ABSTRACTS OF PAPERS

(Presented on December 27, 1939, at the Philadelphia meeting of the Institute)

On the Unbiased Character of Certain Likelihood-Ratio Tests when Applied to Normal Systems. JOSEPH F. DALY, The Catholic University of America.

Consider a random sample of N observations on a set of variates $x^1, \cdots, x^q$, where $x^1, \cdots, x^k$ are assumed to be normally distributed about means which are linear functions $m^i = \Sigma b^i_\sigma x^\sigma$ of the fixed variates $x^{k+1}, \cdots, x^q$. One is sometimes required to decide whether the sample tends to contradict the further hypothesis, H_0, that the coefficients b^i_μ belonging to a certain subset of the fixed variates, say $x^{k+1}, \cdots, x^{k+h}$, have the specific values $b^i_{\mu 0}$. Such a situation occurs, for example, in the generalized analysis of variance. In this paper it is shown that the Neyman-Pearson method of the ratio of likelihoods yields a test of H_0 which is (at least locally) unbiased; in other words, this test is less likely to reject H_0 when the sample is in fact drawn from a normal population in which $b^i_\mu = b^i_{\mu 0}$ than when it is drawn from a normal population in which the b^i_μ are different from but sufficiently close to $b^i_{\mu 0}$. In the special cases $k = 1$ or $h = 1$ the proof goes through even without the restriction that the true b^i_μ be close to $b^i_{\mu 0}$, a result which is also implicit in the papers by P. C. Tang and P. L. Hsu (*Stat. Res. Mem.* Vol. 2).

Similarly with respect to the hypothesis H_I that the deviations $x^i - \Sigma b^i_\sigma x^\sigma$ fall into certain mutually independent sets the λ-test is at least locally unbiased; and it has the additional property that the expected value of any positive integral power of $\sqrt[N]{\lambda}$ is greater when H_I is true than when the sample is drawn from any other normal population.

The Product Seminvariants of the Mean and a Central Moment in Samples. C. C. CRAIG, The University of Michigan.

The method used by the author in calculating the product seminvariants of a pair of central moments in samples is not adapted without modification to the present problem. In the present paper the necessary modification is developed which gives a routine method for the calculation of these sampling distribution characteristics. The calculation is a little heavier than in the previous case but the results for the mean and the second, third, and fourth central moments are given up to the fourth order except in one case in which the weight is 13. It is planned to follow this with a further study of the distribution of Fisher's t in samples from a normal population.

A Method for Minimizing the Sum of Absolute Values of Deviations. ROBERT SINGLETON, Princeton Local Government Survey.

E. C. Rhodes (*Philosophical Magazine*, May 1930) presented a method for the estimation of parameters in a linear regression where it is desired to minimize the sum of absolute values of the deviations. In this paper the structure of the deviation surface is analyzed and a method of steepest descent is developed which for computational purposes is an improvement over Rhodes' method. The process is finite and leads to an exact solution. The method and the formulae used are such as to permit the successive additions of new observations or sets of observations to the original data, or the exclusion of an observation from the original set, and the determination of the parameters for the sets of data so derived, with little additional labor.

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

On Certain Criteria for Testing the Homogeneity of k Estimates of Variance.

C. EISENHART AND FRIEDA S. SWED, University of Wisconsin.

Given k variance estimates $s_1^2, s_2^2, \cdots, s_k^2$ with $n_r s_r^2$, $(r = 1, 2, \cdots, k)$, independently distributed as $\chi^2\sigma_r^2$ for n_r degrees of freedom, tests of the hypothesis, H_0, that $\sigma_r^2 = \sigma^2$, $(r = 1, 2, \cdots, k)$, where σ^2 is unknown, have been based to date on one or the other of the quantities

$$Q_1 = \sum_{r=1}^{k} n_r(s_r^2 - s^2)^2/2s^4$$

$$Q_2 = w \log (ns^2/w) - \sum_{r=1}^{k} w_r \log \{n_r s_r^2/w_r\}$$

where the w_r are weights, $w = \sum_{r=}^{k} w_r$, $n = \sum_{r=1}^{k} n_r$, and $ns^2 = \sum_{r=1}^{k} n_r s_r^2$. A. E. Brandt and W. L. Stevens have advocated the use of Q_1, referring an observed value of Q_1 to the χ^2 distribution for $k - 1$ degrees of freedom. J. Neyman, E. S. Pearson, B. L. Welch, and M. S. Bartlett have advocated tests based on Q_2, Bartlett definitely proposing the use of degrees of freedom as weights, i.e. $w_r = n_r$, and recent work of E. J. G. Pitman and others has shown that unless $w_r = n_r$ tests based on Q_2 are biased. (A statistical test of an hypothesis H is said to be unbiased when the probability of rejecting H by its use is a minimum when H is true; obviously a desirable property.) When $w_r = n_r$ Bartlett has suggested that the distribution of Q_2 can be satisfactorily approximated by referring $Q_2/\left\{1 + \frac{1}{3(k-1)} \cdot \left(\sum_{r=1}^{k} \frac{1}{n_r} - \frac{1}{n}\right)\right\}$ to the χ^2 distribution for $k - 1$ degrees of freedom. In this paper we discuss the adequacy of the χ^2 distribution to describe the distribution of Q_1 and of the adjusted Q_2 when the degrees of freedom, n_r, are small.

