

[Dit is een heel onvolledige transcriptie. De eerste ca 20 blz. waren te vaag om te kunnen scannen voor OCR, en alle figuren zijn weggevallen. Het geeft wel een beeld van mijn eerste pogingen een besliskundige benadering vorm te geven. Het is geloof ik een beetje klunzig allemaal.]

Het nemen van beslissingen: het eenvoudige dichotome geval.

[2002: De eerste 21 bladzijden zijn te licht om te scannen, ik parafraseer ze kort.]

p. 1 Het stuk licht een bepaald type beslissing, het dichotome geval, door op een besliskundige wijze (Raiffa, 1968; Lindley, 1971): diagnostiseren, selecteren, klassificeren.

p. 2 casus: 'geschikte' van 'ongeschikte' studenten scheiden, veronderstellend dat de dichotomie op zich zinvol zou zijn, quod non.

p. 3

Schetst de uitgangssituatie met twee verdelingen over de doelvariabele of discriminantfunctie of what not voor de beide onderscheiden groepen. Grijpt dus terug op klassificeren zoals behandeld in mijn 1968 werkstuk (groep onderwijsresearch, TH Eindhoven), en de figuur zoals in het NRC-artikel gebruikt, en hierboven afgebeeld.

p. 4. Introductie term 'cesuur', en vier mogelijke uitkomsten voor iedere individuele beslissing (geschikt/ongeschikt en conform/onjuist beoordeeld).

a) geschikt maar ongeschikt geoordeeld: verlies A

b) geschikt en geschikt geoordeeld: verlies B = 0

c) ongeschikt en ongeschikt geoordeeld: verlies C = 0

a) ongeschikt maar geschikt geoordeeld: verlies D

p. 5 Over de term 'verlies.' Lindley, 1971, p. 125. Schaal voor verlies bepalen, van 0 tot 1.

p. 6.

blijkt 'geschikt' blijkt 'ongeschikt'

ingedeeld bij
'geschikten'

0

$D = 1$

ingedeeld bij
'ongeschikten'

$A = 2/3$

0

TABEL 1. Verliezen van de mogelijke uitkomsten.

p. 7. Neemt Rorer c.s. 1966 erbij.

FIGUUR 4. De proportie leerlingen binnen de subgroep met een bepaalde test score x die in werkelijkheid tot de groep 'geschikten' behoren.

p. 8. Gevraagde cesuur is punt waarvoor geldt $pA = (1-p)D$. Hier is p de proportie geschikten gegeven toetsscore x (of, als ik me vergis, net omgekeerd). Het plotje op p. 4 geeft de constructie van die proportie over horizontaal de toetsscore.

p. 9. Een illustratie voor selectie bij numerus fixus in het w.o.

p. 10. Gaat over definitie van geschiktheid.

p. 11. Data ontleend aan CBS-statistieken, geslaagd voor kandidaats, opgesplitst naar 'lotingsklasse'.

p. 12. Dat levert per eindexamencategorie de proportie 'geschikten' op, die stijgt van 0,7 via 0,82, 0,85, 0,9 tot 0,96. Op deze blz. geplot, evenals de beide frequentieverdelingen (voor geschikten resp. ongeschikten), verdelingen die elkaar dus niet snijden.

p. 13. geeft uit de losse pols voor- en nadelen van de beide soorten foute beslissing in dit casus, de faculteit als actor.

p. 14. idem, de student als actor.

p. 15. voor een student bepalen van de selectieparameter. Interessant geval waar de student een cesuur bepaalt ...

p. 16. De staatssecretaris kan ook de verhouding van verliezen bepalen, vanuit zijn standpunt bezien.

p. 17. Een numerus fixus is geen vrije situatie, de toegepaste cesuur zal dus doorgaans niet de optimale zijn voor het geval er geen nf zou zijn vastgesteld maar er wel geselecteerd zou worden.

