AN APPLICATION OF THE PRINCIPLE OF INCLUSION AND EXCLUSION

by

Merril W. Hume

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

MASTER OF SCIENCE

in

Statistics

February 1964

Blacksburg, Virginia

TABLE OF CONTENTS

Table of Contents (cont.)

I. INTRODUCTION

The importance of statistics expressible as maxima is well known. Throughout the literature one finds that the upper percentage points for a number of these statistics have been tabulated, usually by various special methods. There is, however, a general approach often useful in providing good approximations. This general procedure is often referred to as the principle of inclusion and exclusion. A representative selection of the various statistical applications of this principle in relation to statistics expressible as maxima is presented in this paper.[1/] It is hoped that such a review will provide the necessary methodology for the use of this procedure in other areas where it may be of value.

The principle of inclusion and exclusion is a very general method for the calculation of upper percentage points, say λ_{α}, for statistics expressed as maxima. Coupled with the Bonferroni inequality, this procedure allows one to

---

[1/] Statistics expressed as minima can be handled in a similar manner, hence will not be discussed in this paper.

establish upper and lower limits to λ_α. The upper bound approximation to λ_α requires only the knowledge of the distribution(s) of the variates under consideration. The lower bound, however, requires also the joint distribution(s) of pairs of the variates. Since the joint distribution is often difficult to evaluate, an approximation technique may be necessary. A detailed discussion of such an approximation with guidelines for its applicability to statistics other than those discussed is presented (Chapter VI).

Two alternative methods for the determination of upper percentage points for statistics expressed as maxima are discussed: Whittle's lower bound approximation and the assumption of independence. It is pointed out that Whittle's lower bound is stronger than that of Bonferroni only under certain conditions. The assumption of independence leads to approximately the same result as Bonferroni.

The upper percentage point, λ_α, of a statistic expressed as a maximum is most generally used as a test for a sample outlier. The following sequential test procedure is appropriate for all the situations considered in this paper. This procedure insures a controlled overall level of significance of $\leq \alpha$.

Given a ranked set of n variates, say $z_{(1)}, z_{(2)}, \ldots, z_{(n)}$ with $z_{(1)}$ being the largest, compute λ_α. Compare $z_{(1)}$ to λ_α. If $z_{(1)}$ is significant, one then excludes $z_{(1)}$ from the set and calculates a new λ_α based on n-1 variates. Compare $z_{(2)}$ with the new λ_α; and so on, until a non-significant value is found.

1.1 The principle of inclusion and exclusion

Consider the n events $A_1, A_2, \ldots, A_n$ and denote the probability of the occurrence of one or m events by p with appropriate subscripts, thus

$$p_i = \Pr\{A_i\}, \quad p_{i,j} = \Pr\{A_i A_j\}, \quad p_{i,j,k} = \Pr\{A_i A_j A_k\}, \ \ldots$$

For the sum of the $\binom{n}{r}$ possible p's with r different subscripts we write S_r, that is

$$S_1 = \sum_{i=1}^{n} p_i \ , \quad S_2 = \sum_{i<j} p_{i,j} \ , \quad S_3 = \sum_{i<j<k} p_{i,j,k} \ , \ \ldots$$

Then the probability $P_{m,n}$ of the realization of at least m events out of n is given by (see, for example, Feller, 1950)

$$P_{m,n} = \sum_{\ell=0}^{n-m} (-1)^{\ell} \binom{m+\ell-1}{m-1} S_{m+\ell} \ . \tag{1.1}$$

We shall be concerned with various statistical applications of this theorem. Let $x_1, x_2, \ldots, x_n$ be a set of random variables not necessarily independent or identically distributed, and $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ be the same variates arranged in descending order of magnitude. We now write $p_{i,j,\ldots,k}(\lambda)$ for the joint probability of the r events $x_i > \lambda$, $x_j > \lambda$, ..., $x_k > \lambda$. Correspondingly, $P_{m,n}(\lambda)$ will be the probability that $x_{(m)}$ exceeds λ. In the special case when the joint distribution of the x's is a symmetric function of the x's, that is, the union of any ℓ events has the same probability for any ℓ events, we have

$$S_r(\lambda) = \binom{n}{r} p_{1,2,\ldots,r}(\lambda)$$

and (1.1) takes the form

$$P_{m,n}(\lambda) = \Pr(x_{(m)} > \lambda) = \sum_{\ell=0}^{n-m} (-1)^{\ell} \binom{n}{m+\ell} \binom{m+\ell-1}{m-1} p_{1,2,\ldots,(m+\ell)}(\lambda) \quad . \tag{1.2}$$

In particular, for the most important case m=1 we have simply

$$\Pr(x_{(1)} > \lambda) = \sum_{\ell=1}^{n} (-1)^{\ell-1} \binom{n}{\ell} p_{1,2,\ldots,\ell}(\lambda) \quad . \tag{1.3}$$

If the events are not only symmetric but also independent, (1.3) reduces to

$$\Pr(x_{(1)}>\lambda) = \sum_{\ell=1}^{n}(-1)^{\ell-1}\binom{n}{\ell}[p_1(\lambda)]^{\ell}$$

$$= 1 - (1-p_1(\lambda))^{\ell} \quad ; \qquad (1.4)$$

a result expected from elementary probabilistic arguments. Applications of equation (1.3) are discussed in this paper.

The upper α significance points of the LHS of equation (1.3) may be determined by setting the RHS equal to α and solving for λ_α. For sufficiently large values of λ_α a first approximation to the LHS is given by the first term on the right, with later terms serving as successive correction terms. Closer approximations to the upper percentage points of statistics expressible as maxima may be obtained if the second and possibly later terms on the RHS can be evaluated.

It is usually quite difficult to evaluate the second and later terms of the RHS, consequently various approximation techniques are needed. Various such approximation techniques appear throughout the literature. The validity of such techniques and their applicability to other situations will be investigated in the later part of this paper.

II. THE EXTREME DEVIATE FROM THE SAMPLE MEAN

Halperin et al. (1955)[2/] and David (1956) have presented similar approaches to the problem of determining approximations to the upper percentage points of the extreme deviate from the sample mean for a sample drawn from a normal population. The results lead to tables which may be employed for an outlier test. These tables may also be used as a basis for multiple significance tests of a set of k means arising from independent samples on normal populations with the same variances but possibly different means. Consideration of the principle of inclusion and exclusion and the related inequality of Bonferroni were used to attain these results.

2.1 The case of known variance

Let $x_1, x_2, \ldots, x_n$ be a random sample of n observations drawn from a normal population with standard

---

[2/] This paper deals with the studentized maximum absolute deviate in normal samples. However, due to the symmetry of the normal distribution, the approach employed may readily be applied to the extreme studentized deviate.

deviation σ. Consider

$$z_i = \frac{x_i - \bar{x}}{\sigma} \qquad i = 1, 2, \ldots, n \quad .$$

Then we have

$$E\{z_i\} = 0 \quad ,$$

$$\mathrm{Var}\{z_i\} = (n-1)/n \quad ,$$

$$\mathrm{Cov}\{z_i, z_j\} = -1/n \quad , \qquad i \neq j$$

and the z_i's $(i=1,\ldots,n)$ have a n-variate joint normal distribution of rank n-1.

Now, at the significance level α, define the positive number, λ_α, by

$$\Pr\{z_1 < \lambda_\alpha,\ z_2 < \lambda_\alpha,\ \ldots,\ z_n < \lambda_\alpha\} = 1 - \alpha \tag{2.1}$$

or equivalently

$$\Pr\{z_{max} > \lambda_\alpha\} = \alpha \quad . \tag{2.2}$$

Define A_i as the event $z_i > \lambda_\alpha$, $i = 1,\ldots,n$. Then from the principle of inclusion and exclusion, we have

$$\Pr\{z_{max} > \lambda_\alpha\} = \binom{n}{1}\Pr(A_1) - \binom{n}{2}\Pr(A_1A_2) + \ldots + (-1)^{n-1}\Pr(A_1A_2\ldots A_n) \quad . \tag{2.3}$$

By Bonferroni's inequalities (Feller, 1950) we have

$$\binom{n}{1}\Pr(A_1) > \Pr\{z_{max} > \lambda_\alpha\} > \binom{n}{1}\Pr(A_1) - \binom{n}{2}\Pr(A_1A_2) \tag{2.4}$$

since the partial sums of (2.3) alternate about the total sum.

