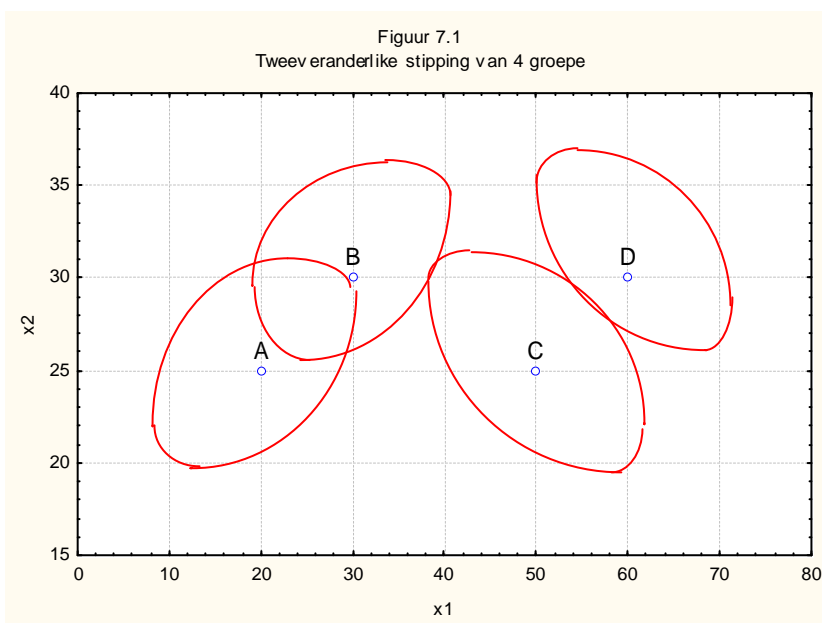


## HOOFSTUK 7

### Meerveranderlike effekgrootte-indekse

Dit is selde dat daar slegs een afhanklike veranderlike betrokke is by 'n studie. In Hoofstuk 3 se Voorbeeld A is daar BDI, POMS\_S en POMS\_B, by Voorbeeld E is daar 5 cholesterol metings asook 'n sistoliese- en diastoliese bloeddrukmetings. Baie keer word elke veranderlike afsonderlik ontleed en kan effekgrootte-indekse daarvoor bepaal word. Dis egter ook sinvol om die veranderlikes as 'n eenheid te beskou waarin nie net die afsonderlike gemiddeldes en SA's 'n rol speel nie, maar ook die interkorrelasies tussen die veranderlikes. So 'n ontleding staan bekend as *meerveranderlik* waarop ook effekgrootte-indekse gebaseer kan word.



Figuur 7.1 gee Olejnik & Algina (2000) se Figuur 2 weer en word vervolgens gebruik om te verduidelik hoe 'n meerveranderlike gestandaardiseerde verskil in gemiddeldes geïnterpreteer kan word. Dit is 'n voorstelling van 'n spreidingsdiagram van die twee veranderlikes  $x_1$  en  $x_2$  in steekproewe uit populasies A, B, C en D. Die punte op die diagram stel elke groep se

gemiddelde  $(\bar{x}_1, \bar{x}_2)$  voor, terwyl die ellipse só is dat die meeste (sê 90%) van die groep se metings daarbinne val. By groepe A en B is daar gevolglik 'n positiewe korrelasie tussen  $x_1$  en  $x_2$ , terwyl dit negatief is by groepe C en D. Let op die oorvleueling by groepe A en B en prakties geen oorvleueling by C en D. Die feit dat die ellipse omtrent ewegroot is, suggereer gelyke SA's by al die groepe vir  $x_1$  en so ook vir  $x_2$ . As groepe A en B se gemiddeldes  $\bar{x}_1$  en  $\bar{x}_2$  vergelyk wil word, gee die *Mahalanobis-afstand*  $D$  die analoog van die eenveranderlike gestandaardiseerde verskil:

$$D = \sqrt{\frac{1}{1-r_p^2} (\hat{\delta}_1^2 + \hat{\delta}_2^2 - 2r_p \hat{\delta}_1 \hat{\delta}_2)} \quad (7.1)$$

waar

$\hat{\delta}_1$  en  $\hat{\delta}_2$  die gestandaardiseerde verskille volgens (4.3) is:

$$\hat{\delta}_1 = \frac{\bar{x}_{A1} - \bar{x}_{B1}}{s_{p1}} \quad \text{en} \quad \hat{\delta}_2 = \frac{\bar{x}_{A2} - \bar{x}_{B2}}{s_{p2}},$$

met  $\bar{x}_{A1}$  en  $\bar{x}_{B1}$  die gemiddeldes van  $x_1$  by groepe A en B en  $s_{p1}$  die saamgevoegde SA vir  $x_1$  oor groepe A en B is, ens.

Verder is  $r_p$  die saamgevoegde binne-groepe korrelasie:

$$r_p = \frac{s_{p12}}{s_{p1}s_{p2}},$$

waar  $s_{p12}$  die saamgevoegde kovariansie tussen  $x_1$  en  $x_2$  is.

Gestel in die voorbeeld van die data in figuur 7.1, dat die saamgevoegde SA vir  $x_1$  8 is en dié van  $x_2$  3 is en verder dat  $r_p$  vir groepe A en B 0,5 en vir C en D - 0,5 is. Uit die figuur is dit dan duidelik dat as groepe A en B vergelyk wil word is

$\hat{\delta}_1 = (30 - 20) / 8 = 1,25$  en  $\hat{\delta}_2 = (30 - 20) / 3 = 1,67$ , terwyl

$$D^2 = \frac{1}{1 - 0,5^2} (1,25^2 + 1,67^2 - 2 \times 0,5 \times 1,25 \times 1,67) = 2,264 / 0,75 \\ = 3,019,$$

dus  $D = 1,74$ .

As groepe C en D vergelyk word, is  $\hat{\delta}_1$  en  $\hat{\delta}_2$  dieselfde as tevore maar omdat  $r_p = -0,5$  is, geld dat  $D = 2,93$ .

Drie dinge is belangrik om op te let:

- Eerstens is die waarde van D groter as die waardes van elk van  $\hat{\delta}_1$  en  $\hat{\delta}_2$ , want dit gee die euklidiese afstand tussen die 2-dimensionele gemiddeldes op 'n skaal wat beide deur die SA's van en korrelasie tussen  $x_1$  en  $x_2$  bepaal word, terwyl  $\hat{\delta}$  die verskil tussen een-dimensionele gemiddeldes in SA-eenhede is.
- Tweedens is daar 'n groter skeiding tussen die punte by groepe C en D as by A en B, al is die gemiddelde punte ewe ver vanmekaar af. Dit kan toegeskryf word aan die feit dat A en B positiewe korrelasies tussen  $x_1$  en  $x_2$  het, terwyl dié van C en D negatief is. Die waardes van D reflekteer dan ook hierdie verskynsel, want by e.g. geval word  $2 r_p \hat{\delta}_1 \hat{\delta}_2$  afgetrek van die positiewe  $\hat{\delta}_1^2 + \hat{\delta}_2^2$ , terwyl dit by C en D bygetel word (omdat  $r_p < 0$ ). Dus waar  $D = 1,74$  is by A en B, is dit heelwat hoër, nl. 2,93, by C en D.
- Derdens speel behalwe die teken, ook die grootte van  $r_p$  'n rol.

In die volgende paragraaf word Mahalanobis-afstand D veralgemeen na meer as 2 veranderlikes. Verder aan in die hoofstuk word in die geval van meer as 2 groepe na effekgrootte-indekse gekyk om omnibus-effekte asook kontras-effekte te bepaal.

## 7.1 Vergelyking van twee groepe, $m$ veranderlikes

In die inleiding het ons vir illustrasie-doeleindes die geval beskou waar 2-veranderlike gemiddeldes vergelyk word tussen steekproewe uit twee populasies.

Gestel egter dat twee  $m$ -veranderlike populasies volledig waargeneem word en dat

$$\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_m) \quad (7.2)$$

die vektor is van effekgrootte-indekse vir veranderlikes  $x_1, x_2, \dots, x_m$  by beide populasies.

Laat verder

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \dots & \rho_{1m} \\ \rho_{12} & 1 & \cdot & \cdot & \cdot \\ \vdots & \cdot & \cdot & \cdot & \cdot \\ \rho_{1m} & \cdot & \cdot & \cdot & 1 \end{pmatrix} \quad (7.3)$$

die gemeenskaplike korrelasie-matriks van beide populasies wees van interkorrelasies tussen  $x_1, x_2, \dots, x_m$ .