p. 18. Hoe zou een en ander komen te liggen wanneer de kandidaten met de hoogste cijfers worden toegelaten. De averechtse toepassing resulteert er voor het getallenvoorbeeld in het casus in dat deze toelatingsprocedure ten onrechte toelaten 5 keer zo erg vindt als ten onrechte afwijzen. In 2002 zou ik zeggen dat dat nog niet zo gek is geschoten, dus aan deze analyse valt geen argument voor gewogen loting te ontleen. Heb ik ook nooit gewild, dacht ik.

p. 19. beschouwing over 'rendement' van verschillende mogelijkheden voor een toelatingsprocedure bij nf . Om er iets zinvols over te zeggen moeten verschillende mogelijke uitkomsten gekwantificeerd worden, bijv. in guldens. Het volstaat niet om alleen naar aantallen 'geslaagden' te kijken. Ik moet dit nog eens terugzien, want ik mis iets in die argumentatie.

p. 20. Cesuurbepaling bij hantering van drempelverlies: meer algemeen. Behandelt het probleem in termen van een criteriumvariabele in de toekomst, dat is dus geheel iets anders dan in de literatuur over criteriumgerefeerd toetsen waarbij het 'criterium' de ware beheersing is.

p. 21. Uitgangssituatie is dan weer de bekende tabel van vier mogelijke uitkomsten van de individuele beslissing: b) en c) terechte beslissingen, a) ten onrechte afgewezen, d) ten onrechte toegelaten. Verlies (of nut) corresponderend B C A of D.

Pikant is de stelling in het stuk 'dat het altijd mogelijk is om twee van de uitkomsten op 0 te stellen, en dat daar een bepaalde vrije keus in is.' Ik heb daar in de marge bijgeschreven dat dat ONJUIST is, en dat dat ook de reden is dat het stuk, dat tot blz. 31 was gekomen, niet is afgerond. Niet alleen pikant, ik moet nagaan hoe dat precies zit, of hier een logische fout is gemaakt, en of die in later werk nog zijn sporen heeft

nagelaten (in de TOR-artikelen?). Dus even in vet markeren deze passage.

[Blz. 22 t/m 31 is gescand, dus in de tekst vanaf dit punt]

[het nemen van beslissingen p.22]

De proportie personen die 'geschikt' is, is dus $a + b$,
de proportie personen die 'ongeschikt' is, is $c + d$.

Dit zijn proporties zoals ze gelden voor de totale groep personen. Dezelfde proporties kunnen we nu ook uitrekenen voor verschillende deelgroepen, waarbij we in het bijzonder geïnteresseerd zijn in de proportie 'geschikten' voor iedere test score x afzonderlijk. Noem p de proportie personen binnen de groep met een gegeven testscore x (of waarvan de test score binnen een bepaald interval ligt) die 'geschikt' zijn; en evenzo noemen we $1-p$ de proportie bij een bepaalde test score die 'ongeschikt' zijn.

De veronderstelling die we voor onze eenvoudige oplossing van het cesuurprobleem nodig hebben, is dat de proportie p groter wordt bij toenemende test score x , een situatie waarvan in figuur 4 een voorbeeld werd gegeven. Volledigheidshalve zij er nu al vast op gewezen dat er situaties zijn waarin deze veronderstelling niet opgaat, situaties waarin méér dan één cesuur gevonden wordt (zie Rorer c.s. 1966a). Het laatste is bij veel klassificatieproblemen het geval, en wordt in het bijzonder vervelend wanneer het aantal categorieën waarin geklassificeerd moet worden meer dan zeg drie of vier bedraagt (zie Wilbrink 1968).

Laten we eens analyseren wat er gebeurt wanneer we een groep personen die een bepaalde test score x hebben, in haar geheel indelen bij de 'geschikten', dan wel bij de 'ongeschikten'. We spreken af dat personen met een zelfde test score altijd op dezelfde wijze behandeld zullen worden, d.w.z. dat ze of allen bij de groep 'geschikten' gerekend worden, of allen tot de groep 'ongeschikten' maar niet sommigen tot de ene, en anderen tot de andere groep.