Now, since each z_i has a normal distribution with zero mean and variance $(n-1)/n$, and each pair, z_i, z_j, has a bivariate normal distribution with zero means, common variance $(n-1)/n$, and covariance $-1/n$, we can write

$$\Pr(A_1) = \frac{1}{\sqrt{2\pi}} \int_{\lambda_\alpha[n/(n-1)]^{\frac{1}{2}}}^{\infty} \exp - \frac{x^2}{2}\, dx \qquad (2.5a)$$

and

$$\Pr(A_1 A_2) = \iint_{\lambda_\alpha[n/(n-1)]^{\frac{1}{2}}}^{\infty} p(x_1, x_2)\, dx_1\, dx_2 \quad , \qquad (2.5b)$$

where $p(x_1 x_2)$ is a bivariate normal density with the variates involved having zero mean, unit variances, and correlation $-1/(n-1)$.

By equating the LHS of (2.4) to α and solving for λ_α, we clearly get an upper limit for λ, say λ_U. Similarly, equating the RHS of (2.4) to α and solving for λ_α we get a lower limit for λ_α, say λ_L.

This first term approximation, λ_U, for the determination of upper percentage points of the extreme deviate from the sample mean for a sample drawn from a normal parent was first suggested by McKay (1935). Its remarkable accuracy was established only when the exact values were tabulated (for $n \leq 25$, Nair (1948b) and Grubbs (1950)). However, this accuracy might well be anticipated from (2.3) since $\Pr(A_1 A_2)$ may be expected to be small for large values of λ_α. The correlation between pairs of the z's is negative so that the joint probability of their exceeding a large value of λ_α is quite small.

2.2 The case of unknown variance; independent estimate

When σ is not known but is estimated by s_ν where $\nu\, s_\nu^2/\sigma^2$ is a chi-square variate with ν degrees of freedom and is independent of $x_1, x_2, \ldots, x_n$, we note that

$$\Pr\{\frac{x_1-\bar{x}}{s_\nu} < \lambda_\alpha, \ \ldots, \ \frac{x_n-\bar{x}}{s_\nu} < \lambda_\alpha\} = 1 - \alpha$$

is equivalent to

$$\Pr\{\max \frac{x_i-\bar{x}}{s_\nu} > \lambda_\alpha\} = \alpha \quad . \tag{2.6}$$

Forming the ratio of the standard normal variate, $\frac{x_i - \bar{x}}{\sqrt{(n-1)/n}\ \sigma}$, to the chi variate, $\sqrt{\nu}\ s_\nu/\sigma$, over the root of its degrees of freedom, $\sqrt{\nu}$, we have $\frac{x_i - \bar{x}}{\sqrt{(n-1)/n}\ s_\nu}$ which is clearly the usual Student's t variate with ν degrees of freedom.

It follows that we may write equation (2.6) as

$$\Pr\{\max \frac{x_i - \bar{x}}{\sqrt{(n-1)/n}\, s_\nu} > \frac{\lambda_\alpha}{\sqrt{(n-1)/n}}\} = \alpha \quad .$$

Now the above arguments for the case of σ known hold with the exception that λ_α is everywhere replaced by $\lambda_\alpha[n/(n-1)]^{\frac{1}{2}}$ and that the integration must be taken over the distribution of Student's t. The analogous result to (2.5) is

$$\Pr(A_1) = \int_{\lambda_\alpha[n/(n-1)]^{\frac{1}{2}}}^{\infty} p(t)dt$$

and

$$\Pr(A_1 A_2) = \iint_{\lambda_\alpha[n/(n-1)]^{\frac{1}{2}}}^{\infty} p(t_1,t_2)dt_1 dt_2 \quad ,$$

where p(t) is the usual Student's t density with ν degrees of freedom and

$$p(t_1,t_2) = \frac{1}{2\pi\sqrt{1-\rho^2}}[1 + \frac{t_1^2-2\rho t_1 t_2+t_2^2}{\nu(1-\rho^2)}]^{-\frac{1}{2}(\nu+2)}$$

with $\rho = -1/(n-1)$ (This bivariate t density is derived by Dunnett and Sobel (1954).).

2.3 The case of unknown variance, pooled estimate

A refinement has been proposed for the problem of detecting outlying observations when, as in section 2.2, in addition to the normal sample $x_1, \ldots, x_n$ at hand an independent mean-square estimate s_ν^2 of the common variance is available. The refinement makes use of the variance estimate s^2 from the sample and for this reason has certain desirable optimal properties. The test statistic employs the same numerator as before, i.e., max $x_i - \bar{x}$ but with s_ν in the denominator replaced by the pooled estimate. Let

$$S^2 = (n-1)s^2 + \nu\, s_\nu^2$$

and

$$b_i = (x_i - \bar{x})/S \quad .$$

The statistic under consideration will now be

$$b = \max b_i = \frac{x_{max} - \bar{x}}{S} \quad . \tag{2.7}$$

It should be noted that S^2, the pooled estimate of common variance, has not been divided by its degrees of freedom.

In order to obtain significance points of b the distribution of b_i and the joint distribution of b_i and b_j are needed. An extremely complicated derivation of essentially the same distributions has been developed by Doornbos, Kesten and Prins (1956). However, a rather neat development of these distributions is given by Quesenberry and David (1961) who make use of the properties of sufficiency and completeness.

The density function of b_i and the joint density function of b_i and b_j are respectively:

$$f(b_i) = \left[\frac{n}{n-1}\right]^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(n+\nu-1))}{\sqrt{\pi}\ \Gamma(\frac{1}{2}(n+\nu-2))}\left[1 - \frac{nb_i^2}{n-1}\right]^{\frac{1}{2}(n+\nu-4)}$$

$$-\left[\frac{n-1}{n}\right]^{\frac{1}{2}} \leq b_i \leq \left[\frac{n-1}{n}\right]^{\frac{1}{2}} \tag{2.8}$$

and

$$f(b_i b_j) = \left[\frac{n}{n-2}\right]^{\frac{1}{2}}\left(\frac{n+\nu-3}{2\pi}\right)\left[1 - \frac{n-1}{n-2}b_i^2 - \frac{2b_i b_j}{n-2} - \frac{n-1}{n-2}b_j^2\right]^{\frac{1}{2}(n+\nu-5)} \tag{2.9}$$

over the ellipse

$$\frac{n-1}{n-2}b_i^2 + \frac{2b_i b_j}{n-2} + \frac{n-1}{n-2}b_j^2 \leq 1$$

and zero elsewhere.

With the distribution of b_i and the joint distribution of b_i and b_j as given by (2.8) and (2.9) respectively, significance points of b as defined by (2.7) may be determined. For a given value of n and ν let λ_α be the required α significance point of b. By the principle of inclusion and exclusion coupled with Bonferroni's inequality, we have for any λ_α

$$nPr(b_i > \lambda_\alpha) - \binom{n}{2} Pr(b_i > \lambda_\alpha, b_j > \lambda_\alpha) \leq Pr(b > \lambda_\alpha) \leq nPr(b_i > \lambda_\alpha) \quad . \quad (2.10)$$

For sufficiently large λ_α, the RHS serves as an upper-bound first approximation to $Pr(b > \lambda_\alpha)$ and the LHS as a lower-bound second approximation. Values of λ_α are found by setting the respective side of (2.10) equal to α and solving.

III. THE SLIPPAGE PROBLEM

A problem closely related to the outlier problem just outlined is often treated under the heading of slippage tests for means of normal distributions. These tests consider the problem of how to compare k categories, such as k types of fertilizers, k machines, etc., so as to decide on the basis of a random sample of observations from each category whether or not the categories are equal, and if not which one has "slipped" or is the "best".

Let $x_1, x_2, \ldots, x_k$ be k random variables with x_i distributed normally with mean μ_i and variance σ^2. The sample

$$\begin{array}{cccc} x_{11}, & x_{12}, & \ldots, & x_{1n_1} \\ x_{21}, & x_{22}, & \ldots, & x_{2n_2} \\ \vdots & \vdots & & \vdots \\ x_{k1}, & x_{k2}, & \ldots, & x_{kn_k} \end{array} \tag{3.1}$$

where x_{ij} is the j^{th} observation on x_i, is drawn for the purpose of testing the hypothesis

$$H_o: \quad \mu_1 = \mu_2 = \dots = \mu_k = \mu$$

against the alternatives (3.2)

$$H_{ai}: \quad \mu_1 = \dots = \mu_{i-1} = \mu_{i+1} = \dots = \mu_k = \mu$$

$$\mu_i = \mu + \Delta \quad , \qquad \Delta > 0 \quad ,$$

for i, μ and Δ unspecified. Note that for testing purposes the difference in this and the outlier testing situation is that here there are n_i observations on the x_i instead of just one.

