STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 107 (M 42)

Rekenschema voor m rangschikkingen met ontbrekende waarnemingen

Ph. van Elteren.

1953

MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Rapport S 107 (M 42)

Rekenschema voor m rangschikkingen met ontbrekende waarnemingen.

door

Ph. van Elteren.

In dit rapport wordt een rekenschema beschreven voor een generalisatie van de methode der m rangschikkingen van FRIEDMANN, indien er waarnemingen ontbreken. Deze methode, waarbij ook meer dan één waarneming per hokje toegelaten wordt en waarover een publicatie van de hand van Ph. van Elteren en A.Benard in voorbereiding is, zetten wij hier uiteen met behulp van een voorbeeld betreffende.een enquête over geluidshinder. In dit rapport behandelen wij alleen het geval, dat er waarnemingen ontbreken. Het algemene schema, waarbij ook meer dan één waarneming per hokje toegelaten wordt, vindt men beschreven in rapport ZW 1952-022 van het Mathematisch Centrum.

Bij de enquête die hier als voorbeeld dient, zijn constructies vergeleken, die "geluidshinder" veroorzaken. Zes verschillende enquëtrices hebben alle constructies of een gedeelte ervan vergeleken en rangnummers toegekend op grond van de door hen geconstateerde "geluidshinder".

De resultaten van het onderzoek zijn gegeven in schema 1. In dit schema vindt men de rangnummers verminderd met het gemiddelde van de rangnummers der rangschikking waarin zij voorkomen.

<u>Schema 1</u>

		Constructies									Aantal "ties" van 1	Aantal "ties" van 2	$96K_\mu$ [1]
		1	2	3	4	5	6	7	8	9			
Enquêtrices	I	$+\frac{1}{2}$	$+2\frac{1}{2}$	$+1\frac{1}{2}$	$+3\frac{1}{2}$		$-1\frac{1}{2}$	$-\frac{1}{2}$	$-2\frac{1}{2}$	$-3\frac{1}{2}$	8	0	72
	II	+2	-3		-2		0	+3	+1	-1	7	0	64
	III	+2	0	+1	+3	+4	-1	$-2\frac{1}{2}$	-4	$-2\frac{1}{2}$	7	1	71
	IV	+1	-2		$-\frac{1}{2}$	$-\frac{1}{2}$		+2			5	0	48
	V	+1	0		-1		+2	-2	-3	+3	7	0	64
	VI	$+\frac{1}{2}$	$-2\frac{1}{2}$				$+1\frac{1}{2}$	$+2\frac{1}{2}$	$-\frac{1}{2}$	$-1\frac{1}{2}$	6	0	56
<u>Kolom-totalen:</u> $(\tilde{u}_1, \ldots, \tilde{u}_9)$		+7	-5	$+2\frac{1}{2}$	+3	$+3\frac{1}{2}$	+1	$+2\frac{1}{2}$	-9	$-5\frac{1}{2}$			

Wij hebben in dit schema, zoals gebruikelijk, de (gereduceerde) rangnummers behorend tot dezelfde rangschikking in één rij geplaatst.

Wij berekenen de getallen K_μ in de laatste kolom als volgt:

Zij: $g_{\mu 1}$ =aantal "ties van 1", d.w.z. aantal rangnummers dat maar eenmaal in de μ^e rij voorkomt

$g_{\mu 2}$ =aantal "ties van 2", d.w.z. aantal paren gelijke rangnummers in de μ^e rij

etc.,

en k_μ =aantal rangnummers voorkomend in de μ^e rij, dan is:

$$K_\mu = \frac{k_\mu^3 - g_{\mu 1} - 2^3 g_{\mu 2} - 3^3 g_{\mu 3} \cdots}{12\, k_\mu (k_\mu - 1)}$$

----

[1]) De factor 96 dient alleen, om geen gebroken getallen te schrijven.

Dus in ons geval:

$$K_1 = \frac{8^3 - 8}{12.\,8.\,7} = \frac{63}{84} = \frac{3}{4} \quad ,$$

$$K_3 = \frac{8^3 - 7 - 2^3}{12.\,8.\,7} = \frac{3}{4} - \frac{1}{96} = \frac{71}{96} \qquad \text{etc.}$$

Vervolgens berekenen wij de covarianties $\sigma_{\nu\nu'}$ van de kolomtotalen. Dit geschiedt als volgt: Zij $k_{\mu\nu}=1$, als het hokje (μ,ν) (μ^e rij, ν^e kolom) in het schema een rangnummer bevat, en $k_{\mu\nu}=0$ als het hokje leeg is [2]). Dan is:

$$\sigma_{\nu\nu'} = -\sum_{\mu=1}^{m} k_{\mu\nu}\, k_{\mu\nu'}\, K_\mu \qquad (\nu \neq \nu')$$

Zo vinden wij bijvoorbeeld:

$$\sigma_{12} = -\frac{72 + 64 + 71 + 48 + 64 + 56}{96} = -\frac{375}{96}$$

en

$$\sigma_{34} = -\frac{72 + 71}{96} = -\frac{143}{96}$$

Verder worden de varianties van de kolomtotalen:

$$\sigma_{\nu\nu} = \sum_{\mu=1}^{m} k_{\mu\nu}\,(k_\mu - k_{\mu\nu})\, K_\mu \ .$$

B.v.:

$$\sigma_1^2 = \frac{7.72 + 6.64 + 8.71 + 4.48 + 6.64 + 5.56}{96} = \frac{2312}{96}$$

en

$$\sigma_3^2 = \frac{7.72 + 8.71}{96} = \frac{760}{96} \ .$$

Ter controle heeft men:

$$\sum_{\nu'=1}^{n} \sigma_{\nu\nu'} = \sum_{\nu=1}^{n} \sigma_{\nu\nu'} = 0 \ .$$

De resultaten vindt men in schema 2.

