hoe men aan een patroon van 0/1 itemscores op de k items van een test zou kunnen zien dat iemand.
 - a. heeft gegist op alle items;
 - b. heeft afgekeken bij de moeilijkste items;
 - c. leed aan testangst in het begin van de test;
 - d. last heeft gehad van concentratieverlies gaandeweg de test.
21. Hoe ziet bij elk van de vier gevallen in de vorige opdracht de persoons-responsfunctie er ongeveer uit?
22. Voer de volgende opdrachten uit/geef antwoord op de vraag
 - a. Geef een beschrijving van het begrip populatie-onafhankelijkheid.
 - b. Zijn metingen volgens de klassieke testtheorie populatie-onafhankelijk?
 - c. Noem enkele toepassingen van tests waarvoor populatie-onafhankelijkheid een handige eigenschap is.
23. Noem zowel voor de klassieke testtheorie als de item-responstheorie enkele voordelen.

Antwoorden Oefenvragen Hoofdstuk 7

1. Indien blijkt dat een model een goede beschrijving geeft, dan volgt daaruit dat de meeteigenschappen van dit model ook in concrete, praktische toepassingen van de test gelden. Dit heet 'meten bij implicatie'. Het wordt afgezet tegen de klassieke testtheorie.
2. Populatie-onafhankelijkheid komt erop neer dat iemands meetwaarde, verkregen met een gemakkelijke test, kan worden vergeleken met de meetwaarde van iemand die een moeilijkere test heeft gemaakt. Bij de klassieke meetmethode is het meten populatie-afhankelijk. Dit betekent dat de betrouwbare score en dus ook de totaalscore afhankelijk zijn van het moeilijkheidsniveau van de gebruikte test.
3. De exacte 'giskans' hangt dus niet alleen af van het aantal alternatieven A , maar ook van inhoudelijke kenmerken van het item. Daarom spreekt men van pseudokansniveau en niet van giskans.
4. δ staat voor de moeilijkheid van de item. γ komt in het Rasch-model niet voor. In dit model geldt dat alle items uit de test hetzelfde discriminerend vermogen (α) hebben.
5. Ruwe scores, geschatte scores, standaarddeviaties, betrouwbaarheidsintervallen.
6. Metingen volgens de modellen van Rasch en Birnbaum, en het drie-parameter logistische model, plaats op een metrische schaal (interval-, verschil- en rationiveau). De modellen van Mokken impliceren ordeningen van meetwaarden.
7. In het model van monotone homogeniteit is de ordening van personen populatie-onafhankelijk. De items zijn dit hierbij niet.
8. Het model van dubbele monotonie lijkt wel wat op het Rasch-model, want de item-responsfuncties onder dat model zijn ook stijgend en snijden elkaar ook niet.
9. Personen worden gemeten op een ordinale schaal. Mensen kunnen namelijk worden geordend op θ met behulp van hun testcores, X .
10. In beide modellen (monotone homogeniteit en dubbele monotonie) wordt verondersteld dat de item-responsfuncties monotoon niet-dalend zijn, maar in het model van dubbele monotonie mogen deze functies elkaar bovendien niet snijden.
11. De item-responstheorie is een verfijning van de klassieke testtheorie omdat het ermee rekening houdt dat de test voor de ene waarde van θ betrouwbaarder is dan voor de andere. De item-responstheorie biedt dus de mogelijkheid een lokale betrouwbaarheid te bepalen. In de klassieke testtheorie wordt de betrouwbaarheid overal op de schaal even groot verondersteld.
- 12a. Verschil-, interval- of ratioschaal
- 12b. Intervalschaal
- 12c. Ordinale schaal
- 12d. Verschilschaal
13. Door de tussentijdse schattingen van de persoonsparameter, θ , worden de volgende items

voor het adaptief testen op basis van iemands score gekozen.

14. Het karakter van θ is niet observeerbaar zoals een itemscore of een ruwe score. Daarentegen is θ een zogenaamde latente variabele die in principe als onbekende uit een vergelijking wordt geschat waarvan de specifieke vorm wordt bepaald door de itemscores die werkelijk op de test zijn behaald. Hierdoor is het voor onderzoekers lastig om er duidelijke uitspraken over te doen.
- 15a. Met odds wordt bedoeld de verhouding van de kans op een positief antwoord en de kans op een negatief antwoord op hetzelfde tem als voor een vaste meetwaarde.
- 15b. Zie formule 7.9. Als bijvoorbeeld, $O_{ig} = 4$, dan betekent dit dat persoon i item g in vier gevallen goed beantwoordt en in één geval fout.
16. Een testinformatiefunctie kan bijvoorbeeld aangeven aan welke kant de schaal betrouwbaarder is. Met een testinformatiefunctie bij adaptief testen kan worden gekeken of de respondent aan de hoge of lage kant van de schaal zit. Zit de persoon aan de hoge kant, dan krijgt hij moeilijkere vragen. Zit hij aan de lage kant, dan krijgt hij makkelijkere vragen.
17. De doelinformatiefunctie moet aangeven dat de test op alle mogelijke punten even betrouwbaar is. De functie moet dus ongeveer één rechte lijn hebben.
18. Voordelen:
 1. Per respondent wordt een nauwkeurige meting verricht.
 2. Respondenten worden geconfronteerd met tests die op hun niveau zijn afgestemd. Dit voorkomt concentratieverlies of frustratie.
 3. De adaptieve testprocedure is geautomatiseerd en daardoor objectief.
 4. De testtijd is relatief kort, waardoor meer respondenten in dezelfde tijd kunnen worden onderzocht of dezelfde hoeveelheid tijd kan worden besteed aan meer tests.
 5. Door de volledige automatisering kan een snelle terugkoppeling van de resultaten aan de respondent plaatsvinden.
 6. Testprestaties die zijn verkregen door middel van verschillende tests kunnen met elkaar worden vergeleken. Dit is het gevolg van de eigenschap van populatie-onafhankelijk meten, die het mogelijk maakt om θ -waarden die zijn verkregen met verschillende tests via equivalering op dezelfde schaal af te beelden.