Vir die vergelyking van  $m$ -veranderlike gemiddeldes tussen twee populasies kan die algemene Mahalanobis-afstand vir populasies gebruik word, nl.

$$D = \sqrt{\boldsymbol{\delta} \mathbf{R}^{-1} \boldsymbol{\delta}'} \quad (7.4)$$

Hierdie behels matriks-berekening wat moeilik met die hand gedoen kan word. 'n Makliker formule vir berekening is:

$$D^2 = \frac{(N_A + N_B)^2 (1 - \Lambda)}{N_A N_B \Lambda} \quad (7.5)$$

waar  $N_A$  en  $N_B$  die populasiegroottes is en die grootheid  $\Lambda$  Wilks se *lambda* is, soos gegee word as resultaat van 'n meerveranderlike variansie-analise

(MANOVA), waar twee gemiddelde vektore vergelyk word, en gegee word deur enige statistiese rekenaarpakket soos SPSS, SAS of STATISTICA.

In die geval waar ewekansige steekproewe van groottes  $n_A$  en  $n_B$  uit twee populasies getrek word, kan  $D$  beraam word met  $\tilde{D}$  waar

$$\tilde{D}^2 = \frac{(n_A + n_B - 2)(1 - \hat{\Lambda}) \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}{\hat{\Lambda}}, \quad (7.6)$$

of

$$\tilde{D}^2 = (n_A + n_B - 2) \hat{U}(1) \left( \frac{1}{n_A} + \frac{1}{n_B} \right).$$

Hier is  $\hat{\Lambda}$  gebaseer op die steekproewe se data. Hierdie beramer is egter sydig en ons stel vervolgens die beramer  $\hat{D}$  voor wat benaderd onsydig is en waar

$$\hat{D}^2 = \left[ (n_A + n_B - m - 3) \frac{1 - \hat{\Lambda}}{\hat{\Lambda}} - m \right] \left( \frac{1}{n_A} + \frac{1}{n_B} \right), \quad (7.7)$$

of

$$\hat{D}^2 = \left[ (n_A + n_B - m - 3) \hat{U}(1) - m \right] \left( \frac{1}{n_A} + \frac{1}{n_B} \right).$$

Let op dat indien  $\hat{D}^2$  negatief is word die waarde as 0 geneem, siende dat  $D^2 \geq 0$  per definisie.

Let op dat  $\hat{U}^{(1)} = \frac{1 - \hat{\Lambda}}{\hat{\Lambda}}$  die *Hotelling-statistiek* is waarvan die waarde gewoonlik saam met Wilks se lambda ( $\hat{\Lambda}$ ) deur die bg. rekenaarpakkette gerapporteer word. Verder kan die sg. *Hotelling* se  $T^2$  verkry word uit

$$T^2 = (n_A + n_B - 2) \frac{1 - \hat{\Lambda}}{\hat{\Lambda}} = (n_A + n_B - 2) \hat{U}(1),$$

sodat  $D$ ,  $\tilde{D}$  en  $\hat{D}$  ook uit  $U^{(1)}$ ,  $\hat{U}^{(1)}$  of  $T^2$  bereken kan word.

'n Benaderde  $(1-\alpha)100\%$  VI vir  $D$  kan ook bereken word. Dit volg daaruit dat

$c \cdot \frac{1-\hat{\Lambda}}{\hat{\Lambda}}$  benaderde nie-sentraal- $F$  verdeeld is met niesentraliteitsparameter

$nsp_D = \frac{n_A n_B}{(n_A + n_B)} D^2$  en  $c$  'n konstante. Meer besonderhede hieroor volg in Bylae

B.

'n SAS-program **VI\_D** is beskikbaar op die webblad van die handleiding en het as insette nodig:  $\hat{\Lambda}$ ,  $n_A$ ,  $n_B$ ,  $m$  en  $\alpha$ .

### Voorbeeld 7.1

Beskou Voorbeeld A, Hoofstuk 3. Gestel die eksperimentele- en kontrolegroepe moet vergelyk word t.o.v. die gemiddelde vektore van BDI se voor-, na- en opvolgtoetse. Die MANOVA wat m.b.v. STATISTICA gedoen is, lewer die volgende resultate op:

$$\hat{\Lambda} = 0,1513 \quad ; \quad F(3;46) = 86,02 \quad (p < 0,0001).$$

Met  $n_A = n_B = 25$  en  $m = 3$  ; volg dat  $\hat{U}^{(1)} = \frac{1-\hat{\Lambda}}{\hat{\Lambda}} = 5,61$  en

$$\tilde{D} = \sqrt{48 \times 5,61 \left( \frac{1}{25} + \frac{1}{25} \right)} = 4,64$$

en

$\hat{D} = \sqrt{(44 \times 5,61 - 3)(0,04 + 0,04)} = 4,42$ , wat kleiner is as  $\tilde{D}$  omdat dit benaderd onsydig is.

Toepassing van die program VI\_D lewer die 95% VI vir  $D$  :  $(3,44; 5,60)$ .

Dit beteken dus dat die Mahalanobis-afstand tussen die twee groepe se 3-dimensionele gemiddelde vektore so klein as 3,44 en so groot soos 5,60 kan wees met 'n 95% waarskynlikheid □

### 7.1.1 Riglynwaardes vir $D$

Omdat  $D$  'n 'afstand' is tussen 2 gemiddelde vektore wat 'geweeg' word in terme van die interkorrelasies van  $x_1, \dots, x_m$ , word dit op 'n ander skaal as  $\delta_1, \delta_2, \dots, \delta_m$ , die individuele effekgrootte-indekse vir die  $m$  veranderlikes, gemeet. Net in baie spesiale gevalle is daar verbande tussen  $D$  en  $\delta_1, \dots, \delta_m$ . Vir die eenvoud vir bespreking, laat ons ons slegs beperk tot die geval van 2 veranderlikes, waar  $D$  deur (7.1) gegee word.

In die ekstreme geval van  $r_p = 0$ , vereenvoudig  $D$  na  $\sqrt{\hat{\delta}_1^2 + \hat{\delta}_2^2}$ , maar weens geen korrelasie tussen  $x_1$  en  $x_2$ , kan elkeen maar apart beskou word en is 'n meerveranderlike ontleding oorbodig. Neig  $r_p$  na 1 toe, neig  $D$  na oneindig, maar in sulke gevalle verklaar  $x_1$  omtrent al die variansie van  $x_2$  sodat net een van  $x_1$  of  $x_2$  ontleed hoef te word. In die geval waar  $\hat{\delta}_1 = \hat{\delta}_2$  en waar  $r_p$  na 1 neig, neig  $D$  na  $\hat{\delta}_1 = \hat{\delta}_2$ . Tabel 7.1 gee vir geselekteerde waardes van  $\hat{\delta}_1 = \hat{\delta}_2$  en  $r_p$  die waardes van  $D$ , waaruit dit duidelik is dat:

- $D > \hat{\delta}_1 = \hat{\delta}_2$
- $D$  groter word na gelang  $r_p$  kleiner word
- Die maksimum  $D$  bereik word as  $r_p$  naby  $-1$  is. Netsoos by die geval waar  $r_p$  naby 1 is, hoef by die geval waar  $r_p$  naby  $-1$  is, net een van  $x_1$  of  $x_2$  ontleed te word omdat hulle omtrent mekaar se variansies volledig verklaar.

Waar tabel 7.1 aanduidings van moontlike riglynwaardes kan gee as ons weet wat  $r_p$  se waardes is en dat  $\hat{\delta}_1$  en  $\hat{\delta}_2$  omtrent gelyk is, is dit in die algemeen met  $m$  veranderlikes nie meer moontlik nie.

