STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 53 (M 27)

Toetsing van de gelijkheid van twee regressiecoëfficiënten indien de afwijkingen normaal verdeeld zijn met gelijke spreidingen.

MATHEMATISCH CENTRUM,
2de Boerhaavestraat 49,
A m s t e r d a m (O).

Statistische Afdeling.
S 53 (M 27)

Toetsing van de gelijkheid van twee regressiecoëfficiënten indien de afwijkingen normaal verdeeld zijn met gelijke spreidingen.[1])

Zij gegeven, dat van vier grootheden ξ_1, η_1, ξ_2 en η_2 de grootheden ξ_1 en ξ_2 foutloos kunnen worden waargenomen, terwijl bij de waarnemingen van η_1 en η_2 onderling onafhankelijk verdeelde fouten [2]) worden gemaakt, die alle een normale verdeling bezitten [3]) met gemiddelde 0 en onbekende, maar steeds dezelfde, spreiding σ.

Verder zij gegeven, dat η_1 en η_2 lineair van ξ_1 resp. ξ_2 afhankelijk zijn, dus dat geldt

$$\eta_1 = \beta_1 \xi_1 + \alpha_1$$
$$\eta_2 = \beta_2 \xi_2 + \alpha_2,$$

waarin de parameters α_1, β_1, α_2 en β_2 onbekend zijn.

Gevraagd wordt nu, op grond van een aantal waarnemingsparen

$(x'_1, y'_1), \ldots, (x'_n, y'_n)$ van (ξ_1, η_1)

en $(x''_1, y''_1), \ldots, (x''_m, y''_m)$ van (ξ_2, η_2)

de hypothese te toetsen, dat de twee regressiecoëfficiënten β_1 en β_2 gelijk zijn:

---

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

2) Onder "fouten" vallen in dit geval ook toevallige afwijkingen van ander karakter, bv. physiologische afwijkingen.

3) De stochastische grootheid $\underline{x}$ bezit een normale verdeling met gemiddelde μ en spreiding σ, indien voor iedere a geldt:

$$P[\underline{x} \leq a] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}\frac{(u-\mu)^2}{\sigma^2}} du.$$

H_0: $\beta_1 = \beta_2$.

Om deze hypothese te toetsen, worden schattingen b_1 en b_2 van β_1 en β_2 berekend volgens de formules:

$$b_1 = \frac{\sum_{i=1}^{n} (y_i' - \bar{y}')(x_i' - \bar{x}')}{\sum_{i=1}^{n} (x_i' - \bar{x}')^2}$$

en

$$b_2 = \frac{\sum_{j=1}^{m} (y_i'' - \bar{y}'')(x_i'' - \bar{x}'')}{\sum_{j=1}^{m} (x_i'' - \bar{x}'')^2},$$

waarin

$$\bar{y}' = \frac{1}{n} \sum_{i=1}^{n} y_i' \quad , \quad \bar{x}' = \frac{1}{n} \sum_{i=1}^{n} x_i', \text{ enz.}$$

Deze schattingen worden verkregen door toepassing van de methode der kleinste kwadraten, die in dit geval overeenkomt met de methode der meest aannemelijke schattingen (Engels: method of maximum likelihood). Op dezelfde wijze kan men als schatting voor σ^2 afleiden:

$$s^2 = \frac{\sum_{i=1}^{n} \left\{ y_i' - b_1 (x_i' - \bar{x}') - \bar{y}' \right\}^2 + \sum_{j=1}^{m} \left\{ y_j'' - b_2 (x_j'' - \bar{x}'') - \bar{y}'' \right\}^2}{(n-2)+(m-2)} =$$

$$= \frac{\sum_{i=1}^{n} (y_i' - \bar{y}')^2 - b_1^2 \sum_{i=1}^{n} (x_i' - \bar{x}')^2 + \sum_{j=1}^{m} (y_j'' - \bar{y}'')^2 - b_2^2 \sum_{j=1}^{m} (x_i'' - x'')^2}{n+m-4}$$

