Statistical Design of Load Balance Traffic Studies

Ole A. Pedersen
GTE Automatic Electric Laboratories, Northlake, Illinois, U.S.A.

ABSTRACT

This paper deals with the experimental design problem of deciding how much data is needed for a traffic study to determine whether or not dial office administration objectives related to the maintenance of load balance of equipment groups are being met with an assurance that the risks of drawing wrong conclusions from the data are under control. Procedures are presented for planning tests based on the use of the studentized range of the hypothesis that the range of equipment group loads or other traffic characteristics lies within prescribed limits. Least squares estimates of the standard deviation of random effects are obtained by use of analysis of variance techniques. Range methods of estimating the standard deviation are also presented for the purpose of reducing the data processing needed for a study. Examples are given to illustrate the methods.

1. INTRODUCTION

At the Second International Teletraffic Congress held at The Hague in 1958, D. N. Barnes presented a paper on statistical methods for administration of dial offices [1]. The emphasis in the paper was on the application of quality control methods such as those used to control manufacturing processes. This paper deals with the experimental design problem of determining how much data is needed for a traffic study to determine whether or not load balance objectives are being met with an assurance that the risks of drawing wrong conclusions from the data are under control.

Section 2 presents a statement of the experimental design problem.

In Section 3 an analysis of variance is done for a mixed two-way cross-classification of data with one observation per combination and interaction whose model includes terms for fixed effects corresponding to equipment group loads and random effects corresponding to traffic variations. An interaction term allows for the possibility of group traffic variations being dependent on group load. The purpose of the analysis is to define the estimate of the error variance to be used to scale the data for statistical tests of the range of group loads.

The simplest test of the range of group loads occurs for two groups. Then only two group means are compared and the study can be planned by making use of the properties of Student's t-test. The planning of such studies is discussed in Section 4. There still remains the problem of planning studies for the comparison of more than two means.

In Section 5 the studentized range is used as a basis for the comparison of more than two means. Procedures are presented for using tables of significance of the studentized range to modify the planning of studies on two groups to make them applicable to the planning of studies on more than two groups.

The amount of data processing for load balance traffic studies is considerable. The bulk of the arithmetic work is in estimating the variance of random effects by least squares methods. Procedures are presented for obtaining range method estimates of the standard deviation that require less calculation effort. Section 6 presents the range method of estimating variance for a one-way classification of data. Section 7 presents the range method of estimating standard deviation for a two-way classification of data.

Section 8 is an example of the analysis of a planned load balance study.

The methods of this paper are applicable to sequential analysis of load balance study data [5][14]. Sequential analysis can arrive at a decision much sooner and with substantially fewer observations than equally reliable test procedures based on a predetermined number of observations. The introduction of loss or utility functions associated with the outcomes of traffic studies takes the planning of these studies into the area of statistical decision making wherein the economic justification for conducting studies can be explicitly considered [9]. Such a development would further the application of the operations research approach to telephone plant measurements espoused by A. Elid on at the Munich Teletraffic Congress in 1970 [3].

2. STATEMENT OF PROBLEM

Mathematically stated, we have n equipment groups whose loads or other traffic characteristics have a range w. The administrative objective is to keep the value of w less than a prescribed value \( w_0 \). A load balance traffic study is to be made to test the hypothesis \( H_0 \) that \( w \) is less than \( w_0 \) versus the alternative hypothesis \( H_1 \) that \( w \) is greater than or equal to \( w_0 \). The problem is to determine the number of hours' data needed to test the hypothesis with the assurances that:

1. \( H_0 \) is true because \( w = w_1 \) where \( w_1 \leq w_0 \), then the probability of concluding that \( H_1 \) is true is not greater than a prescribed value \( \alpha \); and, conversely,

2. \( H_1 \) is true because \( w = w_2 \) where \( w_2 > w_0 \), then the probability of concluding that \( H_0 \) is true is not greater than a prescribed value \( \beta \).

If the first type of error is committed, then unnecessary load balancing is done. If the second type of error is committed, then necessary load balancing is left undone and telephone customers may receive poor service. For the rest of this paper, \( w_1 \) equals zero.
The planning of load balance traffic studies is done in terms of the normalized range $W$ defined by

$$W = \frac{w}{\sigma}$$

where $\sigma$ is the standard deviation of the error in measuring the range. Analysis of variance (ANOVA) methods are used in the next section to develop an estimate of $\sigma$.

