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.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org