**Tabel 7.1:**  
**Waardes van Mahalanobis-**  
**afstand as  $k=2$ ,  $m=2$ ,**

	r								
	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1
<b>0.1</b>	0.45	0.32	0.26	0.22	0.20	0.18	0.17	0.16	0.15
<b>0.2</b>	0.89	0.63	0.52	0.45	0.40	0.37	0.34	0.32	0.30
<b>0.3</b>	1.34	0.95	0.77	0.67	0.60	0.55	0.51	0.47	0.45
<b>0.4</b>	1.79	1.26	1.03	0.89	0.80	0.73	0.68	0.63	0.60
<b>0.5</b>	2.24	1.58	1.29	1.12	1.00	0.91	0.85	0.79	0.75
<b>0.6</b>	2.68	1.90	1.55	1.34	1.20	1.10	1.01	0.95	0.89
<b>0.7</b>	3.13	2.21	1.81	1.57	1.40	1.28	1.18	1.11	1.04
<b>0.8</b>	3.58	2.53	2.07	1.79	1.60	1.46	1.35	1.26	1.19
<b>0.9</b>	4.02	2.85	2.32	2.01	1.80	1.64	1.52	1.42	1.34
<b>1</b>	4.47	3.16	2.58	2.24	2.00	1.83	1.69	1.58	1.49

	r									
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
<b>0.1</b>	0.14	0.13	0.13	0.12	0.12	0.12	0.11	0.11	0.11	0.10
<b>0.2</b>	0.28	0.27	0.26	0.25	0.24	0.23	0.22	0.22	0.21	0.21
<b>0.3</b>	0.42	0.40	0.39	0.37	0.36	0.35	0.34	0.33	0.32	0.31
<b>0.4</b>	0.57	0.54	0.52	0.50	0.48	0.46	0.45	0.43	0.42	0.41
<b>0.5</b>	0.71	0.67	0.65	0.62	0.60	0.58	0.56	0.54	0.53	0.51
<b>0.6</b>	0.85	0.81	0.77	0.74	0.72	0.69	0.67	0.65	0.63	0.62
<b>0.7</b>	0.99	0.94	0.90	0.87	0.84	0.81	0.78	0.76	0.74	0.72
<b>0.8</b>	1.13	1.08	1.03	0.99	0.96	0.92	0.89	0.87	0.84	0.82
<b>0.9</b>	1.27	1.21	1.16	1.12	1.08	1.04	1.01	0.98	0.95	0.92
<b>1</b>	1.41	1.35	1.29	1.24	1.20	1.15	1.12	1.08	1.05	1.03

## 7.2 Effekgrootte van kontras-effekte by $m$ veranderlikes en $k$ groepe

Soos in Hoofstuk 6, kan meer as 2 groepe in terme van kontraste vergelyk word. 'n *Kontras* vergelyk nou populasie-gemiddelde vektore met mekaar en word gedefinieer as:

$$\psi = c_1\mu_1 + c_2\mu_2 + \dots + c_k\mu_k \quad , \quad (7.8)$$



waar

$$\boldsymbol{\mu}_1 = (\mu_{i1}, \mu_{i2}, \dots, \mu_{im})$$

en

$$\boldsymbol{\psi} = (\psi_1, \psi_2, \dots, \psi_m) ,$$

terwyl  $c_1, c_2, \dots, c_k$  die kontrasgewigte is van die  $k$  populasies.

### Voorbeeld 7.2

In Voorbeeld E, Hoofstuk 3 kan die gemiddelde vektore van cholesterol-waardes (chol\_0-chol\_4) vir die hartpasiënte wat geen of min oefeninge doen, vergelyk word met dié wat matig of baie oefen met die kontras:

$$\boldsymbol{\psi} = 0,5\boldsymbol{\mu}_1 + 0,5\boldsymbol{\mu}_2 - 0,5\boldsymbol{\mu}_3 - 0,5\boldsymbol{\mu}_4 ,$$

waar  $\mu_i$  die gemiddeldes van die 5 cholesterol-waardes as komponente het, terwyl  $\psi$  die kontraste van elke cholesterol-waarde as komponente het.

□

As effekgrootte-indeks vir  $\boldsymbol{\psi}$  kan die Mahalanobis afstand  $D_{\boldsymbol{\psi}}$  gebruik word, waar:

$$D_{\boldsymbol{\psi}}^2 = \frac{N(1 - \Lambda_{\boldsymbol{\psi}}) \left( \sum_{i=1}^k \frac{c_i^2}{N_i} \right)}{\Lambda_{\boldsymbol{\psi}}} , \quad (7.9)$$

en waar soos tevore  $N_i$  die populasiegroottes is en  $N = N_1 + \dots + N_k$  , terwyl  $\Lambda_{\boldsymbol{\psi}}$  die Wilks se lambda is vir die kontras  $\boldsymbol{\psi}$  .

Wanneer  $k$  ewekansige steekproewe uit die populasies getrek word, is die beramer (Kline, 2004b : 10):  $\tilde{D}_{\boldsymbol{\psi}}$  , met

$$\tilde{D}_{\boldsymbol{\psi}}^2 = \frac{(n-k)(1 - \hat{\Lambda}_{\boldsymbol{\psi}}) \left( \sum_{i=1}^k \frac{c_i^2}{n_i} \right)}{\hat{\Lambda}_{\boldsymbol{\psi}}} , \quad (7.10)$$

waar  $n = n_1 + \dots + n_k$  en  $\hat{\Lambda}_\psi$  nou op die steekproef-data gebaseer is, en die resultaat van 'n MANOVA is waar  $\psi = 0$  getoets word.

'n Benaderde  $(1 - \alpha)100\%$  VI vir  $D_\psi$  kan verkry word m.b.v. dieselfde metode as in paragraaf 7.1. Die SAS-program **VI\_D-kontras** is beskikbaar op die handleiding se webblad en het die volgende insette nodig:

$$\hat{\Lambda}_\psi, m, n_1, \dots, n_k, c_1, \dots, c_k \text{ en } \alpha.$$

### Voorbeeld 7.3:

In Voorbeeld 7.2 is die MANOVA om  $\psi = 0$  te toets se resultate:

$$\hat{\Lambda}_\psi = 0,765; \quad F(5;42) = 2,58 (p = 0,0404).$$

Verder is  $n = 50$ ,  $k = 4$ ,  $n_1 = 19$ ,  $n_2 = 20$ ,  $n_3 = 9$ ,  $n_4 = 2$

en

$c_1 = 0,5$ ,  $c_2 = 0,5$ ,  $c_3 = -0,5$  en  $c_4 = -0,5$ , sodat

$$\begin{aligned} \tilde{D}_\psi^2 &= \frac{(50-4)(1-0,765)\left(\frac{0,25}{19} + \frac{0,25}{20} + \frac{0,25}{9} + \frac{0,25}{2}\right)}{0,765} \\ &= \frac{1,929}{0,765} = 2,521, \end{aligned}$$

dus  $\tilde{D}_\psi = 1,588$

Die 95% VI vir  $D_\psi$  hier is:  $(0; 2,248)$

(die ondergrens kon nie bereken word nie, dus word dit as nul geneem).

□

### 7.3 Meerveranderlike omnibus effek

In die eenveranderlike geval in Hoofstuk 5 is die effekgrootte-indeks  $\eta^2$  ingevoer om die omnibus effek van verskille in  $k$  populasiegemiddeldes te meet. Dit was

die proporsie van die afhanklike veranderlike se variansie wat toegeskryf kan word aan populasie-lidmaatskap.

Omdat Wilks se lambda  $\Lambda$  beskou kan word as die proporsie van die veralgemeende variansie van 'n vektor van  $m$  veranderlikes wat *nie* aan populasie-lidmaatskap toegeskryf kan word nie (sien Kline, 2004b: 4), is die voor die handliggende effekgrootte-indeks *Wilks se veralgemeende korrelasieverhouding*:

$$\eta_{meerv}^2 = 1 - \Lambda . \quad (7.11)$$

Cohen(1988) stel ook 'n veralgemening van die effekgrootte  $f^2$  by meervoudige lineêre regressie voor wat op  $\Lambda$  gebaseer is, nl.

$$f_r^2 = \Lambda^{-1/r} - 1 \quad (7.12)$$

waar 
$$r = \sqrt{\frac{m^2(k-1)^2 - 4}{m^2 + (k-1)^2 - 5}} . \quad (7.13)$$

### Opmerkings:

1. By meervoudige lineêre regressie is  $m = 1$ , gevolglik word  $r = 1$  en verder  $\Lambda = 1 - R^2$ , sodat  $f^2 = \frac{1}{1 - R^2} - 1 = \frac{R^2}{1 - R^2}$  (kyk Hoofstuk 5, paragraaf 5.2.3).
2. Tabel 7.2 kan gebruik word om  $r$  af te lees vir geselekteerde  $k$ - en  $m$ -waardes, terwyl Tabel 7.3 dan gebruik kan word om waardes van  $f_r^2$  af te lees vir gegewe  $\Lambda$ - en  $R^2$ -waardes.
3. Anders as sy eenveranderlike ewebeeld  $f^2$ , hang  $f_r^2$  van  $k$  en  $m$  af sodat dit moeilik is om riglyne te stel vir wanneer  $f_r^2$  klein, medium of groot is.

