STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 145 (M 56)

Een sequente toets van Student voor één steekproef.

Uittreksel uit S. RUSHTON [3].

MATHEMATISCH CENTRUM
2e Boerhaavestraat 49,
A m s t e r d a m - O.
Statistische afdeling
Rapport S 145(M 56)

Een sequente toets van Student voor één steekproef. [1] [2]

1. Inleiding

Stel de stochastische grootheid $\underline{x}$ bezit een normale verdeling met onbekend gemiddelde μ en met onbekende spreiding σ en men wil, met behulp van een sequente methode, de hypothese $\delta \overset{\mathrm{df}}{=} \frac{\mu}{\sigma} = \delta_0$ toetsen, waarbij δ_0 een gegeven getal is. Daartoe kiest men twee waarden δ_1 en δ_2 voor δ $(\delta_1 < \delta_2)$ zodat als $\delta_1 < \delta < \delta_2$ het voor het onderzoek van weinig belang is of men tot de beslissing $\delta \geq \delta_0$ of tot de beslissing $\delta < \delta_0$ komt, terwijl de beslissing $\delta \geq \delta_0$ onjuist is als $\delta \leq \delta_1$ is en de beslissing $\delta < \delta_0$ onjuist is als $\delta \geq \delta_2$ is.
Stel verder

α = de kans om te besluiten tot $\delta \geq \delta_0$ als $\delta = \delta_1$,
β = de kans om te besluiten tot $\delta < \delta_0$ als $\delta = \delta_2$.

Als nu $\phi(t|\delta, n)$ de verdelingsdichtheid van de toetsingsgrootheid $\underline{t}$ van Student is, berekend op grond van n onafhankelijke waarnemingen van $\underline{x}$, dan wordt de toets als volgt uitgevoerd.

Stel

$$l_n(t|\delta_1, \delta_2) = \frac{\phi(t|\delta_2, n)}{\phi(t|\delta_1, n)} \tag{1.1}$$

dan gaat men door met het nemen van waarnemingen zolang voldaan is aan

$$\lg \frac{\beta}{1-\alpha} < \lg l_n(t|\delta_1, \delta_2) < \lg \frac{1-\beta}{\alpha}. \text{ [3]} \tag{1.2}$$

Zodra niet meer aan (1.2) voldaan is, houdt men op met het nemen van waarnemingen en besluit men tot $\delta \geq \delta_0$ als

$$\lg l_n(t|\delta_1, \delta_2) \geq \lg \frac{1-\beta}{\alpha}$$

----

1) Uittreksel uit: S. RUSHTON [3].
2) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.
3) Met lg wordt steeds de natuurlijke logarithme bedoeld.

is en tot $\delta < \delta_0$ als

$$\lg l_n(t \mid \delta_1, \delta_2) \leq \lg \frac{\beta}{1-\alpha}$$

is.

2. Een benadering voor $\lg l_n(t \mid \delta_1, \delta_2)$.

Als $x_1, x_2, \ldots x_n$ onderling onafhankelijke waarnemingen van $\underline{x}$ zijn en

$$(2.1) \quad \bar{x} = \frac{1}{n} \sum_i x_i ,$$

$$(2.2) \quad s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 ,$$

$$(2.3) \quad \underline{t} = \frac{\underline{\bar{x}}}{\underline{s}} \sqrt{n} ,$$

dan bezit $\underline{t}$ een niet centrale t-verdeling met $(n-1)$ vrijheidsgraden en parameter δ , d.w.z. de verdelingsdichtheid is (zie b.v. [2]):

$$(2.4) \quad \phi(t \mid \delta, n) = \frac{\Gamma(n)\, e^{-\frac{n(n-1)\delta^2}{2(n-1+t^2)}}}{2^{\frac{n-2}{2}}\, \Gamma\left(\frac{n-1}{2}\right) \sqrt{\pi(n-1)}} \left(\frac{n-1}{n-1+t^2}\right)^{\frac{n}{2}} G_{n-1}(-\delta u) ,$$

waarin

$$(2.5) \qquad u = t \sqrt{\frac{n}{n-1+t^2}}$$

en

$$(2.6) \qquad G_n(x) = \int_0^\infty \frac{x^n}{n!}\, e^{-\frac{1}{2}(x+x)^2}\, dx.$$ [4)]

Nu is het in de praktijk eenvoudiger om met u te werken i.p.v met t . Uit (2.5) volgt

$$(2.7) \qquad u = \frac{\sum_i x_i}{\sqrt{\sum_i x_i^2}}$$

en uit (2.4) en (1.1):

$$(2.8) \quad \lg l_n(t \mid \delta_1, \delta_2) = g_n(\delta_2 u) - g_n(\delta_1 u) - \tfrac{1}{2} n (\delta_2^2 - \delta_1^2) ,$$

waarin

----

4) In het artikel van RUSHTON (zie voetnoot 1) wordt voor deze functie de notatie $\mathcal{H}h_n(x)$ gebruikt. Tabellen van deze functie zijn b.v. te vinden in [1] .

(2.9) $$g_n(x) = \frac{1}{2}x^2 + \lg G_{n-1}(-x)$$

Met behulp van tabellen van $G_n(x)$ is de toets dus exact uitvoerbaar.

Het is echter eenvoudiger om met één der volgende benaderingen voor $g_n(x)$ te werken:

(2.10) $$g_n(x) \approx \frac{1}{4}x^2 + x\sqrt{n},$$

(2.11) $$g_n(x) \approx \frac{1}{4}x^2 + x\sqrt{n}\left(1 - \frac{1}{4n}\right).$$

(2.12) $$g_n(x) \approx \frac{1}{4}x^2 + x\sqrt{n}\left(1 - \frac{1}{4n} + \frac{x^2}{24n}\right)$$

waaruit men benaderingen voor $\lg \ell_n(t \mid \delta_1, \delta_2)$ vindt.

Als men $\delta_1 = -\delta_2$ neemt dan vindt men met behulp van (2.10), (2.11) en (2.12) resp.

(2.13) $$\lg \ell_n(t \mid \delta_1, \delta_2) \approx 2u\,\delta_2\sqrt{n},$$

(2.14) $$\lg \ell_n(t \mid \delta_1, \delta_2) \approx 2u\,\delta_2\sqrt{n}\left(1 - \frac{1}{4n}\right),$$

(2.15) $$\lg \ell_n(t \mid \delta_1, \delta_2) \approx 2u\,\delta_2\sqrt{n}\left(1 - \frac{1}{4n}\right) + \frac{\delta_2^3 u^3}{12\sqrt{n}}.$$

In de praktijk kan men nu zo te werk gaan, dat men eerst met de benadering (2.11) werkt en (2.12) of (2.13) gebruikt als men in de buurt van één der grenzen $\lg \frac{1-\beta}{\alpha}$ of $\lg \frac{\beta}{1-\alpha}$ komt.

Een grafische methode is niet beschikbaar. Bij iedere stap moet de waarde van u uit alle beschikbare waarnemingen opnieuw berekend worden.

Literatuur

[1] AIREY, J.R., Tables of the Hh-functions, British Association Mathematical Tables, 1 (1931).

[2] JOHNSON, N.L. and B.L. WELCH, Applications of the non-central t-distribution, Biometrika 31 (1940), p. 362

[3] RUSHTON, S., On a sequential t-test, Biometrika 37 (1950), 326-333.