APPROXIMATE CONFIDENCE INTERVAL FOR LINEAR FUNCTIONS OF MEANS OF K POPULATIONS WHEN THE POPULATION VARIANCES ARE NOT EQUAL

By SAIBAL KUMAR BANERJEE
Indian Statistical Institute, Calcutta

SUMMARY. It is shown that given k-samples of n_i units from populations $N_i\ (m_i, \sigma_i^2)$ $(i=1,2..,k)$ a confidence interval for any linear function $\sum_1^k c_i m_i$ of population means with confidence coefficient not less than any pre-assigned probability α is possible in terms of sample estimates of population means and variances and tabulated values of Student's t-table.

Given k samples of n_i $(i = 1, 2, ..., k)$ units from k normal populations $N_i(m_i, \sigma_i^2)$, with sample estimates of population means and variances $\bar{x}_i$ and $s_i^2 (i = 1, 2, ..., k)$ a confidence interval for any linear function of the population means with approximate confidence co-efficient is possible. Briefly the method is indicated for the case of two samples.

Let P denote the probability of the event

$$(\bar{x}_1+\bar{x}_2-m_1-m_2)^2 < \frac{t_1^2 s_1^2}{n_1} + \frac{t_2^2 s_2^2}{n_2} \qquad \dots (1)$$

where t_1 and t_2 correspond to t-values of Student's t-table satisfying the relation

$$\frac{1}{\sqrt{\nu_i}} \cdot \frac{1}{B\left(\frac{\nu_i}{2}, \frac{1}{2}\right)} \int_{-t_i}^{t_i} \left(1+\frac{t^2}{\nu_i}\right)^{-\frac{\nu_i+1}{2}} dt = \alpha \qquad \dots (2)$$

$$(\nu_i = n_i - 1; \quad i = 1, 2)$$

Since for fixed s_1^2 and s_2^2

$$\frac{(\bar{x}_1+\bar{x}_2-m_1-m_2)^2}{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \qquad \dots (3)$$

is distributed as a χ^2 variate with 1 d.f.,

$$P = \int_0^\infty \int_0^\infty f_1(s_1^2, \sigma_1^2, n_1) f_2(s_2^2, \sigma_2^2, n_2)\, g(s_1^2, s_2^2) ds_1^2 ds_2^2 \qquad \dots (4)$$

where

$$f_1(s_1^2, \sigma_1^2, n_1) = \text{frequency function of } s_1^2$$
$$f_2(s_2^2, \sigma_2^2, n_2) = \text{frequency function of } s_2^2$$

and

$$g(s_1^2, s_2^2) = \int_0^{\frac{t_1^2 s_1^2/n_1 + t_2^2 s_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}} \frac{1}{2} \cdot \frac{1}{\Gamma(\frac{1}{2})} \cdot e^{-\frac{\chi^2}{2}} \left(\frac{\chi^2}{2}\right)^{-\frac{1}{2}} d\chi^2.$$

Now $\int_0^z e^{-y} y^{-\frac{1}{2}} dy$ is an upward convex function of z and therefore

$$g(s_1^2, s_2^2) \geqslant \omega_1 \int_0^{t_1^2 s_1^2/\sigma_1^2} f(\chi^2) d\chi^2 + \omega_2 \int_0^{t_2^2 s_2^2/\sigma_2^2} f(\chi^2) d\chi^2 \qquad \dots (5)$$

where
$$f(\chi^2) = \frac{1}{2}\frac{1}{\Gamma(\frac{1}{2})}\cdot e^{-\frac{\chi^2}{2}}\left(\frac{\chi^2}{2}\right)^{-\frac{1}{2}}$$

and
$$\omega_i = \frac{\sigma_i^2/n_i}{\sigma_1^2/n_1+\sigma_2^2/n_2}, \qquad (i = 1, 2).$$

As
$$\int_0^\infty\int_0^\infty f_1(s_1^2, \sigma_1^2, n_1)f_2(s_2^2, \sigma_2^2, n_2)\left\{\int_0^{t_i^2s_i^2/\sigma_i^2} f(\chi^2)d\chi^2\right\}ds_1^2ds_2^2 = \alpha \qquad (i = 1, 2) \quad \ldots (6)$$

$$\text{Prob}\left\{(\bar{x}_1+\bar{x}_2-m_1-m_2)^2 < \frac{t_1^2s_1^2}{n_1}+\frac{t_2^2s_2^2}{n_2}\right\} \geqslant \omega_1\alpha+\omega_2\alpha = \alpha \quad \ldots (7)$$

so that
$$\text{Prob}\left\{\bar{x}_1+\bar{x}_2-\sqrt{\sum_1^2\frac{t_i^2s_i^2}{n_i}} \leqslant m_1+m_2 \leqslant \bar{x}_1+\bar{x}_2+\sqrt{\sum_1^2\frac{t_i^2s_i^2}{n_i}}\right\} \geqslant \alpha. \quad \ldots (8)$$

Also it can be readily shown that if c_1 and c_2 are known constants
$$\text{Prob}\left\{\sum_1^2 c_i\bar{x}_i-\sqrt{\sum_1^2\frac{t_i^2c_i^2s_i^2}{n_i}} \leqslant \sum_1^2 c_im_i \leqslant \sum_1^2 c_i\bar{x}_i+\sqrt{\sum_1^2\frac{t_i^2c_i^2s_i^2}{n_i}}\right\} \geqslant \alpha. \quad \ldots (9)$$

Further extending to k populations it can be shown that
$$\text{Prob}\left\{\sum_1^k c_i\bar{x}_i-\sqrt{\sum_1^k\frac{t_i^2c_i^2s_i^2}{n_i}} \leqslant \sum_1^k c_im_i \leqslant \sum_1^k c_i\bar{x}_i+\sqrt{\sum_1^k\frac{t_i^2c_i^2s_i^2}{n_i}}\right\} \geqslant \alpha. \quad (10)$$

For the case of two populations if $c_1 = 1$ and $c_2 = -1$ in (9) the following relation is established :
$$\text{Prob.}\left\{|\bar{x}_1-\bar{x}_2-(m_1-m_2)| < \sqrt{\frac{t_1^2s_1^2}{n_1}+\frac{t_2^2s_2^2}{n_2}}\right\} \geqslant \alpha$$

which is the two-means problem. Cochran and Cox at page 92 of the book "Experimental Designs" (1950 edition) suggest an approximate result for this case which in the notation of the present paper is
$$|\bar{x}_1-\bar{x}_2-(m_1-m_2)| < \left(\frac{t_1s_1^2}{n_1}+\frac{t_2s_2^2}{n_2}\right)\Big/\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}.$$

Since
$$\left(\frac{t_1s_1^2}{n_1}+\frac{t_2s_2^2}{n_2}\right)^2 < \left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)\left(\frac{t_1^2s_1^2}{n_1}+\frac{t_2^2s_2^2}{n_2}\right),$$

if $n_1 \neq n_2$ this result gives a stronger result than the present result. To the best of the knowledge of the author he has not seen the result suggested as approximation proved in any published literature nor the approximate nature of the result spelt out. The approximate nature of the present result is to the effect that for all values of (σ_1^2, σ_2^2) or $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)$ the true value will be covered with probability not less than pre-assigned probability α.

REFERENCES

COURANT, R. (1957): *Differential and Integral Calculus*, Volume II, Blackie and Son Limited, London and Glasgow.

COX and COCHRAN (1950): *Experimental Designs*, Wiley Publication in Statistics, New York.

Paper received : October, 1958.

Paper revised : March, 1960.