

Discussie-rubriek

EEN EMPIRISCH ONDERZOEK NAAR DE BRUIKBAARHEID VAN VEREENVOUDIGDE GEWICHTEN IN MULTIPELE REGRESSIEANALYSES.

S.J. Pijl
M. Nijse¹

Instituut voor Orthopedagogiek; Rijksuniversiteit Groningen.

1. Inleiding.

Wanneer een onderzoeker de beschikking heeft over de scores van n personen op k predictoren X_i ($i = 1, \dots, k$), welke op interval - of rationiveau gemeten zijn, dan kan hij deze scores gebruiken om de scores op een criteriumvariabele Y , welke eveneens op interval - of rationiveau gemeten is, te voorspellen.

De voorspelde Y_i -score ($\hat{Y}_i$) wordt berekend m.b.v.

$$\hat{Y}_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \dots + b_k X_{ik} \quad (1)$$

$$= b_0 + \sum_{i=1}^k b_i X_{ii}$$

$$\text{--waarin } b_0 = \bar{Y} - \sum_{i=1}^k b_i \bar{X}_i ;$$

b_i het gewicht is waarmee variabele X_i gewogen wordt.

De b_i - gewichten worden zodanig gekozen dat

$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n \left[Y_i - \left(b_0 + \sum_{i=1}^k b_i X_{ii} \right) \right]^2 \quad (2)$$

een minimumwaarde aanneemt (de kleinste kwadraten - eigenschap).

1. Met dank aan Drs. J.M.F. ten Berge, Drs. A. Boomsma, Drs. W.v.d Burg en Prof.Dr. W. Molenaar voor hun waardevolle opmerkingen.

Vergelijking (1) staat bekend als de regressievergelijking; de b_i gewichten worden regressiegewichten genoemd. De correlatie tussen de voorspelde en werkelijk behaalde Y - scores, $r_{\hat{Y}Y}$, staat bekend als de multiële correlatiecoëfficiënt R . Omdat $\hat{Y}_i$ een lineaire transformatie is van de scores op de k predictoren X_i , kan men R beschouwen als een maat voor de samenhang tussen de gewogen scores op de k predictoren X_i en het criterium Y .

Omdat de grootte van de b_i - gewichten onder meer beïnvloed wordt door de grootte van de standaarddeviaties van de betreffende predictoren, worden de b_i - gewichten gewoonlijk gestandaardiseerd:

$$\beta_i = b_i \left(\frac{s_i}{s_Y} \right) \quad (3)$$

waarin β_i het gestandaardiseerde gewicht van predictor i is;

s_i de standaarddeviatie van predictor i is;

s_Y de standaarddeviatie van criterium Y is.

Deze gestandaardiseerde gewichten geven aan welke relatieve betekenis t.a.v. andere predictoren men aan de betreffende predictor moet hechten.

De grootte van de b_i - (en β_i -) gewichten wordt tevens beïnvloed door toevalsfluctuaties in de steekproefwaarden (sampling error) en meetfouten als gevolg van het ontbreken van een perfecte betrouwbaarheid (measurement error). Beide soorten fouten leiden ertoe dat R een overschatting van de multiële correlatie in de populatie is. Zou men de gevonden b_i - gewichten willen gebruiken om voor een nieuwe steekproef hetzelfde criterium met dezelfde predictoren te voorspellen, dan "krimpt" R in meerdere of mindere mate (het shrinkage effect). De mate van krimp wordt bepaald door het aantal personen t.o.v. het aantal predictoren.

Een zuivere schatter van de populatie - R is

$$R_{\text{pop}} = \sqrt{1 - (1 - R^2) \left(\frac{n-1}{n-k} \right)} \quad (4)$$

NUNNALLY (1967) beveelt aan om steekproeven ter grootte van minimaal 100 personen te gebruiken wanneer men 2 predictoren heeft. Bij 9 of 10 predictoren zou men steekproeven van minimaal 300 à 400 personen moeten gebruiken.

2. Enkele resultaten van eerder verricht onderzoek.

Sinds lang is bekend dat in de meeste gevallen (afhankelijk van de streekproefgrootte, het aantal predictoren en de intercorrelaties van de predictoren) regressiegewichten vereenvoudigd kunnen worden zonder dat dit noodzakelijkerwijs gepaard gaat met een drastische vermindering van het percentage verklaarde variantie en dus daling van de waarde van R (GUILFORD, 1956; LAWSHE & SHUCKER, 1959; WESMAN & BENNETT, 1959; SCHMIDT, 1971, 1972; CLAUDY, 1972; EINHORN & HOGARTH, 1975). De vereenvoudigingen bestaan meestal uit het gebruik van gewichten welke afgerond zijn tot 1 of 2 decimalen of gelijke gewichten.

