Een vergelijking van methoden voor enkelvoudige variantieanalyse zonder de vooronderstelling van gelijke populatievarianties

TECHNISCHE HOGESCHOOL EINDHOVEN

Onderafdeling der Wiskunde

VAKGROEP KANSREKENING, STATISTIEK EN OPERATIONS RESEARCH

Memorandum COSOR 79-03

Een vergelijking van methoden voor enkelvoudige variantieanalyse zonder de vooronderstelling van gelijke populatievarianties

door

P.S.P.J. Werter

Stage Statistische Analyse o.l.v. Prof.dr. R. Doornbos
Drs. J.B. Dijkstra

Eindhoven, februari 1979

Nederland

Een vergelijking van methoden voor enkelvoudige variantieanalyse zonder de vooronderstelling van gelijke populatievarianties

door

P.S.P.J. Werter

Samenvatting

Uit de vergelijking van vier statistische methoden, en wel de methoden van Welch en van Brown-Forsythe en de eerste- en tweede orde methode van James, is de tweede orde methode van James als de beste naar voren gekomen om de nulhypothese te toetsen dat populatiegemiddelden gelijk zijn. Verondersteld is dat de waarnemingen onderling onafhankelijk normaal verdeeld met verwachting μ_i en variantie σ_i^2. Aan de populatievarianties worden geen eisen opgelegd.

1. Inleiding

Om de populaties te toetsen op gelijke populatiegemiddelden is in de literatuur geen exacte toets beschikbaar.
Aangezien wel enkele benaderingen bekend zijn om deze nulhypothese te toetsen, zijn in de stage deze statistische methoden met elkaar vergeleken qua gerealiseerde onbetrouwbaarheid en qua onderscheidend vermogen. We beperken ons hierbij wel tot kleine steekproeven. De waarnemingen worden verondersteld onderling onafhankelijk normaal verdeeld te zijn met verwachting μ_i en variantie σ_i^2 waarbij de populatievarianties verschillend kunnen zijn. In paragraaf 2 komen vier statistische methoden aan de orde. In de daarop volgende twee paragrafen worden deze methoden met elkaar vergeleken. In de laatste paragrafen worden daaruit conclusies getrokken en een modificatie van de tweede orde methode van James gepresenteerd. Als bijlage wordt de procedure Jamestest toegevoegd die gebaseerd is op deze modificatie.

2. Vier statistische methoden

2.1. De methode van Brown en Forsythe ([1])

Aangezien de ANOVA-F toetsingsgrootheid, die gebruikt wordt om bij gelijke populatievarianties de populaties op gelijkheid van populatiegemiddelde te toetsen, door de eis dat niets bekend is over de populatievarianties, nu niet toepasbaar is, hebben Brown en Forsythe aan dit probleem tegemoet willen komen door in de ANOVA-F toetsingsgrootheid de noemer te vervangen door een noemer waarvan de verwachting gelijk is aan de verwachting van de teller wanneer alle populatiegemiddelden gelijk zijn.

De ANOVA-F toetsingsgrootheid is

$$\underline{F} = \frac{\sum_i n_i(\underline{x}_{i.} - \underline{x}_{..})^2/(k - 1)}{\sum_i (n_i - 1)\underline{s}_i^2/(N - k)}$$

waarbij

$$N = \sum_i n_i$$

$$\underline{x}_{i.} = \sum_j \underline{x}_{ij}/n_i$$

$$\underline{x}_{..} = \sum_i \sum_j \underline{x}_{ij}/N$$

$$\underline{s}_i^2 = \sum_j (\underline{x}_{ij} - \underline{x}_{i.})^2/(n_i - 1)$$

en $\underline{x}_{ij}$ is de j-de waarneming uit de i-de populatie met $j = 1,\ldots,n_i$ en $i = 1,\ldots,k$.

De Brown-Forsythe toetsingsgrootheid wordt nu

$$\underline{BF} = \frac{\sum_i n_i(\underline{x}_{i.} - \underline{x}_{..})^2}{\sum_i (1 - n_i/N)\underline{s}_i^2} .$$

De verdeling van deze toetsingsgrootheid $\underline{BF}$ kan benaderd worden door de F-verdeling met voor teller en noemer respectievelijk (k - 1) en f vrijheidsgraden waarbij f gedefinieerd is door de Satterthwaite benadering ([6]) voor het aantal vrijheidsgraden

$$1/\underline{f} = \sum_i \underline{c}_i^2/(n_i - 1)$$

waarbij

$$\underline{c}_i = \frac{(1 - n_i/N)\underline{s}_i^2}{\sum_i (1 - n_i/N)\underline{s}_i^2} .$$

Brown en Forsythe verwerpen de nulhypothese dat alle populatiegemiddelden gelijk zijn als $\underline{BF} > F_f^{k-1}$.

