STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

SM 82 r

Bepaling van een betrouwbaarheidsgebied voor de variatiecoëfficiënt van een normale verdeling op grond van een steekproef.

C. van Eeden.

1960

Receptuur

Bepaling van een betrouwbaarheidsgebied voor de variatiecoëfficiënt van een normale verdeling op grond van een steekproef *)

A confidence region for the coefficient of variation of a normal distribution.

Laat $x_1, \ldots, x_n$ een steekproef uit een normale verdeling voorstellen met onbekend gemiddelde μ en onbekende spreiding σ. Gevraagd wordt op grond van deze waarnemingen een betrouwbaarheidsgebied te bepalen voor $\gamma = \frac{\sigma}{\mu}$.

Exacte methode

Een *exact* betrouwbaarheidsgebied kan gevonden worden met behulp van tabellen van N. L. Johnson and B. L. Welch (1940). Daartoe gaat men uit van de grootheid

$$(1) \qquad t = \frac{\bar{x}}{s}\sqrt{n},$$

waarin

$$(2) \qquad \begin{cases} \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \\ s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right\}. \end{cases}$$

De grootheid $\underline{t}$ bezit een niet-centrale t-verdeling met niet-centraliteits-parameter $\delta = \frac{\sqrt{n}}{\gamma} = \frac{\mu}{\sigma}\sqrt{n}$ en aantal vrijheidsgraden $f = n - 1$.

Op grond van de voor $\underline{t}$ gevonden waarde t kan men nu, met behulp van de tabellen van Johnson and Welch, een betrouwbaarheidsinterval voor δ bepalen. Uit de relatie $\delta = \frac{\mu}{\sigma}\sqrt{n}$ volgt dan een betrouwbaarheidsgebied voor $\gamma = \frac{\sigma}{\mu}$.

Het tweezijdige betrouwbaarheidsinterval voor δ met onbetrouwbaarheid α wordt

$$(3) \qquad t - \lambda(f, t, \tfrac{1}{2}\alpha)\sqrt{1 + \frac{t^2}{2f}} \leqq \delta \leqq t + \lambda(f, -t, \tfrac{1}{2}\alpha)\sqrt{1 + \frac{t^2}{2f}},$$

*) Rapport S 261 (M81) van de Statistische Afdeling van het Mathematisch Centrum, Amsterdam.

waarin $\dot{t}$ de voor t gevonden waarde voorstelt en de coëfficiënten $\lambda\,(f,\ t,\ \tfrac{1}{2}\,\alpha)$ en $\lambda\,(f,\ -t,\ \tfrac{1}{2}\,\alpha)$ berekend kunnen worden met behulp van tabel IV van J o h n s o n and W e l c h. Hierbij gaat men als volgt te werk. Men neemt die bladzijde van tabel IV waarvoor $\varepsilon = \tfrac{1}{2}\,\alpha$; voor de berekening van $\lambda\,(f,\ t,\ \tfrac{1}{2}\,\alpha)$ neemt men $t_0 = t$ en voor de berekening van $\lambda\,(f,\ -t,\ \tfrac{1}{2}\,\alpha)$ neemt men $t_0 = -t$. Vervolgens berekent men

$$(4)\quad \begin{cases} \text{als } \dfrac{t_0^2}{2f} \leqq 0{,}75 & y' = \dfrac{t_0}{\sqrt{2f}}\left(1 + \dfrac{t_0^2}{2f}\right)^{-\frac{1}{2}} = \dfrac{t_0}{\sqrt{2f + t_0^2}} \\[2ex] \text{als } \dfrac{t_0^2}{2f} \geqq 0{,}75 & y = \begin{cases} \left(1 + \dfrac{t_0^2}{2f}\right)^{-\frac{1}{2}} = \sqrt{\dfrac{2f}{2f + t_0^2}} \text{ als } t_0 > 0 \\[2ex] -\left(1 + \dfrac{t_0^2}{2f}\right)^{-\frac{1}{2}} = -\sqrt{\dfrac{2f}{2f + t_0^2}} \text{ als } t_0 < 0. \end{cases} \end{cases}$$

