UDC 519.27

Selectuur

Een grafische methode voor het berekenen van de toetsingsgrootheid van K e n d a l l's rangcorrelatietoets [1])

A graphical method for computing the test statistic of K e n d a l l's rank correlationtest.

Laten $(x_1, y_1), \ldots, (x_n, y_n)$ n waarnemingsparen voorstellen, dan wordt de toetsingsgrootheid S van K e n d a l l's rangcorrelatietoets als volgt gedefiniëerd: de paren (x_i, y_i) en (x_j, y_j) met $i < j$ geven tot S een bijdrage

$$(1) \qquad s_{ij} = \begin{cases} +1 \text{ als } (x_i - x_j)(y_i - y_j) > 0 \\ 0 \text{ als } (x_i - x_j)(y_i - y_j) = 0 \\ -1 \text{ als } (x_i - x_j)(y_i - y_j) < 0 \end{cases}$$

en

$$(2) \qquad S = \sum_{i<j} \sum s_{i,j}.$$

Is P resp. Q het aantal paren (i, j) met $i < j$ waarvoor $s_{i,j}$ positief resp. negatief is, dan is

$$S = P - Q.$$

Voor een uitvoeriger beschrijving van de toets en de wijze waarop S berekend kan worden verwijzen wij naar M. G. K e n d a l l (1955) en C. v a n E e d e n en R. K o r s w a g e n (1959).

Een grafische methode ter berekening van S, afkomstig van S. D. H o l m e s, werd voor het eerst beschreven door P. S a n d i f o r d (1928). S a n d i f o r d en H o l m e s gebruikten deze grafische methode voor het schatten van de product-moment correlatiecoëfficiënt in kleine steekproeven. H. D. G r i f f i n (1958) toonde aan dat de door S a n d i f o r d en H o l m e s gebruikte schatting identiek is met K e n d a l l's τ voor het geval dat in geen van beide waarnemingsreeksen gelijken voorkomen. Tevens gaf G r i f f i n een uitbreiding van de grafische methode voor het geval er wel gelijken zijn.

We geven eerst een beschrijving van de methode voor het geval dat *in geen van beide waarnemingsreeksen gelijken voorkomen.*

In beide reeksen worden de waarnemingen vervangen door hun rangnummers naar opklimmende grootte en de kolommen worden in een zodanige volgorde geplaatst dat één van de twee reeksen in de natuurlijke volgorde staat. Laat nu s voorstellen het aantal snijpunten verkregen door ieder rang-

[1]) Rapport S 267 van de Statistische Afdeling van het Mathematisch Centrum, Amsterdam.

nummer uit de eerste reeks te verbinden met het daaraan gelijke rangnummer in de tweede reeks. Dan is

$$S = \tfrac{1}{2}n\,(n-1) - 2s. \tag{3}$$

Voorbeeld

Laat de volgende twee reeksen rangnummers gegeven zijn

x:	3	5	1	6	2	4	7
y:	1	6	3	7	4	2	5.

Na rangschikking van de kolommen op zodanige wijze, dat de x-reeks in de natuurlijke volgorde komt te staan, krijgen we

x:	1	2	3	4	5	6	7
y:	3	4	1	2	6	7	5.

Het verdient aanbeveling de lijnen zodanig te trekken dat een lijn niet door het snijpunt van twee andere gaat.

Het aantal snijpunten s, dat men het eenvoudigst vindt door turven tijdens het trekken der lijnen, is in ons voorbeeld 6, dus $S = \frac{1}{2}.7.6 - 2.6 = 9$.

In het geval dat *een van de twee reeksen gelijke waarnemingen* bevat gaat men als volgt te werk:

De kolommen worden in zodanige volgorde geplaatst, dat de reeks zonder gelijken in de natuurlijke volgorde komt te staan. De waarnemingen in de reeks zonder gelijken worden vervangen door hun rangnummers naar opklimmende grootte en men berekent twee waarden s_1 en s_2 van s als volgt: in de tweede reeks worden de waarnemingen in opklimmende grootte vervangen door de rangnummers $1, \ldots, n$ en wel zodanig dat

1) binnen iedere groep gelijken de tot die groep behorende rangnummers in de natuurlijke volgorde staan. Dit geeft de waarde s_1 voor s,
2) binnen iedere groep gelijken de tot die groep behorende rangnummers tegengesteld aan de natuurlijke volgorde staan. Dit geeft de waarde s_2 voor s.

Dan is

$$2s = s_1 + s_2. \tag{4}$$

Komen in *beide reeksen gelijken* voor dan berekent men vier waarden s_1, s_2, s_3, s_4 van s. Hiertoe plaatst men eerst de kolommen in zodanige volgorde

dat $x_1 \leqq \ldots \leqq x_n$; gelijke waarnemingen worden hierbij in een willekeurige volgorde gezet. In beide reeksen worden nu de waarnemingen in opklimmende grootte vervangen door de rangnummers 1, . . ., n en wel zodanig dat

1) in beide reeksen binnen iedere groep gelijken de tot die groep behorende rangnummers in de natuurlijke volgorde staan. Dit geeft de waarde s_1 voor s,
2) in beide reeksen binnen iedere groep gelijken de tot die groep behorende rangnummers tegengesteld aan de natuurlijke volgorde staan. Dit geeft de waarde s_2 voor s (nadat de kolommen zodanig verwisseld zijn dat één van de twee reeksen in de natuurlijke volgorde staat).