2. De eerste en tweede orde methoden van James ([2])

Wanneer de populatievarianties bekend zijn, kunnen we de nulhypothese toetsen met de toetsingsgrootheid

$$\underline{z} = \sum_i \omega_i(\underline{x}_{i.} - \bar{\underline{x}})^2 \equiv \sum_i \omega_i \underline{x}_{i.}^2 - (\sum_i \omega_i \underline{x}_{i.})^2/\omega$$

waarbij

$$\omega_i = \frac{1}{\sigma_i^2/n_i}\ ,\ \omega = \sum_i \omega_i \text{ en } \bar{\underline{x}} = \sum_i \omega_i \underline{x}_{i.}/\omega\ .$$

Onder de nulhypothese is deze toetsingsgrootheid $\underline{z}$ $\underline{\chi}^2$-verdeeld met $k - 1$ vrijheidsgraden d.w.z. $P(\underline{z} \leq \chi^2) = p$ waarbij χ^2 kritieke waarde is bij gegeven onbetrouwbaarheid $\alpha = 1 - p$. Wanneer de steekproeven groot zijn mogen we de grootheden ω_i en ω wel vervangen door de benaderingen $\underline{w}_i = \frac{1}{\underline{s}_i^2/n_i}$ en $\underline{w} = \sum_i \underline{w}_i$ waarbij $\underline{s}_i^2$ een schatter is voor σ_i^2.

Met deze benaderingen kunnen we dan toch de nulhypothese toetsen. Aangenomen was echter wel dat de steekproeven groot zijn. Omdat we uitgegaan zijn van kleine steekproeven mogen we dit nu echter niet zomaar doen.

James heeft hiervoor een oplossing gevonden door de omzetting van ω_i en ω in w_i en w en de kritieke waarde χ^2 te vervangen door een kritieke waarde $h(\alpha)$ bij onbetrouwbaarheid α, zodanig dat

$$P(\sum_i \underline{w}_i \underline{x}_{i.}^2 - (\sum_i \underline{w}_i \underline{x}_{i.})^2/\underline{w} \leq h^*(\alpha)) = \alpha\ .$$

James geeft voor de grootheid $h(\alpha)$ twee representaties: de ene tot de orde (-1) in ν_i $(= n_i - 1)$, $i = 1,\dots,k$:

$$h^*(\alpha) = \chi^2[1 + \frac{3\chi^2 + (k+1)}{2(k^2 - 1)} \sum_i \frac{1}{\nu_i}(1 - \frac{\underline{w}_i}{\underline{w}})^2 + O(\nu^{-2})]$$

waarbij χ^2 de kritieke waarde is bij gegeven α en $(k - 1)$ vrijheidsgraden.
De hieruit afgeleide toets wordt nu: verwerp de nulhypothese als

$$\sum_i \underline{w}_i \underline{x}_{i.}^2 - (\sum_i \underline{w}_i \underline{x}_{i.})^2/\underline{w} > h(\alpha)$$

waarbij $h(\alpha)$ gelijk is aan $h^*(\alpha)$ zonder de term $O(\nu^{-2})$.

De andere representatie gaat tot de orde (-2) in $\nu_i = (n_i - 1)$, $i = 1,\dots,k$

$$h^*(\alpha) = \chi^2 + \tfrac{1}{2}(3\chi_4 + \chi_2)\sum_i \frac{1}{\nu_i}(1 - \frac{w_i}{w})^2 + \{\tfrac{1}{16}(3\chi_4 + \chi_2)^2(1 - \frac{k-3}{\chi^2})(\sum_i \frac{1}{\nu_i}(1 - \frac{w_i}{w})^2)^2 +$$

$$+ \tfrac{1}{2}(3\chi_4 + \chi_2)[(8R_{23} - 10R_{22} + 4R_{21} - 6R_{12}^2 + 8R_{12}R_{11} - 4R_{11}^2) +$$