Onze deelgroep bestaat voor een deel p uit 'geschikte' personen. Het verlies wanneer een geschikte persoon bij de 'ongeschikten' wordt ingedeeld is volgens afspraak A. Beschouwen we deze deelgroep als 'ongeschikt', dan boeken we wat de daarin voorkomende 'geschikte' personen betreft een relatief verlies van p maal A. Op dezelfde wijze vinden we dat het verlies van deze beslissing voor wat betreft de in deze

deelgroep voorkomende 'ongeschikte' personen gelijk is aan 1p maal C (als men C liever 'winst' dan 'verlies' noemt, dan is daar geen enkel bezwaar tegen). Totaal verlies bij de beslissing om deze deelgroep in te delen bij de 'ongeschikten' is dan:

$$p A + (1p) C.$$

Op dezelfde wijze kunnen we 'berekenen' wat het verlies is wanneer deze deelgroep in haar geheel wordt gerekend te behoren tot de 'geschikten':

$$p B + (1p) R.$$

[het nemen van beslissingen 23]

Het is duidelijk waar dat heen gaat: of we de deelgroep in haar geheel als 'geschikt' of 'ongeschikt' beschouwen hangt er van af welke van de twee verliezen het grootst is.

De deelgroep met eenzelfde testscore x wordt in haar geheel beschouwd als 'geschikt' wanneer

$$pxB + (1p)xD \text{ groter of gelijk is aan } pxA + (1p)xC. \quad (1)$$

De deelgroep met eenzelfde testscore x wordt in haar geheel beschouwd als 'ongeschikt' wanneer

$$p B + (1p) D \text{ kleiner is dan } p A + (1p) C. \quad (2)$$

Wanneer de grootheden A, B, C, en D niet als 'verlies' maar als 'winst' gewaardeerd zijn, dan zijn de relaties (1) en (2) juist omgekeerd.

Uit het bovenstaande volgt dat de optimale cesuur daar komt te liggen waar

$$p B + (1p) D = p A + (1p) C. \quad (3)$$

De cesuur komt te liggen bij die test score x waarvoor geldt dat het ons onverschillig is of de deelgroep personen met deze test score in haar geheel gerekend wordt tot de 'geschikten' of tot de 'ongeschikten'. Het verlies volgend uit de ene beslissing is gelijk aan het verlies volgend uit

de andere beslissing.

We zijn nu in een positie gekomen van waaruit we tot een concrete cesuurbepaling kunnen komen, en wel door 'p' op te lossen uit (3). Immers, als we p hebben bepaald, dat is de proportie 'geschikten' die zich in de groep personen bevindt met een test score waarvoor (3) geldt, dan kunnen we op grond van de empirische gegevens die eerder verzameld zijn die test score x vinden. Hebben we deze empirische gegevens uitgezet op een manier zoals in figuur 4 geïllustreerd, dan kan volstaan worden met het trekken van een horizontale lijn vanuit het punt op de p-as dat we met behulp van (3) berekend hebben, en uit het snijpunt van deze lijn met de curve een lijn te laten zakken op de test score as: het zo gevonden punt wordt de gezochte cesuur. Een en ander is in het eerste hoofdstuk al geïllustreerd.

Zolang de waarden A, B, C, en D nog niet bepaald zijn, kan de berekening van p nog niet uitgevoerd worden. In het volgende zullen we laten zien welke drastische vereenvoudigingen mogelijk zijn voor het bepalen van de waarden van deze A, B, C en D.

[het nemen van beslissingen 24]

Uit (3) kunnen we p berekenen: (het vermenigvuldigingsteken laten we voortaan achterwege):

$$pB + (1-p)D = pA + (1-p)C$$
$$pB - pD - pA + pC = C - D$$

$$p = (C - D) / (C - D + B - A) \quad (4)$$

De gevonden uitdrukking heeft een aantal bijzondere eigenschappen, die handig uitgebuit kunnen worden om op eenvoudige wijze tot een goede waardebepaling voor de verschillende uitkomsten te komen. Daartoe zullen we (4) anders gaan schrijven. Boven en beneden de streep delen door (B - A) geeft:

