STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 73 (M 30)

Een parametervrije toets voor k steekproeven

T.J. Terpstra.

1952

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49
Amsterdam (O)
Statistische Afdeling

Februari 1952

S 73 (M 30)

Een parametervrije toets voor k steekproeven.[1)]

Door T.J. Terpstra.

Gegeven zijn k onafhankelijke steekproeven van k stochastische grootheden:

$x_{1,1}, x_{1,2}, \ldots, x_{1,n_1}$ van de grootheid $\underline{x}_1$,[2)]

$x_{2,1}, x_{2,2}, \ldots, x_{2,n_2}$ van de grootheid $\underline{x}_2$,

$x_{k,1}, x_{k,2}, \ldots, x_{k,n_k}$ van de grootheid $\underline{x}_k$.

De uitgebreidheden van de steekproeven zijn dus $n_1, n_2, \ldots, n_k$ en de eerste index van de waarneming $x_{i,j}$ geeft aan uit welke steekproef deze waarneming afkomstig is, terwijl de tweede index het nummer der waarneming binnen die steekproef aangeeft.

De hypothese H_0, die wij wensen te toetsen luidt, dat de waarnemingen van dezelfde stochastische grootheid afkomstig zijn. Anders uitgedrukt: H_0 houdt in, dat de stochastische grootheden $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k$ onderling onafhankelijk verdeeld zijn en alle dezelfde waarschijnlijkheidsverdeling bezitten.

Om deze hypothese H_0 te toetsen, maken we gebruik van een toetsingsgrootheid $\underline{U}^2$, die als volgt gevormd wordt.
Allereerst berekenen we het aantal waarnemingen uit de tweede steekproef, dat kleiner is dan $x_{1,1}$. (bij gelijkheid tellen wij ½ in plaats van 1). We noemen dit aantal $U_{1,2}(x_{1,1})$.
Vervolgens het aantal waarnemingen uit de tweede steekproef, dat kleiner is dan de waarde $x_{1,2}$ uit de eerste steekproef. Dit aantal noemen we $U_{1,2}(x_{1,2})$. Deze aantallen bepalen we eveneens met betrekking tot $x_{1,3}, x_{1,4}, \ldots, x_{1,n_1}$ en vinden dan $U_{1,2}(x_{1,3}), U_{1,2}(x_{1,4}), \ldots, U_{1,2}(x_{1,n_1})$.

---

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

2) Stochastische grootheden worden door onderstreping aangeduid.

Op dezelfde wijze bepalen we de grootheden $U_{1,3}, U_{1,4}, \ldots, U_{1,k}$, door de eerste steekproef met alle volgende steekproeven te vergelijken.

Daarna worden de grootheden $U_{2,3}, U_{2,4}, \ldots, U_{2,k}$ bepaald door de tweede steekproef met alle volgende steekproeven te vergelijken.

Hetzelfde doen we met betrekking tot de derde steekproef, etc.

De laatste grootheid, die we op deze wijze vormen is $U_{k-1,k}$.

Uit de grootheid $U_{1,2}$ berekenen we vervolgens $u_{1,2}$ met

$$u_{1,2} = U_{1,2} - \tfrac{1}{2}n_1 n_2,$$

waarin n_1 en n_2 de uitgebreidheden van de eerste en van de tweede steekproef zijn.

Evenzo berekenen we

$$u_{1,3} = U_{1,3} - \tfrac{1}{2}n_1 n_3,$$

$$u_{1,4} = U_{1,4} - \tfrac{1}{2}n_1 n_4,$$

-.-.-.-.-.-.-.-.-.-.-.

$$u_{k-1,k} = U_{k-1,k} - \tfrac{1}{2}n_{k-1}\, n_k.$$

Uit deze laatste grootheden wordt de toetsingsgrootheid $U_k^{\,2}$ gevormd volgens

$$U_k^2 = \frac{12}{n_1+n_2+\ldots n_k+1} \,.$$

$$\cdot\Big\{ \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \frac{n_1+n_2+\ldots+n_{i-1}+n_{i+1}+\ldots+n_{j-1}+n_{j+1}+\ldots+n_k+1}{n_i n_j} u^2_{i,j}$$

$$- 2 \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \sum_{j'=j+1}^{k} \frac{1}{n_i} u_{i,j}\, u_{i,j'}$$

$$-2 \sum_{j=3}^{k} \sum_{i=2}^{j-1} \sum_{i'=1}^{i-1} \frac{1}{n_j} u_{i,j}\, u_{i',j}$$

$$+2 \sum_{i=2}^{k-1} \sum_{h=1}^{i-1} \sum_{j=i+1}^{k} \frac{1}{n_i} u_{h,i}\, u_{i,j} \Big\} = 12 \sum_{i<j} \frac{u_{i,j}^2}{n_i n_j} - \frac{12}{N+1} \sum_i \frac{u_i^2}{n_i}$$

