MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

1953-49(1)

An Example of FIELLER's method for determining confidence limits for the quotient of two means

by
Gerda Klerk-Grobben

1953.

I Introduction.

E.C.FIELLER (1944) developed a method for determining confidence limits for the ratio of the means of two normally (but not necessarily independently) distributed random variables. To this end he uses statistics s_{11}, s_{12} and s_{22} (here denoted by a_{11}, a_{12}, and a_{22}), which he calls estimates of the variances and of the covariance of these random variables. We here give a general description of his methods and show, by means of an example, that the method need not be confined to the use of estimates of the variances and covariance mentioned.

II Notations.

The random character of a variable is denoted by underlining the symbol, thus $\underline{x}$.

"x and y are $N(\xi, \eta; \sigma_{11}, \sigma_{12}, \sigma_{22})$ distributed", means: $\underline{x}$ and $\underline{y}$ have a two dimensional normal distribution with means ξ and η, variances σ_{11} and σ_{22} and covariance σ_{12}.

The probability that a random variable $\underline{x}$ assumes a value $x \leqq a$ will be denoted by $P[\underline{x} \leqq a]$.

For the variance of a random variable $\underline{x}$ we shall use the symbols $\sigma^2_{\underline{x}}$ or $\sigma^2\{\underline{x}\}$.

III General formulation of the method.

We suppose that to determine confidence limits for $\alpha = \frac{\xi}{\eta}$, observations $\underline{w}_1, \ldots, \underline{w}_k$ are made, which enable us to find five functions of $\underline{w}_1, \ldots, \underline{w}_k$, say $\underline{X}$, $\underline{Y}$, $\underline{a}_{11}$, $\underline{a}_{12}$, and $\underline{a}_{22}$, satisfying the following conditions:

1) $\underline{X}$ and $\underline{Y}$ are $N(\xi, \eta; \sigma_{11}, \sigma_{12}, \sigma_{22})$ distributed with unknown parameters (we will suppose that not $\xi = 0$ and $\eta = 0$ is true);

2) the functions
$$\underline{z} \overset{\text{def}}{=} \eta \underline{X} - \xi \underline{Y}$$
and
$$\underline{s}_z \overset{\text{def}}{=} +\sqrt{\eta^2 \underline{a}_{11} - 2\eta\xi \underline{a}_{12} + \xi^2 \underline{a}_{22}}$$
are independently distributed;

3) for some known integer f the random variable
$$\frac{f\, \underline{s}^2_z}{\sigma^2_{\underline{z}}},$$
with
$$\sigma^2_{\underline{z}} = \eta^2 \sigma_{11} - 2\eta\xi\sigma_{12} + \xi^2 \sigma_{22},$$
has a χ^2-distribution with f degrees of freedom.

From the conditions it follows that $\frac{\underline{Z}}{\underline{S}_Z}$ has Student's distribution with f degrees of freedom, that is to say: from a table of Student's distribution we can determine a value t_ε such that:

$$P\left[\left|\frac{\underline{Z}}{\underline{S}_Z}\right| \leq t_\varepsilon\right] = 1-\varepsilon.$$

Thus:

$$P\left[\frac{|\eta\underline{X} - \xi\underline{Y}|}{\sqrt{\eta^2\underline{a}_{11} - 2\eta\xi\underline{a}_{12} + \xi^2\underline{a}_{22}}} \leq t_\varepsilon\right] = 1-\varepsilon,$$

or, since $t_\varepsilon > 0$ if $\varepsilon < 1$,

$$P\left[(\eta\underline{X} - \xi\underline{Y})^2 \leq t_\varepsilon^2\ (\eta^2\underline{a}_{11} - 2\eta\xi\underline{a}_{12} + \xi^2\underline{a}_{22})\right] = 1-\varepsilon,$$

which can be written as follows:

(1) $$P\left[(\underline{X}^2 - t_\varepsilon^2\underline{a}_{11}) - 2\frac{\xi}{\eta}(\underline{XY} - t_\varepsilon^2\underline{a}_{12}) + \frac{\xi^2}{\eta^2}(\underline{Y}^2 - t_\varepsilon^2\underline{a}_{22}) \leq 0\right] = 1-\varepsilon$$

(supposing $\eta \neq 0$).

