A LOWER BOUND TO THE PROBABILITY OF STUDENT'S RATIO

By SAIBAL KUMAR BANERJEE
Indian Statistical Institute, Calcutta

If the parent population is normal, Student's ratio $\frac{\bar{x}-m}{s/\sqrt{n}}$ follows the t distribution

$$\frac{1}{\sqrt{n-1}} \cdot \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)} \cdot \left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}} dt.$$

An examination of the behaviour of Student's ratio for samples drawn from non-normal populations was made by Bartlett (1935), Geary (1936), and Gayen (1949). From consideration of estimation by confidence interval, however, a lower bound to the probability of the event $\bar{x}+\frac{ts}{\sqrt{n}} \geqslant m \geqslant \bar{x}-\frac{ts}{\sqrt{n}}$ may be of some practical value. Starting from the approach of Tchebysheff's lemma it is possible to derive an expression for the lower bound to the probability of such an event. The expression for the lower bound as worked out, involves t, n and B_2-coefficient of the parent population and applies to samples drawn from any parent population, continuous or discontinuous, having finite fourth cumulant. The proof rests upon a simple lemma.

Lemma : *Let x be a variable with mean m and variance σ^2. The probability $P(k^2)$ of the inequality $x \geqslant m-k^2$, satisfy the relation*

$$P(k^2) \geqslant \frac{k^4}{\sigma^2+k^4}, \qquad \ldots \quad (1)$$

whatever k^2 may be.

Let $x_1, x_2, \ldots, x_n$ be a sample of size n drawn at random with replacement from a population with mean, second cumulant and fourth cumulant denoted respectively by m, k_2 and k_4. Let a variate u be defined as

$$u = \frac{t^2 s^2}{n} - (\bar{x}-m)^2, \qquad \ldots \quad (2)$$

where

$$\bar{x} = \frac{\sum_1^n x_i}{n},$$

$$s^2 = \frac{1}{n-1}\left\{\sum_1^n x_i^2 - n\bar{x}^2\right\} \text{ and } t^2 > 1.$$

From (1) we have

$$\text{prob}\left\{ u \geqslant 0\right\} \geqslant \frac{\{E(u)\}^2}{V(u)+\{E(u)\}^2} \geqslant \frac{\{E(u)\}^2}{E(u^2)} \qquad \ldots \quad (3)$$

We have
$$E(u) = (t^2-1)\frac{k_2}{n} \qquad \ldots \quad (4)$$

$$E(u^2) = E\left[\frac{t^4s^4}{n^2}+(\bar{x}-m)^4-\frac{2t^2s^2}{n}(\bar{x}-m)^2\right] \qquad \ldots \quad (5)$$

Since
$$\left.\begin{aligned} E(s^4) &= \frac{k_4}{n}+\frac{2k_2^2}{n-1}+k_2^2 \\ E(\bar{x}-m)^4 &= \frac{k_4}{n^3}+\frac{3k_2^2}{n^2} \\ E\{s^2(\bar{x}-m)^2\} &= \frac{k_4}{n^2}+\frac{k_2^2}{n}. \end{aligned}\right\} \qquad \ldots \quad (6)$$

$$E(u^2) = \frac{t^4}{n^2}\left(\frac{k_4}{n}+\frac{2k_2^2}{n-1}+k_2^2\right)+\frac{k_4}{n^3}+\frac{3k_2^2}{n^2}-\frac{2t^2}{n}\left(\frac{k_4}{n^2}+\frac{k_2^2}{n}\right)$$

$$= \frac{k_4}{n^3}(t^2-1)^2+\frac{k_2^2}{n^2}(t^2-1)^2+\frac{2k_2^2}{n^2}\left\{\frac{t^4}{n-1}+1\right\}$$

$$= \frac{k_2^2(t^2-1)^2}{n^2}\left[\frac{B_2-3}{n}+1+\frac{2}{(t^2-1)^2}\left\{\frac{t^4}{n-1}+1\right\}\right], \qquad \ldots \quad (7)$$

where $B_2 = \frac{k_4}{k_2^2}+3.$

From (3), (4) and (7) we have

$$\text{prob}\ \{u \geqslant 0\} \geqslant \frac{1}{\frac{B_2-3}{n}+1+\frac{2}{(t^2-1)^2}\left\{\frac{t^4}{n-1}+1\right\}} \qquad \ldots \quad (8)$$