Beschouwen wij niet alleen de gevonden waarden $y_1', \ldots, y_n'$, $y_1'', \ldots, y_m''$, maar de verzameling van alle mogelijke waarden, die deze grootheden kunnen aannemen, dan bezitten zij op deze verzameling een waarschijnlijkheidsverdeling, die afhankelijk is van de (exact waargenomen) waarden $x_1', \ldots, x_n'$, $x_1'', \ldots, x_m''$. De schattingen b_1, b_2 en s^2 zijn dan ook stochastisch, hetgeen door onderstreping aangegeven wordt. Zij zijn verder onderling onafhankelijk verdeeld, zoals bv. door Mood [1] bewezen wordt, en $\underline{b}_1$ en $\underline{b}_2$ bezitten beide een normale verdeling met β_1 resp. β_2 als gemiddelden en spreidingen

$$\frac{\sigma}{\sqrt{\sum_i (x_i' - \bar{x}')^2}} \quad \text{resp.} \quad \frac{\sigma}{\sqrt{\sum_j (x_i'' - \bar{x}'')^2}} .$$

Verder bezit $(n+m-4)\underline{s}^2/\sigma^2$ een χ^2-verdeling met (n+m-4) vrijheidsgraden. Hieruit volgt, dat de stochastische grootheid

$$\underline{t} = \frac{\underline{b}_1 - \underline{b}_2 - (\beta_1 - \beta_2)}{\underline{s}\sqrt{\frac{1}{\sum_i (x_i' - \bar{x}')^2} + \frac{1}{\sum_j (x_i'' - \bar{x}'')^2}}}$$

verdeeld is volgens de verdeling van Student met (n+m-4) vrijheidsgraden (zie bv. Cramèr [2] p.237) Indien $\beta_1 - \beta_2 = 0$ is, gaat deze grootheid over in

$$\underline{t} = \frac{\underline{b}_1 - \underline{b}_2}{\underline{s}\sqrt{\frac{1}{\sum_i (x_i' - \bar{x}')^2} + \frac{1}{\sum_j (x_i'' - \bar{x}'')^2}}},$$

die dus, onder de hypothese $\beta_1 = \beta_2$, de bovengenoemde Student-verdeling bezit. Is $\beta_1 \neq \beta_2$, dan zullen bepaalde van nul verwijderd liggende waarden van t een grotere waarschijnlijkheidsdichtheid bezitten dan wanneer $\beta_1 = \beta_2$ is. Voor de toetsing van deze hypothese gebruikt men daarom in het tweezijdige geval (als zowel $\beta_1 > \beta_2$ als $\beta_1 < \beta_2$ als alternatieve mogelijkheid wordt toegelaten) een <u>tweezijdige kritieke zône</u> van de vorm

$$|\underline{t}| \geq t_0,$$

waarbij t_0, behorend bij een bepaalde onbetrouwbaarheidsdrempel α, opgezocht kan worden in tabellen, vermeld in onderstaande litteratuurlijst. Indien slechts alternatieve mogelijkheden van de vorm $\beta_1 < \beta_2$ resp. $\beta_1 > \beta_2$ worden toegelaten gebruikt men de éénzijdige kritieke zônes

$$\underline{t} > t_1, \quad \text{resp.} \quad \underline{t} < -t_1,$$

waarin t_1 in dezelfde tabellen kan worden gevonden als t_0 door te zoeken in een tabel voor de tweezijdige toets met onbetrouwbaarheidsdrempel 2α. Het bij de tabellen vermelde aantal vrijheidsgraden (vaak aangegeven door ν of n' of n-1) is in dit geval gelijk aan n+m-4.

Litteratuur:

[1] A.M. Mood, Introduction to the theory of statistics, Mc Graw-Hill, New York-Toronto-London, 1950, hoofdstuk 13 (regressietheorie) en p. 425 (tabel).

[2] H. Cramer, Mathematical methods of statistics, Princeton Un. Press, Princeton 1946.

[3] M.G. Kendall, The advanced theory of statistics I, Griffin and Co., London 1947, p. 440-441 (tabel).