STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 145 (M 50)

Toets voor een generalisatie van het probleem van m rangschikkingen.

T.J. Terpstra

MATHEMATISCH CENTRUM,
2e Boerhaavestraat 49
Amsterdam - O.

S 14? (M 50)

Toets voor een generalisatie van het probleem van m rangschikkingen [1].

door T.J.Terpstra.

Gegeven zijn m waarnemers, die ieder steekproeven nemen van k stochastische grootheden.
De stochastische grootheden, waarvan de r^e waarnemer ($1 \leq r \leq m$) steekproeven neemt, duiden we aan met $\underline{x}_1^{(r)}, \ldots\ldots, \underline{x}_k^{(r)}$. De uitgebreidheden van de steekproeven noemen we $n_1^{(r)}, \ldots\ldots, n_k^{(r)}$.
De hypothese H_0 houdt nu in, dat voor iedere r de variabelen $\underline{x}_1^{(r)}, \ldots\ldots, \underline{x}_k^{(r)}$ dezelfde waarschijnlijkheidsverdeling bezitten.
Deze hypothese H_0 wensen we te toetsen tegen de alternatieve hypothese H_1, inhoudende dat voor iedere r de variabelen $\underline{x}_1^{(r)}, \ldots\ldots, \underline{x}_k^{(r)}$ niet dezelfde waarschijnlijkheidsverdeling bezitten, terwijl een overeenstemming bestaat tussen de m waarnemers, wat betreft de rangschikking naar opklimmende grootte van de variabelen $\underline{x}_1^{(r)}, \ldots\ldots, \underline{x}_k^{(r)}$.
De statistische grootheid $\underline{S}$, waarmee de hypothese H_0 tegen de alternatieve hypothese H_1 wordt getoetst, wordt op de volgende wijze gevormd.
Voor de r^e waarnemer worden de $n_h^{(r)}+n_j^{(r)}$ waarnemingen van elk paar steekproeven $\underline{x}_{h,1}^{(r)}, \ldots, \underline{x}_{h,n_h^{(r)}}^{(r)}$ en $\underline{x}_{j,1}^{(r)}, \ldots\ldots, \underline{x}_{j,n_j^{(r)}}^{(r)}$ ($1 \leq h < j \leq k$) naar opklimmende grootte gerangschikt en van rangnummers voorzien. Dit zijn de getallen 1 t/m $n_h^{(r)}+n_j^{(r)}$ als alle waarnemingen verschillend zijn; indien t waarnemingen aan elkaar gelijk zijn, krijgen zij alle als rangnummer het gemiddelde van de rangnummers, die ze gekregen zouden hebben als ze verschillend geweest waren. Voor de h^e steekproef wordt vervolgens de som van de rangnummers $\underline{R}_{h,j}^{(r)}$ bepaald. Uit deze grootheid vormen we de grootheid $\underline{\tilde{\mathcal{U}}}_{h,j}^{(r)}$ volgens $\underline{\tilde{\mathcal{U}}}_{h,j}^{(r)} = \underline{R}_{h,j}^{(r)} - \frac{1}{2} n_h^{(r)} (n_h^{(r)} + n_j^{(r)} + 1)$.
De toetsingsgrootheid $\underline{S}$ wordt nu gedefiniëerd als

$$\underline{S} = \sum_{r<s} \frac{\underline{S}_{r,s}^2}{\sigma_{r,s}^2} \quad ,$$

met

$$\underline{S}_{r,s} = 4 \sum_{h<j} \underline{\tilde{\mathcal{U}}}_{h,j}^{(r)} \cdot \underline{\tilde{\mathcal{U}}}_{h,j}^{(s)}$$

---

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

en

$$\sigma_{r.s}^{2} = \frac{1}{9}\left\{\sum_{h<j} n_h^{(r)} n_j^{(r)} \left(n_h^{(r)} + n_j^{(r)} + 1\right) \; n_h^{(s)} n_j^{(s)} \left(n_h^{(s)} + n_j^{(s)} + 1\right) + \right.$$

$$\left. 6 \sum_{h<i<j} n_h^{(r)} n_i^{(r)} n_j^{(r)} \cdot n_h^{(s)} n_i^{(s)} n_j^{(s)} \right\} .$$

In het bijzondere geval m= 2 , is de toetsingsgrootheid het kwadraat van de genormeerde toetsingsgrootheid voor de gegeneraliseerde rangcorrelatietoets van Kendall, beschreven in memorandum S 145 (M 49).

De variabele $\underline{S}$ bezit nu de eigenschap, dat ze onder de hypothese H_0 en voor grote waarden van k bij benadering een χ^2-verdeling bezit met $\nu = \frac{1}{2} m(m-1)$ graden van vrijheid.

Indien de alternatieve hypothese H_1 juist is, bezit $\underline{S}$ gemiddeld grotere waarden dan onder de hypothese H_0 . De hypothese H_0 wordt daarom verworpen ten gunste van H_1, indien de uit de waarnemingen bepaalde waarde van $\underline{S}$ groter is dan de kleinste waarde S_α, welke voldoet aan $P[\underline{S} \geq S_\alpha \mid H_0] \leq \alpha$.

Voor grote waarden van k is S_α bij benadering gelijk aan de waarde χ^2_α, welke gegeven wordt door

$$P\left[\underline{\chi}^2_{\frac{1}{2}m(m-1)} \geq \chi^2_\alpha \mid H_0\right] = \alpha$$

en gemakkelijk bepaald kan worden uit een tabel of nomogram van de χ^2- verdeling.

De overschrijdingskans k^* van de uit de steekproeven bepaalde waarde van S wordt gedefiniëerd door $k^* = P[\underline{S} \geq S \mid H_0]$ en kan voor grote k bij benadering worden bepaald uit een tabel van de χ^2- verdeling volgens

$$k^* = P\left[\chi^2_{\frac{1}{2}m(m-1)} \geq S \mid H_0\right] .$$

Litteratuur :

1. M.G.Kendall, Rank correlation methods, London 1952.
2. W.J.Dixon and F.J.Massey Jr., Introduction to statistical analysis, Mc.Graw-Hill Book Comp., 1951.