$$p = \frac{(C/D)/(B/A)}{(C/D)/(B/A) + 1} \quad (5)$$

Noem

$$L = (CD)/(BA). \quad (6)$$

We kunnen p nu schrijven als

$$p = L / (L + 1) . \quad (7)$$

We zien uit (7) dat de waarde van p niet verandert zolang L niet verandert. Een prettige eigenschap van L (vergelijking (6)) is dat haar waarde gelijk blijft onder een groot aantal mogelijke veranderingen in de waarden van A, B, C, en D. Wanneer zowel van C als van D een konstante k afgetrokken wordt, blijft L gelijk. Evenzo kunnen we van B en A tegelijk eenzelfde konstante 1 aftrekken (of er bij optellen) zonder de waarde van L te veranderen. Tenslotte kunnen alle vier waarden A, B, C, en D met eenzelfde constante m vermenigvuldigd worden zonder L te veranderen. Laten we eens zien in hoeverre dergelijke manipulaties ons probleem kunnen helpen oplossen.

Trek van C zowel als D de constante $k=C$ af, en noem $D_k = D_C = D'$:

$$L = (CD)/(BA) = (0D')/(BA) = D'/(BA). \quad (8)$$

Trek van B zowel als A de konstante $1=B$ af, en noem $A_1 = A_B = A'$:

$$L = D'/(BA) = D'/(0-A') = D'/A' = D'/A'. \quad (9)$$

Laten we de accenten nu weg, dan hebben we:

$$L = D/A \text{ in herschaalde waarden voor A en D.} \quad (10)$$

Uit (10) kunnen we concluderen dat inderdaad voor het cesuurprobleem het voldoende is om het verlies van iemand ten onrechte als 'geschikt' beschouwen relatief af te wegen tegen het verlies iemand ten onrechte als 'ongeschikt' te beschouwen.

[het nemen van beslissingen 25]

Tenslotte, maar dat is verder niet van direct praktisch belang, kunnen we D en A in (10) nog vermenigvuldigen met een geschikte konstante m zodanig dat mA (of mD, al naar gelang A dan wel D het grootste verlies of

de kleinste winst is) gelijk is aan 1. Stel dat A het grootste verlies is, dan vermenigvuldigen we A en D met de konstante m zodanig dat $mA = 1$, en noemen we $mD = D''$:

$$L = D/A = MD/mA = D'/1 = D.$$

Laten we wederom het accent weg, dan hebben we simpel

$$L = D \text{ (in wederom herschaalde waarde voor D),} \quad (11)$$

zodat

$$p = D/(D+I) \text{ (uit vergelijking M).} \quad (12)$$

Op dezelfde wijze als we hier hebben laten zien dat het cesuurprobleem opgelost kan worden door het tegen elkaar afwegen van de verliezen van de 'onjuiste uitkomsten', omdat $L = D/A$ gekozen kan worden, kunnen we ook laten zien dat het cesuurprobleem op te lossen is door andere mogelijke uitkomsten tegen elkaar af te wegen. Immers, we kunnen evengoed $D=0$ kiezen, i.p.v. $C=0$; of $A=0$, i.p.v. $B=0$; of zowel $D=0$ als $A=0$, i.p.v. $C=0$ en $B=0$. Er zullen zeker selectie of diagnostiek problemen zijn waarin het makkelijker is om bijv. B tegen D af te wegen dan A tegen D. Als we dat doen, dan vinden we

$$L = D/B. \quad (13)$$

D is het verlies horend bij ten onrechte als 'geschikt' beschouwde personen, en is een negatieve waarde t.o.v. de 'winst' B behorend bij terecht toegelatenen; door het minteken in (13) kunnen we ofwel de winst negatief nemen, zodat twee 'verliezen' met elkaar vergeleken kunnen worden, ofwel het verlies kan negatief genomen worden, zodat twee 'winsten' tegen elkaar vergeleken worden.

Eenvoudiger is misschien om de verlies en winst waarde op een schaal voor absolute waarden te zetten, door $L = |D|/|B|$ te kiezen. Zie de volgende figuur.

FIGUUR . Vergelijking van verlies D en winst B.