U. S. Nair and D. J. Bishop have given theoretical evidence which suggests that when $n_r \geq 2$, $(r = 1, 2, \cdots, k)$, Bartlett's adjusted Q_2 may be expected to conform to the χ^2 distribution reasonably well in the neighborhood of the 5% and 1% levels. Using 1000 samples of 4 for which $n_r s_r^2/(n_{r+1})$ has been tabulated by W. A. Shewhart in Table D, Appendix II of his "Economic Control of Quality of Manufactured Product," 200 values of Q_1 and Q_2 (with adjustment) were calculated and compared with the χ^2 distribution for $k - 1$ degrees of freedom. Two cases were studied: Case I, $k = 5$ and $n_1 = n_2 = \cdots = 3$; Case II, $k = 3$ and $n_1 = n_2 = 3$ while $n_3 = 9$. As measured by the Chi-Square Goodness of Fit Test, using 11 degrees of freedom, the fits were good in all four instances. In Case I, for Bartlett's adjusted Q_2 the test led to $.80 < P < .90$, and to $.70 < P < .80$ for the Brandt-Stevens Q_1; in Case II, the fits were poorer with $.50 < P < .70$ for Bartlett's criterion and $.10 < P < .20$ for the Brandt-Stevens. However, an examination of the *descending* cumulative distributions showed that in all instances these criteria exhibited a deficiency of large values of χ^2, with the deficiency, in general, more marked in the case of the Brandt-Stevens test. Consequently, when one uses significance levels for these criteria obtained by means of the χ^2 approximation advocated, one is in reality using a level of significance slightly less than that professed. The discrepancy is not great, however, and is on the safe side, i.e. one will reject H_0 falsely in the long run less often than one professes to be doing. Without doubt, however, one will also detect the falsehood of H_0 when $\sigma_r^2 \neq \sigma_t^2$, for at least one pair of values of r and t, $r \neq t$, less often in the long run by the use of these approximate significance levels than if the true levels were used, but we have no definite evidence at present on this point. A somewhat disquieting feature is that the agreement between the χ^2 values yielded by the two criteria becomes worse as one proceeds toward larger values of χ^2 in terms of either quantity. Thus, of 8 samples which Q_2 would have rejected at the 5% level in Case I, only 4 of these would have been rejected by Q_1, and Q_2 would have passed 3 samples of the 7 rejected by Q_1. Thus it appears that, if one wishes to work with a given chance of rejecting H_0 falsely, one should choose one of these criteria and then stick to it in future applications. For large values of the n_r the two criteria tend to equivalence, so the choice between them is of interest mainly for small n_r, but cannot be made with full information until more is known about the bias, if any, of the Brandt-Stevens test, and the relative power of the two tests with regard to alternatives to H_0.

On a Test Whether Two Samples are from the Same Population. A. WALD AND J. WOLFOWITZ, Columbia University and Brooklyn, New York.

Let X and Y be two independent random variables about whose distributions nothing is known except that they are continuous. Let $x_1, x_2, \cdots, x_m$ be a set of m independent observations on X and let $y_1, y_2, \cdots, y_n$ be a set of n independent observations on Y. The null hypothesis to be tested is that the distributions of X and Y are identical.

Let the set of $m + n$ observations be arranged in order of magnitude, thus: $z_1, z_2, \cdots, z_{m+n}$. Replace z_i by v_i $(i = 1, 2, \cdots, m + n)$ where $v_i = 0$ if z_i is a member of the set of x's and $v_i = 1$, if z_i is a member of the set of y's. Since the null hypothesis states only that the distributions of X and Y are identical without specifying them in any other way, the distribution of the statistic U used for testing the null hypothesis must be independent of this common distribution of X and Y. It can easily be shown that the statistic U must be a function only of the sequence $v_1, v_2, \cdots, v_{m+n}$.