$$(\text{[illegible]}_{23} - \text{[illegible]}R_{22} + 2R_{21} - 2R_{12}^2 + 4R_{12}R_{11} - 2R_{11}^2)(\chi_2 - 1) +$$

$$\text{[illegible]}\ 4R_{12}R_{11} - 2R_{12}R_{10} - 4R_{11}^2 + 4R_{11}R_{10} - R_{10}^2)(3\chi_4 - 2\chi_2 - 1)] +$$

$$R_{\text{[illegible]}} - R_{\text{[illegible]}} + 3R_{21} - \text{[illegible]}_{20})(5\chi_6 + 2\chi_4 + \chi_2) +$$

$$\frac{\text{[illegible]}}{6}(R_{12}^2 \ \text{[illegible]}\ 4R_{23} + 6R_{22} - 4R_{21} + R_{20})(35\chi_8 + 15\chi_6 + 9\chi_4 + 5\chi_2) +$$

$$\frac{\text{[illegible]}}{6}(\text{[illegible]}R_{\text{[illegible]}} + 4R_{21} - R_{20} + 2R_{12}R_{10} - 4R_{11}R_{10} + R_{10}^2)(9\chi_8 - 3\chi_6 - 5\chi_4 - \chi_2) +$$

$$\frac{1}{4}(-R_{22} + R_{11}^2)(27\chi_8 + 3\chi_6 + \chi_4 + \chi_2) +$$

$$\frac{1}{4}(R_{23} - R_{12}R_{11})(45\chi_8 + 9\chi_6 + 7\chi_4 + 3\chi_2)\} + O(\nu^{-3})$$

waarbij

$$R_{st} = \sum_i \left(\frac{1}{\nu_i}\right)^s \left(\frac{w_i}{w}\right)^t \quad \text{en} \quad \chi_{2s} = \chi^{2s}/(k-1)(k+1)\ldots(k+2s-3)$$

[illegible] [illegible]tieke waarde is bij gegeven α.

[illegible] verwerp de nulhypothese, dat alle populaties gelijke ge- [illegible] de [illegible] als

$$\sum_i \text{[illegible]}_i \underline{x}_{i.}^2 - \left(\sum_i \underline{w}_i \underline{x}_{i.}\right)^2 / \underline{w} > h(\alpha)$$

[illegible] $h(\alpha)$ gelijk is aan $h^*(\alpha)$ zonder de term $O(\nu^{-3})$.

[illegible] van James zullen we aanduiden met James-1 voor de eerste orde en [illegible] tweede orde methode.

[illegible] van Welch ([3])

[illegible] James is Welch uitgegaan van de toetsingsgrootheid $\sum_i \underline{w}_i(\underline{x}_{i.} - \bar{\underline{x}})^2$ waarbij $\bar{\underline{x}} = \sum \underline{w}_i \underline{x}_{i.}/\underline{w}$.

[illegible] heeft hij een statistische grootheid $\underline{v}^2$ afgeleid gedefinieerd door

$$\underline{v}^2 = \frac{\sum_i \underline{w}_i(\underline{x}_{i.} - \bar{\underline{x}}_{.})^2/(k-1)}{[1 + \frac{2(k-2)}{k^2-1}\sum_i \frac{1}{\nu_i}(1 - \frac{\underline{w}_i}{\underline{w}})^2]} .$$

Onder de nulhypothese geldt in eerste benadering, d.w.z. tot de orde (-1) [illegible], $i = 1,\ldots,k$ dat $P(\underline{v}^2 > F_p) = \alpha$ waarbij F_p kritieke waarde is van een

F-verdeelde statistische grootheid bij onbetrouwbaarheid α en aantal vrijheidsgraden van teller en noemer respectievelijk $f_1 = k - 1$ en

$$f_2 = [\frac{3}{(k^2 - 1)} \sum_i \frac{1}{\nu_i}(1 - \frac{w_i}{w})^2]^{-1} .$$

Welch verwerpt nu de nulhypothese als $\underline{v}^2 > F_p$.

