

Het gebruik van de toets van Kruskal & Wallis bij normale verdelingen met ongelijke varianties.

Jan B. Dijkstra, Technische Hogeschool, Eindhoven

Paul S.P.J. Werter, Hollandse Signaalapparaten, Hengelo

Samenvatting.

De toets van Kruskal & Wallis is niet ontwikkeld voor het op gelijkheid toetsen van gemiddelden bij normale verdelingen met ongelijke varianties. Toch lijkt het op het eerste gezicht aantrekkelijk om deze methode hiervoor te gebruiken. In het onderstaande wordt nagegaan of dat verstandig is.

Inleiding.

Beschouw k steekproeven met steekproefgrootte n_i voor $i = 1, \dots, k$.

De waarnemingen zijn x_{ij} voor $j = 1, \dots, n_i$ en aan iedere waarneming wordt het bijbehorende rangnummer R_{ij} toegekend. Bij gelijke waarnemingen worden de rangnummers gemiddeld.

De toetsingsgrootte van Kruskal & Wallis (1952) wordt nu gegeven door:

$$K = \frac{12}{N(N+1)} \sum_{i=1}^k n_i (\bar{R}_i - \bar{R})^2.$$

Hierbij geldt $N = \sum_{i=1}^k n_i$ en $\bar{R} = \frac{N+1}{2}$.

$\bar{R}_i$ stelt het gemiddelde van de rangnummers in de i -de groep voor.

Met behulp van K kan men nu de nulhypothese H_0 toetsen dat alle steekproeven afkomstig zijn uit dezelfde populatie. Eenvoudig is in te zien dat de toetsingsgrootte in het bijzonder gevoelig is voor onderlinge verschuivingen van de mediaan. Bij symmetrische verdelingen hebben verschillen in de varianties hoegenaamd geen invloed op K . Daarom lijkt het aantrekkelijk om de toets van

Kruskal & Wallis te gebruiken voor de nulhypothese H_0' van gelijke populatiegemiddelden μ_i voor normale verdelingen met ongelijke varianties σ_i^2 . De symmetrie van de normale verdeling maakt immers dat de mediaan en het gemiddelde samenvallen, zodat K behoorlijk gevoelig zal zijn voor afwijkingen van deze nulhypothese.

De verdeling van K onder de nulhypothese H_0

Als alle steekproeven groot zijn, dan geldt volgens Hajek en Sidak (1967):

$$\underline{K} \approx \chi^2_{k-1}$$

Voor kleinere steekproeven zijn in elementaire statistiek-boeken tabellen opgenomen met de kritieke waarde van K voor zekere k , n_i en onbetrouwbaarheid α . Naast het gebruik van deze tabellen en de χ^2 -verdeling is er volgens Wallace (1959) ook nog toetsing mogelijk met de β -verdeling, en wel als volgt:

$$\underline{K} \approx \frac{p}{q}$$

Hierbij geldt $p = (k-1)d/2$ en $q = (N-k)d/2$.

De hulpgrootte d wordt gegeven door:

$$d = 1 - \frac{6}{5} \frac{N+1}{N-1} \frac{1}{\frac{6}{5} + \frac{N}{1-T}}$$

$$\text{Met } T = \frac{N(N+1)}{2(k-1)(N-k)} \left(\sum_{i=1}^k \frac{1}{n_i} - \frac{k^2}{N} \right)$$

Andere toetsen voor de nulhypothese H_0

Het op gelijkheid toetsen van populatiegemiddelden bij normale verdelingen gebeurt doorgaans met de klassieke enkelvoudige variantieanalyse. Een vooronderstelling bij deze toets is echter de gelijkheid van de populatievarianties. Simulatiestudies van Brown & Forsythe (1974) en Ekbohm (1976) hebben duidelijk aangetoond dat afwijkingen van deze vooronderstelling de kans op een fout van de eerste soort behoorlijk oncontroleerbaar maken.