**Tabel 7.2: Waardes vir r vir  
gegewe k en m**

		m											
		1	2	3	4	5	6	7	8	9	10	15	20
k	2	1	1	1	1	1	1	1	1	1	1	1	1
	3	1	2	2	2	2	2	2	2	2	2	2	2
	4	1	2	2.43	2.65	2.76	2.83	2.87	2.90	2.92	2.94	2.97	2.98
	5	1	2	2.65	3.06	3.32	3.49	3.61	3.69	3.75	3.79	3.90	3.94
	6	1	2	2.76	3.32	3.71	4.00	4.21	4.36	4.47	4.56	4.79	4.88
	7	1	2	2.83	3.49	4.00	4.39	4.69	4.92	5.10	5.24	5.62	5.78
	8	1	2	2.87	3.61	4.21	4.69	5.08	5.39	5.63	5.83	6.40	6.64
	9	1	2	2.90	3.69	4.36	4.92	5.39	5.77	6.08	6.34	7.12	7.47
	10	1	2	2.92	3.75	4.47	5.10	5.63	6.08	6.46	6.78	7.78	8.25
	11	1	2	2.94	3.79	4.56	5.24	5.83	6.34	6.78	7.16	8.38	8.99
	12	1	2	2.95	3.83	4.63	5.35	5.99	6.56	7.05	7.48	8.93	9.68
	13	1	2	2.95	3.85	4.68	5.44	6.12	6.74	7.28	7.76	9.43	10.34
	14	1	2	2.96	3.87	4.73	5.51	6.23	6.89	7.47	8.00	9.89	10.95
	15	1	2	2.97	3.89	4.76	5.57	6.32	7.01	7.64	8.21	10.30	11.52
	20	1	2	2.98	3.94	4.87	5.76	6.61	7.42	8.18	8.90	11.82	13.82
	30	1	2	2.99	3.97	4.94	5.89	6.82	7.73	8.62	9.48	13.35	16.50

Tabel 7.3: Waardes van  $f_r^2$  vir gegewe waardes van  $r$ ,  $\Lambda$  (of  $R^2$ )

		Wilks se $\Lambda$														
		0.99	0.96	0.92	0.9	0.88	0.85	0.8	0.75	0.7	0.6	0.5	0.4	0.3	0.2	0.1
		$R^2$														
$r$		0.01	0.04	0.08	0.1	0.12	0.15	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1		0.01	0.04	0.09	0.11	0.14	0.18	0.25	0.33	0.43	0.67	1	1.5	2.33	4	9
2		0.01	0.02	0.04	0.05	0.07	0.08	0.12	0.15	0.2	0.29	0.41	0.58	0.83	1.24	2.16
2.5		0	0.02	0.03	0.04	0.05	0.07	0.09	0.12	0.15	0.23	0.32	0.44	0.62	0.9	1.51
3		0	0.01	0.03	0.04	0.04	0.06	0.08	0.1	0.13	0.19	0.26	0.36	0.49	0.71	1.15
3.5		0	0.01	0.02	0.03	0.04	0.05	0.07	0.09	0.11	0.16	0.22	0.3	0.41	0.58	0.93
4		0	0.01	0.02	0.03	0.03	0.04	0.06	0.07	0.09	0.14	0.19	0.26	0.35	0.5	0.78
4.5		0	0.01	0.02	0.02	0.03	0.04	0.05	0.07	0.08	0.12	0.17	0.23	0.31	0.43	0.67
5		0	0.01	0.02	0.02	0.03	0.03	0.05	0.06	0.07	0.11	0.15	0.2	0.27	0.38	0.58
5.5		0	0.01	0.02	0.02	0.02	0.03	0.04	0.05	0.07	0.1	0.13	0.18	0.24	0.34	0.52
6		0	0.01	0.01	0.02	0.02	0.03	0.04	0.05	0.06	0.09	0.12	0.16	0.22	0.31	0.47
6.5		0	0.01	0.01	0.02	0.02	0.03	0.03	0.05	0.06	0.08	0.11	0.15	0.2	0.28	0.43
7		0	0.01	0.01	0.02	0.02	0.02	0.03	0.04	0.05	0.08	0.1	0.14	0.19	0.26	0.39
7.5		0	0.01	0.01	0.01	0.02	0.02	0.03	0.04	0.05	0.07	0.1	0.13	0.17	0.24	0.36
8		0	0.01	0.01	0.01	0.02	0.02	0.03	0.04	0.05	0.07	0.09	0.12	0.16	0.22	0.33
9		0	0	0.01	0.01	0.01	0.02	0.03	0.03	0.04	0.06	0.08	0.11	0.14	0.2	0.29
10		0	0	0.01	0.01	0.01	0.02	0.02	0.03	0.04	0.05	0.07	0.1	0.13	0.17	0.26
12		0	0	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.04	0.06	0.08	0.11	0.14	0.21
15		0	0	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.05	0.06	0.08	0.11	0.17

Effek: Klein

Medium

Groot

Wanneer met  $k$  steekproewe gewerk word, kan  $\eta_{meerv}^2 = 1 - \Lambda$  beraam word deur:

$$\hat{\eta}_{meerv}^2 = 1 - \hat{\Lambda} \left( \frac{u+v}{v} \right)^r, \quad (7.14)$$

deur van 'n resultaat van Cohen & Nee (1984) gebruik te maak. Hier is

$$u = m(k-1), \quad v = wr + 1 - m(k-1)/2, \quad \text{met } w = n - (m+k)/2 - 1. \quad (7.15)$$

Substitusie van (7.14) in (7.12) gee die beramer

$$\hat{f}_r^2 = \hat{\Lambda}^{-1/r} \left( \frac{v}{u+v} \right) - 1 \quad . \quad (7.16)$$

Soortgelyk aan die F-statistiek in (5.12) by meervoudige lineêre regressie, is die statistiek

$$F(u, v) = (\hat{\Lambda}^{-1/r} - 1) \frac{v}{u} \quad , \quad (7.17)$$

onder die nulhipotese van geen verskil in gemiddelde vektore, benaderd F-verdeel met u en v vryheidsgrade (Cohen, 1988) . As die nulhipotese nie waar is nie, volg verder dat  $F(u, v)$  nie-sentraal-F verdeel is met u en v vryheidsgrade en niesentraliteitsparameter

$$nsp = (\Lambda^{-1/r} - 1)(u + v + 1) \quad . \quad (7.18)$$

Soos in paragraaf 5.2.5, kan 'n  $(1-\alpha)100\%$  VI bepaal word vir  $\eta_{meerv}^2$  en  $f_r^2$  as volg:

- Bepaal soos in Bylae B 'n  $(1-\alpha)100\%$  VI vir nsp:  $(nsp_O, nsp_B)$  .
- Uit (7.18) volg:  $\Lambda = \left( 1 + \frac{nsp}{u+v+1} \right)^{-r}$  en  $f_r^2 = \frac{nsp}{u+v+1}$  . (7.19)
- Deur substitusie van  $nsp_O$  en  $nsp_B$  in (7.19), word die VI's  $[\eta_{meerv}^2(O), \eta_{meerv}^2(B)]$  en  $[f_r^2(O), f_r^2(B)]$  vir  $\eta_{meerv}^2$  en  $f_r^2$  onderskeidelik bepaal.

Hierdie  $(1-\alpha)100\%$  VI vir  $\eta_{meerv}^2$  en  $f_r^2$  kan bereken word deur die SAS-program **VI\_eta2\_meerv** op die handleiding se webblad te gebruik.

Ander effekgrootte-indekse gebaseer op Wilks se  $\Lambda$  ( $\tau^2$ ), Hotelling-Lawley se zeta-statistiek ( $\zeta^2$ ) en op Pillai se statistiek ( $\xi^2$ ), word deur Huberty (1994 : 194) gegee.

Omdat ook 'n benaderde  $(1-\alpha)100\%$  VI bereken kan word vir zeta ( $\zeta^2$ ), bespreek ons kortliks ook net hierdie indeks (kyk Steyn & Ellis, 2009).

