

Oefenvragen bij Hoofdstuk 5

1. Welke rol van de klinisch-intuïtieve interpretatie spelen in de verwerking van testgegevens?
2. Wat is een 'checklist'?
3. Noem de drie manieren waarop de reacties op geprecodeerde items gescoord kunnen worden.
4. De giscorrectieformule [5.1] wordt uitgelegd voor vierkeuze-items. Doe deze uitleg eens helemaal over voor driekeuze-items.
5. Leg uit wat het verschil is tussen de twee toestanden waarin personen zich kunnen bevinden als zij gissen naar het goede antwoord en de drie typen antwoorden die dit kan opleveren.
6. Wat wordt verstaan onder partiële kennis?
7. Hoe kan partiële kennis de giscorrectie in formule [5.1] beïnvloeden?
8.
 - a. Met welke factor neemt de variantie van de gecorrigeerde testcores gebaseerd op driekeuze-items toe ten opzichte van de ongecorrigeerde testscore?
 - b. Hoe zit dit bij vierkeuze-items? En bij vijfkeuze-items?
 - c. Welke trend valt te zien in de resultaten bij de vragen a en b?
9. Wat is het bezwaar tegen het toenemen van de variantie van de testscore als gevolg van giscorrectie?
10. Leg kort uit waarom de auteurs uiteindelijk van mening zijn dat giscorrectie niet zo'n goed idee is.
11. Wat is 'on-line testing'? Wat is de zwakke plek van deze methode?
12. Welke conclusie trekken de auteurs uit het onderzoek dat is gedaan naar de invloed van de wijze van testen, via 'paper and pencil' of de computer, op de testprestatie?
13. Wat is de kern van adaptief testen?
14. Hoe verschilt adaptief testen van testen met behulp van de klassieke standaardtest?
15. Wat wordt bedoeld als wordt gezegd dat vanuit de geteste persoon gezien de kans op het goed maken van een item 0.5 is?
16. Leg uit dat het bij adaptief testen niet 'oneerlijk' is dat de ene persoon veel gemakkelijker items krijgt voorgelegd dan de andere.
17. Wanneer stopt de computer bij adaptief testen met het aanbieden van items?
18. Wat is nu nog de belangrijkste hindernis bij het testen via het internet?
19. Wat is een ruwe score?
20. Wat is het verschil tussen een bewerkte score en een normscore?
21. Zijn voor een zinvol testgebruik normen altijd noodzakelijk? Licht het antwoord toe.
22. In welk type testtoepassing kan men volstaan met ruwe scores?
23. Bespreek de bezwaren tegen het IQ-begrip.
24. Wat zijn percentielscores?
25. Stel, de scores op een test zijn normaal verdeeld met een gemiddelde gelijk aan 10 en een variantie gelijk aan 9. Bereken met behulp van uw statistiekboek de percentielscores van iemand met een testscore van 12 en iemand anders met een testscore van 6.
26. Leg uit hoe het komt dat een verschil tussen percentielscores die corresponderen met testcores rond het midden van de normaalverdeling anders geïnterpreteerd moet worden dan een even groot verschil dat verder van het midden aflight.

27. Zijn z-scores normaal verdeeld? Verklaar het antwoord.
28. Welk percentage van de normaal verdeling ligt tussen x en $(x + 3S_x)$?
29. Bereken wat het effect is van onevenredig veel gemakkelijke items op de verdeling van de ruwe scores op een intelligentietest.
30. Wat is een nadeel van genormaliseerde standaardscores?