3. ANALYSIS OF VARIANCE

Considering a series of measurements on, say, traffic load made on $n$ equipment groups in a study period of $m$ busy hours, we obtain data $x_{ij}$ on group $i$ in busy hour $j$. We assume that $x_{ij}$ is expressible as

$$x_{ij} = \mu + a_i + bj + c_{ij} + z_{ij}$$

where

- $\mu$ is a constant independent of $i$ and $j$;
- $a_i$ is a constant characterizing group $i$;
- $b_j$ is a random variable characterizing busy hour $j$ and having expectation zero and variance $\sigma_{b}^2$;
- $c_{ij}$ is a random variable with expectation zero and variance $\sigma_{c}^2$ accounting for interaction of the group $i$ characteristic with busy hour $j$; and
- $z_{ij}$ is a random variable independently distributed with expectation zero and variance $\sigma_{z}^2$.

The interaction term $c_{ij}$ in the model allows for the possibility of the amplitude of group $i$ variations being dependent on the characteristic $a_i$. All random variables are assumed to have the normal distribution. Furthermore, it is assumed that

$$\sum_{i=1}^{n} a_i = 0 \quad \sum_{i=1}^{n} c_{ij} = 0.$$

Let

$$X_i = \sum_{j=1}^{m} x_{ij} \quad X_j = \sum_{i=1}^{n} x_{ij}$$

Then estimates of $\mu$, $a_i$ and $b_j$ are obtained from

$$\hat{\mu} = \frac{X}{nm} \quad \hat{a}_i = \frac{X_i}{m-\mu} \quad \hat{b}_j = X_j/n-\mu$$

If $a_i$ and $a_o$ are, respectively, the smallest and largest values of the $a_i$'s, then the range $w$ of the $a_i$'s is

$$w = a_o - a_1.$$

If $\hat{a}_i$ and $\hat{a}_o$ are, respectively, the smallest and largest values of the $\hat{a}_i$'s, then the estimate $\hat{w}$ of $w$ is

$$\hat{w} = \hat{a}_o - \hat{a}_1.$$

Let sums of squared deviations (SS) be defined for:

- groups by $SS_a = \sum_{i=1}^{n} x_i^2/n - X^2/nm$;
- hours by $SS_b = \sum_{j=1}^{m} X_j^2/n - x^2/nm$;
- residuals by $SS_{ab} = \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij}^2/m - \sum_{i=1}^{n} X_i^2/m - \sum_{j=1}^{m} X_j^2/n + X^2/nm$; and
- the sum of the above by $SS = \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij}^2/nm$.

The expectations of the SS's are [2]:

$$ESS_a = (n-1)(\sigma_z^2 + \sigma_c^2 + m \sum a_i^2)$$

$$ESS_b = (m-1)(\sigma_z^2 + n \sigma_b^2)$$

$$ESS_{ab} = (n-1)(m-1)(\sigma_z^2 + \sigma_c^2)$$

$$ESS = (n-1)(m-1)(\sigma_z^2 + \sigma_c^2)$$

The important thing to notice in these equations is that $ESS_a$ and $ESS_{ab}$ are functions of the variance $\sigma_z^2$ defined by

$$\sigma_z^2 = \sigma_z^2 + \sigma_c^2.$$
Thus, the variance ratio $F$ becomes in terms of $MS$'s:

$$F = \frac{MS_a}{MS_{ab}}.$$  

The results of the analysis of variance (ANOVA) calculations are typically presented as shown in Table 1.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>groups</td>
<td>(n-1)</td>
<td>SS_a</td>
<td>MS_a</td>
<td>F</td>
</tr>
<tr>
<td>hours</td>
<td>(m-1)</td>
<td>SS_b</td>
<td>MS_b</td>
<td></td>
</tr>
<tr>
<td>residuals</td>
<td>(n-1)(m-1)</td>
<td>SS_{ab}</td>
<td>MS_{ab}</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>(nm-1)</td>
<td>SS</td>
<td>MS</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 1. ANOVA table for least squares estimate of variance.

$MS_{ab}$ is an estimate of the variance $\sigma^2$ of the error encountered in measuring the range. For notational simplicity we shall let the estimate of $\sigma$ be designated by $s$ where

$$s = \frac{MS_{ab}^{1/2}}{s}.$$  

The number $v$ of degrees of freedom in the estimate of $s$ is

$$v = (n-1)(m-1).$$

An estimate $\hat{w}$ of the normalized range $W$ is provided by

$$\hat{w} = \frac{\hat{w}}{s}.$$  

4. THE RANGE OF TWO GROUPS

Except for the case where only two equipment groups are being studied, the $F$-test is not a test of the range $w$ of the group characteristics as a whole. We shall use instead, the studentized range statistic $Q$ defined by

$$Q = m^{1/2}\frac{\hat{w}}{s} = m^{1/2}\frac{\hat{w}}{s}$$

and used in Tukey's method of multiple comparison of means in the analysis of variance [2].