Dit word gedefinieer as

$$\zeta^2 = \frac{U^{(s)}}{s + U^{(s)}} , \quad (7.20)$$

waar  $U^{(s)}$  die Hotelling-Lawley grootheid (kyk Bylae C) by  $m$ -veranderlike populasies is en  $s = \min(m, k-1)$ .

'n Beramer wat benaderd onsydig is vir  $\zeta^2$ , word gegee deur (kyk Bylae C.1):

$$\hat{\zeta}_1^2 = \frac{(n-k-m-1)\hat{U}^{(s)} - m(k-1)}{ns + (n-k-m-1)\hat{U}^{(s)} - m(k-1)} , \quad (7.21)$$

waar  $\hat{U}^{(s)}$  die Hotelling-Lawley-statistiek is – gebaseer op  $k$  steekproewe uit die  $m$ -veranderlike normaalpopulasies. Hierdie statistiek word gewoonlik saam met Wilks se lambda gegee as uitvoer van 'n MANOVA as 'n rekenaarpakket soos SAS, SPSS of STATISTICA gebruik word.

'n Benaderde  $(1-\alpha)100\%$  VI vir  $\zeta^2$  kan bereken word soortgelyk aan diè van  $\eta_{meerv}^2$  en  $f_r^2$ , deur die SAS-program **VI\_zeta\_kwadr1** op die handleiding se webblad. Dui hierdie interval met  $(\zeta_{01}^2, \zeta_{B1}^2)$  aan (kyk Bylae C.1).

Beide die beramer en VI is geskik om veral by kleiner steekproewe te gebruik.

Vir 'n groot, word 'n asimptoties-onsydige beramer (kyk Bylae C.2) verkry as:

$$\hat{\zeta}_2^2 = \frac{(n-k)\hat{U}^{(s)} - m(k-1)}{ns + (n-k)\hat{U}^{(s)} - m(k-1)} . \quad (7.22)$$

Beide die twee beramers vir  $\zeta^2$  kan negatief wees as  $\hat{U}^{(s)}$  baie klein is. Omdat  $\zeta^2$  teoreties nie-negatief is, neem ons dan die beraamde waarde as nul.

'n Asimptotiese  $(1-\alpha)100\%$  VI vir  $\zeta^2$  kan verkry word m.b.v. 'n SAS-program (**VI\_zeta\_kwadr2**) wat op die handleiding se webblad beskikbaar is. Dui hierdie interval aan met  $(\zeta_{02}^2, \zeta_{B2}^2)$  (kyk Bylae C.2).

#### Voorbeeld 7.4:

Beskou Voorbeeld F, Hoofstuk 3. Die resultate van 'n MANOVA op die veranderlikes S\_Cho, S\_Tri, HDL\_C en LDL\_C om die 3 aktiwiteitsgroepe te vergelyk, was:

$n = 1362$ ,  $\hat{\Lambda} = 0,88$ ,  $F(8; 2712) = 22,28$  ( $p < 0,0001$ ). Verder

$\hat{U}^{(2)} = 0,136$ ,  $F(8; 2710) = 22,95$  ( $p < 0,0001$ ).

$$r = \sqrt{\frac{4^2(3-1)^2 - 4}{4^2 + (3-1)^2 - 5}} = \sqrt{\frac{60}{15}} = 2, \quad u = 4(3-1) = 8, \quad w = 1362 - (4+3) / 2 = 1357,5,$$

$$v = 1357,5(2) + 1 - 4(2) / 2 = 2712.$$

$$\hat{f}_r^2 = 0,88^{-1/2} \left( \frac{2712}{8+2712} \right) - 1 = 1,066(0,997) - 1 = 0,063.$$

95% VI vir  $f_r^2$ : (0,045; 0,084) - met SAS-program **VI\_eta2\_meerv** bepaal.

$$\hat{\eta}_{meerv}^2 = 1 - 0,88 \left( \frac{8+2712}{2712} \right)^2 = 1 - 0,88(1,006) = 0,115.$$

95% VI vir  $\eta_{meerv}^2$ : (0,084; 0,149) - met SAS-program **VI\_eta2\_meerv** bepaal.

Beraming van  $\zeta^2$ : Omdat  $n$  baie groot is, neem



$$\zeta^2 = \frac{1359 \times 0,136 - 8}{1362 \times 1359 \times 0,136 - 8} = \frac{176,8}{2724 \times 176,8}$$

$$= 0,0609 \quad ,$$

met 95% VI benaderd as: (0,045 ; 0,079) wat beteken dat  $\zeta^2$  so klein as 0,045 en so groot as 0,079 kan wees met 95% waarskynlikheid □

#### 7.4 Effekgrootte-indekse by kanoniese korrelasies

In paragraaf 5.2 is  $\rho_{y.A}^2$  en  $R_{y.A}^2$  as effekgroottes beskou om die proporsie van  $y$  se variansie wat deur 'n meervoudige lineêre regressie-verband

$$\hat{y} = a + b_1x_1 + b_2x_2 + \dots + b_ux_u \quad ,$$

verklaar word in 'n populasie en ewekansige steekproewe respektiewelik. Die notasie 'A' is gebruik vir die versameling van voorspellers  $x_1, x_2, \dots, x_u$ , terwyl  $b_1, b_2, \dots, b_u$  die regressiekoëffisiënte of -gewigte wat só gekies is dat die korrelasie  $\rho_{y.A}$  of  $R_{y.A}$  tussen  $y$  en  $\hat{y}$  'n maksimum is. Die konstante  $a$  speel geen rol by die bepaling van  $\rho_{y.A}^2$  en  $R_{y.A}^2$  nie en kan as nul geneem word, wat daarop neerkom dat met  $y - \bar{y}$ ,  $x_1 - \bar{x}_1$ , ...,  $x_u - \bar{x}_u$  gewerk word i.p.v. met  $y$ ,  $x_1, \dots, x_u$ .

Gestel dat i.p.v. net een kriterium veranderlike  $y$ , daar 'n versameling B is bestaande uit  $y_1, y_2, \dots, y_v$ . Die vraag ontstaan nou hoe die verband tussen die twee versamelings van veranderlikes bepaal kan word?

Die eerste stap is om by elkeen van A en B die *kanoniese veranderlikes* of hoofkomponente te bepaal. Dit kom daarop neer dat vir A 'n lineêre kombinasie van  $x_1, x_2, \dots, x_u$ , nl.  $A_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1u}x_u$  bepaal word wat die maksimum proporsie variansie van die totale variansie in A verklaar. Daarna word 'n tweede

lineêre kombinasie  $A_2$  gevind wat die maksimum proporsie van die oorblywende variansie in A verklaar. So word voortgegaan tot 'n lineêre kombinasie  $A_u$  gevind is. Hierdie lineêre kombinasies van  $x_1, \dots, x_u$  staan bekend as *kanoniese veranderlikes* en daar is net soveel as wat daar oorspronklike veranderlikes is. Op dieselfde wyse kan vir versameling B die eerste kanoniese veranderlike  $B_1 = b_{11}y_1 + b_{12}y_2 + \dots + b_{1v}y_v$  en daarna  $B_2, \dots, B_v$  gevind word. Die korrelasies van  $x_1, \dots, x_u$  elk met  $A_1$ , staan bekend as die *beladings* (of struktuur-koëffisiënte) van  $x_1, \dots, x_u$  op die eerste kanoniese veranderlike. Die gemiddeld van die som van kwadrate van die beladings gee dan die *proporsie variansie* van versameling A wat deur  $A_1$  verklaar word en word deur  $VA_1$  aangedui. So kan  $VA_2, \dots, VA_u$  en ook  $VB_1, \dots, VB_v$  verkry word.

Die tweede stap is om die eerste paar kanoniese veranderlikes  $A_1$  en  $B_1$  elk te 'roteer' sô dat die korrelasie tussen hulle 'n maksimum is. Rotasie beteken dat die stelle gewigte  $a_{11}, \dots, a_{1u}$  en  $b_{11}, \dots, b_{1v}$  gemanipuleer word om *kanoniese gewigte* te kry wat sô is dat die proporsie variansies  $VA_1$  en  $VB_1$  onveranderd bly. Hierdie maksimum korrelasie heet die *eerste kanoniese korrelasie* en word aangedui deur  $\lambda_1$ . So kan  $\lambda_1, \dots, \lambda_s$  ook verkry word as die kanoniese korrelasies tussen die pare  $(A_2, B_2), \dots, (A_s, B_s)$ , waar s die minimum van u en v is.

