S 168 (M 69)

Een exacte toets tegen kettingcorrelatie
Van M. Ogawara (1951)

Stichting Mathematisch Centrum
Statistische Afdeling
1955

MATHEMATISCH CENTRUM,
2e Boerhaavestraat 49,
A m s t e r d a m - O.
S 168 (M 69)

Een exacte toets tegen kettingcorrelatie van M.OGAWARA (1951) [1)]

We gaan uit van de veronderstelling dat de grootheden $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_{2n+1}$ een simultane normale verdeling bezitten met $\mathcal{E}\{\underline{y}_i\} = \mu$ en $\sigma^2\{\underline{y}_i\} = \sigma^2$ en dat er tussen de grootheden $\underline{y}_i$ een correlatie kan bestaan van de vorm

$$(1) \qquad (\underline{y}_t - \mu) = \rho(\underline{y}_{t-1} - \mu) + \underline{v}_t \;,$$

waarin $\underline{v}_t$ onderling onafhankelijk $N(0, \sigma^2(1-\rho^2))$ verdeeld zijn en $|\rho| < 1$. Een correlatie van deze gedaante noemt men eerste orde kettingcorrelatie (Engels: first order serial correlation) Kettingcorrelatie kan men bv. verwachten bij tijdreeksen. Bovenstaand model wordt dan vaak gebruikt indien men kan veronderstellen dat de correlatie tussen twee opeenvolgende waarnemingen belangrijk sterker is dan tussen verder van elkaar gelegen waarnemingen.

M.OGAWARA (1951) leidt nu een exacte toets voor de hypothese $\rho = 0$ in bovenstaand model af, door de voorwaardelijke simultane verdeling van $\underline{y}_2, \underline{y}_4, \ldots, \underline{y}_{2n}$ te beschouwen onder voorwaarde van de gevonden waarden $y_1, y_3, \ldots, y_{2n+1}$ van de variabelen met oneven index. Onder deze voorwaarde zijn de grootheden $\underline{y}_{2s}$ $(s=1,\ldots,n)$ namelijk onderling onafhankelijk normaal verdeeld, onafhankelijk van de waarde van ρ.

Gemiddelde en variantie van $\underline{y}_{2s}$ kunnen wij als volgt afleiden:

$$(\underline{y}_{2s} - \mu) = \rho(y_{2s-1} - \mu) + \underline{v}_{2s}$$

en

$$(y_{2s+1} - \mu) = \rho(\underline{y}_{2s} - \mu) + \underline{v}_{2s+1} \;,$$

of

$$\rho^2(\underline{y}_{2s} - \mu) = \rho(y_{2s+1} - \mu) - \rho\,\underline{v}_{2s+1} \,.$$

Tellen we de linker- en rechterleden van de eerste en derde gelijkheid op dan ontstaat:

$$(1+\rho^2)(\underline{y}_{2s} - \mu) = \rho(y_{2s-1} + y_{2s+1}) - 2\rho\mu + \underline{v}_{2s} - \rho\,\underline{v}_{2s+1} \,,$$

dus

$$(2) \qquad \underline{y}_{2s} = \mu\{1 - 2\rho(1+\rho^2)^{-1}\} + \rho(1+\rho^2)^{-1}(y_{2s-1} + y_{2s+1}) + (1+\rho^2)^{-1}(\underline{v}_{2s} - \rho\,\underline{v}_{2s+1}).$$

Hieruit volgt

$$\mathcal{E}\{\underline{y}_{2s} | y_1, y_3, \ldots, y_{2n+1}\} = \mu\{1 - 2\rho(1+\rho^2)^{-1}\} + \rho(1+\rho^2)^{-1}(y_{2s-1} + y_{2s+1})$$

----

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

en

$$\sigma^2\{\underline{y}_{2s} \mid y_1, y_3, \ldots, y_{2n+1}\} = (1+\rho^2)^{-2}\left[\sigma^2\{\underline{v}_{2s}\} + \rho^2\sigma^2\{\underline{v}_{2s+1}\}\right] = (1-\rho^2)(1+\rho^2)^{-1}\sigma^2.$$

Vergelijking (2) kunnen we schrijven als:

(3) $$\underline{y}_{2s} = \gamma_0 + \gamma_1 z_s + \underline{v}'_{2s} \qquad (s = 1, \ldots, n),$$

waarin dus:

(4) $$\gamma_0 = \mu\{1 - 2\rho(1+\rho^2)^{-1}\}$$

is,

(5) $$\gamma_1 = 2\rho(1+\rho^2)^{-1}$$

en

(6) $$z_s = 2^{-1}(y_{2s-1} + y_{2s+1}),$$

terwijl de stochastische grootheden $\underline{v}'_{2s}$ (s = 1,... ,n) onderling onafhankelijk $N(0, \sigma^2(1-\rho^2)(1+\rho^2)^{-1})$ verdeeld zijn (we beschouwen y_{2s-1} en y_{2s+1} en dus ook z_s als gegeven en dus als niet-stochastische grootheden).