WAINER (1976) stelt in een publicatie met als ondertitel "it don't make no nevermind", het wegingsprobleem aan de orde. Zich met name basierend op EINHORN & HOGARTH (1975) pleit WAINER voor het gebruik van gelijke gewichten ("unit weights"). De geringere accuraatheid van de voorspelling bij het gebruik van gelijke gewichten weegt ten volle op tegen de voordelen: eenvoudig te bepalen, geen beïnvloeding door sampling error, geen gebruik van vrijheidsgraden en, dank zij deze eigenschappen, een grote robuustheid.

Hoe gering het verlies aan accuraatheid is blijkt uit het gelijke gewichten theorema: "When k linearly independent predictor variables x_i ($i = 1, \dots, k$) with zero mean and unit variance are used to predict a variable Y , which is also scaled to zero mean and unit variance, and when the population values of the standardised least squares regression coefficients are β_i ($i = 1, \dots, k$), then the expected loss of variance explained using equal (.5) weights will be less than $k/96$ if all β_i 's are uniformly distributed on the interval $[-.25, .75]$. This expected loss is diminished considerably if the x_i 's are not independent". WAINER's uiteindelijke conclusie doet, zeker gezien het feit dat de bewijsvoering van het gelijke gewichten theorema op een aantal punten te wensen overlaat², wel erg absoluut aan:

"When you are interested solely in prediction, it is a very rare situation that calls for regression weights which are unequal".³

2. De genoemde factor $k/96$ zou $k/48$ moeten zijn; dit verlies moet dan nog gezien worden als gemiddeld maximum verlies "in the long run".

3. In WAINER & THISSEN (1976) wordt het theorema iets genuanceerder gepresenteerd.

3. Beschrijving en enkele resultaten van eigen onderzoek.

Wij hebben op gegevens (verzameld in de vijfde en zesde klas lagere school), welke betrekking hadden op schoolvorderingen, intelligentie, prestatiemotivatie, aantal doublures lagere school, leeftijd en beroep vader WAINER's uitspraken gebruikt om het schooltype voortgezet onderwijs, klas 2 resp. klas 3, dat de leerlingen zonder zittenblijven hadden bereikt, te voorspellen. (De schooltypen waren MAVO, HAVO, Atheneum en Gymnasium.) Tevens hebben wij de suggestie van o.a. GUILFORD (1956) om het aantal decimalen van de regressiegewichten te beperken tot 4 resp. 2 resp. 1 in praktijk gebracht. De voornaamste beschrijvende statistische maten zijn vermeld in Tabel 1 en 2. De resultaten welke betrekking hebben op de multiële regressies zijn vermeld in Tabel 3.

Tabel 1. Gemiddelden (M), standaarddeviaties (S), scheefheden (Sch) en medianen (Md) van de gebruikte variabelen.

	M	S	Sch	Md
1. Schoolvorderingentoetsen ISI-test	42.8	12.22	0.28	42.0
2. Verbale intelligentie ISI-test	37.5	6.42	0.16	37.6
3. Figuurlijke intelligentie ISI-test	37.6	7.76	-0.48	39.0
4. Verbale intelligentie GALO-test	33.9	8.43	-0.28	36.0
5. Numerieke intelligentie GALO-test	52.7	10.23	-0.24	53.0
6. Figuurlijke intelligentie GALO-test	35.2	7.63	0.01	35.0
7. IQ GALO-test	117.2	9.90	-0.18	117.5
8. IQ ISI-test	111.7	12.23	0.11	112.0
9. P-schaal PMTK	46.6	28.94	0.08	46.5
10. Leeftijd in maanden	135.7	4.57	0.59	135.0
11. Beroep vader	2.8	1.43	0.04	3.0
12. Schoolvorderingentoets Taal	115.3	8.86	-0.68	116.7
13. Schoolvorderingentoets Rekenen	65.3	9.75	-1.23	67.0
14. Aantal doublures lagere school	0.2	0.41	2.71	0.0
15. Schooltype klas 2	2.1	1.10	0.33	2.0
16. Verbale intelligentie ISI- en GALO-test	71.4	13.52	-0.07	72.0
17. Schooltype klas 3	2.2	1.07	0.23	2.0
18. Figuurlijke intelligentie ISI- en GALO-test	72.8	13.78	-0.22	73.5

Tabel 2: Intercorrelaties van de variabelen van Tabel 1. (decimaalpunten weggelaten).

