

STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 168 (M 61)

Een parameter vrije toets tegen verloop voor groepen waarnemingen.



Een parameter vrije toets tegen verloop voor groepen waarneming n.¹⁾

Wij beschouwen k onderling onafhankelijke stochastische grootheden $\alpha_1, \alpha_2, \dots, \alpha_k$. Van de grootheid α_i zijn n_i onderling onafhankelijke waarnemingen $\alpha_{i,1}, \dots, \alpha_{i,n_i}$ gegeven ($i=1, \dots, k$).

De hypothese H_0 die wij willen toetsen luidt, dat $\alpha_1, \dots, \alpha_k$ dezelfde waarschijnlijkheidsverdeling bezitten, terwijl de alternatieve hypothesen inhouden, dat de stochastische grootheden $\alpha_1, \dots, \alpha_k$ een stijgend of dalend verloop vertonen, gedefinieerd door:

$$(1) \quad \sum_{i < j} \left\{ P[\alpha_i < \alpha_j] - P[\alpha_i > \alpha_j] \right\} \neq 0.$$

De toetsingsgrootte $\underline{W}$ wordt als volgt gedefiniëerd. Stel voor

$i < j$ is $\underline{U}_{i,j}$ het aantal paren waarnemingen $(\alpha_{i,\lambda}, \alpha_{j,\mu})$ met $\alpha_{i,\lambda} < \alpha_{j,\mu}$ vermeerderd met de helft van het aantal paren $(\alpha_{i,\lambda}, \alpha_{j,\mu})$ met $\alpha_{i,\lambda} = \alpha_{j,\mu}$ ($\lambda \leq n_i, \mu \leq n_j$).²⁾ Stel verder

$$(2) \quad \underline{W}_{i,j} \stackrel{\text{def}}{=} 2 \left\{ \underline{U}_{i,j} - \mathcal{E}(\underline{U}_{i,j} | H_0) \right\} = 2 \underline{U}_{i,j} - n_i n_j,$$

dan is

$$(3) \quad \underline{W} \stackrel{\text{def}}{=} \sum_{i < j} \frac{\underline{W}_{i,j}}{n_i n_j}.$$

Als onder de $N = \sum_i n_i$ waarnemingen h groepen gelijke waarnemingen optreden en als t_ℓ het aantal waarnemingen in de ℓ^e groep is ($\ell=1, 2, \dots, h$) dan geldt

$$(4) \quad \mathcal{E}[\underline{W} | t_1, t_2, \dots, t_h; H_0] = 0.$$

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.

2) $\underline{U}_{i,j}$ is dus de toetsingsgrootte van de toets van WILCOXON, toegepast op de waarnemingen van α_i en α_j en $\underline{U}_{j,i} = n_i n_j - \underline{U}_{i,j}$.

3) De hier beschreven toets is een wijziging van een door T.J. TERPSTRA [4] ontwikkelde toets voor dit probleem. Hij gebruikt de toetsingsgrootte $\underline{T} = \sum_{i < j} \underline{W}_{i,j}$. Door de toetsingsgrootte $\underline{W}$ te gebruiken bereikt men, dat de toets bruikbaar is voor een klasse van alternatieve hypothesen, die niet afhangt van de verhoudingen der steekproefgrootten (zie ook [5]).

$$(5) \left\{ \begin{array}{l} \sigma^2[\underline{W} | t_1, t_2, \dots, t_h; H_0] = \\ \frac{\{N^3 - \sum t_l^3 - 3(N^2 - \sum t_l^2)\} \sum \frac{(k+1-2i)^2}{n_i} - \{2(N^2 - \sum t_l^2) - 3N(N^2 - \sum t_l^2)\} \sum_{i < j} \frac{1}{n_i n_j}}{3N(N-1)(N-2)} \end{array} \right.$$

Treden er geen gelijke waarnemingen op, dan is $t_l = 1$ voor iedere l en:

$$(6) \quad \sigma^2[\underline{W} | H_0] = \frac{1}{3} \left\{ \sum_i \frac{(k+1-2i)^2}{n_i} + \sum_{i < j} \frac{1}{n_i n_j} \right\}.$$

Als H_0 juist is, is de grootheid $\underline{W}$ voor grote waarden van N bij benadering normaal verdeeld met gemiddelde en variantie volgens (4) en (5).