[noot 2002: op de verticale schaal staan $|B|$ en $|D|$, het zijn projecties van resp. B en D op de horizontale schaal rechts en links.]

Vergelijking met andere benaderingen van dit cesuur probleem.

De noodzakelijke voorwaarde voor iedere rationele methode van cesuurbepaling is dat een onderzoek gedaan is naar de validiteit van de te gebruiken test of andere gegevens voor de beslissingen die men moet nemen. Op technieken en rationale van dergelijk valideringsonderzoek kan in het kader van dit rapport niet ingegaan worden; verwezen wordt naar bijv. Cronbach (1971). Resultaten van valideringsonderzoek worden vaak gegeven in de vorm van de correlatie tussen de test (op basis waarvan voorspellingen gedaan worden) en de criteriumvariabele (die het te voorspellen gedrag 'meet'; bijv. al dan niet afstuderen). Wanneer we zoeken naar een optimale cesuurbepaling dan gaat het vrijwel nooit om het zo goed mogelijk kunnen voorspellen van een criterium score (bijv. studieprestaties over een aantal jaren), maar gaat het er om dat de winst of het verlies verbonden aan de mogelijke uitkomsten voorspeld worden. Cronbach & Gleser (1965) zeggen het zo, dat het er niet om gaat een bepaald criterium te voorspellen, maar het geëvalueerde criterium te kunnen voorspellen. Dit is equivalent aan het specificeren van een verliesfunctie (zie Novick & Jackson 1974), waarvoor t.a.v. de hier gepresenteerde cesuurbepaling gekozen is voor drempelverlies. Als we drempelverlies als uitgangspunt nemen, dan kunnen de resultaten van het valideringsonderzoek altijd weergegeven worden in de vorm van twee afzonderlijke frequentieverdelingen: één voor de groep personen die op de criterium variabele beneden het gestelde minimum scoorde, en één voor de groep personen die een score op de criterium variabele bereikte die gelijk was aan het minimum of er boven lag. Voorbeelden daarvan werden gegeven in figuur 3 en figuur 5. Dergelijke frequentieverdelingen zijn een directe en onveranderde weergave van de gegevens verkregen in het onderzoek. Hoewel dergelijke frequentieverdelingen er vaak onregelmatig uitzien, is het in de meeste gevallen aannemelijk te veronderstellen dat er bepaalde theoretische verdelingen aan ten grondslag liggen (normaal verdeling, binomiaal verdeling, etc.); dan kunnen ruwe frequentieverdelingen 'glad getrokken' worden door ze te vervangen door een dergelijke theoretische verdeling. Of ook: als de frequentieverdeling gebaseerd is op een steekproef uit een of andere grotere groep (populatie) kan men de verdeling in de populatie vaak beschouwen als een bepaalde theoretische verdeling. Zie voor een grondige behandeling Kendall & Stuart 1969.

[het nemen van beslissingen 27]

Welnu, in de literatuur gegeven procedures voor het bereiken van een optimale cesuur worden vrijwel zonder uitzondering gegeven in termen van veronderstelde theoretische verdelingen. Daarmee bereikt men wel elegante oplossingen voor allerlei problemen die samenhangen met de te leggen cesuur, maar vereenvoudigt er het cesuurprobleem zelf in het geheel niet mee. Integendeel: de in principe zeer eenvoudige oplossing zoals in dit rapport gegeven, wordt door anderen vaak langs ingewikkelde en slechts voor insiders te volgen weg bereikt. Wanneer men bovendien vertrekt vanuit niet geheel dezelfde veronderstellingen wordt het erg lastig de verschillende gegeven oplossingen te herkennen als feitelijk identiek (voorbeeld daarvan de oplossingen onafhankelijk van elkaar door Alf & Dorfman 1967 enerzijds, en Rorer c.s. 1966 a,b anderzijds gegeven; beide zullen besproken worden).