We kunnen nu de volgende twee gevallen onderscheiden

1. f is een in de tabel voorkomende waarde ($f = 4\,(1)\,9,\ 16,\ 36,\ 144$). Dan berekent men de grootheid $\lambda\,(f,\ t_0,\ \tfrac{1}{2}\,\alpha)$ door lineaire interpolatie naar y' (of y) in de bij f behorende kolom.
2. f komt niet in de tabel voor. Dan kiest men twee waarden f_1 en f_2 uit de tabel zo dicht mogelijk bij f met $f_1 < f < f_2$ en berekent $\lambda\,(f_1,\ t_0,\ \tfrac{1}{2}\,\alpha)$ en $\lambda\,(f_2,\ t_0,\ \tfrac{1}{2}\,\alpha)$ beide door lineaire interpolatie naar y' (of y). De grootheid $\lambda\,(f,\ t_0,\ \tfrac{1}{2}\,\alpha)$ wordt dan gevonden uit $\lambda\,(f_1,\ t_0,\ \tfrac{1}{2}\,\alpha)$ en $\lambda\,(f_2,\ t_0,\ \tfrac{1}{2}\,\alpha)$ door lineaire interpolatie naar $\dfrac{12}{\sqrt{f}}$.

Uit het interval (3) volgt, na deling door $\sqrt{n}$, een betrouwbaarheidsinterval voor $\dfrac{1}{\gamma} = \dfrac{\mu}{\sigma}$. De grenzen van dit interval kan men ook schrijven als functie van de in de steekproef gevonden variatie-coëfficiënt $c = \dfrac{s}{\bar{x}}$. Dit geeft

$$(5)\quad \frac{1}{c}\left[1 - \lambda\,(f,t,\tfrac{1}{2}\,\alpha)\sqrt{\frac{c^2}{n} + \frac{1}{2f}}\right] \leqq \frac{1}{\gamma} \leqq \frac{1}{c}\left[1 + \lambda\,(f,-t,\tfrac{1}{2}\alpha)\sqrt{\frac{c^2}{n} + \frac{1}{2f}}\right] \text{ met } t = \frac{\sqrt{n}}{c}.$$

Indien dit interval de waarde 0 niet bevat, dus als beide grenzen positief of beide grenzen negatief zijn, volgt uit (5) het interval

$$(6)\quad \frac{c}{1 + \lambda\,(f,-t,\tfrac{1}{2}\,\alpha)\sqrt{\dfrac{c^2}{n} + \dfrac{1}{2f}}} \leqq \gamma \leqq \frac{c}{1 - \lambda\,(f,\ t,\ \tfrac{1}{2}\,\alpha)\sqrt{\dfrac{c^2}{n} + \dfrac{1}{2f}}} \text{ met } t = \frac{\sqrt{n}}{c}$$

voor γ.

Als in (5) de ondergrens negatief en de bovengrens positief is, dan is het betrouwbaarheidsgebied voor γ geen interval, maar het gebied

$$(7) \qquad \left\{\begin{array}{l} \gamma \geqq \dfrac{c}{1 + \lambda\,(f, -t, \tfrac{1}{2}\,\alpha)\sqrt{\dfrac{c^2}{n} + \dfrac{1}{2f}}} \\[2ex] \gamma \leqq \dfrac{c}{1 - \lambda\,(f, t, \tfrac{1}{2}\,\alpha)\sqrt{\dfrac{c^2}{n} + \dfrac{1}{2f}}} \end{array}\right. \qquad \text{met } t = \frac{\sqrt{n}}{c}.$$

De bovenbeschreven betrouwbaarheidsgebieden zijn behoudens interpolatiefouten exact. Met behulp van de tabellen van Johnson en Welch kan men deze dus vinden voor iedere $f \geqq 4$ en voor die waarden van ε, die in de tabel voorkomen.

Benaderingen

Voor grote waarden van f zijn een aantal benaderingen bekend voor het betrouwbaarheidsinterval voor δ. Zo'n benadering kan men als volgt vinden:

Het betrouwbaarheidsinterval voor δ bestaat uit alle bij toetsing op grond van de waarnemingen $x_1, \ldots, x_n$ niet verworpen waarden δ_0. De ondergrens δ_1 en de bovengrens δ_2 van het interval voldoen dus aan

$$(8) \qquad \left\{\begin{array}{l} \mathrm{P}\,[\underline{t} \geqq t \mid f, \delta_1] = \tfrac{1}{2}\,\alpha, \\ \mathrm{P}\,[\underline{t} \geqq t \mid f, \delta_2] = 1 - \tfrac{1}{2}\,\alpha. \end{array}\right.$$