If $Q$ is greater than some critical value $Q_a(n,v)$ defined for a level of significance $\alpha$, then it is considered to be statistically significant. By this is meant that, if the hypothesis $H_0$ that $w$ equals zero is true, then, for $n$ groups and $v$ degrees of freedom in the estimate of $\sigma$, the probability of $Q$ being equal to or greater than $Q_a(n,v)$ is $\alpha$. If $Q$ is both statistically significant and greater than some prescribed value $Q_0$ defined by

$$Q_0 = m^{1/2}\frac{W_0}{s}$$

where $W_0$ corresponds to an unacceptable degree of unbalance, then it is concluded that the range of the group characteristics is excessive and that the groups are unbalanced. A short table of values of $Q_0, w_0(n,v)$ is presented in Table 2. More extensive tables for $\alpha$ equal to 0.01, 0.05 and 0.10 may be found in [7].

### Table 2. Critical values of studentized range

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>$Q_{0.10}$ ($n,v$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.01 4.59 5.39 5.93 6.33 6.65</td>
</tr>
<tr>
<td>4</td>
<td>2.75 4.07 4.73 5.17 5.50 5.76</td>
</tr>
<tr>
<td>6</td>
<td>2.63 3.83 4.43 4.83 5.13 5.36</td>
</tr>
<tr>
<td>12</td>
<td>2.52 3.62 4.16 4.51 4.78 4.99</td>
</tr>
<tr>
<td>20</td>
<td>2.44 3.46 3.95 4.27 4.51 4.70</td>
</tr>
<tr>
<td>60</td>
<td>2.36 3.31 3.75 4.04 4.25 4.42</td>
</tr>
</tbody>
</table>

The planning of load balance studies on two equipment groups is based on the comparison of two group means for groups having unknown but equal variances. The studentized range test is a generalization of this test to a comparison of the extreme values of the means for more than two groups.

For two groups, the critical value of $Q$ is $Q_a(2,v)$ where $v$ equals the degrees of freedom in the estimate of $\sigma$ used in calculating $Q$. If $t_a(v)$ and $F_a(n-1,v)$ are, respectively, the critical values for a double-sided $t$-test and an $F$-test at levels of significance $\alpha$, then

$$Q_a(2,v) = \frac{t_a(v)}{\sqrt{2F_a(n-1,v)}}.$$  

The second of these distributions enables us to plan a study with the use of the power function of the $F$-distribution for $n=2$. The tables of the $t$- and $F$-distributions commonly found in statistical textbooks are tables of $t_a(v)$ and $F_a(n-1,v)$[2][10].

The non-centrality parameter $\phi^2$ for the $F$-distribution when comparing the means of two groups is

$$\phi^2 = \frac{\alpha}{\phi^2} = \frac{\alpha}{\phi^2}.$$  

The variance ratio $F$ provides an estimate $\theta^2$ of $\phi^2$ given by

$$F = 1 + n\theta^2$$

where, when comparing the means of two groups,

$$\theta^2 = \frac{\alpha}{\phi^2} = \frac{\alpha}{\phi^2}.$$  

Since $F$ is a random variable, $\theta$ also is a random variable. It can also serve as an estimate of the parameter $\phi$ which is a measure of $W$. A critical value of $\phi$ can be associated with any critical value of $F$. If $\phi$ equals zero, then $F$ is characterized by having the $F$-distribution. On the other hand, if $\phi$ is greater than zero, then $F$ has the "non-central $F$-distribution." When $\phi$ equals zero, the critical value of $F$ at the $\alpha$ level of significance equals $F_0(n-1,v)$.