Omdat die *eerste kanoniese korrelasie*  $\lambda_1$  gebaseer is op  $A_1$  en  $B_1$  wat die maksimum proporsies variansies uit hulle onderskeie versamelings verklaar, is dit 'n goeie maatstaf van verband tussen versamelings A en B. Die kwadraat  $\lambda_1^2$  gee dan die proporsie variansie wat  $A_1$  van  $B_1$  verklaar, ook die *eiewaarde* genoem.

Omdat  $A_1$  en  $B_1$  elk slegs gedeeltes van totale variansies in A en B verklaar (die proporsies  $VA_1$  en  $VB_1$ ), kan  $\lambda_1$  nie werklik gebruik word om die verband tussen die versamelings  $x_1, \dots, x_u$  en  $y_1, \dots, y_v$  te meet nie. Dit gee eerder 'n soort partiële korrelasie – die korrelasie tussen  $A_1$  en  $B_1$  sonder in ag neming van  $A_2, \dots, A_s$  en  $B_2, \dots, B_s$ . (die sg. “redundancy” of oortolligheid).

'n Beter maatstaf is die *oortolligheid van B gegee A* wat deur Stewart & Lowe (1968) gedefinieer word as:

$$\bar{R}_{B.A}^2 = \sum_{i=1}^s \lambda_i^2 VB_i = \sum_{i=1}^s \lambda_i^2 \left( \sum_{j=1}^v \ell_{ij}^2 \right) / v, \quad (7.23)$$

met  $\ell_{ij}$  die belading van  $y_j$  op  $B_i$ .

In hierdie formulering word versameling B as die kriterium veranderlikes en A as die voorspellers beskou sodat  $\bar{R}_{B.A}^2$  die *oortolligheid van die kriterium-versameling gegee die voorspellers*, gee. Indien A en B se rolle omruil, word

$$\bar{R}_{A.B}^2 = \sum_{i=1}^s \lambda_i^2 VA_i = \sum_{i=1}^s \lambda_i^2 \left( \sum_{j=1}^u \ell_{ij}^2 \right) / u, \quad (7.24)$$

met  $\ell_{ij}$  die belading van  $x_j$  op  $A_i$ .

'n Ander metode om  $\bar{R}_{B.A}^2$  te verkry is:

$$\bar{R}_{B.A}^2 = \sum_{i=1}^v R_i^2 / v \quad (7.25)$$

waar  $R_i$  die meervoudige korrelasie koëffisiënt is van  $x_1, \dots, x_u$  op  $y_i$ . Vandaar die notasie  $\bar{R}_{B.A}^2$ . Netso volg

$$\bar{R}_{A.B}^2 = \sum_{i=1}^u R_i^2 / u, \quad (7.26)$$

waar  $R_i$  die meervoudige korrelasie koëffisiënt is van  $y_1, \dots, y_v$  op  $x_i$ .

Die oortolligheid  $\bar{R}_{B.A}^2$  en  $\bar{R}_{A.B}^2$  word vir populasies sowel as steekproewe gedefinieer en omdat al verskil die data is wat gebruik word (volledige populasie-data of steekproef data), onderskei ons nie daartussen nie en volstaan met een notasie vir albei.

Om al die begrippe en die betekenis daarvan te illustreer, bespreek ons dit aan die hand van die volgende voorbeeld. Kanoniese korrelasies en oortolligheid kan baie moeilik met die hand bereken word, en omdat ons gewoonlik van 'n rekenaarpakket gebruikmaak om 'n kanoniese analise te doen, gee ons die resultate soos bv. deur STATISTICA verkry, in die voorbeeld.

**Voorbeeld 7.5:**

In Voorbeeld A, Hoofstuk 3 wil ons die verband tussen die BDI se voor-, na- en opvolgtoetse (BECKPRE, BECKPO1 en BECKPO2) aan die eenkant (B) en die van POMS\_A (TENSPRE, TENSPOS1 en TENSPOS2) en POMS\_D (DEPRE, DEPPOS1 en DEPPOS2) aan die ander kant (A), bepaal. Die resultate van 'n kanoniese analise wat met STATISTICA gedoen is, is as volg:

(a)

Canonical Analysis Summary (kdklerk_total scales)		
Canonical R: .81938		
Chi <sup>2</sup> (18)=70.097 p=0.0000		
Include condition: group=101 or group=103		
	Left Set	Right Set
N=50		
No. of variables	3	6
Variance extracted	100.000%	77.4621%
Total redundancy	48.4597%	38.5355%
Variables:	1	BECKPRE TENS PRE
	2	BECKPO1 DEPPRE
	3	BECKPO2 TENSPOS1
	4	DEPPOS1
	5	TENSPOS2
	6	DEPPOS2

(b)

Eigenvalues (kdklerk_total scales) Include condition: group=101 or group=103			
Root	Root 1	Root 2	Root 3
Value	0.671381	0.242453	0.183382

(c)

Factor Structure, left set (kdklerk_total scales) Include condition: group=101 or group=103			
Variable	Root 1	Root 2	Root 3
BECKPRE	0.348103	0.739241	-0.576496
BECKPO1	0.825877	0.336229	0.452633
BECKPO2	0.982646	-0.158999	-0.095525

(d)

Variance Extracted (Proportions), left set (kdklerk_total scales) Include condition: group=101 or group=103		
Factor	Variance extracted	Reddncy.
Root 1	0.589615	0.395856
Root 2	0.228269	0.055345
Root 3	0.182116	0.033397

(e)

Factor Structure, right set (kdklerk_total scales) Include condition: group=101 or group=103			
Variable	Root 1	Root 2	Root 3
TENSPRE	0.148037	0.627317	-0.444477
DEPPRE	0.311308	0.732679	-0.542715
TENSPOS1	0.772995	-0.102487	0.129677
DEPOS1	0.875580	0.369798	0.232106
TENSPOS2	0.776935	-0.181875	-0.112816
DEPOS2	0.877712	-0.006656	-0.323257

(f)

Variance Extracted (Proportions), right set (kdklerk_tota)		
Include condition: group=101 or group=103		
Variable	Variance extractd	Reddncy.
Root 1	0.476166	0.319688
Root 2	0.185120	0.044883
Root 3	0.113335	0.020784

(g)

Canonical Weights, left set (kdklerk_total scales)			
Include condition: group=101 or group=103			
Variable	Root 1	Root 2	Root 3
BECKPRE	0.059731	0.784921	-0.692047
BECKPO1	0.243237	0.799263	1.171768
BECKPO2	0.792069	-0.949809	-0.739669

(h)

Canonical Weights, right set (kdklerk_total scales)			
Include condition: group=101 or group=103			
Variable	Root 1	Root 2	Root 3
TENS PRE	-0.248088	0.461440	-0.12587
DEPPRE	0.170837	0.358266	-0.62371
TENSPOS1	0.378857	-0.855018	-0.34040
DEPPOS1	0.282165	1.130572	1.42870
TENSPOS2	0.180081	0.349648	0.18841
DEPPOS2	0.346035	-0.889216	-1.04977

Die tabel by (a) gee die eerste kanoniese korrelasie  $\lambda_1 = 0,819$ . Die tweede en derde kanoniese korrelasies is  $\lambda_2 = \sqrt{0,2425} = 0,492$  en  $\lambda_3 = \sqrt{0,1834} = 0,428$ , uit tabel (b) verkry, wat  $\lambda_1^2$ ,  $\lambda_2^2$  en  $\lambda_3^2$  gee. Let op dat  $s = 3$ , die minimum van die hoeveelheid veranderlikes in elke groep. Daarom is daar slegs 3 kanoniese korrelasies. Die BDI-groep se eerste kanoniese veranderlike  $B_1$  het die proporsie van 0,59 van die 3 veranderlikes verklaar (kyk tabel (d)) wat verkry is as die som van kwadrate van die eerste kolom van tabel (c), gedeel deur 3:  $VB_1 = (0,348^2 + 0,826^2 + 0,983^2) / 3 = 0,590$ . Tabel (c) gee die beladings van elke veranderlike op die kanoniese veranderlikes. Netso is  $VA_1 = 0,476$  uit tabel (f), die proporsie variansie wat  $A_1$  van die 6-POMS-veranderlikes verklaar. Die

kanoniese korrelasie  $\lambda_1 = 0,819$  is dus die korrelasie tussen  $A_1$  en  $B_1$  wat elkeen op sy beurt maar proporsies van 0,476 en 0,590 van hulle onderskeie groepe veranderlikes verklaar.