2.4. In zijn artikel bewijst Welch dat zijn methode equivalent is met de methoden van James tot de orde (-1) in ν_i, $i = 1,\ldots,k$. Vandaar dat in deze stage Welch vergeleken is met James-1 om te zien in hoeverre deze equivalentie opgaat voor kleine steekproeven.
Bovendien is de vergelijking met James-2 van belang omdat verwacht mag worden dat een tweede orde benadering betere resultaten zal geven.
Uit de structuur van de formules van Welch en Brown-Forsythe valt op te merken dat de gemiddelden van de verschillende steekproeven door Brown-Forsythe gewogen worden met de factoren n_i, terwijl Welch als wegingsfactoren de factoren $w_i = n_i/s_i^2$ gebruikt. Voor grote verschillen in steekproefgemiddelden, waarbij de afwijkende steekproefgemiddelden optreden bij de kleine steekproefvarianties, mag verwacht worden dat dit meer een vergrotende invloed heeft op de toetsingsgrootheid $\underline{v}^2$ bij de methode van Welch dan op de toetsingsgrootheid $\underline{BF}$ bij de methode van Brown en Forsythe. Het omgekeerde mag verwacht worden wanneer de grote verschillen in steekproefgemiddelden zo optreden dat de afwijkende steekproefgemiddelden voorkomen bij de grote steekproefvarianties.

3. Simulatieonderzoek naar de gerealiseerde onbetrouwbaarheid van de verschillende methoden

Om de nauwkeurigheid te onderzoeken waarmee tevoren vastgestelde onbetrouwbaarheden van 10%, 5% en 1% worden gerealiseerd is een simulatieonderzoek verricht aan de hand van 2500 herhalingen. Dit is als volgt gedaan: voor iedere set(n_i,μ_i,σ_i), $i = 1,\ldots,k$ zijn in totaal N $(= \sum_i n_i)$ waarnemingen gegenereerd volgens de Box-Müller techniek ([5]).
Bij de berekening van de gerealiseerde onbetrouwbaarheid op de Burroughs B7700 is gebruik gemaakt van de volgende procedures ([7], [9]):

CHISQUARESTATISTIC
FISHERPROBABILITY
NORMALARRAY .

De ingestelde onbetrouwbaarheid duiden we aan met nominale onbetrouwbaarheid. Wanneer alle steekproefgemiddelden gelijk aan nul zijn vinden we de volgende gerealiseerde onbetrouwbaarheid: (de met (*) en (**) aangeduide percentages zijn gerealiseerde onbetrouwbaarheden waarvan de relatieve frequentie meer dan $2\sqrt{pq/2500}$ respectievelijk $3\sqrt{pq/2500}$ boven de ingestelde onbetrouwbaarheid uitvalt).

Onderzoeken we in hoeverre de gerealiseerde onbetrouwbaarheid voor de verschillende methoden rond de nominale onbetrouwbaarheid van 10%, 5% en 1% fluctueert, dan vinden we de volgende resultaten:

nominale onbetrouwbaarheid	Brown-Forsythe	James-1	James-2	Welch
10%	7.72-12.48%	9.32-16.56%	8.88-11.12%	9.28-13.08%
5%	3.32-7.24%	4.60-10.40%	3.92-6.12%	4.48-6.88%
1%	0.44-2.12%	0.84-4.36%	0.68-1.96%	0.76-2.36%

Δ := gerealiseerde onbetrouwbaarheid - nominale onbetrouwbaarheid.
We kunnen tabel 1 (zie blz. 7) ook onderzoeken op de ongelijkheden

$$2\sqrt{pq/2500} \leq \Delta \leq 3\sqrt{pq/2500} \quad \text{en} \quad 3\sqrt{pq/2500} \leq \Delta \;,$$

om te zien hoe goed de methoden zich gedragen.

	$2\sqrt{pq/2500} \leq \Delta \leq 3\sqrt{pq/2500}$	$3\sqrt{pq/2500} \leq \Delta$
Brown-Forsythe	9	10
James-1	5	27
James-2	3	1
Welch	11	7

Uit deze resultaten blijkt de tweede orde methode van James het minst naar boven af te wijken van de nominale onbetrouwbaarheid. De tweede orde methode van James is te prefereren boven de andere drie methoden.
Onderzoeken we in hoeverre de eerste orde methode van James equivalent is met de methode van Welch dan blijkt dat beide methoden niet als equivalent mogen worden beschouwd voor kleine steekproeven. Voor kleine steekproeven levert de methode van Welch beduidend betere resultaten.