From the last formula it is clear that the values, which, substituted for $\frac{\xi}{\eta}$ in (1), satisfy the inequality between square brackets and form a confidence interval for $\alpha = \frac{\xi}{\eta}$, corresponding to the confidence level $1-\varepsilon$.

In many cases, according to condition 3) and the definitions of σ_Z and $\underline{S}_Z$, the functions $\underline{a}_{11}$, $\underline{a}_{12}$ and $\underline{a}_{22}$ will be estimates of the variances and covariance, $_{11}$, σ_{12} and σ_{22}, of $\underline{X}$ and $\underline{Y}$. However, this is not necessary as will be clear from the following example.

IV <u>Confidence limits for the slope of a straight line, if both variables are subject to errors, according to A.WALD.</u>

A.WALD in his paper starts with the following assumptions:

1) $\underline{u}$ and $\underline{v}$ are $N(0,0;\sigma_{11},\sigma_{12},\sigma_{22})$ distributed (σ_{11}, σ_{12} and σ_{22} unknown).

2) For the variables ξ and η, satisfying

$$\xi = \alpha\eta + \beta,$$

observations $(x_1,y_1),\ldots,(x_{2m},y_{2m})$ are made, such that

$$\underline{x}_i = \xi_i + \underline{u}_i,$$
$$\underline{y}_i = \eta_i + \underline{v}_i,$$

where, for each i, $\underline{u}_i$ and $\underline{v}_i$ are distributed like $\underline{u}$ and $\underline{v}$, and two pairs $(\underline{u}_i,\underline{v}_i)$ and $(\underline{u}_j,\underline{v}_j)$ $(i \neq j)$ $(i,j = 1,\ldots,2m)$ are independent.

3) Independent of the values of the observations the pairs $(\underline{x}_i,\underline{y}_i)$ can be renumbered and divided into two

equal groups, $(x_1,y_1),\ldots,(x_m,y_m)$ and $(x_{m+1},y_{m+1}),\ldots$ $\ldots,(x_{2m},y_{2m})$, such that all of $\xi_1,\ldots\xi_m$ are smaller than $\xi_{m+1},\ldots\xi_{2m}$ (or such that all of $\eta_1,\ldots,\eta_m$ are smaller than $\eta_{m+1},\ldots\eta_{2m}$).

The observations x_i and y_i here take the place of $\underline{w}_1,\ldots,\underline{w}_k$ in the preceding section.

We now define:

$$\underline{\bar{x}}_1 = \frac{1}{m}\sum_1^m \underline{x}_i \text{ and } \underline{\bar{x}}_2 = \frac{1}{m}\sum_{m+1}^{2m} \underline{x}_i,$$

where $\underline{\bar{x}}_1$ and $\underline{\bar{x}}_2$ are independently distributed, according to assumption 3), and likewise $\underline{\bar{y}}_1,\underline{\bar{y}}_2,\underline{\bar{u}}_1,\underline{\bar{u}}_2,\underline{\bar{v}}_1,\underline{\bar{v}}_2$, $\bar{\xi}_1$, $\bar{\xi}_2$, $\bar{\eta}_1$ and $\bar{\eta}_2$.

Taking for the five functions $\underline{X},\underline{Y},\underline{a}_{11},\underline{a}_{12}$ and $\underline{a}_{22}$ the following expressions:

$$\underline{X} = \underline{\bar{x}}_1 - \underline{\bar{x}}_2, \qquad \underline{Y} = \underline{\bar{y}}_1 - \underline{\bar{y}}_2,$$

$$\underline{a}_{11} = \frac{1}{m(m-1)}\left[\sum_{i=1}^{m}(\underline{x}_i-\underline{\bar{x}}_1)^2 + \sum_{i=m+1}^{2m}(\underline{x}_i-\underline{\bar{x}}_2)^2\right],$$

$$\underline{a}_{12} = \frac{1}{m(m-1)}\left[\sum_{1}^{m}(\underline{x}_i-\underline{\bar{x}}_1)(\underline{y}_i-\underline{\bar{y}}_1) + \sum_{m+1}^{2m}(\underline{x}_i-\underline{\bar{x}}_2)(\underline{y}_i-\underline{\bar{y}}_2)\right],$$

$$\underline{a}_{22} = \frac{1}{m(m-1)}\left[\sum_{1}^{m}(\underline{y}_i-\underline{\bar{y}}_1)^2 + \sum_{m+1}^{2m}(\underline{y}_i-\underline{\bar{y}}_2)^2\right];$$

it is not difficult to show that the conditions of paragraph III are fulfilled and thus a confidence interval for α (estimate: $\underline{a} = \frac{\underline{X}}{\underline{Y}}$) can be determined as is explained in that paragraph.