Let $\phi_a(n,1,v)$ be the critical value of $\phi$ for a non-central $F$-distribution with degrees of freedom $(n-1)$ and $v$ for which the probability of $F$ being equal to or greater than $F_0(n-1,v)$ is $(1-\phi)$. $(1-\phi)$ is the power of the test of $H_0$ at the level of significance $\alpha$. Table 3 is a short table of values of $\phi_a(n,1,v)$ derived from curves and tables in [8][11][12].
\[ \phi_B(1,v) = \frac{\phi_B(1,\nu)}{\nu} \]

\[ \alpha = \beta / v \]

\[ \begin{array}{cccccccc}
0.10 & 0.10 & 2.54 & 2.36 & 2.28 & 2.20 & 2.15 & 2.10 & 2.07 \\
0.05 & 2.87 & 2.66 & 2.58 & 2.47 & 2.42 & 2.36 & 2.33 & 0.01 \\
0.10 & 3.21 & 3.09 & 2.99 & 2.92 & 2.85 & 2.81 & & \\
0.05 & 2.76 & 2.65 & 2.50 & 2.41 & 2.33 & 2.29 & & \\
0.05 & 3.08 & 2.92 & 2.78 & 2.67 & 2.58 & 2.55 & & \\
0.10 & 3.70 & 3.48 & 3.32 & 3.18 & 3.07 & 3.03 & & \\
0.01 & 3.84 & 3.48 & 3.17 & 2.97 & 2.81 & 2.73 & & \\
0.05 & 4.33 & 3.82 & 3.49 & 3.15 & 3.06 & 2.98 & & \\
0.01 & 4.98 & 4.47 & 4.06 & 3.79 & 3.55 & 3.47 & & \\
\end{array} \]

**TABLE 3.** Critical values of the non-centrality parameter for non-central F-distributions having 1 and \( \nu \) degrees of freedom, \( \phi_B(1,\nu) \) and type II error probability \( \beta \).

When planning a study on two groups it is necessary to specify beforehand, with the values of \( \alpha \), \( \beta \) and \( \sigma \), the range objective \( w_0 \) within which balance is to be maintained. The minimum number \( m \) of hours needed for the study is the smallest integral value of \( m \) satisfying the inequality

\[ m^2 w_0 / 2 \sigma^2 \geq \phi_B(1,\nu) \]

A lower bound \( m_0 \) on \( m \) is provided by

\[ m_0 = 2 \phi_B(1,\nu) \]

As an example, consider the determination of the minimum number of hours needed for a study on two equipment groups each of which have been engineered to switch approximately 3000 calls per hour. The range of the calls offered to the equipment groups should not be greater than 10% of the average \( \mu \). It is given for planning purposes that \( \alpha = 0.10 \), \( \beta = 0.05 \) and \( \sigma = 0.10 \). We have that \( \mu = 3000 \) and \( w_0 / \mu = 0.10 \) so that \( w_0 = 300 \). From Table 3, \( \phi_B(1,\nu) = 2.33 \). Therefore the number \( m \) of hours needed for the study is not less than \( m_0 = 1.95 \times 2 \) hours. For \( m = 5 \) hours \( m^2 w_0 / 2 \sigma^2 = 3.73 \) and is greater than \( \phi_B(1,\nu) \) so that \( m = 300 \). Therefore, 5 hours are sufficient for the study. It is not possible to determine from Table 3 whether less than 5 hours would be sufficient for the study.

5. **THE RANGE OF MORE THAN TWO GROUPS**

The procedure for planning studies on only two groups is applicable to the situation in which interest lies in comparing the means of a particular pair of a larger number \( n \) of groups. In this situation, it is necessary that the two groups of interest be selected before any data has been looked at. Furthermore, if the error variance \( \sigma^2 \) is the same for all groups, then the estimate \( \bar{\sigma}^2 \) can be based on \( (n-1)(m-1) \) instead of on \( (m-1) \) degrees of freedom and thereby possibly allow the duration of the study to be shortened. In this case, the minimum number \( m \) of hours needed for the study is the smallest integral value of \( m \) satisfying the inequality

\[ m^2 w_0 / 2 \sigma^2 \geq \phi_B(1,(n-1)(m-1);\nu) \]

The extension of the procedure to the case where we wish to test the equality of \( m \) group means is simple. All that is required is an adjustment of the value of the standard deviation \( \sigma \) in the calculations. The adjustment reflects the fact that, in testing the range, a comparison is made of the means of two out of \( n \) groups that have been selected after an examination of the data.