The procedures presented for outlier tests can with slight changes be used for slippage tests when the k category sizes are equal. When the n_i are nearly equal the approach is still of value for approximate tests.

3.1 The relation to the outlier problem

The extension of the outlier test to that of the slippage test will be illustrated for the b statistic of Quesenberry and David (1961) which was established in the previous section. From the sample given by (3.1) we wish to test the hypothesis that the array is from a common normal parent against the alternative that the i^{th} subsample $(x_{i1}, x_{i2}, \dots, x_{in_\ell})$ is from a normal parent with a larger mean, where i is unspecified.

Let $$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

and s_t^2 be an independent mean-square estimate of error with t degrees of freedom.

3.1a <u>Equal sample sizes</u>

For the important special case that all k categories have equal subsample sizes, i.e., $n_1 = n_2 = \ldots = n_k = m$, the statistic

$$\max_i \frac{\sqrt{m}\,(\bar{x}_i - \bar{x})}{S_1}$$

where

$$S_1^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} (x_{ij} - \bar{x})^2 + ts_t^2$$

is distributed as b. The significance points of this statistic can be obtained from tables of b with the parameters n and ν replaced by $n = k$ and $\nu = k(m-1) + t$. This slippage test possesses all the desirable properties of the corresponding outlier test.

3.1b Unequal sample sizes

If the n_i are not all equal but are approximately so, the tables of b can be used as approximate tests. Let

$$\bar{x}_w = \frac{1}{k} \sum_{i=1}^{k} \sqrt{n_i}\, \bar{x}_i \quad , \qquad N = \sum_{i=1}^{k} n_i \quad ,$$

then the statistic

$$\max_i \frac{\sqrt{n_i}\, \bar{x}_i - \bar{x}_w}{s_1}$$

serves as an approximate test based on tables of b with the parameters of the table being $n = k$ and $\nu = N + t - k$.

3.2 A multiple decision procedure

Paulson (1952a) develops the slippage problem as a multiple decision procedure which has certain optimum properties. To test the hypothesis given by (3.2) he introduces the set of k+1 decisions $(D_0, D_1, \ldots, D_k)$, corresponding to the acceptance of $H_0, H_1, \ldots, H_k$. The problem then is to select one member of the set on the basis of a sample of n observations from each of the k categories. The procedure is optimal subject to the following restrictions:

1 - When H_0 is true, D_0 should be selected with probability $1 - \alpha$.

2 - The decision procedure must be invariant if a constant is added to all the observations or if all the observations are multiplied by a positive constant.

3 - The decision procedure must be symmetric in the sense that the probability of making a correct decision when the i^{th} alternative is correct is the same for all $i = 1, \ldots, k$.

Under these conditions for optimality, the procedure maximizing the probability of a correct decision when one of the k distributions has slipped to the right is:

$$\text{if} \quad n(\bar{x}_M - \bar{x})/S > \lambda_\alpha \quad \text{select} \quad D_M$$

$$\text{if} \quad n(\bar{x}_M - \bar{x})/S \leq \lambda_\alpha \quad \text{select} \quad D_0$$

where

$$\bar{x}_M = \bar{x}_{max} \qquad (M = 1, 2, \ldots, k) \quad ,$$

$\bar{x}$ is the grand mean ,

S^2 is the <u>total</u> sum of squares

$$\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x})^2 \quad , \quad \text{and}$$

λ_α is a constant determined by condition 1, that is, λ_α is the upper α significance point of

$$n(\bar{x}_M - \bar{x})/S \quad .$$

For the proof that this decision procedure is an optimum one, the reader is referred to the original paper by Paulson (1952a). This test statistic is, however, essentially the same as that developed by Quesenberry and David (1961) as previously discussed in section 2.3.

3.3 Comparing k-1 categories to a control

A further paper by Paulson (1952b) considers the slippage problem of determining the best of k categories when comparing k-1 experimental categories with a standard or control. The experimental categories can be classified into two groups: one group consisting of those categories which are superior to the control and a second group of those experimental categories which are inferior to or at most equal to the control. In such a situation it is usually desirable to have special protection against the selection of an experimental category as best when it actually is inferior to the control. This will be accomplished by requiring that the statistical procedure used provides a special assurance that the control will be selected as best if the second group consists of all the k-1 experimental categories. When this is the case, the control will be selected with probability $\geq 1-\alpha$.

Consider each of the k categories containing n observations which are assumed to be normally and independently distributed with mean μ_i $(i = 1,\ldots,k)$ and common <u>known</u> variance σ^2. For convenience, let i=1 be the control category. Now the best category is defined to be the one associated with the largest value of μ_i.

The statistical procedure proposed for the selection of the best category is:

> If $\max(\bar{x}_i-\bar{x}_1) \geq \lambda_\alpha\sigma\sqrt{2/n}$ select the category whose estimated mean is $\bar{x}_i$; otherwise select the control.

The selection procedure will be complete upon determining λ_α. It is possible to tabulate the values of λ_α as a function of α and k, however, an approximate method with a definite limit of error is available.

Let A_i represent the event $\bar{x}_i - \bar{x}_1 \geq \lambda_\alpha\sigma\sqrt{2/n}$. We have by the principle of inclusion and exclusion

$$\Pr\{\max(\bar{x}_i-\bar{x}_1)\geq\lambda_\alpha\sigma\sqrt{2/n}\} = \sum_{i=2}^{k}\Pr(A_i)-\sum_{i<j}\sum\Pr(A_iA_j)+\ldots \quad ,$$

where the probabilities involved are to be calculated for the case when all k means are equal. Making use of Bonferroni's inequality and due to symmetry when the means are equal, this becomes

$$(k-1)\Pr(A_2)>\Pr\{\max(\bar{x}_i-\bar{x}_1)\geq\lambda_\alpha\sigma\sqrt{2/n}\}>(k-1)\Pr(A_2)-\binom{k-1}{2}\Pr(A_2A_3) \ . \tag{3.3}$$

Since $(\bar{x}_2-\bar{x}_1)$ and $(\bar{x}_3-\bar{x}_1)$ have a bivariate normal distribution with correlation equal ½, we obtain

$$\Pr(A_2) = \frac{1}{\sqrt{2\pi}}\int_{\lambda_\alpha}^{\infty} e^{-\frac{1}{2}t^2}\,dt \quad ,$$

and

$$\Pr(A_2A_3) = \frac{1}{\pi\sqrt{3}}\iint_{\lambda_\alpha}^{\infty} e^{-\frac{2}{3}(x^2-xy+y^2)}\,dxdy \quad .$$

Upper and lower limits for λ_α can be obtained by setting the LHS and RHS of (3.3) respectively to α and solving. When the upper limit for λ_α is used, it follows that $\Pr\{\max(\bar{x}_i-\bar{x}_1) \geq \lambda_\alpha\sigma\sqrt{2/n}\}$ will be less than α by an amount which cannot exceed $\frac{1}{2}(k-1)(k-2)\Pr(A_2A_3)$. This later term can be evaluated from tables of the volumes under the normal bivariate surface. This type of approximation appears to yield good results for values of α which ordinarily are of interest if k is not too large, say ≤ 6.

This procedure is completely analogous to that of the outlier test described in section 2.1. For the case where the variance is not known, approximations can be readily developed in a similar manner to the previous section.

3.4 Categorical frequencies

The problem of testing whether the largest observed categorical frequency in a multinomial distribution is significantly large is of great importance. This problem falls into the class of slippage tests since the null hypothesis of equal categorical probabilities is the most likely beginning hypothesis in this situation. One desires to classify n objects (or situations) into k categories and then to determine whether the number of objects occurring most frequently in a category is significant. This extreme categorical frequency is usually of most interest since it is often valuable for further study. Kozelka (1956) has provided a simple asymptotic approximation to the upper percentage points of this distribution.