Een exacte toets met redelijk onderscheidend vermogen, die gebaseerd is op de F-verdeling, bestaat niet voor de hier behandelde nulhypothese. Scheffé (1944) heeft aangetoond dat er voor het geval $k = 2$ geen symmetrische t-toets te vinden is, en dit bewijs kan gegeneraliseerd worden tot een F-toets voor $k > 2$. Symmetrie wil in dit verband zeggen dat de waarde van de toetsingsgrootte ongevoelig is voor permutaties binnen de steekproeven. Asymmetrische toetsen maken doorgaans weinig efficiënt gebruik van de beschikbare gegevens als de steekproefgroottes aanzienlijke verschillen vertonen. Daarom zijn er al verscheidene benaderende toetsen voorgesteld. Vermoedelijk de meest bekende methoden hiervoor zijn van James (1951), Welch (1951), Brown & Forsythe (1974) en Scheffé (1959). Bij een vergelijking van deze toetsen door Dijkstra & Werter (1981) komt de methode van James er qua controle over de onbetrouwbaarheid het beste van af. De onderscheidende vermogens ontlopen elkaar niet veel, zodat het interessant lijkt om de toets van James te vergelijken met die van Kruskal & Wallis voor normale verdelingen met ongelijke varianties.

De nominale en de gerealiseerde onbetrouwbaarheid.

De formules bij de toets van James zijn nogal bewerkelijk en de lezer wordt dan ook verwezen naar zijn oorspronkelijke artikel waarin James laat zien hoe een toetsingsgrootte en de bijbehorende kritieke waarde uit de steekproeven kunnen worden gevonden. In het bedoelde artikel worden twee toetsen behandeld, maar hier wordt alleen aandacht besteed aan wat James aanduidt als zijn tweede orde methode.

Een variant op op deze toets die direct de overschrijdingskans geeft, kan men vinden in het reeds genoemde artikel van Dijkstra & Werter.

Voor een vergelijking van de toetsen van James en Kruskal & Wallis is een simulatie-onderzoek gedaan naar de gerealiseerde onbetrouwbaarheid. Er zijn hiervoor 42 patronen beschouwd voor k en $\sigma_1^2, \dots, \sigma_k^2$ die samen pretenderen ongeveer datgene te dekken dat men in de praktijk zou kunnen ontmoeten. Iedere simulatie berust op 2500 herhalingsen. De nulhypothese van gelijke populatiegemiddelden geldt steeds en de steekproeven worden volgens de methode van Box en Muller (1958) gegenereerd uit normale verdelingen met verschillende σ_1^2 . De resultaten staan in de volgende tabel, waarin ieder interval gegeven is door het kleinste en het grootste verwerpspercentage van de 42 beschouwde gevallen.

nominaal	James	Kruskal & Wallis	
		met χ^2	met β
10%	8.88-11.12	5.12-16.92	2.04-10.08
5%	3.92-6.12	1.92-9.36	0.92-5.96
1%	0.68-1.96	0.16-2.28	0.20-1.56
gerealiseerde onbetrouwbaarheid			

Uit deze tabel kan men de volgende conclusies trekken:

1. Kruskal & Wallis met χ^2 verwerpt de ware nulhypothese voor sommige patronen van k en $\sigma_1^2, \dots, \sigma_k^2$ te vaak en is dus ongeschikt voor deze toepassing.
2. Kruskal & Wallis met β is nogal conservatief bij dit gebruik. Het toepassen hiervan is dus wel gerechtvaardigd, maar het onderscheidend vermogen zal vermoedelijk te wensen overlaten. In de volgende paragraaf wordt hierop nog teruggekomen.

Het is goed nog even stil te staan bij de oorzaak van de geconstateerde afwijkingen. De toets van Kruskal & Wallis is ontworpen voor de nulhypothese:

H_0 : Alle K steekproeven komen uit dezelfde verdeling.

De toepassing is hier echter anders. Voor normale verdelingen met ongelijke populatievarianties wordt getoetst:

H_0' : $\mu_1 = \dots = \mu_k$

Het slechte gedrag van de toets van Kruskal & Wallis met χ^2 zal zeker niet alleen veroorzaakt worden door de verschillen tussen H_0 en H_0' .

Vanwege de symmetrie van de normale verdeling kan namelijk in die richting wel enige robuustheid verwacht worden. Veel storender zullen de soms nogal kleine steekproeven zijn, waaraan in dit onderzoek aandacht is besteed. Als alle n_i groot zijn dan valt de toets van Kruskal & Wallis met χ^2 misschien nog wel mee.

Het onderscheidend vermogen.