The adjustment factor is \( q_\alpha(n,\nu) \) defined by

\[ q_\alpha(n,\nu) = \frac{Q_\alpha(n,\nu)}{Q_\alpha(2,\nu)} \]

Table 4 presents values of \( q_\alpha(n,\nu) \) calculated from values of \( Q_\alpha(n,\nu) \) as in Table 2. For a range study, the minimum number \( m \) of hours needed for a study is the smallest integral value of \( m \) satisfying the relationship

\[ m^2 w_0 / 2 \sigma^2 \geq q_\alpha(n,\nu) \]

A lower bound \( m_0 \) on \( m \) is provided by

\[ m_0 = 2 q_\alpha(n,\nu) \]

**TABLE 4. Scale factors \( q_\alpha(n,2) \) for converting critical values of the studentized range for 2 groups to the critical values of the studentized range for \( n \) groups.**

As an example, consider the determination of the minimum number of hours needed for a study on four equipment groups each of which have been engineered to switch approximately 3000 calls per hour. The administration objective is to keep the range \( w_0 \) of the calls offered to the groups within 10% of their average \( \mu \). It is given for planning purposes that \( \alpha = 0.10 \), \( \beta = 0.05 \) and \( \sigma = 0.10 \). We have that \( \mu = 3000 \) and \( w_0 / \mu = 0.10 \) so that \( w_0 = 300 \). From Tables 3 and 4, \( q_\alpha(4,\nu) = 1.47 \) and \( q_\alpha(4,\nu) = 1.39 \). Therefore, the number \( m \) of hours needed for the study is not less than \( m_0 = 3.78 \times 4 \) hours. In order to determine the minimum number of hours
needed for the study, the following table is constructed in which \( m = 3(m-1) \):

<table>
<thead>
<tr>
<th>( v )</th>
<th>( 8 )</th>
<th>( 12 )</th>
<th>( 20 )</th>
<th>( 60 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( q_{0.10} (4,v) )</td>
<td>3.67</td>
<td>5.0</td>
<td>7.67</td>
<td>21.0</td>
</tr>
<tr>
<td>( q_{0.10} (3,v;0.05) )</td>
<td>1.46</td>
<td>1.44</td>
<td>1.42</td>
<td>1.40</td>
</tr>
<tr>
<td>( q_{0.10} (1,v;0.05) )</td>
<td>2.58</td>
<td>2.47</td>
<td>2.42</td>
<td>2.36</td>
</tr>
<tr>
<td>( q_{0.10} (2,v;0.05) )</td>
<td>1.97</td>
<td>1.59</td>
<td>1.24</td>
<td>0.72</td>
</tr>
</tbody>
</table>

The value of \( w_2/2a \) equals 1.67 and exceeds the value of \( m \) greater than 3.67 and less than 5.0. Interpolating in the table in terms of \( m \), it is found that \( m = 4.65 \). 6 hours are needed for the study.

6. RANGE ESTIMATION OF VARIANCE FOR A ONE-WAY CLASSIFICATION OF DATA

A considerable amount of data processing is required to conduct a load balance study. Most of the effort is directed towards obtaining an estimate of the residual or error variance \( \sigma^2 \). Range estimates of \( \sigma^2 \) are obtained more easily. We shall begin with the estimation of variance for a one-way classification of data [6][7].

Consider a sample of \( n \) independent identically distributed random variables having the normal distribution with mean \( \mu \) and variance \( \sigma^2 \). Let \( W_n \) be the range of the sample and \( W_n \) be the normalized range, i.e.,

\[
W_n = \frac{w_n}{\sigma}
\]

If the expected value \( EW_n \) of \( W_n \) is \( d_n \), then an unbiased estimate \( \delta \) of \( \sigma \) is obtained from

\[
\delta = \frac{w_n}{d_n}.
\]

Let \( V_n \) be the variance of \( W_n \), i.e.,

\[
V_n = E(W_n-d_n)^2
\]

The unbiased range estimate \( s_w^2 \) of \( \sigma^2 \) is

\[
s_w^2 = 2V_n w_n^2 / d_n^2.
\]

We shall use \( s_w \) instead of \( \delta \) as an estimate of \( \sigma \) to conform with the procedure of using \( s \) as an estimate of \( \sigma \). The quantity \( s_w \) is an unbiased estimate of \( \sigma \); however, \( s \) is not an unbiased estimate of \( \sigma \).