In de vorm (3) herkennen we het lieaire regressiemodel, waarbij z_s, het gemiddelde van de voorafgaande en volgende waarde van $\underline{y}_{2s}$ (dus $2^{-1}(y_{2s-1} + y_{2s+1})$), als onafhankelijke variabele fungeert.

Daar aan alle voorwaarden voldaan is (onderling onafhankelijk normaal verdeelde grootheden $\underline{v}'_{2s}$ met gelijke varianties) kunnen we in dit regressiesysteem met de methode der kleinste kwadraten meest aannemelijke schattingen voor de onbekenden γ_0 en γ_1 vinden:

(7) $$\hat{\underline{\gamma}}_0 = \underline{\bar{y}} - \hat{\underline{\gamma}}_1 \bar{z},$$

(8) $$\hat{\underline{\gamma}}_1 = \sum_1^n (\underline{y}_{2s} - \underline{\bar{y}})(z_s - \bar{z}) \Big/ \sum_1^n (z_s - \bar{z})^2,$$

waarin

$$\bar{y} = \sum_1^n y_{2s}/n \qquad \text{en} \qquad \bar{z} = \sum_1^n z_s/n.$$

Ook geeft de regressieanalyse (variantieanalyse) exacte toetsen voor lineaire hypothesen over γ_0 en γ_1. In het bijzonder interesseert ons de hypothese $\rho = 0$; in model (3) betekent dit $\gamma_1 = 0$. Voor het toetsen van deze hypothese gebruiken we dan als toetsingsgrootheid:

(9) $$\underline{F}_{1,n-2} = (n-2)\hat{\underline{\gamma}}_1^2 \Sigma(z_s - \bar{z})^2 / \Sigma(\underline{y}_{2s} - \hat{\underline{\gamma}}_0 - \hat{\underline{\gamma}}_1 z_s)^2 = (n-2)\hat{\underline{\gamma}}_1^2 \Sigma(z_s - \bar{z})^2 / \left[\Sigma(\underline{y}_{2s} - \underline{\bar{y}})^2 - \hat{\underline{\gamma}}_1^2 \Sigma(z_s - \bar{z})^2\right],$$

die onder de hypothese $\gamma_1 = 0$ een F-verdeling (Fisher) met 1 en n - 2 vrijheidsgraden bezit, of

(10) $$\underline{t}_{n-2} = \sqrt{\underline{F}_{1,n-2}} = \hat{\underline{\gamma}}_1 \Big/ \sqrt{\left[\Sigma(\underline{y}_{2s} - \underline{\bar{y}})^2 - \hat{\underline{\gamma}}_1^2 \Sigma(z_s - \bar{z})^2\right] \Big/ \left[(n-2)\Sigma(z_s - \bar{z})^2\right]},$$

die onder de hypothese $\gamma_1 = 0$ een t-verdeling (Student) met (n-2) vrijheidsgraden bezit (zie stelling 4.1 van H.B. Mann (1949) of memorandum S 73 (M 34) met $n_1 = n_2 = \ldots . n_h = 1$).

De toets van OGAWARA kan tot hogere orde kettingcorrelatie worden gegeneraliseerd (zie M.OGAWARA (1951) en ook E.J.HANNAN (1955)). Een andere uitbreiding gaf E.J.HANNAN voor lineaire regressieproblemen (zie E.J.HANNAN (1955) en memorandum S 168 (M 70)).

Literatuur

M. OGAWARA (1951) A note on the test of serial correlation coefficients.
Annals, 22, pp. 115-118.

E.J. HANNAN (1955) Exact tests for serial correlation.
Biometrika, 42, pp. 133-143.

H.B. MANN (1949) Analysis and design of experiments, hoofdstuk 4.
New York, Dover Publications, Inc.

S 73 (M 34) Toets voor de hypothese dat een regressiecoefficiënt nul is, wanneer de afwijkingen van de regressielijn normaal verdeeld zijn met gelijke spreidingen.
Memorandum van de Statistische Afdeling van het Mathematisch Centrum te Amsterdam (19..)