Since $u = \frac{t^2s^2}{n}-(\bar{x}-m)^2$, from (8) we have,

$$\text{prob}\left\{\bar{x}+\frac{ts}{\sqrt{n}} \geqslant m \geqslant \bar{x}-\frac{ts}{\sqrt{n}}\right\} \geqslant \frac{1}{\frac{B_2-3}{n}+1+\frac{2}{(t^2-1)^2}\left\{\frac{t^4}{n-1}+1\right\}} \qquad \ldots (9)$$

The expression for the lower bound of the probability of the inequality $\bar{x}+\frac{ts}{\sqrt{n}} \geqslant m \geqslant \bar{x}-\frac{ts}{\sqrt{n}}$ as derived in (9) depends upon t, n and B_2 coefficient of the parent population. Numerical values of the lower bound of the probability for $t = 3$, $n = 4, 6, 8, 10, 12, 20$ and 30, and $B_2=1, 2, 3, 4, 5, 6, 8$, and 10 are given in the following Table.

LOWER BOUND OF PROBABILITY OF THE INEQUALITY $\bar{x}+\frac{3s}{\sqrt{n}} \geqslant m \geqslant \bar{x}-\frac{3s}{\sqrt{n}}$

(values worked out from expression (9) taking $t = 3$)

B_2 coefficient of the parent population	sample size						
	4	6	8	10	12	20	30
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0.727	0.830	0.875	0.899	0.914	0.939	0.951
2	0.615	0.729	0.789	0.825	0.849	0.897	0.921
3	0.533	0.650	0.718	0.762	0.793	0.850	0.894
4	0.471	0.587	0.659	0.708	0.744	0.823	0.868
5	0.421	0.535	0.609	0.661	0.700	0.791	0.844
6	0.381	0.491	0.566	0.620	0.662	0.761	0.821
8	0.320	0.422	0.496	0.552	0.596	0.707	0.778
10	0.276	0.370	0.441	0.497	0.542	0.660	0.740

From the Table it is seen that if $n = 8$, probability of the inequality $\bar{x}+\frac{3s}{\sqrt{n}} \geqslant m \geqslant \bar{x}-\frac{3s}{\sqrt{n}}$ is greater than .60, even if B_2 be as high as 5.0. If, however, B_2 be 4.0 or less, the probability of the inequality exceeds or equals 0.659. It is further seen from the Table that if $n \geqslant 12$ confidence statement regarding the population mean in the form $\bar{x} \pm \frac{3s}{\sqrt{n}}$ may be made with a confidence coefficient of the order of 0.700 if B_2 be less than or equal to 5.0.

The expression for the lower bound suggests some lines of investigation and these are being investigated. In this connection it may be stated with certain amount of reservation, that if prior knowledge of B_2 or rather an upper bound of the numerical value of B_2 coefficient of the parent population is available, it may be possible to test a suggested value m_0 for the sample mean, on the basis of a sample of size n and sample values $x_1, x_2 \ldots, x_n$.

REFERENCES

BARTLETT, M. S. (1935): The effect of non-normality on the t-distribution. *Proc. Camb. Phil. Soc.*, **31**, 223-231.

GEARY, R. C. (1936): The distribution of student's ratio for non-normal sample. *Supp. J. Roy. Stat. Soc.*, **3**, 178-184.

GAYEN, A. K. (1949): Distribution of student's t in random samples of any size drawn from non-normal universes. *Biometrika*, **36**, 353-369.

USPENSKY, J. V. (1937): *Introduction to Mathematical Probability*. McGraw Hill Book Company.

Paper received : August, 1956.