Een benadering voor δ_1 en δ_2 kan men dus vinden als men een benadering heeft voor de percentielen $t\,(f, \delta, \alpha)$ van de niet-centrale t-verdeling, waarbij dus

$$(9) \qquad \mathrm{P}\,[\underline{t} \geqq t\,(f, \delta, \alpha) \mid f, \delta] = \alpha.$$

Benaderingen voor $t\,(f, \delta, \alpha)$ zijn te vinden in C. van Eeden (1959). De eerste in par. 3.1 van dit rapport genoemde benadering voor $t\,(f, \delta, \alpha)$ is het geschiktst om benaderingen voor δ_1 en δ_2 te bepalen.
Deze benadering geeft

$$(9) \qquad \left\{\begin{array}{l} \delta_1 \approx t\,\beta_f - u_{\frac{1}{2}\alpha}\sqrt{1 + t^2\,(1 - \beta_f^2)}, \\ \delta_2 \approx t\,\beta_f + u_{\frac{1}{2}\alpha}\sqrt{1 + t^2\,(1 - \beta_f^2)}. \end{array}\right.$$

Hierin is

$$(10) \qquad \beta_f = \frac{\sqrt{2}\,\Gamma\left(\dfrac{f+1}{2}\right)}{\sqrt{f}\,\Gamma\left(\dfrac{f}{2}\right)},$$

terwijl u_ε gedefiniëerd wordt door

$$(11) \qquad P[\underline{u} \geqq u_\varepsilon] = \varepsilon \quad (\underline{u} \text{ is } N(0,1) \text{ verdeeld}).$$

Tabellen van β_f voor $f = 1\,(1)\,20\,(5)\,50\,(10)\,100$ zijn bijv. te vinden in E. S. Pearson and H. O. Hartley (1954, tabel 35).

Uit (9) volgt als benadering voor het interval voor $1/\gamma$

$$(12) \quad \frac{1}{c}\left[\beta_f - u_{\frac{1}{2}\alpha}\sqrt{\frac{c^2}{n} + 1 - \beta_f^2}\right] \leqq \frac{1}{\gamma} \leqq \frac{1}{c}\left[\beta_f + u_{\frac{1}{2}\alpha}\sqrt{\frac{c^2}{n} + 1 - \beta_f^2}\right].$$

Past men de derde benadering uit par. 3.1 van het bovengenoemde rapport toe, dan vindt men als benadering voor het interval voor $1/\gamma$

$$(13) \qquad \frac{1}{c}\left[1 - u_{\frac{1}{2}\alpha}\sqrt{\frac{c^2}{n} + \frac{1}{2f}}\right] \leqq \frac{1}{\gamma} \leqq \frac{1}{c}\left[1 + u_{\frac{1}{2}\alpha}\sqrt{\frac{c^2}{n} + \frac{1}{2f}}\right].$$

Deze benadering, die in het algemeen minder nauwkeurig is dan (12), is bijv. te vinden in A. Hald (1952, p. 302); hij komt neer op het vervangen van $\lambda(f, t, \frac{1}{2}\alpha)$ en $\lambda(f, -t, \frac{1}{2}\alpha)$ in (5) door $u_{\frac{1}{2}\alpha}$.

Voor zeer grote waarden van f zijn de twee benaderingen practisch identiek, daar

$$\lim_{f\to\infty} \beta_f = 1$$
$$\lim_{f\to\infty} f(1 - \beta_f^2) = \tfrac{1}{2}.$$

Voor een uitvoeriger beschrijving van deze benaderingen en hun nauwkeurigheid verwijzen wij naar het bovengenoemde rapport.

Literatuur

[1] van Eeden, C. (1959), Some approximations to the percentage points of the non-central t-distribution, Rapport S 242 van de Statistische Afdeling van het Mathematisch Centrum, Amsterdam.

[2] Hald, A. (1952), Statistical theory with engineering applications, John Wiley and Sons, Inc., New York.

[3] Johnson, N. L. and B. L. Welch (1940), Applications of the non-central t-distribution, Biometrika **31**, 362—389.

[4] Pearson, E. S. and H. O. Hartley (1954), Biometrika tables for statisticians, Vol. I, Cambridge University Press.

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