Let F_1, F_2, ..., F_k be the observed proportions of a sample of n objects falling into k multinomial categories with assumed equal probabilities. Using the well-known multivariate normal approximation to the multinomial one considers in this problem that the observed proportions, F_i, are asymptotically joint multivariate normal variates with means $1/k$, variances $(k-1)/k^2n$ and covariances $-1/k^2n$. Let

$$t_i = \frac{F_i - 1/k}{\sqrt{(k-1)/k^2n}} \qquad i = 1, 2, \ldots, k$$

be the corresponding standardized variable, and let A_i represent the event $t_i > \lambda_\alpha$. We may now write

$$\Pr(t_{max} > \lambda_\alpha) = k\Pr(A_1) - \binom{k}{2}\Pr(A_1A_2) + \ldots \qquad (3.4)$$

which will be familiar from earlier work. Using the first term approximation, it follows from the asymptotic normality that

$$\Pr(t_{max} > \lambda_\alpha) \leq \frac{k}{\sqrt{2\pi}} \int_{\lambda_\alpha}^{\infty} e^{-\frac{1}{2}t^2}\, dt$$

for large λ_α.

Since observed categorical frequencies must necessarily be integers, this first term approximation appears satisfactory even for small values of n. However, this approximation decreases in accuracy for increasing k.

IV. THE LARGEST OF A SET OF ESTIMATED NORMAL VARIANCES

In some cases a test for the equality of the variances of a set of normal distributions, which is especially powerful against the alternative that one of the variances is larger than the others, might be useful. A case in point occurs when one has to select out of a set of machines or out of several processes the one with the highest variability of product. Such a test would then be useful for determining which machine or process needs repair or refinement to insure a more uniform product.

4.a Equal sample sizes

For a set of n independent estimates of the same variance, each based on two degrees of freedom, Fisher (1929) obtained the distribution of the largest estimate as a fraction of the total. Cochran (1941) extended this result to a set of estimates each based on ν degrees of freedom. A completely analogous test concerning the smallest variance estimate has been developed by Doornbos (1956). Proof that the latter two tests have the optimum properties (see section 3.2) as outlined by Paulson (1952a) has been established by Truax (1953).

Suppose we have a sample consisting of k groups

$$\begin{matrix} x_{11}, & \ldots, & x_{1n} \\ \vdots & & \vdots \\ x_{k1}, & \ldots, & x_{kn} \end{matrix}$$

completely independent of one another, normally distributed with assumed common variance σ^2 and with means which may differ from group to group but not within the groups. The variance estimates

$$s_i^2 = \frac{1}{n-1} \sum_{j=1}^{n} (x_{ij} - \bar{x}_{i.})^2 \quad , \qquad (i=1,\ldots,k)$$

where

$$\bar{x}_{i.} = \frac{1}{n} \sum_{j=1}^{n} x_{ij} \quad ,$$

are then distributed independently as $\chi^2\sigma^2/\nu$, χ^2 having $\nu = n-1$ degrees of freedom. Now define

$$a_i = s_i^2 / \sum_{i=1}^{k} s_i^2$$

and

$$a = \max a_i = s_{max}^2 / \sum_{i=1}^{k} s_i^2 \quad .$$

The null hypothesis, H_0, to be tested is that the variances, σ_i^2 $(i=1,\ldots,k)$, of the k populations under consideration are equal, to say σ^2. The alternative hypothesis, H_a, is that the variance of the i^{th} (i unknown) population has slipped to the right, that is

$$H_a: \quad \sigma_1^2 = \ldots = \sigma_{i-1}^2 = \sigma_{i+1}^2 = \ldots = \sigma_k^2 = \sigma^2$$

and

$$\sigma_i^2 = \gamma^2\sigma^2 \quad , \qquad \text{where} \quad \gamma > 1 \quad .$$

Let D_0 be the decision that all the k variances are equal, and let D_i be the decision that D_0 is false and that σ_i^2 has slipped to the right. The statistical procedure for selecting one of the k+1 decisions D_0, D_1, ..., D_k is:

$$\text{if} \quad a \leq \lambda_\alpha \quad \text{select} \quad D_0 \quad ,$$

$$\text{if} \quad a > \lambda_\alpha \quad \text{select} \quad D_i \quad , \qquad \text{where } s_i^2 = s_{max}^2$$

and a has been previously defined. The selection procedure will be complete upon determining λ_α.

Let A_i represent the event $a_i > \lambda_\alpha$. Now the probability that the largest of the k ratios a_i exceeds λ_α may be written

$$Pr(a>\lambda_\alpha) = \binom{k}{1}Pr(A_i)-\binom{k}{2}Pr(A_iA_j)+ \ldots \quad .$$

For upper and lower bound approximations to λ_α, set the RHS or LHS respectively of

$$k\,Pr(A_i) \geq Pr(a > \lambda_\alpha) \geq k\,Pr(A_i) - \binom{k}{2} Pr(A_i A_j)$$

equal to α and solve.

Cochran (1941) has developed the probabilities of the events A_i, $A_i A_j$, and $A_i A_j A_k$ so that it is possible to obtain a highly accurate, three-term approximation to λ_α. However, he points out that the first term upper approximation is quite accurate for all situations normally of interest. He gives

$$Pr(A_i) = \frac{1}{B(\frac{\nu}{2}, \frac{\nu(k-1)}{2})} \int_{\lambda_\alpha}^{1} x^{\frac{\nu}{2}-1} (1-x)^{\frac{\nu(k-1)}{2}-1} dx \quad ,$$

$0 \leq x \leq 1$, which may be evaluated from tables of the incomplete beta-function.

4.b Unequal sample sizes

Doornbos and Prins (1956) have extended the work of the previous section to handle significance tests for the largest of a set of estimated variances which is valid for different sample sizes. When all the variance estimates are based on the same degrees of freedom this test reduces to Cochran's test.

The hypothesis to be tested is as before, i.e., the null hypothesis, H_0, being that the variances of the k populations under consideration are equal with the alternative hypothesis, H_a, being that the variance of the i^{th} population has slipped to the right.

Employing the notation of the previous section with the exception that the variance estimate s_i^2 $(i=1,\ldots,k)$ is now based on ν_i degrees of freedom, one computes the k ratios

$$x_i = \frac{\nu_i s_i^2}{\sum_{i=1}^{k} \nu_i s_i^2} \quad .$$

The following incomplete Beta-integrals are determined:

$$d_i = \frac{1}{B(\frac{\nu_i}{2}, \frac{N-\nu_i}{2})} \int_0^{x_i} x^{\frac{\nu_i}{2}-1} (1-x)^{\frac{N-\nu_i}{2}-1} dx$$

$$= I_{x_i}(\frac{\nu_i}{2}, \frac{N-\nu_i}{2}) \quad , \qquad i = 1, \ldots, k$$

where $N = \sum_{i=1}^{k} \nu_i$. Now the test statistic d is defined as

$$d = \max d_i \quad .$$

The test criterion will be to reject H_0 when d takes a value $d > \alpha/k$.

Upper and lower bounds for the level of significance of this test are determined in the following manner. Let A_i

represent the event $d_i > \lambda_\alpha$ where λ_α is a positive real number. The probability that the largest of the k d_i's exceeds λ_α may be written

$$Pr(d > \lambda_\alpha) = k\,Pr(A_i) - \binom{k}{2} Pr(A_i A_j) + \ldots$$

and the upper and lower bounds to λ_α may be determined in the usual fashion.

4.1 <u>The maximum F-ratio</u>

Let $s_1^2, \ldots, s_k^2$ be independent mean square estimates of an unknown variance σ^2 of a normal population based respectively on ν_t $(t=1,\ldots,k)$ degrees of freedom. Let s^2 be another independent estimate of the same variance based on ν degrees of freedom against which each of the first k estimates is to be tested for significant differences from s^2. In routine analysis of variance where such tests frequently occur, it is customary to test each of the k variance estimates against s^2 and declare separately whether or not there is a significant difference. Forming now for each of the independent s_t^2 the ratio $F_t = s_t^2/s^2$, it is obvious that the largest ratio, $F_{max} = \max s_t^2/s^2$, is more likely to be declared significant than any other variance ratio in the set. It is well known that if there are no real treatment effects and all treatment mean squares are tested by the ordinary F-test at say the 100α percent level of significance, the risk of returning at least one mean square wrongly as significant may be much greater than α. Indeed, for $k = 20$ the largest ratio, F_{max}, would be expected to be significant at the 5% level with the usual F-test. This form of error can be avoided if the percentage points of the largest F-ratio are used. Such percentage points depend on k, the number of independent mean squares s_t^2, on ν_t, their degrees of freedom, and on ν, the degrees of freedom of the error mean square s^2.