Indien de steekproeven gelijk zijn, dus $n_1 = n_2 = \ldots = n_k = n$ gaat deze uitdrukking over in

$$U_k^2 = \frac{12}{n^2(kn+1)} \Big[\{(k-2)n+1\} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} u^2_{i,j}$$

$$- 2n \cdot \Big\{ \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \sum_{j'=j+1}^{k} u_{i,j}\, u_{i,j'} + \sum_{j=3}^{k} \sum_{i=2}^{j-1} \sum_{i'=1}^{i-1} u_{i,j}\, u_{i',j} -$$

$$- \sum_{i=2}^{k-1} \sum_{h=1}^{i-1} \sum_{j=i+1}^{k} u_{h,i}\, u_{i,j} \Big\} \Big] .$$

Voor 3 steekproeven (k=3) vinden we voor U_k^2:

$$u_i = \sum_{h<i} u_{h,i} + \sum_{i<j} u_{i,j} ,$$

$$\underline{U}_3^2 = \frac{12}{n_1+n_2+n_3+1}\left(\frac{n_3+1}{n_1n_2} u_{1,2}^2 + \frac{n_2+1}{n_1n_3} u_{1,3}^2 + \frac{n_1+1}{n_2n_3} u_{2,3}^2 - \frac{2}{n_1}u_{1,2}u_{1,3} - \frac{2}{n_3} u_{1,3}u_{2,3} + \frac{2}{n_2} u_{1,2} u_{2,3}\right)$$

en voor 4 steekproeven, dus k = 4:

$$\underline{U}_4^2 = \frac{12}{n_1+n_2+n_3+n_4+1} \Big\{\left(\frac{n_3+n_4+1}{n_1n_2} u_{1,2}^2 + \frac{n_2+n_4+1}{n_1n_3} u_{1,3}^2 + \frac{n_2+n_3+1}{n_1n_4} u_{1,4}^2 + \frac{n_1+n_4+1}{n_2n_3} u_{2,3}^2 + \frac{n_1+n_3+1}{n_2n_4} u_{2,4}^2 + \frac{n_1+n_2+1}{n_3n_4} u_{3,4}^2\right)$$

$$-2\left(\frac{1}{n_1} u_{1,2} u_{1,3} + \frac{1}{n_1} u_{1,2} u_{1,4} + \frac{1}{n_1} u_{1,3} u_{1,4} + \frac{1}{n_2} u_{2,3} u_{2,4}\right)$$

$$-2\left(\frac{1}{n_3}u_{1,3} u_{2,3} + \frac{1}{n_4}u_{1,4} u_{2,4} + \frac{1}{n_4} u_{1,4} u_{3,4} + \frac{1}{n_4} u_{2,4} u_{3,4}\right)$$

$$+2\left(\frac{1}{n_2} u_{1,2} u_{2,3} + \frac{1}{n_2}u_{1,2} u_{2,4} + \frac{1}{n_3} u_{1,3} u_{3,4} + \frac{1}{n_3} u_{2,3} u_{2,4}\right)\Big\}.$$

Beschouwen wij de verzameling van alle bij het experiment mogelijke uitkomsten, dan bezit $\underline{U}^2$ op deze verzameling een waarschijnlijkheidsverdeling (vandaar de onderstreping van U^2). Indien de hypothese H_0, inhoudende, dat alle waarnemingen uit dezelfde verdeling afkomstig zijn, juist is, bezit de toetsingsgrootheid $\underline{U}^2$ voor grote waarden van $n_1, n_2, \ldots, n_k$ bij benadering een χ^2-verdeling met $\binom{k}{2} = \frac{k(k-1)}{2}$ vrijheidsgraden.[3)]
Indien de hypothese H_0 niet vervuld is bezit $\underline{U}^2$ gemiddeld grotere waarden, dan indien H_0 wel vervuld is.
De hypothese H_0 wordt daarom verworpen, indien de uit de steekproef bepaalde waarde van U^2 groter is dan een kritieke waarde U_α^2, met $P\left[\underline{U}^2 \geq U_\alpha^2\right] = P\left[\underline{\chi}^2_{\frac{k(k-1)}{2}} \geq U_\alpha^2\right] = \alpha$,

waarin α de, vooraf gekozen, onbetrouwbaarheidsdrempel is. De hypothese H_0 wordt dan verworpen met een onbetrouwbaarheid α.
De waarde U_α^2 kan gemakkelijk bepaald worden uit tabellen en nomogrammen van de χ^2-verdeling, evenals de overschrijdingskans k van U^2, die gedefiniëerd wordt door

$$k = P\left[\underline{U}^2 \geq U^2\right] = P\left[\underline{\chi}^2_{\frac{k(k-1)}{2}} \geq U^2\right].$$

---

3) De benadering is voldoende nauwkeurig, indien de aantallen $n_1, n_2, \ldots, n_k$ ieder minstens gelijk zijn aan 10.

Opmerking.

De geldigheid van de toets is, strikt genomen, slechts bewezen voor het geval, dat er geen gelijken optreden.

Tabellen en Nomogrammen.

M.G. Kendall, The Advanced Theory of Statistics, I, 1947, p. 444-446.

H. Cramèr, Mathematical Methods of Statistics, 1946, p. 559.

Statistica 1 (1946), p. 109.