Bij dit onderzoek is H_0' getoetst voor 57 verschillende patronen voor k , σ_1^2 , ..., σ_k^2 en μ_1 , ..., μ_k . Ook hier werden steeds 2500 trekkingen uit normale verdelingen genomen. Bij de analyse worden drie gevallen onderscheiden.

1e geval: $\sigma_1^2 = \sigma_j^2$

nominaal	James	K&W met β
10%	94.72-100	91.92-100
5%	86.84-100	66.08-100
1%	58.80-99.12	17.68-99.92
verwerpingspercentages		

De eerder genoemde bewerkelijkheid van de toets van James betreft de berekening van de kritieke waarde uit de gekozen onbetrouwbaarheid en de steekproef. De toetsingsgrootte zelf is nogal eenvoudig:

$$J = \sum_{i=1}^k \omega_i (\bar{x}_i - \bar{x})^2$$

Met $\omega_i = \frac{n_i}{s_i^2}$, $\omega = \sum_{i=1}^k \omega_i$ en $\bar{x} = \sum_{i=1}^k \omega_i \bar{x}_i / \omega$. Deze formule suggereert

dat het onderscheidend vermogen van de toets van James gering is als extreme gemiddelden optreden bij grote steekproefvarianties s_i^2 . Van die situatie ondervindt de toets van Kruskal & Wallis veel minder hinder omdat daarbij gewogen wordt met n_i in plaats van n_i/s_i^2 .

2e geval: $\sigma_i^2 \neq \sigma_j^2$ en de afwijkende gemiddelden treden op bij grote varianties

nominaal	James	K&W met β
10%	33.60-96.44	35.88-94.20
5%	22.72-91.48	24.12-90.08
1%	7.92-68.72	7.32-77.32
verwerpingspercentages		

3e geval: $\sigma_1^2 \neq \sigma_j^2$ en de afwijkende gemiddelden treden op bij kleine varianties

nominaal	James	K&W met β
10%	77.72-100	29.72-100
5%	60.28-100	14.36-100
1%	25.36-99.84	2.56-99.72
verwerpingspercentages		

Uit het voorgaande kan men de volgende conclusies trekken:

1. Het onderscheidend vermogen (bij het toetsen van H_0) van de toets van James is beter dan die van Kruskal & Wallis met β , tenzij de extreme gemiddelden optreden bij grote varianties. In dat geval ontlopen de beide toetsen elkaar niet veel.
2. Als geen programmatuur voor de toets van James voorhanden is (erg gecompliceerd algoritme; veel rekenwerk), dan is de toets van Kruskal & Wallis met β een redelijk alternatief voor de hier besproken toepassing.

Literatuur

- [1] Kruskal, W.H. en W.A. Wallis
Use of ranks in one-criterion variance analysis
 Journal of American Statistical Association 47 (1952), 583-621
 en 48 (1953), 907-911
- [2] Hajek, J en Z. Sidak
Theory of Rank Tests
 Academia, Praag (1967)
- [3] Wallace, D.L.
Simplified Beta-approximations to the Kruskal-Wallis H-test
 Journal of American Statistical Association 54 (1959), 225-230

- [4] Brown, M.B. en A.B. Forsythe
The small sample behaviour of some statistics which test the equality of several means
 Technometrics 16 (1974), 129-132
- [5] Ekbohm, G.
On testing the equality of several means with small samples
 The Agricultural College of Sweden (1976), Uppsala, 547-553
- [6] Scheffé, H.
On solutions of the Behrens-Fisher problem, based on the t-distribution
 The annals of the Mathematical Statistics 14 (1943), 35-44
- [7] Scheffé, H.
A note on the Behrens-Fisher problem
 The annals of the Mathematical Statistics 15 (1944), 430-432
- [8] James, G.S.
The comparison of several groups of observations when the ratios of the population variances are unknown
 Biometrika 38 (1951), 324-329
- [9] Welch, B.L.
On the comparison of several mean values: an alternative approach
 Biometrika 38 (1951), 330-336
- [10] Dijkstra, Jan B. en Paul S.P.J. Werter
Testing the equality of several means when the population variances are unequal
 Geaccepteerd door Communications in Statistics; binnenkort te verschijnen
- [11] Box, G.E.P. en M.E. Müller
A note on the generation of random normal deviates
 The annals of the Mathematical Statistics 29 (1958), 610-613