Cesurbepaling kan in principe gewoon gebeuren aan de hand van de op basis van validerings onderzoek verkregen frequentie verdelingen; het spreekt vanzelf dat daarbij de nodige voorzichtigheid in acht genomen moet worden: dergelijke gegevens kunnen niet zonder meer op facevalue geaccepteerd worden, evenmin als bij de benaderingen die uitgaan van bepaalde veronderstelde theoretische verdelingen men niet op alle mogelijke manieren de houdbaarheid van dergelijke veronderstellingen zal moeten blijven toetsen. Geen enkel procedurevoorschrift kan ooit toegepast worden zonder een goed inzicht in de aard van de onderhavige beslissingssituatie, en de statistische zwakheden van de gegevens waarop men zich moet baseren.

Rorer, HoffmanI, LaForge, enHsieh. 1966a,b. Utilities as baserate multipliers in the determination of optimum cutting scores for the discrimination of groups of unequal size and variance. Journal of Applied Psychology, volume 50.

In het eerste van twee artikelen besteden deze auteurs aandacht aan de vaak wat al te zorgeloos gehanteerde veronderstelling dat twee van elkaar te onderscheiden groepen even groot zouden zijn, en hun scores niet in varLantie van elkaar verschillen. In vrijwel alle praktische situaties waarin twee groepen van elkaar onderscheiden moeten worden, zal de ene groep groter zijn dan de andere; in niet zelden voorkomende gevallen zal de variantie van de scores voor de ene groep ook aanzienlijk verschillen van die in de andere groep. De consequenties kunnen zeer drastisch zijn: wanneer twee groepen gemiddeld behoorlijk van elkaar

verschillen, kan desondanks een goede klassificatie procedure, dus ook een goede cesuur, vaak niet gevonden worden.

[het nemen van beslissingen 28]

Het laatste is met name het geval wanneer één van de beide groepen zeer klein is t.o.v. de andere. Voorbeelden zouden dan gevonden kunnen worden in de medische en psychologische diagnostiek waar liet vaak gaat om in de populatie als geheel niet vaak voorkomende ziekten. Ook in die gevallen waarin de groep waarbinnen men nog weer wil selecteren al door diverse selecties is heen gekomen, zal zich al snel de situatie kunnen voordoen waarin voor iedere testscore er altijd meer 'geschikten' dan 'ongeschikten' worden gevonden. De toelating tot numerus fixus studierichtingen wordt hierdoor gekenmerkt. Een optimale cesuur (anders dan iedereen tot éézelfde groep rekenen) is dan slechts te vinden in die gevallen waarin het 'verlies' van iemand uit de kleinste deelgroep ten onrechte indelen bij de grootste deelgroep groot is t.o.v. het verlies van de indeling van iemand uit de grootste groep bij de kleinste groep. Het tweede artikel van Rorer c.s., waarvan de titel genoemd is, behandelt dan het cesuurprobleem wanneer beide mogelijke fouten niet als noodzakelijk even 'ernstig' worden beschouwd. Rorer c.s. geven een oplossing van het cesuurprobleem wanneer voor de testscoreverdelingen van beide deelgroepen aangenomen kan worden dat ze normaalverdelingen zijn (met mogelijk verschillende gemiddelden en varianties). In hun eerste artikel geven zij tabellen waarin de optimale cesuur gevonden kan worden voor een groot aantal gevallen; deze tabellen zijn ook bruikbaar wanneer met verschillende 'verliezen' voor de mogelijke uitkomsten gewerkt wordt, via een conversie gegeven op blz 366.

Gegeven 'net veronderstelde normaal verdeeld zijn van scores voor ieder van de beide te onderscheiden deelgroepen, en gegeven de relatieve grootte van beide groepen en een specificatie van de waarden die men toekent aan de vier mogelijke uitkomsten (d.w.z. gegeven A, B, C, en D, zie blz. 21), leiden Rorer c.s. een parameter L af, die bepalend is voor de optimale cesuur. Deze parameter is gelijk

$$L = (CD)/(BA), \quad (14)$$

gelijk aan de parameter L uit onze formule (6) op blz. 24.

Rorer c.s. hebben er met name op gewezen dat L dezelfde waarde houdt onder een groot aantal mogelijke veranderingen in waarden A, B, C en D, echter zonder dit uit te werken tot de eenvoudige techniek voor het toekennen van relatieve 'verliezen' zoals in dit rapport gepresenteerd.