Nadelen:

1. Item-responsmodellen (die nodig zijn om adaptief te kunnen testen) leggen vele restricties op aan de testgegevens.
 2. Er zijn hoge kosten bij de ontwikkeling van een itembank en een daaraan gekoppeld adaptief testsysteem.
 3. Theorieën omtrent psychologische eigenschappen zijn vaak onvoldoende scherp en eenduidig geformuleerd, en het gevolg is dat operationalisering hieruit niet moeiteloos volgen.
19. Als de testprestaties van personen uit verschillende populaties met elkaar worden

vergeleken, kan dit verschil in betrouwbaarheid of validiteit maken dat er sprake is van bevoordeling of benadeling van de een ten opzichte van de ander. Het onderzoek dat zich op itemniveau met deze problematiek bezighoudt, wordt onderzoek naar vraagzuiverheid genoemd, ook wel 'bias' en 'differential item functioning'. Als een onderzoek zuiver is, wil dit dus zeggen dat de verschillende populaties ten opzichte van elkaar geen bevoordeling of benadeling hebben.

20a. De goede antwoorden zullen verspreid zijn over de gehele test. De goede antwoorden zijn hierbij onafhankelijk van de moeilijkheid van de vragen omdat deze persoon toch geen enkel antwoord wist.

20b. De moeilijkere vragen zullen beter gemaakt zijn dan de makkelijke vragen.

20c. De antwoorden op de vragen zullen in het begin slechter zijn geweest dan naarmate de test vorderde.

20d. De antwoorden zullen in het begin van de test erg goed zijn geweest, naarmate de test vorderde werden de antwoorden steeds slechter.

21a. De persoons-responsfunctie zal laag zijn, omdat de persoon geen enkel antwoord zeker wist.

21b. De persoons-responsfunctie zal hoog zijn, omdat de persoon uit zichzelf al een aantal antwoorden wist en de moeilijkste vragen heeft afgekeken.

21c. De persoons-responsfunctie zal in het begin lager zijn geweest dan tegen het einde aan.

21d. De persoons-responsfunctie zal in het begin hoog zijn geweest, dit zal gaandeweg steeds meer afnemen.

22a. Populatie-onafhankelijkheid komt erop neer dat iemands meetwaarde, verkregen met een gemakkelijke test, kan worden vergeleken met de meetwaarde van iemand die een moeilijkere test heeft gemaakt.

22b. Nee, bij de klassieke meetmethode is het meten populatie-afhankelijk. Dit betekent dat de betrouwbare score en dus ook de totaalscore afhankelijk zijn van het moeilijkheidsniveau van de gebruikte test.

22c. Je kunt nu allerlei meetwaarden met elkaar vergelijken, ongeacht de test die hierbij gebruikt is.

23. Voordelen van de item-responstheorie:

1. Item-responsmodellen kunnen als nulhypothese op de testgegevens worden getoetst.
2. Personen en items worden populatie-onafhankelijk gemeten. Dit maakt het mogelijk om de testprestaties van personen die met verschillende tests voor dezelfde eigenschap zijn gemeten op dezelfde schaal af te beelden.
3. De betrouwbaarheid van de meting varieert over de schaal, waardoor aan het licht komt dat de ene persoon met de test nauwkeuriger kan worden gemeten dan de andere.

Voordelen van de klassieke testtheorie:

1. De klassieke testtheorie is eenvoudiger dan de item-responstheorie en daardoor wellicht beter toegankelijk.

2. Weliswaar kan niet direct worden getoetst of de klassieke testtheorie bij de gegevens past, toch is het maar de vraag of de validiteit van tests die volgens deze methode zijn geconstrueerd zoveel minder is dan die van tests geconstrueerd volgens een item-responsmodel.
3. De klassieke testtheorie 'werkt' bij de testconstructie bijna altijd, doordat van elke verzameling items de item-restcorrelaties (dit zijn de correlaties van een item met de totaalscore op de overige $k - 1$ items) kunnen worden gebruikt om eventueel items weg te laten en van de resterende verzameling de betrouwbaarheid kan worden bepaald.