NOTE ON THE VARIATION OF MEANS

By Casper Goffman

Wayne University

In a manufactured product, batch to batch variations may appear, and it may be of interest to be able to compare these variations for different runs. The simplest case is that for which there is normal distribution with the same standard deviation for each batch, but where the mean may vary from batch to batch. The question arises regarding what function of the set of means should be taken as a measure of its variation. Thus, if $x_1, x_2, \cdots, x_n$ are independent random variables, all with the same standard deviation, say $\sigma = 1$, and means $\mu_1, \mu_2, \cdots, \mu_n$, the question is what function $f(\mu_1, \mu_2, \cdots, \mu_n)$ should be taken to measure the variation of the means. We find, in this note, that if $f(\mu_1, \mu_2, \cdots, \mu_n)$ is subjected to four conditions, three of which seem quite natural and the fourth of which, although perhaps not so natural, has a certain appeal, then $f(\mu_1, \mu_2, \cdots, \mu_n) = F(V)$, where V is the sum of squares $\sum_{i=1}^{n} (\mu_i - \bar{\mu})^2$, $\bar{\mu} = \sum_{i=1}^{n} \mu_i/n$. The properties we have in mind are:

(i) $f(\mu_1, \mu_2, \cdots, \mu_n)$ is continuous, nonnegative, and is equal to zero if and only if $\mu_1 = \mu_2 = \cdots = \mu_n$.

(ii) For every $\epsilon > 0$, there is a $\delta > 0$, such that whenever $f(\mu_1, \mu_2, \cdots, \mu_n) < \delta$ then $| \mu_i - \mu_j | < \epsilon$ for every $i, j = 1, 2, \cdots, n$.

(iii) For every $\mu_1, \mu_2, \cdots, \mu_n$ and every h, $f(\mu_1 + h, \cdots, \mu_n + h) = f(\mu_1, \cdots, \mu_n)$.

(iv) If $x_1, x_2, \cdots, x_n; x_1', x_2', \cdots, x_n'$ are normally distributed with standard deviation $\sigma = 1$ and means $\mu_1, \cdots, \mu_n; \mu_1', \cdots, \mu_n'$, respectively, and if $f(\mu_1, \cdots, \mu_n) = f(\mu_1', \cdots, \mu_n')$, then the random variables $u = f(x_1, \cdots, x_n)$ and $v = f(x_1', \cdots, x_n')$ have the same distribution function.

Condition (iv) says that the distribution of the estimate of the variation of means, obtained from samples, depends only upon the measures of the variation of the means, (assuming standard deviation 1) and upon no other aspect of the set of means.

In this connection, we note that the distribution of the sum of squares of n independent variables with means $a_1, a_2, \cdots, a_n$ depends only on the variance of the means, as does the power function [1] of the analysis of variance test.

Theorem. *If $f(\mu_1, \cdots, \mu_n)$ has properties* (i)–(iv), *there is a continuous $F(x)$ such that $F(V) = f(\mu_1, \cdots, \mu_n)$, where $V = \sum_{i=1}^{n} (\bar{\mu} - \mu_i)^2$, $\bar{\mu} = \sum_{i=1}^{n} \mu_i/n$.*

Proof. Let $x_1, \cdots, x_n; x_1', \cdots, x_n'$ be normal, with means $\mu_1, \cdots, \mu_n; \mu_1', \cdots, \mu_n'$ and standard deviation $\sigma = 1$. Suppose

$$\sum_{i=1}^{n} (\mu_i - \bar{\mu})^2 \neq \sum_{i=1}^{n} (\mu_i' - \bar{\mu}')^2. \tag{1}$$

Received 7/10/52.