Het is duidelijk, dat $\underline{W}$ in het algemeen grote positieve waarden zal aannemen als er een stijgend en grote negatieve als er een dalend verloop is; de tweezijdige kritieke zône bestaat dus uit grote waarden van $|\underline{W}|$.

Opmerkingen.

1. Als $k=2$, dus als wij twee stochastische grootheden α_1 en α_2 hebben met resp. n_1 en n_2 waarnemingen, dan geldt:

$$(7) \quad \underline{W} = \frac{2 \underline{U} - n_1 n_2}{n_1 n_2},$$

waarin $\underline{U}$ de toetsingsgrootheid van de toets van WILCOXON is, toegepast op de steekproeven van α_1 en α_2 . Uit (4) en (5) volgt:

$$E[\underline{W} | t_1, t_2, \dots, t_h; H_0] = 0,$$

$$\sigma^2[\underline{W} | t_1, t_2, \dots, t_h; H_0] = \frac{N^3 - \sum t_l^3}{3 n_1 n_2 N(N-1)} \quad . \text{ (zie memorandum S 47 (M 7))}$$

Als er in dit geval geen gelijke waarnemingen optreden, kan men de exacte overschrijdingskans opzoeken in tabellen van de exacte verdeling van $\underline{U}$ (zie b.v. [1]).

2. Als $n_i = 1$ voor iedere i dan is

$$\underline{W} = \sum_{i < j} \text{sgn}(\alpha_j - \alpha_i),$$

waarin

$$\text{sgn } z \stackrel{\text{def}}{=} \begin{cases} 1 & \text{als } z > 0 \\ 0 & \text{als } z = 0 \\ -1 & \text{als } z < 0. \end{cases}$$

Uit (4) en (5) volgt:

$$E[W | t_1, t_2, \dots, t_h; H_0] = 0,$$

$$\sigma^2[W | t_1, t_2, \dots, t_h; H_0] = \frac{2(N^2 - \sum t_\ell^2) + 3(N^2 - \sum t_\ell^2)}{18}$$

De toets is in dit geval identiek met de rangcorrelatie methode van KEDALL voor het geval, dat één der rijen bestaat uit de getallen $1, 2, \dots, k$ (zie memorandum S 47 (M 13)). Treden er in dit geval geen gelijke waarnemingen op en is $k \leq 40$ dan kan men de exacte overschrijdingskans opzoeken in de tabellen van de exacte verdeling van de toetsingsgrootte S van KENDALL (zie [2]).

Is $k \leq 10$ en $t_\ell \leq 3$ voor iedere ℓ dan kan men gebruik maken van de tabellen van de exacte verdeling van S van SILLITTO [3]

Literatuur

- [1] Van der Vaart, H.R., Gebruiksaanwijzing voor de toets van Wilcoxon, Rapport S 32 (M 4) van de Statistische Afdeling van het Mathematisch Centrum.
- [2] Kaarsmaker, L. en A. van Wijngaarden, Tables for use in rankcorrelation, Rapport R 73 van de Rekenafdeling van het Mathematisch Centrum.
- [3] Sillitto, G.P., The distribution of Kendall's coefficient of rankcorrelation in rankings containing ties, Biometrika 34 (1947), 36-40.
- [4] Terpstra, T.J., The asymptotic normality and consistency of Kendall's test against trend when ties are present in one ranking, Proc.Kon.Ned.Ak. 55 (1952).
- [5] Van Eeden, C., A test for the equality of probabilities against a class of specified alternatives, including trend, Rapport S 157(VP 3) van de Statistische Afdeling van het Mathematisch Centrum.