Antwoorden Oefenvragen Hoofdstuk 5

1. Bij de klinisch-intuïtieve benadering dient men zich er bewust van te zijn dat de subjectieve impressies en ideeën geen waarheidsgarantie bezitten, ook al wekken ze bij de psycholoog vaak een gevoel van juistheid. Het is dus van belang om bij het verwerken van de resultaten hier rekening mee te houden.
2. Bij een checklist kun je op een formulier aankruisen wat wel en niet van belang is. Bij gedrag kun je dan bijvoorbeeld aangeven of dit wel of niet voorkomt bij de onderzochte en in welke mate.
3. Handscoreng, zelfscoring, machinale scoring.
4. We vinden X_c (de gecorrigeerde testscore) als volgt. Als $A = 3$, en iemand heeft 6 items fout beantwoord, dan zijn dit dus alle fout gegiste items. De kans op fout gissen is $2/3$. Dan weten we ook dat naar verwachting $1/3$ deel van de items waarop is gegist goed is beantwoord. Dat zijn dan 3 items, ofwel $1/2$ deel van 6. In het algemene geval is dit $1/(A-1)$ deel van het aantal fout beantwoorde items. De formule voor de berekening van X_c verkrijgen we dan als volgt: we nemen het totale aantal goed, X , en trekken daar het deel vanaf dat door gissen goed was; dit is het aantal fout, $k - X$, gedeeld door $A - 1$:
$$X_c = X - \frac{k - X}{A - 1}$$
5. Personen kunnen zich in de volgende twee toestanden bevinden: hij weet met volledige zekerheid wat het goede antwoord is en geeft dat antwoord, of hij heeft geen idee en gist blind met succeskans $1/A$. De drie typen antwoorden die dit op kan leveren zijn: goede antwoorden die het gevolg zijn van kennis, goede antwoorden die het gevolg zijn van blind gissen en foute antwoorden die het gevolg zijn van gissen.
6. Dit houdt in dat de proefpersoon wel iets over de vraag en het antwoord weet, maar niet genoeg om met zekerheid het goede antwoord te geven.
7. Door partiële kennis kan een reële giskans groter of kleiner zijn dan de theoretische, blinde giskans. Een voorbeeld is dat respondenten, ook al weten ze het juiste antwoord niet, sommige afleiders toch als onjuist kunnen wegstrepen doordat ze wel weten dat deze niet juist kunnen zijn. In dit geval geeft formule [5.1] een ondercorrectie: omdat bij sommige items de giskans groter is dan $1/A$, had X_c eigenlijk lager moeten uitvallen.
8.
 - a. 2.25
 - b. 1.78, 1.56
 - c. De factor wordt steeds kleiner.
9. Het ongewenste effect hiervan op andere berekeningen wordt, bijvoorbeeld, zichtbaar als men de gecorrigeerde testscore optelt bij de testcores op andere tests, zoals wel gebeurt bij de deelttests van intelligentietestbatterijen, om een schatting van de algemene intelligentie te maken. De test met scores X_c krijgen dan een groter gewicht, zonder dat eigenschappen die dit zouden kunnen rechtvaardigen, zoals betrouwbaarheid en de validiteit van de meting, zijn toegenomen.
10. De betrouwbaarheid en de validiteit van de testscore bij de giscorrectie niet noemenswaardig toenemen boven het niveau van dat van de ongecorrigeerde testscore (X).