By property (iii), we may suppose that $\sum_{i=1}^{n} \mu_i = \sum_{i=1}^{n} \mu_i' = 0$. Then (1) becomes

$$\sum_{i=1}^{n} \mu_i^2 \neq \sum_{i=1}^{n} \mu_i'^2. \tag{2}$$

Let $\alpha = [\sum_{i=1}^{n} \mu_i^2]^{\frac{1}{2}}$, $\beta = [\sum_{i=1}^{n} \mu_i'^2]^{\frac{1}{2}}$ and suppose $\beta > \alpha$. Let $\epsilon = (\beta - \alpha)/(2n + 1)$. By property (ii), there is a $\delta > 0$ such that if $f(x_1, \cdots, x_n) < \delta$ then $| x_i - x_j | < \sqrt{n}\, \epsilon$ for all $i, j = 1, 2, \cdots, n$. Now, let E be the set of points, $(x_1, \cdots, x_n)$, for which $f(x_1, \cdots, x_n) < \delta$, and let $E_0 \subset E$ be those points of E for which $\sum_{i=1}^{n} x_i = 0$. Then

$$\begin{aligned} P(u < \delta) &= (\tfrac{1}{2}\pi)^{n/2} \int_E \cdots \int e^{-\frac{1}{2}\sum_{i=1}^{n} (\mu_i - x_i)^2} dx_1 \cdots, dx_n \\ &\geqq (\tfrac{1}{2}\pi)^{n/2} \mid E_0 \mid \int_{-\infty}^{\infty} e^{-\frac{1}{2}[(\alpha+n\epsilon)^2+x^2]}\, dx, \end{aligned} \tag{3}$$

$$\begin{aligned} P(v < \delta) &= (\tfrac{1}{2}\pi)^{n/2} \int_E \cdots \int e^{-\frac{1}{2}\sum_{i=1}^{n}(\mu_i'-x_i')^2}\, dx_1, \cdots, dx_n \\ &\leqq (\tfrac{1}{2})^{n/2} \mid E_0 \mid \int_{-\infty}^{\infty} e^{-\frac{1}{2}[(\beta-n\epsilon)^2+x^2]}\, dx, \end{aligned} \tag{4}$$

where $\mid E_0 \mid$ is the $n - 1$ dimensional measure of E_0.

That expressions (3) and (4) hold may be shown as follows: E is the cylinder whose axis is the line $x_1 = x_2 = \cdots = x_n$. The n-tuple integrals may be evaluated by letting x be measured along this axis and by integrating, for each x, over the hyperplane normal to the axis at x, and then by integrating with respect to x. It follows that $P(u < \delta) \geqq (\frac{1}{2}\pi)^{n/2} \mid E_0 \mid \int_{-\infty}^{\infty} \varphi(x)\, dx$ and $P(v < \delta) \leqq (\frac{1}{2}\pi)^{n/2} \mid E_0 \mid \int_{-\infty}^{\infty} \psi(x)\, dx$, where $\varphi(x) \leqq e^{-\frac{1}{2}\sum_{i=1}^{n}(\mu_i-x_i)^2}$ and $\psi(x) \geqq e^{-\frac{1}{2}\sum_{i=1}^{n}(\mu'_i-x'_i)^2}$ for all $(x_1, \cdots, x_n)$ and $(x_1', \cdots, x_n')$ in E whose projections on the axis $x_1 = x_2 = \cdots = x_n$ fall at x. Now, for every such $(x_1, \cdots, x_n)$, $\sum_{i=1}^{n} (\mu_i - x_i)^2 \leqq x^2 + (\alpha + n\epsilon)^2$, since the vector whose components are $\mu_i - x_i$, $i = 1, \cdots, n$, has x as one orthogonal component, the other of which is not greater than the sum of the distances of $(\mu_1, \cdots, \mu_n)$ and $(x_1, \cdots, x_n)$ from the line $x_1 = x_2 = \cdots = x_n$; but this is readily seen not to to exceed $\alpha + n\epsilon$. Similarly, $\sum_{i=1}^{n} (\mu_i' - x_i')^2 \geqq x^2 + (\beta - n\epsilon)^2$. Accordingly, we may take $\varphi(x) = e^{-\frac{1}{2}[(\alpha+n\epsilon)^2+x^2]}$ and $\psi(x) = e^{-\frac{1}{2}[(\beta-n\epsilon)^2+x^2]}$. Moreover,

$$\mid E_0 \mid > 0, \tag{5}$$

for, since $f(x_1, \cdots, x_n)$ is continuous, there is a sphere S of radius less than ϵ, containing $(0, 0, \cdots, 0)$, for every point $(x_1, \cdots, x_n)$ of which $f(x_1, \cdots, x_n) < \delta$. The set $S_0 \subset S$, of points $(x_1, \cdots, x_n) \,\varepsilon\, S$ for which $\sum_{i=1}^{n} x_i = 0$, is a subset of E_0 of positive $n - 1$ dimensional measure. But $\epsilon = (\beta - \alpha)/(2n + 1)$ implies

$\beta - n\epsilon > \alpha + n\epsilon$. It then follows by (3), (4), and (5) that if $\mu_1, \cdots, \mu_n$; $\mu_1', \cdots, \mu_n'$ satisfies (1), then $P(u < \delta) > P(v < \delta)$. Hence, by property 4, $f(\mu_1, \cdots, \mu_n) \neq f(\mu_1', \cdots, \mu_n')$.