[het nemen van beslissingen 29]

Alf & Dorfman, 1967, The classification of individuals into two criterion groups on the basis of a discontinuous payoff function. Psychometrika, volume 32, 115123.

Rorer c.s. (1966a,b) gingen uit van een groep personen die bestond uit leden van twee deelgroepen, waarbij ieder van de beide deelgroepen normaal verdeeld was op de test.

Alf & Dorfman maken de veronderstelling dat de hele groep personen een score verdeling op de test heeft die normaal verdeeld is. Bovendien veronderstellen zij dat er een te voorspellen criterium variabele is, waarop de scores voor de groep als geheel ook normaal verdeeld zijn. Ergens op die criterium variabele is een minimum bepaald, waardoor de groep in twee deelgroepen gesplitst wordt (de 'geschikten' en de 'ongeschikten' bijv.). Deze aannamen zijn praktisch en theoretisch zeer verschillend van die van Rorer c.s., toch is ook bij Alf & Dorfman het resultaat dat voor ieder van de beide criterium groepen ('geschikten, en ongeschikten') een afzonderlijke verdeling op de test variabele aangegeven zou kunnen worden. In het algemeen zullen deze nieuw geconstrueerde verdelingen niet normaal kunnen zijn, omdat verondersteld is dat de score verdeling voor de totale groep normaal was (als dan de gemiddelde test scores voor de beide deelgroepen verschillen, wat vaak het geval zal zijn, en voor iedere deelgroep zou de verdeling normaal zijn, dan zou de som verdeling voor de totale groep tweetoppig zijn, i.t.t. de veronderstelde normaal verdeling voor de gehele groep). Dat betekent dat de resultaten van Rorer c.s., en die van Alf & Dorfman, niet zonder meer op dezelfde situaties van toepassing zullen zijn.

Uit deze aanname van bivariaat normaal verdeeld zijn, leidén Alf & Dorfman een parameter P af, waarvoor geldt:

$$P = (CD) / (CD)(AB) \quad (15)$$

waarbij A, B, C, en D de waarden zijn gegeven aan de vier verschillende

mogelijke uitkomsten zoals eerder beschreven op blz. 21.

Deze parameter P is dan ook identiek aan de in formule (4) gegeven parameter p , de proportie 'geschikten' binnen de subgroep met test scores gelijk aan de cesuur.

De toevoeging van Alf & Dorfman aan de cesuurbepaling zoals in het voorgaande beschreven, bestaat hierin dat, gegeven de nu bepaalde cesuur, en gegeven de bivariaatnormale verdeling van test en criterium variabelen, de proporties a , b , c en d berekend kunnen worden (a is de proportie personen uit de totale groep die 'geschikt' zijn maar op de test beneden de cesuur scoorden; zie blz. 21)

[het nemen van beslissingen 30]

Alf en Dorfman geven een aantal voorbeelden van cesuurbepaling onder deze veronderstellingen, en de resultaten er van ook voor gevallen waarin de uitkomsten A , B , C , en D in dollars geevalueerd kunnen worden. Het is dan mogelijk de totale verwachte 'opbrengst' $E(v)$ van de beslissingsprocedure te schatten:

$$E(v) = aA + bB + cC + dD. \quad (16)$$

Dergelijke schattingen van de totale 'opbrengst' van een beslissingsprocedure (een bepaalde gekozen cesuur), zijn voor de cesuurbepaling zelf niet van nut. Voor de cesuurbepaling immers is alleen maar van belang of het aannemen van een subgroep met een bepaalde test score nog iets toevoegt aan de 'opbrengst' of niet: in het eerste geval wordt die subgroep aangenomen, in het tweede geval niet. Ook al zou de extra opbrengst maar gering zijn, dan zou aannemen beter zijn dan afwijzen; het is dan nog wel mogelijk dat de testkosten de opbrengst niet dekken, maar dat betekent niet dat de cesuur dan opgetrokken zou moeten worden (de testkosten zijn immers al gemaakt), maar dat de test beter niet gebruikt kan worden. Zie voor overwegingen van deze aard vooral Gronbach & Gleser 1965.