For the more general case where there are \( m \) independent samples like the one just considered, let \( w_j \) be the \( j \)th sample range and \( \bar{W}_n \) be the mean range defined by

\[
\bar{W}_n = \frac{1}{m} \sum_{j=1}^{m} W_j/m.
\]

The mean normalized range \( \bar{W}_n \) is \( \bar{W}_n/\sigma \). The expected value \( d_{nm} \) and variance \( V_{nm} \) of \( \bar{W}_n \) are

\[
d_{nm} = d_n \quad V_{nm} = V_n/m.
\]

\( \bar{W}_n \) has approximately the distribution of \( c_{nm} x/\sqrt{V_{nm}} \) where \( c_{nm} \) is a "scale factor," \( x^2 \) has the chi-squared distribution and \( V_{nm} \) is the "equivalent degrees of freedom." Let

\[
r = \frac{d_{nm}^2}{2 V_{nm}}.
\]

then,

\[
c_{nm} = d_{nm}^2 + V_{nm}
\]

and \( V_{nm} \) is approximately given by

\[
V_{nm} = r + 1/4 - 3/16 r^2 + 3/64 r^3.
\]

For large values of \( r \),

\[
c_{nm} \approx d_{nm} \quad V_{nm} \approx r.
\]

The unbiased range estimate \( s_w^2 \) of \( \sigma^2 \)

\[
s_w^2 = \frac{w_n^2}{c_{nm}^2}
\]

The equivalent degrees of freedom \( V_{nm} \) is the degrees of freedom to be associated with \( s_w \) and \( s_w^2 \) in statistical tests.

A load balance study requires the analysis of data having a two-way classification. Therefore, the above procedure requires modification.

7. RANGE ESTIMATION OF VARIANCE FOR A TWO-WAY CLASSIFICATION OF DATA

The procedure for obtaining values of the range for estimating the variance \( \sigma^2 \) in a load balance study is as follows [4][7]: For each equipment group, calculate a group mean \( \bar{x}_i \) given by

\[
\bar{x}_i = \frac{x_i}{m}.
\]

For each observation \( x_{ij} \), calculate a residual \( r_{ij} \) given by

\[
r_{ij} = x_{ij} - \bar{x}_i.
\]

For each hour's data, calculate the range \( W_j \) of the residuals \( r_{ij} \) and then calculate the mean range \( \bar{W}_n \) given by

\[
\bar{W}_n = \frac{1}{m} \sum_{j=1}^{m} W_j/m.
\]

The mean range is the basis of the estimate of \( \sigma^2 \). However, because of correlation between the hourly estimates of the range due to the adjustment of the data to obtain the residual deviations from the group means, the procedure for calculating the scale factor \( c_{nm} \) and equivalent degrees of freedom \( V_{nm} \) must be modified. The modification is in expressing the expectation \( d_{nm} \) and variance \( V_{nm} \) of the normalized mean range \( \bar{W}_n \) as

\[
d_{nm} = (1-1/m) r^2 d_n \quad V_{nm} = (m-1) (1+(m-1) r^2) V_n/m^2
\]

where \( r \) is the correlation between any two of the hourly ranges. The scale factor \( c_{nm} \) and equivalent degrees of freedom \( V_{nm} \) are determined from

\[
c_{nm}^2 = d_{nm}^2 + V_{nm}
\]

\[
V_{nm} = r + 1/4 - 3/16 + 3/64 r^2
\]

where

\[
r = d_{nm}^2/2 V_{nm}.
\]

The correlation coefficient \( r \) is a complex function of \( n \) and \( m \) for which a limited table of values has been calculated. In order to simplify the procedure for calculating the values of \( c_{nm} \) and \( V_{nm} \), an alternative expression for \( V_{nm} \) is used. It is

\[
V_{nm} = V_n/m
\]
Table 6 presents the data and its analysis obtained in a study on the balance of call attempts for four group selectors in a No. 1 EAX system. This is the study planned earlier for which the nominal values of \( \alpha \) and \( \sigma \) were assumed to be, respectively, 3000 and 90. The purpose of the study was to determine whether or not the normalized range \( W \) was greater than \( W_0 = 3.33 \) with risks \( c = 0.10 \) and \( \delta = 0.05 \).

8. AN EXAMPLE OF A TRAFFIC STUDY

Table 6 presents the data and its analysis obtained in a study on the balance of call attempts for four group selectors in a No. 1 EAX system. This is the study planned earlier for which the nominal values of \( \alpha \) and \( \sigma \) were assumed to be, respectively, 3000 and 90. The purpose of the study was to determine whether or not the normalized range \( W \) was greater than \( W_0 = 3.33 \) with risks \( c = 0.10 \) and \( \delta = 0.05 \).

### Table 6: Scale factors \( c_{nm} \) and equivalent degrees of freedom \( v_{nm} \) for the range estimate of the residual standard deviation for two-way classification of data with \( n \) groups and \( m \) hours.