(The proof is performed by substituting:

$$\underline{x}_i = \xi_i + \underline{u}_i, \text{ and}$$
$$\underline{y}_i = \eta_i + \underline{v}_i; \qquad \xi_i = \alpha\eta_i + \beta \text{ .)}$$

If we now compare the expression $\underline{a}_{11}$ with the variance $\sigma^2_{\underline{X}}$ it is seen that we cannot use $\underline{a}_{11}$ as an estimate of $\sigma^2_{\underline{X}}$.

Indeed:

$$\sigma^2_{\underline{X}} = \sigma^2\{\underline{\bar{x}}_1-\underline{\bar{x}}_2\} = \sigma^2\{\bar{\xi}_1-\bar{\xi}_2+\underline{\bar{u}}_1-\underline{\bar{u}}_2\} =$$
$$= \sigma^2\{\underline{\bar{u}}_1-\underline{\bar{u}}_2\} = \frac{2}{m}\sigma_{11}$$

whereas;

$$\underline{a}_{11} = \frac{1}{m(m-1)}\left[\sum_1^m(\xi_i-\bar{\xi}_1+\underline{u}_i-\underline{\bar{u}}_1)^2+\sum_{m+1}^{2m}(\xi_i-\bar{\xi}_2+\underline{u}_i-\underline{\bar{u}}_2)^2\right]=$$

$$= \frac{1}{m(m-1)}\Big[\sum_{1}^{m} (\underline{u}_i-\overline{\underline{u}}_1)^2 + \sum_{m+1}^{2m} (\underline{u}_i-\overline{\underline{u}}_2)^2 +$$

$$+ \sum_{1}^{m} (\xi_i - \overline{\xi}_1)^2 + \sum_{m+1}^{2m} (\xi_i - \overline{\xi}_2)^2 +$$

$$+ 2\sum_{1}^{m} (\xi_i - \overline{\xi}_1)(\underline{u}_i-\overline{\underline{u}}_1) + 2\sum_{m+1}^{2m} (\xi_i - \overline{\xi}_2)(\underline{u}_i-\overline{\underline{u}}_2)\Big].$$

In this expression we recognize the part

$$\frac{1}{m(m-1)}\Big[\sum_{1}^{m} (\underline{u}_i-\overline{\underline{u}}_1)^2 + \sum_{m+1}^{2m} (\underline{u}_i-\overline{\underline{u}}_2)^2\Big]$$

as an ordinary estimate of σ_{11}, but the value of $\underline{a}_{11}$ also depends on the values of ξ_i. Clearly the mathematical expectation of $\underline{a}_{11}$ is larger than σ_x^2, the difference increasing with increasing differences between the ξ_i.

References.

E.C.FIELLER, A fundamental formula in the statistics of biological assay and some applications, Quart.J.Phar., 17(1944), pp.117-123.

A.WALD, The fitting of straight lines if both variables are subject to error, The Annals of Math.Stat., 11 (1940), pp.284-300.

G.KLERK-GROBBEN, Een betrouwbaarheidsgebied voor het quotiënt voor de verwachtingen van twee stochastische grootheden, die een simultane normale verdeling bezitten. Rapport S 90(M36) van de Statistische Afdeling van het Mathematisch Centrum, Amsterdam 1952.

H.J.PRINS, Een betrouwbaarheidsgebied voor het quotiënt van de verwachtingen van twee onafhankelijke normaal verdeelde grootheden met gegeven spreidingsverhouding. Rapport S 90(M36a) van de Statistische Afdeling van het Mathematisch Centrum, Amsterdam 1953.