Het is wel interessant om in gevallen waarin A , B , C , en D gekwantificeerd kunnen worden, en een quota selectie gehanteerd wordt (de 'besten' uit het aanbod worden aangenomen totdat een beschikbaar aantal plaatsen gevuld is), de opbrengst onder deze quota selectie te vergelijken met de opbrengst die mogelijk is wanneer de cesuur niet bepaald wordt door het gestelde quotum, maar optimaal gelegd kan worden.

Alle berekeningen die onder de veronderstellingen van Alf & Dorfman

gemaakt kunnen worden, kunnen uiteraard ook uitgevoerd worden op basis van valideringsonderzoekgegevens zoals uitgezet zijn in frequentieverdelingen van het soort zoals gegeven in figuur 3 en figuur 5. Men kan dan eenvoudig het aantal personen in iedere deelgroep dat zich boven, dan wel beneden de cesuur bevindt tellen, de tellingen omzetten in proporties a, b, e, en d, en vervolgens $E(v)$ berekenen uit (16)

Noot: De cesuur wordt onder de veronderstellingen van Alf & Dorfman berekend door k op te lossen uit

$$z = (y(c) - rk) / \sqrt{1 - r^2} \quad (17)$$

$y(c)$ is de minimum score op de criterium var. (gestandaardiseerd)

r is de correlatie tussen test en criterium variabele

z is de standaard score behorend bij de proportie p. (in (15))

k is de cesuur, ook in standaard score.

[het nemen van beslissingen 31]

Vastenhouw 1973. Optimale rationele selectie: een waardenprobleem. Onderzoek van Onderwijs, jaargang 2, no 4.

Vastenhouw volgt dezelfde benadering als Alf & Dorfman (1967), veronderstelt eveneens een bivariate normaalverdeling tussen test en criteriumvariabele, en zoekt een cesuur die de verwachte opbrengst van de beslissings procedure maximaliseert. De oplossing van Vastenhouw:

Stel

$t = b + d$ (t is dus de proportie toegelatenen,
of de proportie die boven de cesuur zit)

$g = a + b$ (g is de proportie 'geschikten' in de groep)

We weten:

$c = 1 - a - b - d$.

De optimale cesuur moet gevonden worden door t (de proportie van de groep boven de cesuur) te veranderen totdat de verwachte waarde van de beslissings procedure, $E(v)$ een maximum bereikt. We moeten dus een uitdrukking voor $E(v)$ vinden die zich eenvoudig partiel laat differentiëren naar t; immers, door de partiële afgeleide naar t gelijk nul te stellen vinden we de maximale waarde van $E(v)$ voor variabele t.

We maken daarbij gebruik van het feit dat $g=a+b$ bekend verondersteld

wordt, en dus als constante beschouwd kan worden bij het nemen van de afgeleide van $E(v)$ naar t .

$$\begin{aligned}
 E(v) &= aA+bB+cC+dD \\
 &= gAbA+bB+CaCbCdC+dD \\
 &= gAbA+bB+CC+bCbCtC+bC+tDbD \\
 &= gAgCtU+tDbA+bB+bCbD+C \\
 &= g(AC) + t(DC) + b\sim BA+CD) +
 \end{aligned}$$

$$\partial E(v) / \partial t = (DC) + (\partial b / \partial t) (BA+CD) = 0 \quad (18)$$

]

We vinden wederom de ons uit formules (4), en (15), bekende selectieparameter. Vastenhouw geeft een aantal nomogrammen (voor een correlatie van 0,30 en een correlatie van 0,65) waaruit bij een gegeven totaal aantal 'geschikten' g de combinaties van selectieparameter p en proportie toe te laten personen t die $E(v)$ maximaliseren, afgelezen kunnen worden.

(bewijsvoering uitvoeriger dan door Vastenhouw gegeven; notatie aangepast bij de eerder in dit rapport ingevoerde symbolen.)