2	60																		
3	40	41																	
4	47	65	32																
5	46	44	51	50															
6	23	41	60	41	46														
7	51	62	61	77	79	75													
8	64	82	80	58	54	59	72												
9	02	09	-01	07	04	-05	04	06											
10	-05	05	-03	-03	-11	-03	-09	01	05										
11	-33	-27	-08	-20	-20	-08	-21	-23	08	07									
12	73	51	18	53	39	24	49	49	-08	-04	-22								
13	57	27	30	18	51	21	42	37	-06	-05	-24	60							
14	-29	-04	-13	-25	-44	-11	-40	-08	06	31	-09	-18	-33						
15	60	45	39	45	62	33	61	52	04	-07	-35	61	65	-42					
16	57	88	39	93	52	45	77	75	09	01	-25	58	24	-18	50				
17	61	49	35	44	58	34	61	53	02	-04	-33	61	68	-28	98	51			
18	36	46	89	41	54	89	76	78	-03	01	-09	24	29	-14	40	47	39		
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17			

In dit onderzoek werden voor een aantal verschillende combinaties van 2 of 3 predictoren kleinste kwadraten regressieanalyses uitgevoerd.

De hierbij verkregen b-gewichten werden afgerond tot 4 resp. 2 resp. 1 decimaal. De multiële correlaties werd met de afgeronde gewichten opnieuw berekend.

Aangezien de multiële correlatie, berekend met de tot op 4 decimalen afgeronde gewichten, steeds exact dezelfde was als de oorspronkelijke multiële correlatie, wordt deze niet in Tabel 3 vermeld.

De twee andere reeksen multiële correlaties, berekend over de tot op 2 resp. 1 decimaal afgeronde gewichten, zijn in Tabel 3 onder R_2 resp. R_1 terug te vinden.

Tabel 3: Resultaten (voor verklaring zie tekst).

Predictoren	Crit.	n	b		R	R ₂	R ₁	R _w	β		β_w	
7 12	15	74	0593	0501	748	748	747	745	50	37	50	50
8 12	15	74	0343	0618	671	671	609	668	33	45	50	50
7 13	15	74	0581	0554	791	790	790	790	49	45	50	50
8 13	15	74	0369	0646	730	729	654	724	35	53	50	50
12 14	15	52	0605	-9252	625	625	614	617	47	-33	50	-50
13 14	15	52	0611	-6751	653	653	649	630	53	-24	50	-50
7 9	15	171	0670	0005	611	610	610	610	61	01	50	00
8 9	15	171	0465	0004	522	522	000	522	52	01	50	00
7 9 11	15	163	0631	0015	661	659	657	624	57	04	50	-50
8 9 11	15	163	0412	0016	575	574	351	557	46	04	50	-50
7 10 11	15	163	0636	0030	660	659	657	624	57	01	50	-50
8 10 11	15	163	0416	-0123	576	576	351	557	47	-05	50	-50
12 10 11	15	71	0815	-0156	637	636	633	551	60	-07	50	-50
13 10 11	15	71	0772	-0131	649	648	646	557	62	-06	50	-50
1 16 18	17	123	0373	0150	654	651	000	634	44	19	50	50
1 2 3	17	123	0384	0263	634	634	000	611	46	16	50	50
4 5 6	17	123	0267	0517	616	616	583	578	21	46	50	50
7 10 11	17	114	0649	-0018	640	639	635	602	56	-01	50	-50
8 10 11	17	114	0441	-0137	564	564	325	539	48	-06	50	-50

Nadat de variabelen werden omgezet in standaardscores werden - uiteraard over dezelfde predictoren-combinaties - multiële regressieanalyses met gelijke gewichten (.50 of -.50) uitgevoerd. Het teken van een gewicht werd bepaald door het teken van de correlatie tussen de betreffende predictor en het criterium (EINHORN & HOGARTH, 1975). Wanneer de correlatie niet significant verschilde van .00 werd als gewicht 0 gebruikt.

De β -gewichten van de kleinste kwadraten regressieanalyses (β), de β -gewichten van de gelijke gewichten regressieanalyses (β_w) en de gelijke gewichten multiële correlaties (R_w) zijn in Tabel 3 vermeld.

Inspectie van de verschillen tussen R en R_w leert dat in dit onderzoek het verlies aan accuraatheid van de voorspelling veelal gering is.

4. Bespreking van de resultaten en conclusies.

Vooraf dient opgemerkt te worden dat een tweetal effecten invloed gehad kan hebben op onze uitkomsten: de scheve verdeling van een aantal variabelen (zie Tabel 1) en het bestaan van suppressoren. In dit onderzoek zijn de eventuele gevolgen van deze effecten t.a.v. onze bevindingen niet verder uitgewerkt. M.b.t. suppressor-effecten valt op te merken dat deze in het algemeen weinig voorkomen; in ons onderzoek zijn zij vrijwel zeker afwezig.