Verder vindt men s_3 door voor de eerste reeks de rangnummers te nemen die voor de berekening van s_1 werden gebruikt en voor de tweede reeks de rangnummers, die voor de berekening van s_2 werden gebruikt; s_4 wordt gevonden door voor de eerste reeks de rangnummers te nemen die voor de berekening van s_2 werden gebruikt en voor de tweede reeks die welke voor de berekening van s_1 werden gebruikt. In het laatste geval moet men de kolommen weer zodanig verwisselen dan één van de twee reeksen in de natuurlijke volgorde komt te staan.
Nu is

$$4s = s_1 + s_2 + s_3 + s_4. \tag{5}$$

We zullen nu laten zien dat de op deze wijze berekende grootheid S identiek is met die volgens (2) [1]).

We beschouwen eerst het geval dat in geen van beide reeksen gelijken voorkomen. Dan is

$$P + Q = \tfrac{1}{2}n\,(n - 1) \tag{6}$$

en dus

$$S = P - Q = \tfrac{1}{2}n\,(n - 1) - 2Q, \tag{7}$$

zodat we in dit geval moeten bewijzen dat $Q = s$ is.

Laat de kolommen zodanig gerangschikt zijn dat $x_1 < \ldots < x_n$ is. Laat verder in beide reeksen de waarnemingen vervangen zijn door hun rangnummers naar opklimmende grootte en laat, voor $i = 1, \ldots, n$, r_i het rangnummer van y_i voorstellen.

Laat verder, voor $i = 1, \ldots, n$, L_i de verbindingslijn van de rangnummers r_i in de twee reeksen voorstellen, dan is het aantal snijpunten s gelijk aan het aantal paren (i, j) met $i < j$ waarvoor L_i en L_j elkaar snijden. Daar de x-reeks

[1]) Het hier gegeven bewijs wijkt af van het door G r i f f i n gegevene, dat niet geheel duidelijk is.

in de natuurlijke volgorde staat snijden L_i en L_j voor $i < j$ elkaar dan en slechts dan als in de tweede reeks $r_i > r_j$ is. Dus is s gelijk aan het aantal paren (i, j) met $i < j$, waarvoor in de tweede reeks $r_i > r_j$ is en dit is de grootheid Q.

We beschouwen nu het geval dat in één van beide reeksen gelijken voorkomen. Stel

$$S_\lambda = \tfrac{1}{2}n\,(n-1) - 2s_\lambda \quad \lambda = 1, 2, \tag{8}$$

waarin s_1 resp. s_2 de aantallen snijpunten voorstellen bij toekenning van de rangnummers volgens 1) resp. 2). Dan is (volgens het bovengegeven bewijs) S_1 resp. S_2 Kendall's toetsingsgrootheid bij toekenning van de rangnummers volgens 1) resp. 2). In dit geval is het dus voldoende te bewijzen dat

$$S = \frac{S_1 + S_2}{2}.$$

Evenzo is het, voor het geval in beide reeksen gelijken voorkomen, voldoende te bewijzen dat S het gemiddelde is van de 4 waarden $S_1, \ldots, S_4$, die gevonden worden bij toekenning van de rangnummers op de 4 bovenbeschreven manieren.

Dat S gelijk is aan het gemiddelde der S_λ kan men als volgt inzien (zie ook Kendall (1955), p. 37):

Een paar (x_i, y_i), (x_j, y_j) met $i < j$, dat tot S een bijdrage $\neq 0$ geeft, geeft tot ieder der S_λ eenzelfde bijdrage en wel dezelfde bijdrage als tot S.

Voor een paar (x_i, y_i), (x_j, y_j) met $i < j$, dat tot S een bijdrage 0 geeft, geldt

$$x_i = x_j \text{ en (of) } y_i = y_j.$$

De bijdragen van dit paar tot S_λ zijn $+1$ of -1 en wel is het aantal gevallen, waarin $+1$ optreedt, gelijk aan het aantal gevallen, waarin -1 optreedt. Het gemiddelde van de bijdragen, die zo'n paar tot ieder der S_λ geeft, is dus 0 en dus gelijk aan de bijdrage tot S.

Literatuur

[1] van Eeden, C. en R. Korswagen (1959), Handleiding voor de rangcorrelatietoets van Kendall, Rapport S 262 (M83) van de Statistische Afdeling van het Mathematisch Centrum, Amsterdam.

[2] Griffin, H. D. (1958), Graphic computation of tau as a coefficient of disarray, Journ. Amer. Stat. Ass. **53**, 441—447.

[3] Kendall, M. G. (1955), Rank Correlation Methods, Ch. Griffin, London.

[4] Sandiford, P. (1928), A graphical method of estimating R for small groups, Appendix B of „Educational psychology", New York, Longmans, Green and Co.

Constance van Eeden
Statistische Afdeling van het
Mathematisch Centrum, Amsterdam.