----

[2]) In het algemeen: $k_{\mu\nu}$ =aantal rangnummers in het hokje (μ,ν). Zie voor het geval, dat $k_{\mu\nu}>1$ optreedt, rapport ZW 1952-022 van het Mathematisch Centrum.

Schema 2

$96\,\sigma_{\nu\nu'}$

$\nu=1,\dots,9$

$\nu'=1,\dots,9$

$\nu \backslash \nu'$	1	2	3	4	5	6	7	8	9	Kolomtotalen van schema 1 $\tilde{u}_\nu$
1	2312	-375	-143	-319	-119	-327	-375	-327	-327	+7
2	-375	2312	-143	-319	-119	-327	-375	-327	-327	-5
3	-143	-143	1072	-143	-71	-143	-143	-143	-143	$+2\frac{1}{2}$
4	-319	-319	-143	2032	-119	-271	-319	-271	-271	+3
5	-119	-119	-71	-119	760	-71	-119	-71	-71	$+3\frac{1}{2}$
6	-327	-327	-143	-271	-71	2120	-327	-327	-327	+1
7	-375	-375	-143	-319	-119	-327	2312	-327	-327	$+2\frac{1}{2}$
8	-327	-327	-143	-271	-71	-327	-327	2120	-327	-9
9	-327	-327	-143	-271	-71	-327	-327	-327	2120	$-5\frac{1}{2}$
$\tilde{u}_{\nu'}$ Kolomtotalen van schema 1	+7	-5	$+2\frac{1}{2}$	+3	$+3\frac{1}{2}$	+1	$+2\frac{1}{2}$	-9	$-5\frac{1}{2}$	0

De toetsingsgrootheid χ^2 wordt als volgt berekend:
Neem de matrix

$$\begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} & \tilde{u}_1 \\ \vdots & & \vdots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} & \tilde{u}_n \\ \tilde{u}_1 & \cdots & \tilde{u}_n & 0 \end{pmatrix}$$

laat hieruit <u>één</u> rij en één <u>kolom</u> (onverschillig welke, maar niet de laatste rij of de laatste kolom) weg en bereken de determinant Δ_1 van de verkregen matrix.
Neem vervolgens de matrix:

$$\begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix}$$

laat hieruit eveneens één rij en één kolom weg en bereken de determinant Δ_2 van de verkregen matrix.
Dan is $\chi^2 = \left|\frac{\Delta_1}{\Delta_2}\right|$ (de absolute waarde van $\frac{\Delta_1}{\Delta_2}$).
Het is onverschillig wélke rijen en kolommen weggelaten worden; χ^2 is daarvan <u>niet</u> afhankelijk.

In het voorbeeld van de geluidshinder kunnen wij Δ_1 en Δ_2 uit schema 2 als volgt bepalen:
Schrap b.v. kolom 9 en rij 9. Bepaal dan de determinant van het hele schema Δ_1'. Laat dan ook nog de buitenrand (kolom $\tilde{u}_\nu$ en rij $\tilde{u}_\nu$ weg) en bepaal de determinant Δ_2'. Dan is:

$$\Delta_1 = 96^{-7}\Delta_1'$$

$$\Delta_2 = 96^{-8}\Delta_2'$$

Dus is:

$$\chi^2 = \left|\frac{\Delta_1}{\Delta_2}\right| = 96\left|\frac{\Delta_1'}{\Delta_2'}\right| .$$

De berekening van de determinanten Δ_1' en Δ_2' hebben wij niet uitgevoerd.

De nulhypothese van de toets behelst, dat de rangschikkingen onafhankelijk zijn en dat in iedere rij alle permutaties van de rangnummers even waarschijnlijk zijn. Onder deze hypothese zal χ^2 asymptotisch voor grote m (het aantal rangschikkingen) en een aantal weinig ernstige beperkende voorwaarden betreffende de aantallen gelijke waarnemingen en open plaatsen (die wij hier niet vermelden) een χ^2-verdeling met $n-1$ (dus bij ons voorbeeld 8) vrijheidsgraden hebben. Grote waarden van χ^2 zijn kritiek.

De hier beschreven toets is gevoelig voor overeenstemming in de rangschikkingen; men zal dus bij het beschouwde voorbeeld tot verwerping van de nulhypothese komen, als de enquêtrices geneigd zijn, eenzelfde volgorde aan de constructies toe te kennen.

Indien men onafhankelijk van het waarnemingsmateriaal, redenen heeft een bepaalde volgorde te onderstellen, kan men op veel eenvoudiger wijze het materiaal op een verloop ("trend") in die richting te onderzoeken. Zie hiertoe b.v. memorandum S 47 (M 13a) van de Statistische Afdeling van het Mathematisch Centrum.