# A NOTE ON THE ANALYSIS OF VARIANCE WITH UNEQUAL CLASS FREQUENCIES[1]

By Abraham Wald[2]

Let us consider $p$ groups of variates and denote by $m_j$ $(j = 1, \cdots, p)$ the number of elements in the $j$-th group. Let $x_{ij}$ be the $i$-th element of the $j$-th group. We assume that $x_{ij}$ is the sum of two variates $\epsilon_{ij}$ and $\eta_j$, i.e. $x_{ij} = \epsilon_{ij} + \eta_j$, where $\epsilon_{ij}$ $(i = 1, \cdots, m_j; j = 1, \cdots, p)$ is normally distributed with mean $\mu$ and variance $\sigma^2$, and $\eta_j$ $(j = 1, \cdots, p)$ is normally distributed with mean $\mu'$ and variance $\sigma'^2$. All the variates $\epsilon_{ij}$ and $\eta_j$ are supposed to be distributed independently.

The intraclass correlation $\rho$ is given by[3]

$$\rho = \frac{\sigma'^2}{\sigma^2 + \sigma'^2}.$$

Confidence limits for $\rho$ have been derived only in case of equal class frequencies, i.e. $m_1 = m_2 = \cdots = m_p$. In this paper we shall deal with the problem of determining the confidence limits for $\rho$ in the case of unequal class frequencies. Since $\rho$ is a monotonic function of $\frac{\sigma'^2}{\sigma^2}$, our problem is solved if we derive confidence limits for $\frac{\sigma'^2}{\sigma^2}$.

Denote by $\bar{x}_j$ the arithmetic mean of the $j$-th group, i.e.

$$\bar{x}_j = \frac{\sum_{i=1}^{m_j} \epsilon_{ij}}{m_j} + \eta_j. \tag{1}$$

Hence the variance of $\bar{x}_j$ is equal to

$$\sigma^2_{\bar{x}_j} = \frac{\sigma^2}{m_j} + \sigma'^2. \tag{2}$$

Denote $\frac{\sigma'^2}{\sigma^2}$ by $\lambda^2$. Then we have

$$\sigma^2_{\bar{x}_j} = \sigma^2\left(\frac{1}{m_j} + \lambda^2\right) = \frac{\sigma^2}{w_j}, \tag{3}$$

---

[1] The author is indebted to Professor H. Hotelling for formulating the problem dealt with in this paper.

[2] Research under a grant-in-aid from the Carnegie Corporation at New York.

[3] See for instance R. A. Fisher, *Statistical Methods for Research Workers*, 6-th edition, p. 228.

where

$$w_j = \frac{m_j}{1 + m_j\lambda^2}. \tag{4}$$

Now we shall prove that

$$\frac{1}{\sigma^2}\sum_{j=1}^{p}\left[w_j\left(\bar{x}_j - \frac{\sum_{j=1}^{p} w_j\bar{x}_j}{\sum_{j=1}^{p} w_j}\right)^2\right] \tag{5}$$

has the $\chi^2$-distribution with $p - 1$ degrees of freedom. Let

$$y_j = \sqrt{w_j}\,\bar{x}_j \qquad (j = 1, \cdots, p)$$

and consider the orthogonal transformation

$$\begin{aligned} y_1' &= L_1(y_1, \cdots, y_p), \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ y_{n-1}' &= L_{p-1}(y_1, \cdots, y_p), \\ y_p' &= L_p(y_1, \cdots, y_p) = \frac{\sqrt{w_1}\,y_1 + \cdots + \sqrt{w_p}\,y_p}{\sqrt{w_1 + \cdots + w_p}}, \end{aligned}$$

where $L_1(y_1, \cdots, y_p), \cdots, L_{p-1}(y_1, \cdots, y_p)$ denote arbitrary homogenous linear functions subject to the only condition that the transformation should be orthogonal.