11. 'On-line testing': de leerling of student bepaalt zelf het moment waarop hij de studiestof naar zijn idee voldoende beheerst en vraagt de computer om een toets. De computer selecteert uit de itembank een studietoets, presenteert deze aan de student, codeert de reacties, levert een eindbeoordeling aan de student en slaat de toetsgegevens centraal op. Een gevaar van deze toetsen is de beveiliging. Leerlingen kunnen bijvoorbeeld de vragen overschrijven en doorvertellen, ook kunnen ze de toets voor een ander maken.
12. Bij computergestuurde tests is het onmogelijk om items over te slaan zonder een toets in te drukken. Dit kan leiden tot enigszins andere verdelingen van itemscores dan bij conventionele tests (paper and pencil). Verder is het vaak niet mogelijk om terug te gaan naar eerder beantwoorde items of de beantwoording van een item nog even uit te stellen. Veel invloed op de testprestatie lijken deze factoren overigens niet te hebben. Ook verschillen de presentatie en vormgeving van conventioneel gepresenteerde items nogal van de presentatie op het beeldscherm, maar ook hier bleek dat de testvorm geen belangrijke invloed had op de testprestatie.
13. Adaptief testen komt erop neer dat iedere respondent een test krijgt voorgelegd die op zijn of haar niveau is toegesneden. Zo krijg je een beter beeld van het niveau dat de respondent heeft.
14. Bij de klassieke wijze van testen houdt in dat een standaardtest, voor iedereen gelijk, wordt voorgelegd aan de respondenten. Het verschil is dus dat bij adaptief testen de test is aangepast aan de respondent, bij de klassieke wijze is dit niet het geval.
15. Dat de item voor de respondent dus een middelmatige moeilijkheid heeft, want vanuit de persoon gezien is de kans op een goed antwoord even groot als de kans op een fout antwoord.
16. Omdat de metingen gecorrigeerd kunnen worden voor het moeilijkheidsniveau van de test. Dit is de reden waarom de meetwaarden van verschillende personen kunnen worden vergeleken ondanks het feit dat zij tests van uiteenlopend moeilijkheidsniveau hebben gemaakt.
17. Totdat een meetwaarde is geschat die een bevredigende nauwkeurigheid heeft.
18. De beveiliging van de gegevens via internet zullen nog moeten worden verbeterd.
19. Een ruwe score is de testuitslag die wordt uitgedrukt in een basisscore. Deze score heeft dus nog geen verdere bewerking ondergaan. Voorbeelden zijn: het aantal goede antwoorden, het aantal foute antwoorden etc.
20. Bewerkte scores zijn bewerkingen die onafhankelijk van de prestaties van anderen zijn. Een normscore houdt de vergelijking in van de score met de scoreverdeling van de referentiepopulatie.
21. Nee, die zijn niet altijd noodzakelijk. Je kunt ook genoegen nemen met een eenvoudiger vorm van bewerkte scores. Een voorbeeld: soms hoeft men slechts de plaats in de rangschikking binnen de onderzochte groep te weten, in het geval dat de drie besten van een groep van sollicitanten moeten worden uitgezocht.
22. Als men bijvoorbeeld voor onderzoeksdoeleinden geïnteresseerd is in de samenhang tussen testscores en criteriumscores. Hierbij is een omzetting in bewerkte scores niet noodzakelijk.
23.
 - De formule [5.5] suggereert dat ML (mentale leeftijd) en CL (chronologische leeftijd) vergelijkbare grootheden zijn, dit is in werkelijkheid niet het geval.

- Lange tijd vonden psychologen het IQ-begrip belangrijk omdat men ervan uitging dat het constant was. Het IQ vertoont echter bij de ontwikkeling van de intelligentie nogal wat schommelingen.
 - In normale gevallen blijft men bij het berekenen van het IQ boven de hoogste leeftijd waar beneden de test nog discrimineert, steeds delen door deze topleeftijd. Dit is in de regel 15, 16, of 17 jaar. Op hogere leeftijden lopen de prestaties in vele intellectuele functies echter terug. Het is dus duidelijk dat oudere mensen in het nadeel zijn doordat in hun IQ-ratio de noemer gelijk en maximaal blijft.
 - Niet veel tests waarbij IQ-scores gebruikt worden voldoen aan de eis van een evenredig toenemende spreiding bij hogere leeftijden.
24. Een percentielscore geeft aan hoeveel procent van de andere respondenten je achter je laat. Een voorbeeld: 80 percentiel betekent dat je 80% van de andere respondenten achter je laat, je presteert dus beter dan de andere 80%.
 25. Testscore 12 correspondeert bij deze gegevens met percentielscore 75, en testscore 6 met percentielscore 9 (beide percentielscores afgerond).
 26. Bij percentielscores wordt er rekening gehouden met wat mensen behaald hebben en wat ze zullen halen. Zie figuur 5.6 op bladzijde 181. Veel mensen zullen een gemiddelde score behalen (hier liggen de percentielen dus dicht bij elkaar). Er zullen minder mensen zijn die de test heel slecht of heel goed hebben gemaakt (hier liggen de percentielen dus verder uit elkaar).
 27. Nee, dit is niet het geval. Zie voor verduidelijking figuur 5.7 op pagina 183.
 28. 49.87 procent.
 29. Veel mensen zullen erg veel goede antwoorden hebben gegeven, wat waarschijnlijk geen goede afspiegeling is van de werkelijkheid.
 30. Een nadeel is dat sommige van de oorspronkelijke score-eenheden worden 'uitgerekt' en andere 'ineengedrukt' teneinde een verdeling te krijgen die normaal verdeeld is.