<table>
<thead>
<tr>
<th>( n )</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>( c_{nm} )</th>
<th>( v_{nm} )</th>
<th>( c_{nm} )</th>
<th>( v_{nm} )</th>
<th>( c_{nm} )</th>
<th>( v_{nm} )</th>
<th>( d_n )</th>
<th>( d_n^{20}/v_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.00</td>
<td>1.00</td>
<td>1.08</td>
<td>3.70</td>
<td>1.10</td>
<td>8.11</td>
<td>1.11</td>
<td>12.5</td>
<td>1.12</td>
<td>16.9</td>
<td>1.128</td>
<td>0.876</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.35</td>
<td>1.98</td>
<td>1.57</td>
<td>7.49</td>
<td>1.63</td>
<td>16.6</td>
<td>1.65</td>
<td>25.7</td>
<td>1.66</td>
<td>34.7</td>
<td>1.693</td>
<td>1.815</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.58</td>
<td>2.93</td>
<td>1.88</td>
<td>11.12</td>
<td>1.97</td>
<td>24.9</td>
<td>2.00</td>
<td>38.6</td>
<td>2.02</td>
<td>52.3</td>
<td>2.059</td>
<td>2.738</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.75</td>
<td>3.82</td>
<td>2.12</td>
<td>14.42</td>
<td>2.22</td>
<td>32.9</td>
<td>2.26</td>
<td>50.9</td>
<td>2.28</td>
<td>69.1</td>
<td>2.326</td>
<td>3.623</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2.00</td>
<td>4.68</td>
<td>2.43</td>
<td>17.41</td>
<td>2.42</td>
<td>40.4</td>
<td>2.46</td>
<td>62.8</td>
<td>2.48</td>
<td>85.1</td>
<td>2.534</td>
<td>4.466</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2.09</td>
<td>5.48</td>
<td>2.45</td>
<td>21.33</td>
<td>2.58</td>
<td>47.7</td>
<td>2.62</td>
<td>73.9</td>
<td>2.64</td>
<td>100</td>
<td>2.704</td>
<td>5.207</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2.18</td>
<td>7.03</td>
<td>2.69</td>
<td>27.57</td>
<td>2.84</td>
<td>61.4</td>
<td>2.89</td>
<td>95.5</td>
<td>2.91</td>
<td>129</td>
<td>2.970</td>
<td>6.758</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2.25</td>
<td>7.68</td>
<td>2.78</td>
<td>30.11</td>
<td>2.93</td>
<td>67.3</td>
<td>2.98</td>
<td>105</td>
<td>3.00</td>
<td>142</td>
<td>3.078</td>
<td>7.454</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2.31</td>
<td>8.35</td>
<td>2.86</td>
<td>32.72</td>
<td>3.02</td>
<td>73.3</td>
<td>3.07</td>
<td>114</td>
<td>3.09</td>
<td>155</td>
<td>3.173</td>
<td>8.120</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>2.37</td>
<td>8.99</td>
<td>2.94</td>
<td>35.35</td>
<td>3.10</td>
<td>79.1</td>
<td>3.15</td>
<td>123</td>
<td>3.18</td>
<td>167</td>
<td>3.258</td>
<td>8.760</td>
<td></td>
</tr>
</tbody>
</table>

which is obtained by letting \( \rho_w \) equal \( 1/(m-1) \).

The results of the calculations using this approximation are presented in Table 5 for \( n=1 \) and \( m=2,5,10,15,20 \). The values of \( d_n \) and \( r = d_n/2v_n \) are also given. Comparison of the entries in Table 5 with similar tables in [4] and [7] reveals that the use of the approximate expression for \( V_{nm} \) has no significant effect on the values of the scale factors and the equivalent degrees of freedom; the numbers tend to agree with differences of about 2%.

A simple approximation for the values of \( v_{nm} \) in Table 5 is given by

\[
v_{nm} = (n-1)d_n^2/2v_n.
\]

A measure of the relative efficiency \( \eta_{nm} \) of using the range estimates \( d_n^2 \) instead of mean square estimates \( s^2 \) is given by

\[
\eta_{nm} = v_{nm}/(n-1)(m-1) = d_n^2/(2n-1)v_n.
\]

The minimum value of \( \eta_{nm} \) occurs in Table 5 for 20 groups and equals 0.69.