Er werd reeds opgemerkt dat we zelf ~~met~~ ruwe benaderingen van de oorspronkelijke regressiegewichten ~~amper~~ de hoogte van de multiële correlatie beïnvloeden (zie kolommen R_2 en R_1). Volgens Guilford (1965, blz. 422) kunnen we de zeer nauwkeurige regressiegewichten (in 4 decimalen) veelal ongestraft afronden tot op 1 decimaal. Het afronden tot op 2 decimalen levert in dit onderzoek vrijwel geen verlies van het percentage verklaarde variantie op. Bij het drastische afronden tot op 1 decimaal treedt in vrij veel gevallen een onaanvaardbaar groot verlies op. Deze verliezen ontstaan doordat de predictoren, welke gewogen worden met een waarde net onder .05, wegvallen in de regressievergelijking, terwijl deze predictoren, gewogen met de oorspronkelijke gewichten, wel degelijk een behoorlijk deel aan het percentage verklaarde variantie bijdragen.

Samenvattend kunnen wij opmerken dat vereenvoudiging van regressiegewichten tot op 2 decimalen zeer wel mogelijk blijkt; bij verdere vereenvoudiging tot 1 decimaal blijven de verliezen beperkt zolang we niet te maken hebben met het toekennen van een gewicht van 0 aan die variabelen waarvan het niet-afronde gewicht als kleiner dan .05 is.

Het verminderen van de accuraatheid van de voorspelling is evenmin onrustbarend wanneer we gewichten volgens WAINER & THISSEN gebruiken (vgl. R en R_w). Dat het verlies van het percentage verklaarde variantie in een tweetal regressieanalyses groter dan $3/48 = 6\%$ (i.p.v. $3/96 = 3\%$) is, kan veroorzaakt zijn door het niet voldoen aan WAINER's assumptie in het gelijke gewichten theorema dat de β 's uniform verdeeld moeten zijn over het interval .25 - .75. De vrij grote accuraatheid bij het gebruik van gelijke gewichten zet een vraagteken bij het toepassen en vooral het interpreteren van gestandaardiseerde kleinste kwadraten regressiegewichten.

De waarde van een op een regressieanalyse gebaseerde uitspraak als: "Bij het voorspellen van het schooltype voortgezet onderwijs is het GALO-IQ belangrijker dan de score op de Schoolvorderingentoets Taal" wordt uiterst bedenkelijk.

Een onderzoek op één bepaald soort materiaal kan vanzelfsprekend geen uitsluitend geven omtrent de algemene vraag of toepassing van gelijke gewichten even goed is als toepassing van differentiële gewichten. Voortgezet onderzoek met andere soorten gegevens is noodzakelijk. Wanneer echter de predictorenkeuze al gedaan is valt in gevallen waarin voorspellingen moeten worden gedaan, WAINER's aanbeveling zeker te overwegen. Uiteraard biedt WAINER's methode geen soulaas wanneer predictoren geselecteerd moeten worden uit een groter aantal; een vorm van klassieke stapsgewijze regressieanalyse blijft dan de aangewezen weg.

In gevallen waarin WAINER's uitspraken houdbaar blijken te zijn (zoals voor de meeste van onze regressieanalyses het geval is), kunnen de gevolgen ook een verderstrekkende betekenis hebben. Wij denken aan conclusies die getrokken worden uit padanalyses en canonische correlaties. Vooralsnog onttrekken deze mogelijke gevolgen zich echter aan onze waarneming.

5. Literatuur.

CLAUDY, J.G. (1972). A comparison of five variable weighting procedures. Educ. Psychol. Measmt., 32, 311 - 322.

EINHORN, H.J., & HOGARTH, R.M. (1975). Unit weighting scheme for decision making. Org. Behav. Hum. Perform., 13, 171 - 192.

GUILFORD, J.P. (1956). Fundamental statistics in psychology and education; third edition. New York: Mc Graw - Hill.

LAWSHE, C.H., & SHUCKER, R.E. (1959). The relative efficiency of four test weighting methods in multiple prediction.

Educ. Psychol. Measmt., 19, 103 - 114.

NUNNALLY, J.C. (1967). Psychometric theory. New York: Mc Graw - Hill.

SCHMIDT, F.L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. Educ. Psychol. Measmt., 31, 699 - 714.

SCHMIDT, F.L. (1972). The reliability of differences between linear regression weights in applied differential psychology. Educ. Psychol. Measmt., 32, 879 - 886.

WAINER, H. (1976). Estimating coefficients in linear models: it don't make no never-mind. Psychol. Bull., 83, 213 - 217.

WAINER, H., & THISSEN, D. (1976). Three steps toward robust regression. Psychometrika, 41, 9 - 34.

WESMAN, A.G., & BENNETT, G.K. (1959). Multiple regression vs. simple addition of scores in prediction of college grades.

Educ. Psychol. Measmt., 19, 243 - 246.