aantal steekproeven	steekproef grootten	standaard afwijkingen	Brown-Forsythe 10%	5%	1%	James-1 10%	5%	1%	James-2 10%	5%	1%	Welch 10%	5%	1%
4	4,4,4,4	1,1,1,1	7.72	3.48	0.44	12.96**	7.40**	2.32**	10.28	4.64	0.84	9.96	4.52	0.76
		1,2,2,3	9.84	4.80	0.96	13.88**	8.56**	3.12**	11.08	5.84	1.32	11.36*	5.84	1.12
4	4,6,8,10	1,1,1,1	8.08	4.16	0.64	11.44*	6.44**	1.80**	9.96	4.56	1.20	10.28	4.96	1.28
		1,2,2,3	9.56	5.16	1.00	10.00	5.56	1.60**	9.12	4.72	1.00	9.16	4.72	1.00
		3,2,2,1	10.24	5.64	1.24	12.64**	7.48**	3.08**	10.24	5.64	1.52*	10.92	6.32**	1.68**
4	10,10,10,10	1,1,1,1	9.60	4.64	1.24	10.68	5.60	1.24	10.44	5.36	0.88	10.48	5.36	0.92
		1,2,2,3	10.80	6.12*	1.72**	10.40	5.92**	1.28	9.72	5.52	0.84	9.92	5.56	0.92
4	10,15,15,20	1,1,1,1	9.04	4.68	0.92	9.64	5.04	1.28	9.52	4.88	1.12	9.52	4.88	1.16
		1,2,2,3	10.68	5.96*	1.48*	10.40	5.12	1.36	10.16	5.00	1.28	10.24	5.00	1.32
		3,2,2,1	10.12	4.84	1.44*	10.24	5.00	1.16	9.72	4.72	0.96	9.84	4.84	1.00
4	20,20,20,20	1,1,1,1	9.20	4.80	1.12	9.32	4.88	1.00	9.28	4.80	0.92	9.28	4.84	0.92
		1,2,2,3	10.80	5.96*	1.48*	10.04	4.60	0.84	9.96	4.48	0.76	9.96	4.48	0.76
6	4,4,4,4,4,4	1,1,1,1,1,1	8.04	3.32	0.44	15.04**	8.92**	3.44**	9.84	5.28	1.12	11.52*	6.12*	1.44*
		1,1,2,2,3,3	9.44	4.64	1.04	16.56**	10.40**	4.36**	11.12	6.12*	1.96**	13.08**	6.88**	2.36**
6	4,6,8,10,12,14	1,1,1,1,1,1	8.56	4.32	0.60	11.52*	6.80**	2.00**	9.56	5.04	1.28	10.20	6.04*	1.44*
		1,1,2,2,3,3	10.16	5.88*	1.48*	10.76	5.36	1.28	8.88	3.92	0.68	9.48	4.72	0.88
		3,3,2,2,1,1	10.32	5.72	1.48*	12.20**	7.80**	2.76**	9.84	5.40	1.44*	11.12*	6.72**	2.16**
6	10,10,10,10,10,10	1,1,1,1,1,1	10.48	5.12	0.84	11.60*	6.60**	1.72**	11.00	5.84	1.24	11.20*	6.00*	1.36
		1,1,2,2,3,3	12.48**	6.84**	2.12**	12.12**	6.72**	1.56*	11.00	5.76	1.16	11.76*	6.24*	1.32
6	10,10,15,15,20,20	1,1,2,2,3,3	12.52**	7.24**	1.92**	10.36	5.20	0.88	9.76	4.76	0.76	10.04	5.00	0.84
		3,3,2,2,1,1	11.44*	6.60**	1.68**	10.16	5.60	1.24	9.40	4.88	1.08	9.92	5.24	1.20

4. Simulatieonderzoek naar het onderscheidend vermogen van de verschillende methoden

Op dezelfde manier als bij het simulatieonderzoek naar de gerealiseerde onbetrouwbaarheid van de verschillende methoden is voor de verschillende sets (n_i,μ_i,σ_i) het percentage verwerpingen van de nulhypothese berekend.
Het onderzoek vergelijkt nu voornamelijk sets waarvan niet alle gemiddelden gelijk zijn. Wel gaan we weer uit van de tevoren vastgestelde onbetrouwbaarheden van 10%, 5% en 1%. De resultaten voor de verschillende sets bij 2500 herhalingen vinden we in tabel 2 (zie blz. 9).