The exact distributions for obtaining the percentage points of F_{max} for the cases $\nu_t = 1$, $\nu_t = 2$ have been investigated in detail by Nair (1948a) and Finney (1941) respectively. The former case, $\nu_t = 1$, has useful application in the analysis of variance for the important class of factorial experiments of the 2^{ℓ} type.

Hartley (1955) has proposed a simple conservative approximation for the upper percentage points of F_{max} when all the ν_t are equal. We may write

$$Pr(F_{max} > \lambda_\alpha) = k Pr(A_i) - \binom{k}{2} Pr(A_i A_j) + \ldots \tag{4.1}$$

where

$$Pr(A_iA_j) = Pr(\frac{s_i^2}{s^2} > \lambda_\alpha, \frac{s_j^2}{s^2} > \lambda_\alpha) \quad .$$

Using the first term on the RHS of equation (4.1) for an approximate test procedure, we have

$$Pr(F_{max} > \lambda_\alpha) \leq k\,Pr(\frac{s_i^2}{s^2} > \lambda_\alpha) = \alpha$$

where α is the desired level of significance. Clearly, the critical value λ_α is the 100 α/k percentage point of the F distribution with ν_t and ν degrees of freedom. This approximation is also suggested when the ν_t are not all equal, but nearly equal.

David (1956) has found exact results when the ν_t are equal and even. Let

$$Pr\{A_1A_2\ldots A_j\} = Pr\{\frac{s_1^2}{s^2} > \lambda_\alpha, \frac{s_2^2}{s^2} > \lambda_\alpha, \ldots, \frac{s_j^2}{s^2} > \lambda_\alpha\}$$

where λ_α is some positive number to be determined. This joint probability may be written as

$$Pr\{A_1A_2\ldots A_j\} = \int_0^\infty \prod_1^j e^{-\lambda_t x} [1+\lambda_t x+\ldots+\frac{(\lambda_t x)^{\frac{1}{2}\nu_t-1}}{(\frac{1}{2}\nu_t-1)!}]p(x)dx \tag{4.2}$$

where

$$\lambda_t = \nu_t\lambda_\alpha/2\nu \quad , \quad (t=1,\ldots,j) \quad ,$$

and

$$p(x) = 2^{-\nu/2}\,\Gamma^{-1}(\nu/2)x^{\frac{1}{2}\nu-1}\,e^{-\frac{1}{2}x} \qquad (x \geq 0)$$

if the ν_t are even. Evaluation of this expression is accomplished when the integrand has been multiplied out.

Using equation (1.2) with the above result, we have the probability that the m^{th} largest variance ratio exceeds λ_α is

$$\Pr\{\frac{s^2_{(m)}}{s^2}>\lambda_\alpha\} = \sum_{\ell=0}^{k-m}(-1)^{\ell}\binom{k}{m+\ell}\binom{m+\ell-1}{m-1}\Pr\{A_1A_2\ldots A_{(m+\ell)}\} \quad . \tag{4.3}$$

The probability that the largest variance ratio, F_{max}, exceeds λ_α will be (for m=1)

$$\Pr\{F_{max}>\lambda_\alpha\} = \sum_{\ell=1}^{k}(-1)^{\ell-1}\binom{k}{\ell}\Pr\{A_1A_2\ldots A_\ell\}$$

$$= \binom{k}{1}\Pr(A_1)-\binom{k}{2}\Pr(A_1A_2)+\ldots \quad . \tag{4.4}$$

In the usual manner one may set the RHS of (4.4) equal to the desired level of significance of the test, α, and solve for the upper percentage point λ_α. The first term on the RHS of (4.4) will serve as a first term upper approximation with succeeding terms as further correction factors. Upper percentage points for the m^{th} largest variance ratio may also be determined by employing the same procedure to equation (4.3).

V. DISTRIBUTION FREE SLIPPAGE TESTS

5.1 Friedman's method of m rankings

Doornbos and Prins (1958) have applied the principle of inclusion and exclusion to the well known method of m rankings due to Friedman (1937), (see for example, Kendall (1955)), to develop a slippage test. The method of m rankings enables one to investigate whether m observers agree on their opinion about k objects. Each observer is required to rank k objects. The sum of the m ranks of each object will be denoted by $s_1, \ldots, s_k$. The null hypothesis, H_0, of this distribution free slippage test is that the rankings are chosen at random from the collection of all permutations of the numbers 1, ..., k and that they are independent. The alternative hypothesis, H_a, is that one object has a larger probability than the other ones of being ranked high while the other (k-1) objects are ranked in a random order.

The test criterion is to reject H_0 if

$$\max s_i \geq \lambda_\alpha \qquad (i=1,\ldots,k)$$

where λ_α is the smallest integer satisfying

$$\Pr\{s_i \geq \lambda_\alpha\} \leq \alpha/k \quad .$$

The distribution of s_i, under H_0, may be written as

$$\Pr\{s_i=n\} = \sum_{x=0}^{\infty} I_{n-xk-m}\binom{m}{x}\binom{n-xk-1}{m-1}(-1)^x\, k^{-m}$$

$$(i=1,\ldots,k; \quad m\leq n\leq km)$$

where I_y is defined by

$$I_y = 0 \quad \text{if} \quad y < 0$$

$$I_y = 1 \quad \text{if} \quad y \geq 0 \quad .$$

This test procedure is equivalent to setting the upper limit of Bonferroni's inequality to the desired level of significance and solving for λ_α. That is, setting the first term on the RHS of

$$\Pr(s_{max}\geq\lambda_\alpha) = k\Pr(A_i) - \binom{k}{2}\Pr(A_iA_j) + \ldots$$

equal to α and solving, where A_i is defined to be the event $s_i \geq \lambda_\alpha$.

5.2 The Kruskal-Wallis test

The following distribution free k-sample slippage test has been proposed by Doornbos and Prins (1958). They consider the independent variates $u_1, \ldots, u_k$ which have, under the null hypothesis H_0, the same continuous distribution function. A sample of n_i independent observations

u_{ij} $(j=1,\ldots,n_i)$ is taken from each of the k populations. The null hypothesis is to be tested against the alternative H_a that the i^{th} population is stochastically larger than the others, that is

$$H_a: \quad Pr\{u_i > u_j\} > \tfrac{1}{2} \qquad (j \neq i)$$

$$u_j \quad (j=1,\ldots,i-1,\ i+1,\ \ldots,\ k)$$

follow the same continuous distribution, for one unknown value of i.

The test procedure is to rank all of the observations u_{ij} $(i=1,\ldots,k,\ \ j=1,\ldots,n_i)$ and calculate the sum of the ranks (T_i, say) on each population. Since T_i is a linear function of the well known Wilcoxon's test statistic applied to the i^{th} sample and the other k-1 samples combined, its distribution under H_0 is known (see Mann and Whitney (1947)). For each set of values T_i $(i=1,\ldots,k)$ compute under H_0

$$p_i = Pr\{T_i \geq \lambda_\alpha\} \quad .$$

The test criterion will be to reject H_0 if

$$\min p_i \leq \alpha/k \quad .$$

The concluding remarks of section 5.1 apply to this test.

VI. A SECOND TERM APPROXIMATION

In the previous sections of this paper we have shown how the principle of inclusion and exclusion has been applied to various situations in order to obtain upper percentage points of statistics expressible as maxima. The general result for every case has been of the form

$$Pr(z_{max}>\lambda_\alpha) = \binom{n}{1}Pr(A_1)-\binom{n}{2}Pr(A_1A_2)+\ldots+(-1)^{n-1}Pr(A_1A_2\ldots A_n) \quad ,$$

where A_i represents the event $z_i > \lambda_\alpha$ $(i=1,2,\ldots,n)$. Then using Bonferroni's inequality we obtain

$$\binom{n}{1}Pr(A_1) \geq Pr(z_{max}>\lambda_\alpha) \geq \binom{n}{1}Pr(A_1)-\binom{n}{2}Pr(A_1A_2) \quad . \quad (6.1)$$

We have seen that this result allows us to establish upper and lower limits for λ_α, the upper percentage point of z_{max}. This is accomplished by setting the LHS and RHS respectively of (6.1) to the desired level of significance α and solving for λ_α.

It is clear from previous development that the second term on the RHS of (6.1) is quite often difficult to calculate. For this reason many authors suggest using the LHS of (6.1) as an upper limit approximation to λ_α except when great accuracy is desired. This position might easily be appreciated since one can argue that the joint probability of two variates exceeding λ_α (for large λ_α) is small, especially if z_i and z_j are negatively correlated. In fact, there is frequently a simple gauge available for the accuracy of this first term approximation.