Since the mean value of $y_j$ is equal to $\sqrt{w_j}\,(\mu + \mu')$ and the variance of $y_j$ is equal to $\sigma^2$, we obviously have: The mean value of $y_j'$ $(j = 1, \cdots, p - 1)$ is equal to zero, the variance of $y_j'$ $(j = 1, \cdots, p)$ is equal to $\sigma^2$. In order to prove our statement, we have only to show that the expression (5) is equal to $\frac{1}{\sigma^2}(y_1'^2 + \cdots + y_{p-1}'^2)$. If we substitute in (5) $\frac{y_j}{\sqrt{w_j}}$ for $\bar{x}_j$, we get

$$\frac{1}{\sigma^2}\sum_{j=1}^{p}\left\{w_j\left[\frac{y_j^2}{w_j} - 2\frac{y_j}{\sqrt{w_j}}\frac{\sum_{j=1}^{p}\sqrt{w_j}\,y_j}{\sum_{j=1}^{p} w_j} + \left(\frac{\sum_j \sqrt{w_j}\,y_j}{\Sigma w_j}\right)^2\right]\right\}$$

$$= \frac{1}{\sigma^2}\left[\sum_j y_j^2 - 2\frac{(\sum_j \sqrt{w_j}\,y_j)^2}{\sum_j w_j} + \frac{(\sum_j \sqrt{w_j}\,y_j)^2}{\sum_j w_j}\right] \tag{5'}$$

$$= \frac{1}{\sigma^2}\left[\sum_j y_j^2 - \frac{(\sum_j \sqrt{w_j}\,y_j)^2}{\sum_j w_j}\right] = \frac{1}{\sigma^2}\left[\sum_{j=1}^{p} y_j^2 - y_p'^2\right] = \frac{1}{\sigma^2}\left[\sum_{j=1}^{p} y_j'^2 - y_p'^2\right]$$

$$= \frac{1}{\sigma^2}(y_1'^2 + \cdots + y_{p-1}'^2).$$

Since $\frac{\Sigma\Sigma(x_{ij} - \bar{x}_j)^2}{\sigma^2}$ has the $\chi^2$ distribution with $N - p$ degrees of freedom, the expression

$$F = \frac{N - p}{p - 1} \frac{\sum_{j=1}^{p} \left\{ w_j \left( \bar{x}_j - \frac{\Sigma w_j \bar{x}_j}{\Sigma w_j} \right)^2 \right\}}{\Sigma\Sigma(x_{ij} - \bar{x}_j)^2} \tag{6}$$

has the analysis of variance distribution with $p - 1$ and $N - p$ degrees of freedom, where $N = m_1 + \cdots + m_p$. In case $m_1 = m_2 = \cdots = m_p = m$, we have

$$F = \frac{N - p}{p - 1} \frac{\sum_{j=1}^{p} (\bar{x}_j - \bar{x})^2}{\Sigma\Sigma(x_{ij} - \bar{x}_j)^2} \cdot \frac{m}{1 + m\lambda^2} = \frac{1}{1 + m\lambda^2} F^*, \tag{6'}$$

where $\bar{x} = \frac{\Sigma\Sigma x_{ij}}{N}$ and $F^* = \frac{N - p}{p - 1} \frac{m\Sigma(\bar{x}_j - \bar{x})^2}{\Sigma\Sigma(x_{ij} - \bar{x}_j)^2}$.

Hence

$$\lambda^2 = \left( \frac{F^*}{F} - 1 \right) \frac{1}{m}.$$

If $F_1$ denotes the lower and $F_2$ the upper confidence limit of $F$, we obtain for $\lambda^2$ the confidence limits

$$\left( \frac{F^*}{F_1} - 1 \right) \frac{1}{m} \quad \text{and} \quad \left( \frac{F^*}{F_2} - 1 \right) \frac{1}{m}.$$