Vergelijken we de uitkomsten van de vier methoden dan zien we dat geen methode uniform meer onderscheidend vermogen bezit dan de drie andere methoden. Op grond hiervan kan geen voorkeur voor een van de vier worden uitgesproken.
In paragraaf 2.4 werd uit de structuur van de formules van Welch en Brown-Forsythe de verwachting gedestilleerd dat wanneer grote verschillen in steekproefgemiddelden optreden waarbij de extreme steekproefgemiddelden optreden bij de kleine varianties, dit meer een vergrotende invloed heeft op de grootheid v^2 bij de methode van Welch dan op de grootheid BF bij de methode van Brown en Forsythe. Voor het onderzoek naar het onderscheidend vermogen betekent dat dat dan een groter percentage verwerpingen van de nulhypothese wordt gevonden bij Welch dan bij Brown en Forsythe.
In de practijk blijkt dit inderdaad juist te zijn. Ook het geval dat bij grote verschillen in steekproefgemiddelden de extreme steekproefgemiddelden optreden bij de grote varianties, blijkt in de practijk een groter percentage verwerpingen van de nulhypothese voor Brown en Forsythe te geven dan voor Welch; dit alles in overeenstemming met wat verwacht werd.
Ook Ekbohm ([4]) komt in zijn vergelijkingsonderzoek tussen o.a. Welch en Brown-Forsythe tot eens luidende conclusies.

5. Conclusies

Aangezien in het simulatieonderzoek naar de gerealiseerde onbetrouwbaarheid een duidelijke voorkeur voor de tweede orde methode van James gebleken is en in het onderzoek naar het onderscheidend vermogen geen duidelijke voorkeur voor een van de vier methoden kon worden aangegeven, lijkt het het verstandigst bij een keuze uit de vier methoden de tweede orde methode van James te kiezen. In de literatuur komt nog geen vergelijking van de tweede orde methode van James met andere methoden voor om de nulhypothese, dat populaties gelijk populatiegemiddelde hebben, te toetsen.

aantal	steekproef	steekproef	standaard	Brown-Forsythe			James-1			James-2			Welch		
steekproeven	grootten	gemiddelden	afwijkingen	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
4	4,4,4,4	3,0,0,0	1,1,1,1	98.16	93.80	70.32	96.48	91.92	75.48	94.72	86.84	58.80	94.88	86.48	54.56
		5,0,0,½		100	100	99.16	100	99.96	98.48	99.96	99.64	94.76	99.96	99.68	93.44
		3,0,0,0	1,2,2,3	50.28	31.16	9.00	83.52	72.04	44.20	77.72	60.28	25.36	77.92	59.88	23.48
		0,0,0,3		44.04	30.64	11.48	40.08	28.72	14.44	33.60	22.72	7.92	34.04	22.68	7.00
		5,0,0,½		89.36	75.24	36.88	99.64	98.60	91.20	99.20	97.08	76.64	99.36	97.08	72.88
		½,0,0,5		77.48	63.52	74.72	66.04	52.44	31.84	58.48	43.72	19.24	58.96	43.44	17.92
4	4,6,8,10	3,0,0,0	1,1,1,1	99.72	98.80	90.12	98.16	95.40	84.04	97.28	92.88	74.88	97.60	93.52	77.64
		0,0,0,3		100	100	99.32	100	100	99.92	100	100	99.12	100	100	99.64
		3,0,0,0	1,2,2,3	71.88	54.28	21.56	94.48	89.12	70.68	93.48	86.96	64.36	93.60	87.28	64.84
		0,0,0,3		83.92	73.76	49.72	69.16	55.24	27.04	67.12	50.40	19.88	67.08	51.32	20.52
		5,0,0,½		99.64	97.88	82.08	100	99.96	98.52	100	99.88	97.44	100	99.88	97.60
		½,0,0,5		99.72	98.92	94.88	97.04	92.92	76.84	96.44	91.48	68.72	96.52	91.56	69.56
		3,0,0,0	3,2,2,1	47.76	34.80	16.16	41.12	30.00	15.68	36.36	24.12	10.44	37.24	25.76	11.52
		0,0,0,3		81.76	67.04	35.36	99.08	97.04	83.76	98.60	94.64	67.68	98.72	95.40	72.52
		5,0,0,½		82.88	71.20	44.88	72.32	60.60	38.00	67.52	51.64	26.64	69.00	54.28	29.52
		½,0,0,5		99.12	95.88	79.00	100	100	100	100	100	99.84	100	100	99.88
6	4,6,8,10,12,14	3,0,0,0,0,0	1,1,1,1,1,1	99.48	99.16	94.20	97.76	94.72	84.52	96.88	91.60	73.44	97.48	93.76	79.24
		3,0,0,0,0,0	1,1,2,2,3,3	67.00	48.96	19.00	97.08	93.72	80.84	96.04	90.76	73.08	96.60	92.44	76.96
		3,0,0,0,0,0	3,3,2,2,1,1	45.76	33.56	15.52	39.64	29.92	16.96	34.48	23.96	10.08	37.48	27.12	13.36