This simplification requires that

$$\Pr(A_i A_j) \leq \Pr(A_i)\Pr(A_j) \quad . \tag{6.2}$$

The validity of this inequality will be discussed later in this chapter for the various statistics considered previously in this paper.

Combining this last relationship (6.2) with (6.1) we obtain

$$n \Pr(A_i) \geq \Pr(z_{max} > \lambda_\alpha) \geq n \Pr(A_i) - \binom{n}{2} \Pr(A_i)\Pr(A_j) \quad .$$

Now if λ_α is determined such that the probability of every event A_i under the null hypothesis is equal to α/n, we have

$$\alpha \geq \Pr(z_{max} > \lambda_\alpha) \geq \alpha - \frac{n-1}{2n} \alpha^2$$

or for $n \geq 2$

$$\alpha \geq \Pr(z_{max} > \lambda_\alpha) \geq \alpha - \frac{\alpha^2}{2} \quad .$$

As an example of the accuracy of these bounds, when

$\alpha = 0.05,\ 0.01$

$$\left.\begin{matrix}0.05\\0.01\end{matrix}\right\} > \Pr(z_{max} > \lambda_\alpha) > \left\{\begin{matrix}0.04875\\0.00995\end{matrix}\right. .$$

These bounds give the significance level of the test. That is, the probability of rejecting H_0 when H_0 is true when tested against the alternative H_a will lie between α and $\alpha - \alpha^2/2$.

6.1 A unifying statement for the second term approximation

The second term approximation proposed in the previous section,

$$\Pr(A_i A_j) \leq \Pr(A_i)\Pr(A_j) \quad , \tag{6.2}$$

is valid for the various statistics previously considered in this paper with the exception of two cases. These two cases will be discussed in the next section.

Throughout the literature one finds the validity of (6.2) established for various statistics. The proofs encountered are usually of considerable length, making detailed use of the distribution of the statistic under consideration. All these proofs are special cases of the following result which has wide validity although it is not universally true. We introduce this result as a unifying statement.

Statement:

If $z_1, z_2, \ldots, z_n$ are non-negative random variables whose sum is fixed, and if the joint distribution of the z_i's has compound symmetry (i.e., is symmetrical in the z_i's) then

$$\Pr(z_i > c_i,\ z_j > c_j) \leq \Pr(z_i > c_i)\Pr(z_j > c_j) \quad . \tag{6.3}$$

Argument:

$$\Pr(z_i > c_i \mid z_j > c_j) \leq \Pr(z_i > c_i)$$

or

$$\frac{\Pr(z_i > c_i,\ z_j > c_j)}{\Pr(z_j > c_j)} \leq \Pr(z_i > c_i) \quad .$$

Hence

$$\Pr(z_i > c_i,\ z_j > c_j) \leq \Pr(z_i > c_i)\Pr(z_j > c_j) \quad .$$

and for the case of interest in this paper, $c_i = c_j = \lambda$, (6.3) reduces to (6.2).

This argument can be applied directly to Cochran's test (section 4.a), the Friedman test (section 5.1), the Kruskal-Wallis test (section 5.2), and to tests of equality of scores in the method of paired comparisons.

Specific proofs for the validity of (6.2) have been given for:

The extreme deviate from the sample mean

(a) variance known (section 2.1)

(b) pooled variance estimate (section 2.3)

see Doornbos and Prins (1958)

Cochran's test

(a) equal degrees of freedom (section 4.a)

see Dornbos (1956)

(b) unequal degrees of freedom (section 4.b)

see Doornbos and Prins (1956)

Friedman test (section 5.1)

see Doornbos and Prins (1958)

Kruskal-Wallis test (section 5.2)

see Doornbos and Prins (1958)

Paired comparisons, equality of scores

see Huber (1963)

The validity of (6.2) has been well established for the above cases as an approximation technique for the tests presented in this paper. That the result is not always true is demonstrated by the following example due to H. Kesten (Doornbos, 1956) for Cochran's test.

Suppose the variables x_1, x_2, and x_3 can take the values 1 and 2 with probabilities p and 1-p respectively. Defining

$$y_i = \frac{x_i}{x_1+x_2+x_3} \qquad i = 1,2,3 \quad ,$$

and c being some real number $\frac{1}{4} \leq c < \frac{1}{3}$, we obtain

$$\Pr\{y_1 > c\} = 1 - p + p^3$$

and

$$\Pr\{y_1 > c,\ y_2 > c\} = 1 - 2p + p^2 + p^3 \quad .$$

Consequently

$$\Pr\{y_1 > c,\ y_2 > c\} > [\Pr\{y_1 > c\}]^2$$

if $\frac{1}{4} \leq c < \frac{1}{3}$ and $1 > p > \frac{1}{2}(\sqrt{5} - 1)$.

Note that in this example we are restricting the value of c then determining a probability level. This procedure is the exact reverse of the method used in this paper.

6.2 Two exceptional cases

It has been pointed out that the inequality (6.2) is valid for the various statistics previously discussed in this paper with the exception of two cases. These two cases are that of the extreme normal deviate from the sample mean with an independent estimate of variance as the denominator (section 2.2) and the maximum F-ratio (section 4.1). Both of these cases consider statistics comprised of a denominator independent of the numerator. This fact leads to the intuitive argument that the independent denominator may completely override the numerator, causing inequality (6.2) to be invalid for such statistics.

The second term of the Bonferroni inequality for statistics comprised of a denominator independent of the numerator may be considered by the following argument.

Let x and y be two independent random variables with the same distribution; also let s be a random variable independent of x and y with not necessarily the same distribution. Then

$$\Pr(\frac{x}{s} > c,\ \frac{y}{s} > c) > \Pr(\frac{x}{s} > c)\Pr(\frac{y}{s} > c) \quad , \qquad (6.4)$$

where c is some constant. This result is established by writing the inequality (6.4) as

$$\int\{\Pr(x > cs\,|\,s)\}^2 f(s)\,ds > \{\int \Pr(x > cs\,|\,s) f(s)\,ds\}^2 \qquad (6.5)$$

or, equivalently,

$$E\{P(cs)\}^2 > [E\{P(cs)\}]^2$$

where $P(cs) = Pr(x > cs|s)$. Clearly,

$$E\{P(cs)\}^2 - [E\{P(cs)\}]^2 = \text{var } P(cs) > 0 \quad .$$

A more restrictive form of this argument has been developed as a theorem by Halperin (Kimball, 1951).

6.2a <u>The maximum F-ratio</u>

Kimball (1951) has shown that $Pr(A_iA_j) > Pr(A_i)Pr(A_j)$ for the maximum F-ratio of section 4.1. Using the notation of section 4.1 with i = 1, j = 2 for definiteness, he proves

$$Pr(\frac{s_1^2}{s^2} \leq \lambda_\alpha, \frac{s_2^2}{s^2} \leq \lambda_\alpha) > Pr(\frac{s_1^2}{s^2} \leq \lambda_\alpha)Pr(\frac{s_2^2}{s^2} \leq \lambda_\alpha)$$

or, equivalently by theorem I of section 6.3

$$Pr(\frac{s_1^2}{s^2} > \lambda_\alpha, \frac{s_2^2}{s^2} > \lambda_\alpha) > Pr(\frac{s_1^2}{s^2} > \lambda_\alpha)Pr(\frac{s_2^2}{s^2} > \lambda_\alpha) \quad , \qquad (6.6)$$

where s_1^2 and s_2^2 are based on n_1 and n_2 degrees of freedom respectively.

The argument of the previous section may be used to prove inequality (6.6) for $n_1 = n_2 = n$. Since s_1^2, s_2^2, and s^2 are independent and distributed as constant multiples of χ^2 densities, inequality (6.6) may clearly be written

$$\Pr(\chi_1^2 > c\chi^2, \chi_2^2 > c\chi^2) > \Pr(\chi_1^2 > c\chi^2)\Pr(\chi_2^2 > c\chi^2) \quad ,$$

where $c = n\lambda_\alpha/\nu$. Expressed in integral form, this inequality becomes

$$\int_0^\infty f_1^2(x)f(x)dx > \{\int_0^\infty f_1(x)f(x)dx\}^2 \qquad (6.7)$$

where $f_1(x)$ is the integral from zero to cx of a χ^2 density with n degrees of freedom and f(x) is the χ^2 density function with ν degrees of freedom. Since inequality (6.7) is of the same form as inequality (6.5), the proof is established.