As an example of planning a study using the range estimate of \( \sigma \), consider the determination of the minimum number \( m \) of hours needed for a study on the four equipment groups for which it has already been determined that 5 hours are sufficient when the least squares estimate \( \sigma \) is used. Since the range estimate of \( \sigma \) is less efficient than the least squares estimate, a study using the range estimate cannot be shorter than one using the least squares estimate. For a 5-hour study, \( m^2\rho_0/2\sigma = 3.73 \). Referring to Table 5, it is found that \( v_{nm} = 11.2 \) for \( n=4 \) and \( m=5 \). Interpolating in Tables 3 and 4, we get \( v_{nm}(5,11.2;0.05) = 2.49 \) and \( v_{nm}(4,11.2) = 1.44 \). Their product equals 3.59 and is less than \( m^2\rho_0/2\sigma \). Therefore, 5 hours is sufficient for the study using the range estimate of \( \sigma \).

8. AN EXAMPLE OF A TRAFFIC STUDY

Table 6 presents the data and its analysis obtained in a study on the balance of call attempts for four group selectors in a No. 1 EAX system. This is the study planned earlier for which the nominal values of \( \alpha \) and \( \sigma \) were assumed to be, respectively, 3000 and 90. The purpose of the study was to determine whether or not the normalized range \( W \) was greater than \( W_0 = 3.33 \) with risks \( c = 0.10 \) and \( \delta = 0.05 \).

### Table 6A: Example of analysis of traffic study data

<table>
<thead>
<tr>
<th>Group</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>( \hat{b} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hour</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2829</td>
<td>2555</td>
<td>2872</td>
<td>2511</td>
<td>-163.80</td>
</tr>
<tr>
<td>2</td>
<td>3207</td>
<td>2841</td>
<td>3053</td>
<td>2648</td>
<td>81.70</td>
</tr>
<tr>
<td>3</td>
<td>3074</td>
<td>2876</td>
<td>2929</td>
<td>2622</td>
<td>17.20</td>
</tr>
<tr>
<td>4</td>
<td>3030</td>
<td>2841</td>
<td>3159</td>
<td>2532</td>
<td>34.95</td>
</tr>
<tr>
<td>5</td>
<td>3031</td>
<td>2948</td>
<td>3056</td>
<td>2507</td>
<td>29.95</td>
</tr>
</tbody>
</table>

\[
\hat{\omega} = 178.65 + 291.55 = 470.20
\]

### Table 6B: ANOVA table for least squares estimate of \( \sigma \)

<table>
<thead>
<tr>
<th>Group</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>( \hat{w} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hour</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>178.65</td>
<td>43.35</td>
<td>156.25</td>
<td>-291.55</td>
<td>2855.55</td>
</tr>
</tbody>
</table>

\[
\hat{\omega} = 178.65 + 291.55 = 470.20
\]

### Table 6C: Table for range estimate of \( \sigma \)

Table 6A presents the raw data, the mean deviations \( \bar{a} \) and \( \bar{b} \) from the grand mean \( \bar{u} \) and the value of \( \hat{\omega} \). The estimate \( \hat{\omega} \) of the range is found to be 470.20. Table 6B presents an ANOVA table for the estimation of the value of \( \hat{\omega} \).
The square root $s$ of the mean variance of the residuals is 87.17. Therefore the estimate $\hat{W}$ of the normalized range is 5.39. Since $\hat{W}$ is greater than $W_0$, it is concluded that the group selector matrices have unbalanced calling rates. Table 6C illustrates the range estimation of $\sigma$. The body of the table consists of values of $(x_{ij} - a_i - c_j)$ obtained from data in Table 6A. The marginal data $W_i$ are the ranges of the entries in the rows of the table. The mean range $\bar{W}$ equals 175.36. The value of the scale factor $c_{MN}$ obtained from Table 5 is 1.88. Therefore, the range estimate $s_{W}$ of $\sigma$ is 93.3 and the range estimate $W_{W}$ of the normalized range is 5.04. Since $W_{W}$ is greater than $W_0$, it is concluded that the group selector matrices have unbalanced calling rates.

The least squares and range estimates of $\sigma$ are consistent; they differ by 7%. Limited experience with the methods described in this paper resulted in range estimates of $\sigma$ being within $\pm 5\%$ of the least squares' estimates for relatively high degrees of imbalances in group traffic characteristics, i.e. for values of $w/u$ as great as 0.5, with range estimates tending to be smaller than the least squares' estimates. This is the result of interaction in the data. Higher degrees of imbalance result in range estimates tending to become increasingly less than the least squares' estimates. This suggests that the range method provides estimates of $\sigma^2$ instead of $(\sigma^2 + cc')$. The use of a logarithmic transformation of the data usually is sufficient to make the estimates consistent. It should be noted that, if a logarithmic transformation is used, then the range statistic $\hat{W}$ is a measure of $\log a_u/a_L$.

REFERENCES