Die oortolligheid van B gegee A:

$\bar{R}_{B.A}^2 = 0,4846$  wat verkry word uit Tabel (a) en bereken kan word uit (7.24):

$$\begin{aligned}\bar{R}_{B.A}^2 &= 0,6714 \times 0,5896 + 0,2425 \times 0,2283 + 0,1834 \times 0,1822 \\ &= 0,3959 + 0,0554 + 0,0334 \\ &= 0,4847\end{aligned}$$

(let op dat die laaste kolom van (d) reeds die produkte hierbo gee).

Indien meervoudige regressies van al die POMS-veranderlikes op elk van die BDI-veranderlikes gedoen word, lewer dit die meervoudige korrelasies: 0,524 ; 0,723 ; 0,810 sodat  $\bar{R}_{B.A}^2 = (0,524^2 + 0,723^2 + 0,810^2) / 3 = 0,4844$ , die oortolligheid.

Indien die POMS as kriterium geneem word, is  $\bar{R}_{A.B}^2 = 0,3854$  (kyk tabel (a)).  $\square$

## 7.5 Riglyne vir meerveranderlike omnibus-effekte

Soos in die geval van die eenveranderlike omnibus-effek  $\eta^2$  en die beramers daarvan soos in Hoofstuk 6 bespreek, kom die vraag ook hier na vore wanneer is só 'n effek 'groot' of 'klein'? Kline (2004b) benader hierdie probleem deur eers na die geval van 2 meerveranderlike populasies te kyk, waarby  $\eta_{meerv}^2$  die proporsie variansie van die diskriminantfunksie (DF) is wat deur populasie-lidmaatskap verklaar word. Die DF is niks anders as die lineêre kombinasie van die uitkomsveranderlikes  $y_1, \dots, y_m$  wat die maksimum korrelasie met die digotome

groeperingsveranderlike gee nie. Daarom is  $\eta_{meerv}^2 = \lambda_1^2$ , met  $\lambda_1$  die kanoniese korrelasie tussen DF en die groeperingsveranderlike. Die oortoligheid van die  $m$  uitkomsveranderlikes, gegee die digotome groeperingsveranderlike, is dan  $\lambda_1^2 V_1$ , waar  $V_1$  die proporsie variansie van die DF is.

Kline(2004b) voer twee redes aan waarom  $\eta_{meerv}^2$  substansieel groter kan wees as enige individuele  $y_i$  se  $\eta^2$ :

- (a)  $\eta_{meerv}^2$  weerspieël die proporsie van veralgemeende variansie, wat op  $m$  veranderlikes gelyktydig gebaseer is;
- (b) Populasie-lidmaatskap verklaar net 'n gedeelte van die variansie in die DF, terwyl die DF op sy beurt ook net 'n gedeelte van die variansie van die stel uitkomsveranderlikes verklaar (kyk vorige paragraaf: die DF is 'n kanoniese veranderlike, wat geroteer is, met 'n variansie kleiner as die totale variansie van die uitkomsveranderlikes). Dus is  $\eta_{meerv}^2$  'n proporsie van 'n gedeelte van die variansie en moet dus groter wees as die ware proporsie (bv. as DF se variansie 3,5 is en die totale variansie van die uitkomsveranderlikes 5 is, terwyl as die variansie wat aan populasielidmaatskap toegeskryf kan word bv. 2 is, is  $\eta_{meerv}^2 = 2/3,5 = 0,57$ , wat duidelik groter is as wanneer dit as proporsie van die totale variansie van 5 geneem sou word).

Hierdie argumentasie van Kline kan uitgebrei word ook as  $k$  populasies betrokke is. In hierdie geval is daar dan 'n stel van  $k-1$  digotome groeperingsveranderlikes (ook skynveranderlikes genoem) wat populasie-lidmaatskap aandui. Gevolglik kan die verband tussen die  $m$  uitkomsveranderlikes en  $k-1$  groeperingsveranderlikes deur die kanoniese korrelasie  $\lambda_1, \dots, \lambda_s, s = \min(k-1, m)$ , beskryf word. Wilks se lambda kan in terme van kanoniese korrelasies geskryf word as:

$$\Lambda = (1 - \lambda_1^2)(1 - \lambda_2^2) \dots (1 - \lambda_s^2). \quad (7.27)$$



Omdat  $\lambda_1^2, \dots, \lambda_s^2$  elk 'n proporsie variansie van 'n kanoniese veranderlike, gebaseer op die uitkomsveranderlikes, t.o.v. 'n kanoniese veranderlike wat op groeperingsveranderlikes gebaseer is, vang dit nie al die proporsie variansie binne die uitkomsveranderlikes op nie. Gevolglik is die proporsie  $\eta_{meerv}^2$  groter as wanneer dit in terme van die uitkomsveranderlikes se variansie uitgedruk sou word.

Volgens die vorige paragraaf gee die oortolligheid van die *uitkomsveranderlikes*, *gegee die groeperingsveranderlikes* (stel A), die werklike proporsie. Dit kan dan dien as 'n effekgrootte-indeks en is:

$$\bar{R}_{y.A}^2 = \sum_{i=1}^s \lambda_i^2 V_i \quad , \quad (7.28)$$

met  $V_i$  die proporsie variansie deur die  $i$ -de kanoniese veranderlike van  $y_1, \dots, y_m$  verklaar. In die praktyk is dit makliker om eerder  $\bar{R}_{y.A}^2$  te bereken uit:

$$\bar{R}_{y.A}^2 = \sum_{i=1}^m \eta_i^2 / m \quad , \quad (7.29)$$

waar  $\eta_i^2$  die proporsie variansie toe te skryf aan populasie-lidmaatskap vir  $y_i$  is, die gewone effekgrootte-indekse soos in Hoofstuk 6 gedefinieer is. Let op dat  $\eta_i$  niks anders is as die meervoudige korrelasiekoëffisiënt van  $y_i$  op die  $k-1$  digotome groeperingsveranderlikes nie.

Indien met steekproewe gewerk word, word die oortolligheid verkry deur  $\eta_i^2$  met sy beramer  $\hat{\eta}_i^2$  te vervang.

### Voorbeeld 7.6

Beskou weer Voorbeeld 7.4. 'n MANOVA is op die veranderlikes S\_CHOL, S\_TRI, HDL\_C en LDL\_C gedoen om die 3 aktiwiteitsgroepe van mans te vergelyk. Met behulp van STATISTICA is die volgende resultate verkry:  $n=1362$ ,  $k=3$ ,  $m=4$ ,  $\hat{\Lambda}=0,8804$ . Daarna is vir elke veranderlike 'n ANOVA

gedoen met die volgende F-waardes: 60,49; 53,43; 24,59 en 40,70. Die beraming vir  $\eta_{meerv}^2$  was 0,115 (kyk Voorbeeld 7.4).

Gebruik ons (6.11) is die beraming van S\_CHOL se proporsie variansie wat deur populasie lidmaatskap verklaar word:

$$\hat{\eta}_1^2 = \frac{(n-k-2)F/(n-k)-1}{(n-k-2)F/(n-k)+(n-k)/(k-1)} = \frac{1354 \times 60,49 / 1356 - 1}{1354 \times 60,49 / 1356 + 1356 / 2} = 0,0804$$

Netso volg  $\hat{\eta}_2^2 = 0,0716$  ,  $\hat{\eta}_3^2 = 0,0335$  en  $\hat{\eta}_4^2 = 0,0552$ .

Die oortolligheid is dus

$$\bar{R}_{y.A}^2 = \frac{1}{4}(0,0804 + 0,0716 + 0,0335 + 0,0552) = 0,0602.$$

Waar 0,115 die proporsie variansie gee wat die 2 kanoniese veranderlikes kan toeskryf aan populasie-lidmaatskap, gee die oortolligheid eerder die proporsie variansie van die 4 uitkomsveranderlikes wat verklaar word, wat amper die helfte kleiner is as  $\eta_{meerv}^2$ . □

'n Alternatiewe indeks wat ook op  $\Lambda$  gebaseer is, is (Huberty, 1997: 194):

$$\tau^2 = 1 - \Lambda^{1/s}, \tag{7.30}$$

met beramer

$$\hat{\tau}^2 = 1 - \hat{\Lambda}^{1/s} \left( \frac{u+v}{v} \right)^{r/s}, \tag{7.31}$$

waar  $s = \min(m, k-1)$ , r gegee word deur (7.13) en u en v deur (7.15).