Let us now consider the general case that $m_1, \cdots, m_p$ are arbitrary positive integers. First we shall show that the set of values of $\lambda^2$, for which (6) lies between its confidence limits $F_1$ and $F_2$, is an interval. For this purpose we have only to show that

$$f(\lambda^2) \equiv \sum_{j=1}^{p} \left\{ w_j \left( \bar{x}_j - \frac{\Sigma w_j \bar{x}_j}{\Sigma w_j} \right)^2 \right\}$$

is monotonically decreasing with $\lambda^2$. In fact

$$\frac{df(\lambda^2)}{d\lambda^2} = \sum_{j=1}^{p} \frac{dw_j}{d\lambda^2} \left( \bar{x}_j - \frac{\Sigma w_j \bar{x}_j}{\Sigma w_j} \right)^2 - 2 \frac{d}{d\lambda^2} \left( \frac{\Sigma w_j \bar{x}_j}{\Sigma w_j} \right) \left[ \sum_{j=1}^{p} w_j \left( \bar{x}_j - \frac{\Sigma w_j \bar{x}_j}{\Sigma w_j} \right) \right].$$

Since

$$\sum_{j=1}^{p} w_j \left( \bar{x}_j - \frac{\Sigma w_j \bar{x}_j}{\Sigma w_j} \right) = 0,$$

we have

$$\frac{df(\lambda^2)}{d\lambda^2} = \sum_{j=1}^{p} \frac{dw_j}{d\lambda^2} \left( \bar{x}_j - \frac{\Sigma w_j \bar{x}_j}{\Sigma w_j} \right)^2 = \sum_{j=1}^{p} - w_j^2 \left( \bar{x}_j - \frac{\Sigma w_j \bar{x}_j}{\Sigma w_j} \right)^2 < 0,$$

which proves our statement.

Hence the lower confidence limit $\lambda_1^2$ of $\lambda^2$ is given by the root of the equation in $\lambda^2$:

$$F = \frac{N-p}{p-1} \frac{\sum_{j=1}^{p} \left\{ w_j \left( \bar{x}_j - \frac{\Sigma w_j \bar{x}_j}{\Sigma w_j} \right)^2 \right\}}{\Sigma\Sigma(x_{ij} - \bar{x}_j)^2} = F_2 \tag{7}$$

and the upper confidence limit $\lambda_2^2$ of $\lambda^2$ is given by the root of the equation in $\lambda^2$:

$$F = F_1 . \tag{8}$$

Since $f(\lambda^2)$ is monotonically decreasing, the equations (7) and (8) have at most one root in $\lambda^2$. If the equation (7) or (8) has no root, the corresponding confidence limit has to be put equal to zero. If neither (7) nor (8) has a root, we have to reject at least one of the hypotheses:

(1) $x_{ij} = \epsilon_{ij} + \eta_j$.

(2) The variates $\epsilon_{ij}$ and $\eta_j$ ($i = 1, \cdots, m_j$ ; $j = 1, \cdots, p$) are normally and independently distributed.

(3) Each of the variates $\epsilon_{ij}$ has the same distribution.

(4) Each of the variates $\eta_j$ has the same distribution.

The equations (7) and (8) are complicated algebraic equations in $\lambda^2$. For the actual calculation of the roots of these equations, well known approximation methods can be applied making use also of the fact that the left members are monotonic functions of $\lambda^2$. In applying any approximation method it is very useful to start with two limits of the root which do not lie far apart. We shall give here a method of finding such limits.

Denote by $\bar{F}$ the function which we obtain from $F$ (formula (6)) by substituting

$$\bar{w}_j = \frac{l_j}{1 + l_j\lambda^2} \text{ for } w_j \qquad (j = 1, \cdots, p).$$

Let $\bar{f}$ be the function obtained from $f$ by the same process.