Dit is niet zo verwonderlijk als men zich realiseert dat ten tijde van het ontstaan van deze toetsingsmethoden alles nog met tafelrekenmachines moest worden berekend. Zo komt het ook dat de meeste computerpakketten als toetsingsmethode de methode van Welch bevatten.
Uit de stage is gebleken dat beter de tweede orde methode van James toegevoegd kan worden.
Aan de methode van James zijn echter de nadelen verbonden dat de berekening zeer bewerkelijk is en dat bij een gegeven nominale onbetrouwbaarheid de nulhypothese alleen wordt verworpen of geaccepteerd. Voor iedere α moet de berekening opnieuw gebeuren. Dit in tegenstelling tot de methoden van Welch en Brown-Forsythe waarbij voor de verschillende α's de kritieke waarden zo in een F-tabel kunnen worden opgezocht als men de grootheden eenmaal heeft berekend.
In de volgende paragraaf wordt daarom een modificatie van de tweede orde methode van James gegeven die aan dit bezwaar tegemoet komt.

6. Modificatie van de tweede orde methode van James

Een nadeel van de tweede orde methode van James (trouwens ook van de eerste orde methode) is dat na berekening van de toetsingsgrootheid niet in een tabel die α opgezocht kan worden waarvoor de nulhypothese nog net verworpen wordt dan wel geaccepteerd.
Vandaar dat de tweede orde methode van James als volgt gemodificeerd is: los α op uit de volgende vergelijking $p_1 - h(\alpha) = 0$ waarbij

$$\underline{p}_1 = \sum_i \underline{w}_i \underline{x}_{i.}^2 - \Big(\sum_i \underline{w}_i \underline{x}_{i.}\Big)^2/\underline{w} \ .$$

Door de bisectiemethode toe te passen op het interval $[0,1]$ wordt hierdoor in 11 stappen een oplossing α berekend met een nauwkeurigheid van 10^{-3}.
Aan de hand van deze overschrijdingskans α kan de gebruiker van deze methode dan zelf beslissen of hij de nulhypothese verwerpt dan wel accepteert.
Als bijlage bij dit stageverslag is een procedure Jamestest toegevoegd die gebaseerd is op deze modificatie van de tweede orde methode van James.

7. Literatuur

[1] Brown, M.B. en Forsythe, A.B.; The small sample behaviour of some statistics which test the equality of several means. Technometrics 16, 129-132.

[2] James, G.S.; The comparison of several groups of observations when the ratios of the population variances are unknown. Biometrika 38, 324-329.

[3] Welch, B.L.; On the comparison of several mean values: An alternative approach. Biometrika 38, 330-336.

[4] Ekbohm, G.; On testing the equality of several means with small samples. Biometrische Zeitschrift 18, 547-553.

[5] Box, G.E.P. en Müller, M.E.; A note on the generation of random normal deviates. The Annals of Mathematical Statistics, 29, 610-513.

[6] Satterthwaite, F.E.; Synthesis of variance. Psychometrika 6, 309-316.

R.C. Informatie Technische Hogeschool Eindhoven:

[7] Pp. 4.11, aug. 1977: Verdelingsfuncties.

[8] Pp. 4.14, dec. 1976: Variantieanalyse.

[9] Pp. 4.15, aug. 1977: Aselecte Trekkingen.

Procedure Jamestest — Bijlage

Korte functiebeschrijving

Binnen een aantal groepen worden waarnemingen gegeven. Het is niet nodig dat het aantal waarnemingen per groep gelijk is. Berekend wordt de overschrijdingskans waarmee de nulhypothese dat de populatiegemiddelden gelijk zijn nog wordt verworpen. Het is niet noodzakelijk dat de populatievarianties gelijk zijn. Een normale verdeling binnen iedere groep is echter wel vereist.