On the other hand, let $\mu_1, \cdots, \mu_n$; $\mu_1', \cdots, \mu_n'$ be such that

$$V = \sum_{i=1}^{n} (\bar{\mu} - \mu_i)^2 = \sum_{i=1}^{n} (\mu_i' - \bar{\mu}')^2. \tag{6}$$

Suppose $a = f(\mu_1, \cdots, \mu_n)$, $b = f(\mu_1', \cdots, \mu_n')$, and that $a \neq b$. Let C_1 and C_2 be continuous curves joining $(\mu_1, \cdots, \mu_n)$ to $(\mu_1', \cdots, \mu_n')$ such that for every $(\mu_1^{(1)}, \cdots, \mu_n^{(1)}) \; \varepsilon \; C_1$, not an end-point, $\sum_{i=1}^n (\mu_i^{(1)} - \bar{\mu}^{(1)})^2 < V$, and for every $(\mu_1^{(2)}, \cdots, \mu_n^{(2)}) \; \varepsilon \; C_2$, not an end-point, $\sum_{i=1}^n (\mu_i^{(2)} - \bar{\mu}^{(2)})^2 > V$. Since $f(\mu_1, \cdots, \mu_n)$ is continuous, there are points $(\mu_1^{(1)}, \cdots, \mu_n^{(1)}) \; \varepsilon \; C_1$ and $(\mu_1^{(2)}, \cdots, \mu_n^{(2)}) \; \varepsilon \; C_2$ for which

$$f(\mu_1^{(1)}, \cdots, \mu_n^{(1)}) = f(\mu_1^{(2)}, \cdots, \mu_n^{(2)}) = \tfrac{1}{2}(a + b). \tag{7}$$

But (7) contradicts the fact, already established, that $\sum_{i=1}^n (\mu_i - \bar{\mu})^2 \neq \sum_{i=1}^n (\mu_i' - \bar{\mu}')^2$ implies $f(\mu_1, \cdots, \mu_n) \neq f(\mu_1, \cdots, \mu_n)$. We have now proved that $f(\mu_1, \cdots, \mu_n) = f(\mu_1', \cdots, \mu_n')$ if and only if $\sum_{i=1}^n (\mu_i - \bar{\mu})^2 = \sum_{i=1}^n (\mu_i' - \bar{\mu}')^2$. But this is simply another way of saying that there is an $F(x)$ such that $F(V) = f(\mu_1, \cdots, \mu_n)$.

Conversely, it is easy to prove that:

If $F(x)$ *is continuous, monotonically increasing, and* $F(0) = 0$, *then* $f(\mu_1, \cdots, \mu_n) = F(V)$ *has properties* (i)–(iv).

REFERENCE

[1] P. C. Tang, "The power function of the analysis of variance test with tables and illustrations of their use," *Stat. Res. Memoirs*, Vol. 2 (1938), pp. 126–157.

ON MILL'S RATIO FOR THE TYPE III POPULATION

By Des Raj[1]

University of Lucknow

1. Introduction and summary. Mills [1], Gordon [2], Birnbaum [3], and the author [4] have studied the ratio of the area of the standardized normal curve from x to ∞ and the ordinate at x. The object of this note is to establish the monotonic character of, and to obtain lower and upper bounds for, the ratio of the ordinate of the standardized Type III curve at x and the area of the curve from x to ∞. This ratio, as shown by Cohen [5] and the author [6], has to be calculated for several values of x when solving approximately the equations involved in the problem of estimating the parameters of Type III populations from truncated samples. It was found by the author that, for large values of

Received 2/28/52, revised 12/31/52.

[1] Now at the Indian Statistical Institute.