Denote by $\varphi(m, \lambda^2)$ the function which we obtain from $\bar{F}$ by substituting $m$ for $l_1, \cdots, l_p$. We shall first show that $\bar{F}$ is non-decreasing with increasing $l_k$ ($k = 1, \cdots, p$), i.e. $\frac{\partial \bar{F}}{\partial l_k} \geq 0$. For this purpose we have only to show that $\frac{\partial \bar{f}}{\partial l_k} \geq 0$. We have:

$$\frac{\partial \bar{f}}{\partial l_k} = \sum_j \frac{\partial \bar{w}_j}{\partial l_k} \left( \bar{x}_j - \frac{\Sigma \bar{w}_j \bar{x}_j}{\Sigma \bar{w}_j} \right)^2 - 2 \frac{\partial}{\partial l_k} \left( \frac{\Sigma \bar{w}_j \bar{x}_j}{\Sigma \bar{w}_j} \right) \cdot \left[ \Sigma \bar{w}_j \cdot \left( \bar{x}_j - \frac{\Sigma \bar{w}_j \bar{x}_j}{\Sigma \bar{w}_j} \right) \right]$$

$$= \sum_j \frac{\partial \bar{w}_j}{\partial l_k} \left( \bar{x}_j - \frac{\Sigma \bar{w}_j \bar{x}_j}{\Sigma \bar{w}_j} \right)^2 = \frac{1}{(1 + l_k\lambda^2)^2} \left( \bar{x}_k - \frac{\Sigma \bar{w}_j \bar{x}_j}{\Sigma \bar{w}_j} \right)^2 \geq 0.$$

Hence our statement is proved. Denote by $m'$ the smallest and by $m''$ the greatest of the values $m_1, \cdots, m_p$. Then we obviously have

$$\varphi(m', \lambda^2) \leq F \leq \varphi(m'', \lambda^2). \tag{9}$$

Denote by $\lambda_1'^2, \lambda_1''^2, \lambda_2'^2, \lambda_2''^2$ the roots in $\lambda^2$ of the following equations respectively:

$$\varphi(m', \lambda^2) = F_2 ;$$

$$\varphi(m'', \lambda^2) = F_2 ;$$

$$\varphi(m', \lambda^2) = F_1 ; \qquad \varphi(m'', \lambda^2) = F_1 .$$

Since $F$ is monotonically decreasing with increasing $\lambda^2$, on account of (7), (8), and (9) we obviously have

$$\lambda_1'^2 \leq \lambda_1^2 \leq \lambda_1''^2$$

and

$$\lambda_2'^2 \leq \lambda_2^2 \leq \lambda_2''^2.$$

The above inequalities give us the required limits.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

## THE DISTRIBUTION OF QUADRATIC FORMS IN NON-CENTRAL NORMAL RANDOM VARIABLES

BY WILLIAM G. MADOW[1]

The following theorem is the algebraic basis of the theorem of R. A. Fisher and W. G. Cochran which states necessary and sufficient conditions that a set of quadratic forms in normally and independently distributed random variables should themselves be independently distributed in $\chi^2$-distributions.[2]

THEOREM I. *If the real quadratic forms $q_1, \cdots, q_m$, in $x_1, \cdots, x_n$, are such that*

$$\sum_\gamma q_\gamma = \sum_\nu x_\nu^2, \tag{1}$$

*and if the rank of $q_\gamma$ is $n_\gamma$, then a necessary and sufficient condition that*

$$q_\gamma = \sum_\alpha z_\alpha^2, \tag{2}$$

[1] The letters $i, j, \mu, \nu$ will assume all integral values from 1 through $n$, the letter $\gamma$ will assume all integral values from 1 through $m$, $(n \geq m)$, the letter $\alpha$ will assume all integral values from $n_1 + \cdots + n_{\gamma-1} + 1$ through $n_1 + \cdots + n_\gamma$, $(n_0 = 0, n_1 + \cdots + n_m = n')$, the letters $\beta, \beta'$ will assume all integral values from 1 through $n'$, and the letters $r, s$ will assume all integral values from 1 through $n - 1$.

[2] The references are: W. G. Cochran, "The Distribution of Quadratic Forms in a Normal System, with Applications to the Analysis of Covariance," *Proc. Camb. Phil. Soc.*, Vol. 30 (1934), pp. 178–191, and R. A. Fisher, "Applications of 'Student's' Distribution," *Metron*, Vol. 5 (1926), pp. 90–104.