Procedure heading

```
'PROCEDURE'Jamestest(X,GROUP,M,N,NUMBEROFGROUPS,EXCEEDENCE);
'VALUE'M,N,NUMBEROFGROUPS;
'REAL''ARRAY'X[*];
'INTEGER''ARRAY'GROUPS[*];
'INTEGER'M,N,NUMBEROFGROUPS;
'REAL'EXCEEDENCE;
```

Formele parameters

X	bevat bij aanroep de waarnemingen.
GROUP	bevat bij aanroep voor ieder element van x het nummber van de groep waartoe dit element behoort.
M,N	kleinste, respectievelijk grootste, index voor X en Group.
NUMBEROFGROUPS	bevat bij aanroep het aantal groepen. Deze groepen zijn genummerd van 1 tot NUMBEROFGROUPS.
EXCEEDENCE	bevat na afloop de overschrijdingskans.

Methode

afkortingen s_i^2 is de steekproefvariantie binnen de i-de groep

$$f_i = n_i - 1$$

$$w_i = n_i/s_i^2$$

$$w = \sum_i w_i$$

$$R_{st} = \sum_i \left(\frac{1}{f_i}\right)^s \left(\frac{w_i}{w}\right)^t$$

$$\chi_{2s} = \chi^{2s}/(k-1)(k+1)\ldots(k+2s-3)$$ waarbij

k = NUMBEROFGROUPS en χ^2 = CHISQUARESTATISTIC(EXCEEDENCE,k-1,0.001);

$$p_1 = \sum_i w_i (x_{i.} - \Sigma w_i x_{i.}/w)^2$$ en

$$h = \chi^2 + \frac{1}{2}(3\chi_4 + \chi_2)(\sum_i \frac{1}{f_i}(1 - \frac{w_i}{w})^2 + \{\frac{1}{16}(3\chi_4 + \chi_2)^2(1 - \frac{k-3}{2})(\sum_i \frac{1}{f_i}(1 - \frac{w_i}{w})^2)^2 +$$

$$\frac{1}{2}(3\chi_4 + \chi_2)[(8R_{23} - 10R_{22} + 4R_{21} - 6R_{12}^2 + 8R_{12}R_{11} - 4R_{11}^2) +$$

$$(2R_{23} - 4R_{22} + 2R_{21} - 2R_{12}^2 + 4R_{12}R_{11} - 2R_{11}^2)(\chi_2 - 1) +$$

$$\frac{1}{4}(-R_{12}^2 + 4R_{12}R_{11} - 2R_{12}R_{10} - 4R_{11}^2 + 4R_{11}R_{10} - R_{10}^2)(3\chi_4 - 2\chi_2 - 1)] +$$

$$(R_{23} - 3R_{22} + 3R_{21} - R_{20})(5\chi_6 + 2\chi_4 + \chi_2) +$$

$$\frac{3}{16}(R_{12}^2 - 4R_{23} + 6R_{22} - 4R_{21} + R_{20})(35\chi_8 + 15\chi_6 + 9\chi_4 + 5\chi_2) +$$

$$\frac{1}{16}(-2R_{22} + 4R_{21} - R_{20} + 2R_{12}R_{10} - 4R_{11}R_{10} + R_{10}^2)(9\chi_8 - 3\chi_6 - 5\chi_4 - \chi_2) +$$

$$\frac{1}{4}(-R_{22} + R_{11}^2)(27\chi_8 + 3\chi_6 + \chi_4 + \chi_2) +$$

$$\frac{1}{4}(R_{23} - R_{12}R_{11})(45\chi_8 + 9\chi_6 + 7\chi_4 + 3\chi_2)\} .$$

Opgelost wordt EXCEEDENCE in de vergelijking $p_1 - h = 0$.

Opmerkingen.

1. Dit is geen exacte toets; echter beter dan Welchs ALTERNATIVE ten aanzien van een fout van de eerste soort.
2. Elke steekproef moet minstens 2 verschillende elementen bevatten.
3. Door gebruik te maken van de BISECTIE-methode wordt in 11 stappen voor EXCEEDENCE een nauwkeurigheid bereikt van 10^{-3}.

Literatuur

James, G.S.; The comparison of several groups of observations when the ratios of the populationvariances are unknown. (Biometrika 38, 324-329).