6.2b <u>The extreme sample deviate with an independent estimate of variance</u>

The extreme sample deviate with an independent estimate of variance (section 2.2) may be considered by the argument of section 6.2.

Recall that x and y are two correlated random variables distributed normally with means zero and variances σ^2; also σ^2 is estimated by s^2 where $\nu s^2/\sigma^2$ is a chi-square variate with ν degrees of freedom and is independent of x and y. To discuss the validity of

$$\Pr(\frac{x}{s} > c, \frac{y}{s} > c) > \Pr(\frac{x}{s} > c)\Pr(\frac{y}{s} > c) \qquad (6.8)$$

or, equivalently,

$$\int \Pr(x > cs,\ y > cs \mid s) f(s)\, ds > \{\int \Pr(x > cs \mid s) f(s)\, ds\}^2 \quad ,$$

consider

$$\delta = \int \Pr(x > cs,\ y > cs \mid s) f(s)\, ds - \{\int \Pr(x > cs \mid s) f(s)\, ds\}^2 \quad . \tag{6.9}$$

Now for fixed c, $\Pr(\frac{x}{s} > c,\ \frac{y}{s} > c)$ is a monotonic increasing function of ρ (or a monotonic decreasing function of n since $\rho = -1/(n-1)$). It follows that for both c and ν fixed

$$\int \Pr(x > cs,\ y > cs \mid s) f(s)\, ds$$

is a monotonic increasing function of ρ. The difference δ, equation (6.9), is consequently also a monotonic increasing function of ρ. Since

$$\delta > 0 \quad \text{for} \quad \rho > 0 \quad ,$$

$$\delta > 0 \quad \text{for} \quad \rho = 0 \quad ,$$

and

$$\delta = -\{\int \Pr(x > cs \mid s) f(s)\, ds\}^2 \quad \text{for} \quad \rho = -1 \quad ,$$

it follows that $\delta = 0$ for some $\rho = \rho_o$ $(-1 < \rho_o < 0)$ and that

$$\delta > 0 \quad \text{for} \quad \rho > \rho_o \quad .$$

Likewise, if c and ρ (<0) are fixed, δ is a monotone decreasing function of ν, and

$$\delta < 0 \quad \text{for} \quad \nu = \infty \quad .$$

If there exists a ν_o making $\delta = 0$, then again

$$\delta > 0 \quad \text{for} \quad \nu < \nu_o \quad .$$

This argument points out that the inequality (6.8) is valid when $\rho \geq 0$ and also when $\rho < 0$ if the above conditions are met.

6.3 <u>An equivalence theorem</u>

Much of the work encountered in the literature pertaining to this paper has been developed for statistics expressed as minima. The approaches used and the results will hold equally well for statistics expressed as maxima. The following theorem proves this equivalence for the second term approximation discussed in this chapter.

Theorem I:

If $\quad Pr(A_i A_j) \leq Pr(A_i)Pr(A_j)$

then

$$Pr(\bar{A}_i \bar{A}_j) \leq Pr(\bar{A}_i)Pr(\bar{A}_j)$$

where

A_i represents the event $z_i > \lambda_\alpha$,

$\bar{A}_i$ represents the event $z_i \leq \lambda_\alpha$,

and $\quad \lambda_\alpha \geq 0 \quad .$

Proof:

We have $Pr(A_i) = 1 - Pr(\bar{A}_i)$ so

$$Pr(A_i)[1-Pr(A_j)] = Pr(\bar{A}_i)[1-Pr(\bar{A}_i)] \quad . \tag{6.11}$$

Also

$$Pr(A_i) - Pr(A_iA_j) = Pr(\bar{A}_j) - Pr(\bar{A}_i\bar{A}_j) \tag{6.12}$$

wince both sides equal the probability that one of a pair of variates exceeds and the other does not exceed λ_α.

Now subtracting (6.12) from (6.11) and making use of the fact that $Pr(A_i) = Pr(A_j)$, we have

$$Pr(A_iA_j) - Pr(A_i)Pr(A_j) = Pr(\bar{A}_i\bar{A}_j) - Pr(\bar{A}_i)Pr(\bar{A}_j)$$

which proves the theorem.

VII. TWO ALTERNATIVE APPROACHES

The previous sections of this paper have dealt with statistics expressed as maxima by the principle of inclusion and exclusion. Upper and lower limits for the upper percentage points of such statistics were then established by using the Bonferroni inequality. Two alternate approaches will now be discussed and compared to this method.

7.1 Whittle's approximation

The following theorem developed by Whittle (1960) may be used for determining a lower bound to $Pr(z_{max} > \lambda_\alpha)$. This theorem is very general, applying to all the statistics considered in this paper.

Theorem II:

If $z_1, z_2, \ldots, z_n$ is a set of variates of which we require only that second moments exist, then

$$Pr(z_{max} > \lambda_\alpha) \geq \frac{\{\sum_{i=1}^{n} Pr(z_i > \lambda_\alpha)\}^2}{\sum_{i=1}^{n} \sum_{j=1}^{n} Pr(z_i > \lambda_\alpha, z_j > \lambda_\alpha)} \qquad (7.1)$$

where

$$Pr(z_i > \lambda_\alpha, z_j > \lambda_\alpha) = Pr(z_i > \lambda_\alpha) \quad \text{for all } i = j .$$

For the case of interest in this paper, when the joint distribution of the z's is a symmetrical function of the z's, this becomes

$$\Pr(z_{max} > \lambda_\alpha) \geq \frac{n\{\Pr(z_1 > \lambda_\alpha)\}^2}{\Pr(z_1 > \lambda_\alpha) + (n-1)\Pr(z_1 > \lambda_\alpha,\ z_2 > \lambda_\alpha)} \quad . \tag{7.2}$$

Proof:

Let $x_1, x_2, \ldots, x_n$ be a set of variates of which we require only that second moments exist. Then if $F(x)$ is the joint distribution function

$$E\{c - \sum_{i=1}^{n} x_i\}^2 = \int \cdots \int \{c - \sum x_i\}^2 \, d\,F(x) \geq 0 \quad ,$$

where c is any real constant and the integration is taken over the range of the x's.

Now suppose there is probability $1 - Q$ that all x_i are zero. Excluding the point $x = 0$ from the integration, we obtain

$$Q\,c^2 - 2\,c\,E(\Sigma\, x_i) + E(\Sigma\, x_i)^2 \geq 0 \quad .$$

Solving this quadratic in c, we obtain the general inequality

$$Q \geq \frac{[E(\Sigma\, x_i)]^2}{E(\Sigma\, x_i)^2} \tag{7.3}$$

since the discriminant must be imaginary due to the inequality.

Now suppose we have another set of variates y_i and define

$$x_i = 0 \quad \text{if} \quad y_i \leq \lambda_\alpha$$
$$= 1 \quad \text{if} \quad y_i > \lambda_\alpha$$

so that

$$E(x_i) = E(x_i^2) = \Pr(y_i > \lambda_\alpha) = P_i \quad \text{(say)} \qquad (7.4)$$

and

$$E(x_i x_j) = \Pr(y_i > \lambda_\alpha,\ y_j > \lambda_\alpha) = P_{ij} \quad ,\ i \neq j \quad \text{(say)} .$$

Then the inequality (7.3) becomes

$$Q = \Pr(\text{some } y_i > \lambda_\alpha) = \Pr(y_{max} > \lambda_\alpha)$$

$$\geq \frac{\left(\sum_{i=1}^{n} P_i\right)^2}{\sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij}}$$

where $P_{ii} = P_i$ by (7.4).

To compare the two lower bounds so far established, i.e.,

$$\Pr(z_{max} > \lambda_\alpha) \geq \begin{cases} \dfrac{(\sum_i P_i)^2}{\sum\sum_{ij} P_{ij}} & \text{by Whittle} \\ \\ \sum_i P_i - \sum_{i<j}\sum P_{ij} & \text{by Bonferroni} \end{cases}$$

assume

$$\frac{(\sum_i P_i)^2}{\sum\sum_{ij} P_{ij}} > \sum_i P_i - \sum_{i<j}\sum P_{ij} \quad . \tag{7.5}$$

This may be rewritten as

$$(\sum_i P_i)^2 > \{\sum_i P_i + 2 \sum_{i<j}\sum P_{ij}\}\{\sum_i P_i - \sum_{i<j}\sum P_{ij}\}$$

which reduces to

$$(\sum_i P_i)^2 > (\sum_i P_i)^2 + \sum_{i<j}\sum P_{ij}\{\sum_i P_i - 2 \sum_{i<j}\sum P_{ij}\} \quad .$$

This leads to the inequality

$$\sum_i P_i < 2 \sum_{i<j}\sum P_{ij}$$

which must hold for the inequality (7.5) to be true. Hence, the lower bound of Whittle is stronger than that of Bonferroni iff

$$\sum_i P_i < 2 \sum_{i<j}\sum P_{ij}$$

or

$$P_i < (n-1)P_{ij}$$

for the case of the joint distribution of the z's being a symmetrical function of the z's.