Uit (7.27) volg dus dat  $\Lambda^{1/s}$  die geometriese gemiddelde is van die  $(1 - \lambda_i^2)$ 'e.

Dit geld dat

$$\tau^2 < \eta_{meerv}^2 .$$

Omdat

$$\frac{1}{s} U^{(s)} = \frac{1}{s} \sum_{i=1}^s \frac{\lambda_i^2}{1-\lambda_i^2} , \text{ die rekenkundige gemiddelde van } \frac{\lambda_i^2}{1-\lambda_i^2} \text{ 's, is:}$$

$$\zeta^2 = \frac{1}{s} \sum_{i=1}^s \frac{\lambda_i^2}{1-\lambda_i^2} / \left( n + \frac{1}{s} \sum_{i=1}^s \frac{\lambda_i^2}{1-\lambda_i^2} \right) . \quad (7.32)$$

Hoewel  $\zeta^2$  ook van  $\lambda_1, \dots, \lambda_s$  afhang, is daar dus 'n ingewikkelder verband as by  $\eta_{meerv}^2$  en  $\tau^2$ .

Die heel eenvoudigste funksie van  $\lambda_1, \dots, \lambda_s$  is gebaseer op Pillai se statistiek (Huberty, 1994: 194):

$$\xi^2 = \frac{1}{s} \sum_{i=1}^s \lambda_i^2 , \quad (7.33)$$

met as spesiale geval Roy se statistiek

$$\theta = \lambda_1^2 . \quad (7.34)$$

Let op dat waar  $\xi^2$  die gemiddeld van die  $\lambda_i^2$ 's is, is die oortoligheid volgens (7.28) 'n *geweeegde gemiddelde* van die  $\lambda_i^2$ 's met gewigte die proporsie variansies  $V_i$ 's deur die kanoniese veranderlikes verklaar. Indien die variansie deur die eerste kanoniese veranderlike verklaar,  $V_i$ , naby is aan 1, is  $\theta$ ,  $\xi^2$  en  $\bar{R}_{y.A}^2$  omtrent dieselfde, en kan enige een as 'n effekgrootte-indeks gebruik word.

Omdat elkeen van die indekse  $\eta_{meerv}^2$ ,  $\zeta^2$ ,  $\tau^2$ ,  $\xi^2$  en  $\theta$  funksies is van die kanoniese korrelasies op 'n ander manier, en almal tussen 0 en 1 lê, kan dit as verskillende proporsies beskou word, hoe groter die waarde hoe groter die effek. Almal gee egter proporsies in terme van die s kanoniese veranderlikes wat na rotasie maksimum korrelasies met die groepeeringsveranderlike gee. Hierdie kanoniese veranderlikes vang egter nie al die variansie wat in die m

uitkomsveranderlikes is, op nie, behalwe wanneer  $m \leq k-1$ . In sulke gevalle is  $s = m$  en verklaar  $\lambda_1, \dots, \lambda_s$  al die variansie van  $y_1, \dots, y_m$ .

Opsommend kan gesê word dat dit moeilik is om riglyne te gee vir wanneer die omnibus effek 'klein' of 'groot' is. Die redes is:

- Die indekse  $\eta_{meerv}^2$ ,  $\tau^2$ ,  $\zeta^2$ ,  $\xi^2$  en  $\theta$  is op die MANOVA-statistieke van Wilks, Hotelling-Lawley, Pillai en Roy gebaseer, wat hoewel almal funksies is van die kanoniese korrelasies  $\lambda_1, \dots, \lambda_s$ , dit *verskillende* funksies is. Dus is almal se waardes verskillend.
- Die kanoniese korrelasies meet die verband tussen die  $s$  diskriminantfunksies op die  $m$  uitkomsveranderlike gebaseer, en die stel van  $k-1$  digotome groeperingsveranderlikes. Dit gee nie direk die verband met die  $m$  uitkomsveranderlikes nie en dus is die proporsies  $\eta_{meerv}^2$ ,  $\tau^2$ ,  $\zeta^2$ ,  $\xi^2$  en  $\theta$  almal te groot.

Die oortoligheid  $\bar{R}_{y.A}^2$  gee egter 'n proporsie variansie waar dit gekompenseer is vir die gedeeltes van die variansies wat deur die diskriminantfunksies verklaar word. Daarom stel ons voor dat  $\bar{R}_{y.A}^2$  as die *primêre effekgrootte-indeks* gebruik word om proporsie van die uitkomsveranderlikes gesamentlik, wat toegeskryf kan word aan populasie-lidmaatskap, te meet. Omdat dit die gemiddelde van die eenveranderlike  $\eta^2$ 'e is, sou dieselfde riglyne as in Hoofstuk 6 gebruik kon word. Omdat só 'n gemiddelde deur een baie groot  $\eta^2$ -waarde beïnvloed kan word, moet hierdie riglyne egter met groot omsigtigheid bejeën word.

### **Voorbeeld 7.7:**

Beskou Voorbeeld 7.4. Die volgende tabelle gee die resultate soos met STATISTICA verkry:

Multivariate Tests of Significance (dvdwesthuizen) Sigma-restricted parameterization Effective hypothesis decomposition Include condition: geslag=1						
Effect	Test	Value	F	Effect df	Error df	p
Intercept	Wilks	0.03053	10766.06	4	1356	0.00
	Pillai's	0.96947	10766.06	4	1356	0.00
	Hotelling	31.75828	10766.06	4	1356	0.00
	Roy's	31.75828	10766.06	4	1356	0.00
Akt_grp	Wilks	0.88044	22.28	8	2712	0.00
	Pillai's	0.11982	21.62	8	2714	0.00
	Hotelling	0.13550	22.95	8	2710	0.00
	Roy's	0.13330	45.22	4	1357	0.00

Test of SS Whole Model vs. SS Residual (dvdwesthuizen) Include condition: geslag=1											
Dependent Variable	Multiple R	Multiple R <sup>2</sup>	Adjusted R <sup>2</sup>	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
S_CHO	0.285917	0.081749	0.080397	1827693	2	913846.6	20529796	1359	15106.55	60.49342	0.000000
S_TRI	0.269990	0.072895	0.071530	1461133	2	730566.6	18583351	1359	13674.28	53.42632	0.000000
HDL_C	0.186886	0.034926	0.033506	30022	2	15010.8	829543	1359	610.41	24.59143	0.000000
LDL_C	0.237713	0.056508	0.055119	952893	2	476446.7	15910235	1359	11707.31	40.69651	0.000000

Tevore is verkry dat beraming vir

$$\eta_{meerv}^2 : \hat{\omega}_{meerv}^2 = 0,115, \text{ met } 95\% \text{ VI as } (0,084; 0,149)$$

$$\zeta^2 : \hat{\zeta}^2 = 0,0609, \text{ met } 95\% \text{ VI as } (0,045; 0,079).$$

Vir Wilks se lambda as 0,8804 volg die beramer van  $\tau^2$ :

$$\begin{aligned} \hat{\tau}^2 &= 1 - 0,8804^{1/2} \\ &= 0,0617. \end{aligned}$$

Uit Pillai se statistiek as 0,1198 volg die beraming van  $\xi^2$  as

$$\hat{\xi}^2 = \frac{1}{2} \times 1198 = 0,0599, \text{ terwyl Roy se statistiek as } 0,1333 \text{ die beraming } \hat{\theta} = 0,1333$$

gee.

Uit die tweede tabel volg direk die waardes van  $\hat{\eta}_i^2$  as die aangepaste  $R^2$ -waardes, d.i. 0,0804 ; 0,0715 ; 0,0335 en 0,0551, soos tevore in Voorbeeld 7.4 bereken. Die gemiddeld lewer die oortolligheid

$$\bar{R}_{y.A}^2 = 0,0602 .$$

Al hierdie waardes gee proporsies van variansies van S\_CHOL, S\_TRIG, HDL\_C en LDL\_C wat toegeskryf kan word aan die aktiwiteitsgroepe. Die oortolligheidsindeks lewer 0,06 op wat ook naastenby deur  $\hat{\zeta}^2$ ,  $\hat{t}^2$ ,  $\hat{\xi}$  gegee word en dus in terme van die riglyne in Hoofstuk 6, dui dit op 'n medium effek.

□