7.2 The assumption of independence

When all the variates $z_1, z_2, \ldots, z_n$ are assumed to be independent and identically distributed, we have immediately

$$\Pr(z_{max} > \lambda_\alpha) = 1 - \Pr(\text{all } z\text{'s} \leq \lambda_\alpha)$$

$$= 1 - (1-\alpha_o)^n \qquad (7.6)$$

where $\alpha_o = \Pr(z_i > \lambda_\alpha)$. Setting equation (7.6) equal to α, we can solve for α_o, hence for λ_α. We have

$$\alpha = 1 - (1-\alpha_o)^n$$

so

$$\alpha_o = 1 - (1-\alpha)^{1/n} \quad .$$

For large n and small α we have $(1-\alpha)^{1/n} \doteq 1 - \alpha/n$, hence $\alpha_o \doteq \alpha/n$. Using this approximation, α_o is overestimated causing the equality of equation (7.6) to become an inequality ($<$).

The resulting equation to be solved for the critical value λ_α, i.e.,

$$\Pr(z_i > \lambda_\alpha) = \alpha/n$$

is seen to be exactly the same result as reached by the usual method of setting the upper bound of Bonferroni's inequality equal to α.

REFERENCES

Cochran, W. G., "The distribution of the largest of a set of estimated variances as a fraction of their total," Ann. Eugen., London, Vol. 11 (1941), pp. 47-52.

David, H. A., "On the application to statistics of an elementary theorem in probability," Biometrika, Vol. 43 (1956), pp. 85-91.

David, H. A., "Multiple decisions and multiple comparisons," Virginia Polytechnic Institute, Technical Report No. 39, 1958.

Doornbos, R., "Significance of the smallest of a set of estimated normal variances," Statist. Neerlandica, Vol. 10 (1956), pp. 117-126.

Doornbos, R. and H. J. Prins, "Slippage tests for a set of Gamma-variates," Indag. Math., Vol. 18 (1956), pp. 329-337.

Doornbos, R. and H. J. Prins, "On slippage tests," Indag. Math., Vol. 20 (1958), I, pp. 38-46; II, pp. 47-55.

__________ Proc. Kon. Ned. Ak. van Wetensch., Vol. 61, Ser. A (1958), III, pp. 438-447.

Eisenhart, C., M. W. Hastay, and W. A. Wallis, editors, Selected Techniques of Statistical Analysis, McGraw-Hill Book Co., New York, 1947, pp. 384-394.

Feller, W., "An Introduction to Probability and Its Applications," John Wiley and Sons, New York, 1950, Chapt. 4.

Halperin, M., S. W. Greenhouse, J. Cornfield, and Julia Zolokar, "Tables of percentage points for the studentized maximum absolute deviate in normal samples," J. Amer. Stat. Assn., Vol. 50 (1955), pp. 185-195.

Hartley, H. O., "Some recent developments in analysis of variance," Comm. Pure and Appl. Math., Vol. 8 (1955), pp. 47-72.

Kimball, A. W., "On dependent tests of significance in the analysis of variance," Ann. of Math. Stat., Vol. 22 (1951), pp. 600-602.

Kozelka, R. M., "Approximate upper percentage points for extreme values in multinomial sampling," Ann. of Math. Stat., Vol. 27 (1956), pp. 507-512.

Paulson, E., "On the comparison of several experimental categories with a control," Ann. of Math. Stat., Vol. 23 (1952), pp. 239-246.

Paulson, E., "An optimum solution to the k-sample slippage problem for the normal distribution," Ann. Math. of Stat., Vol. 23 (1952), pp. 610-616.

Quesenberry, C. P. and H. A. David, "Some tests for outliers," Biometrika, Vol. 48, 3 and 4, (1961), pp. 379-390.

Sarhan, A. E. and G. B. Greenberg, editors, "Contributions to Order Statistics," John Wiley and Sons, New York, 1962, Chapt. 9.

Whittle, P., personal communication to H. A. David, dated April 4, 1960, Cambridge, England.

Whittle, P., Sur la distribution du maximum I'un polynome trigonometrique a coefficients aleatoires," Colloquium proceedings, "Le Calcul des Probabilites et ses Applications, Centre National de la Recherche Scientifique, Paris, 1959.

RELATED REFERENCES

Dunnett, C. W. and M. Sobel, "A bivariate generalization of Student's t-distribution with tables for certain special cases," Biometrika, Vol. 41 (1954), pp. 153-169.

Finney, D. J., "The joint distribution of variance ratios based on a common error mean square," Ann. Eugen., Lond., Vol. II (1941), pp. 136-140.

Fisher, R. A., "Tests of significance in harmonic analysis," Proc. Roy. Soc. A, Vol. 125 (1929), pp. 54-59.

Friedman, M., "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," J. Amer. Statist. Assn., Vol. 32 (1937), pp. 675-699.

Grubbs, F. E., "Sample criteria for testing outlying observations," Ann. of Math. Stat., Vol. 21 (1950), pp. 27-58.

Huber, P. J., "A remark on a paper of Trawinski and David entitled: "Selection of the best treatment in a paired-comparison experiment"," Ann. Math. Stat., Vol. 34 (1963), pp. 92-94.

Kendall, M. G., "Rank Correlation Methods," Sec. Ed., London, 1955, Chapt. 6-7.

Kruskal, W. H. and W. A. Wallis, "Use of ranks in one-criterion variance analysis," J. Amer. Stat. Assn., Vol. 47 (1952), pp. 583-621.

Mann, H. B. and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," Ann. Math. Stat., Vol. 18 (1947), pp. 50-60.

McKay, A. T., "The distribution between the extreme observation and the sample mean in samples of n from a normal universe," Biometrika, Vol. 27 (1935), pp. 466-472.

Nair, K. R., "The studentized form of the extreme mean square test in the analysis of variance," Biometrika, Vol. 35 (1948a), pp. 16-31.

Nair, K. R., "The distribution of the extreme deviate from the sample mean and its studentized form," Biometrika, Vol. 35 (1948b), pp. 118-144.

Truax, D. R., "An optimum slippage test for the variances of normal distributions," Ann. Math. Stat., Vol. 24 (1953), pp. 669-674.

ACKNOWLEDGEMENTS

The author wishes to express his most sincere appreciation to Dr. Herbert A. David for his invaluable guidance and advice during the course of this work.

Gratitude is extended to Dr. John G. Saw, who kindly consented to read this thesis in manuscript form. Appreciation is expressed to Dr. Boyd Harshbarger for his assistance and encouragement to the graduate student.

To , the author is indebted for the preparation of the final typewritten copies.

The author wishes, also, to express his gratitude for the general financial assistance from the National Institutes of Health.

The vita has been removed from
the scanned document

ABSTRACT

This thesis is concerned with an application of the principle of inclusion and exclusion and with related approximation techniques. These procedures are extensively employed for developing test criteria based on statistics expressible as maxima.

Upper percentage points of a number of such statistics have been tabulated by various special methods. However, the application of the principle of inclusion and exclusion, coupled with the Bonferroni inequalities, is often useful in providing good approximations. An extensive review of this method is presented in this report.

This procedure allows one to establish upper and lower limits to upper percentage points, say λ_α, of statistics expressible as maxima. The upper bound approximation to λ_α requires only the knowledge of the distribution(s) of the variates under consideration. The lower bound, however, requires also the joint distribution(s) of pairs of the variates. Since the joint distribution is often difficult to calculate, an approximation technique may be necessary. A detailed discussion of such an approximation with guidelines

for its applicability to statistics other than those discussed is presented.

Two alternative methods for the determination of upper percentage points for statistics expressed as maxima are discussed: Whittle's lower bound approximation and the assumption of independence. It is pointed out that Whittle's lower bound is stronger than that of Bonferroni only under certain conditions. The assumption of independence leads to approximately the same result as Bonferroni.