A subsequence $v_s, v_{s+1}, \cdots, v_{s+r}$ (where r may also be 0) is called a run if $v_s = v_{s+1} = \cdots = v_{s+r}$ and if $v_{s-1} \neq v_s$ when $s < 1$ and if $v_{s+r} \neq v_{s+r+1}$ when $s + r < m + n$. The statistic U defined as the number of runs in the sequence $v_1, v_2, \cdots, v_{m+n}$ seems a suitable statistic for testing the null hypothesis. A difference in the distribution functions of X and Y tends to decrease U. Hence the critical region is defined by the inequality $U < u_0$, where u_0 depends only on m, n, and the level of significance adopted. If $m \leq n$ and $P\{U = c\}$ is the probability that $U = c$, then:

$$P\{U = 2K\} = \frac{2({}^{m-1}C_{k-1} \cdot {}^{n-1}C_{k-1})}{{}^{m+n}C_m}, \qquad (K = 1, 2, \cdots, m),$$

$$P\{U = 2K - 1\} = \frac{({}^{m-1}C_{k-1}\,{}^{n-1}C_{k-2} + {}^{m-1}C_{k-2}\,{}^{n-1}C_{k-1})}{{}^{m+n}C_m}, \qquad (K = 2, 3, \cdots, m + 1).$$

The mean of U is:

$$\frac{2mn}{m + n} + 1.$$

The variance of U is:

$$\frac{2mn(2mn - m - n)}{(m + n)^2(m + n - 1)}.$$

If $\frac{m}{n} = \alpha$ (a positive constant) and $m \to \infty$, the distribution of U converges to the normal distribution.

The Distribution of Quadratic Forms In Non-Central Normal Random Variables. WILLIAM G. MADOW, Washington, D. C. (Presented to the Institute under a slightly different title)

Let the distribution of a sum of non-central squares of normally and independently distributed random variables which have the unit variances be called the χ'^2 distribution. It is proved that if a set of quadratic forms have a sum which is the sum of the squares of their variables, then a necessary and sufficient condition that the quadratic forms be independently distributed in χ'^2 distributions is that the rank of the sum of quadratic forms be equal to the sum of the ranks of the quadratic forms. Furthermore, the constants on which the χ'^2 distributions depend may be obtained by substituting the values about which the variables are taken for the variables themselves in the quadratic forms. Roughly speaking the theorem states that if a set of quadratic forms satisfy the conditions of the Fisher-Cochran theorem when the true means vanish, then the set of quadratic forms will be independently distributed in χ'^2 distributions when the true means do not vanish.

Some Theoretical Aspects of the Use of Transformations in the Statistical Analysis of Replicated Experiments. W. G. Cochran, Iowa State College.

The device of transforming the data to a different scale before performing an analysis of variance has recently been recommended by a number of writers for replicated experiments in which the original data show a markedly skew distribution. The use of transformations to obtain an approximate analysis has been supported mainly on the grounds that in the transformed scale the true experimental error variance is approximately the same on all plots. This paper considers the relation of the method of transformations to a more exact analysis. Discussion is confined to the $\sqrt{x}$ and $\sin^{-1}\sqrt{x}$ transformations, which appear to receive the most frequent use in practice.

To obtain an exact analysis, it is necessary to specify (*i*) how the expected value on any plot is obtained from unknown parameters representing the treatment and block (or row and column) effects (*ii*) how the observed values on the plots vary about the expected values. If the latter variation follows the Poisson law, (a case to which the square root transformation has been considered appropriate), the equations of estimation by maximum likelihood take the form

$$\sum_c \left(\frac{x - m}{m}\right)\frac{\partial m}{\partial c} = 0, \tag{1}$$

where x is the observed and m the expected value on any plot, c is a typical unknown parameter, and the summation extends over all plots whose expectations involve c. As the number of parameters is usually large (e.g. 16 in a 6 x 6 Latin square), these equations are laborious to solve; moreover, the question of obtaining small-sample tests of significance is difficult. It is shown that if a particular form can be assumed for the prediction formula in (*i*), namely that $\sqrt{m}$ is a linear function of the treatment and block (or row and column) constants, the equations of estimation may be reduced to the simpler form

$$\sum_c 4(r' - \sqrt{m}) = 0, \tag{2}$$

where $r' = \frac{1}{2}\left(\sqrt{m} + \frac{x}{\sqrt{m}}\right)$ is a function closely related to the square root of x. It follows that the statistical analysis in square roots, with some slight adjustments, coincides with the maximum likelihood solution, provided that the above form can be assumed for the prediction formula. The appropriateness of this form in practice is briefly considered and a "goodness of fit" test by χ^2 is developed. A numerical example is worked as an illustration and indicates that a good approximation is obtained by the transformation alone even with very small numbers per plot. The corresponding theory is also discussed for the inverse sine transformation, which applies where the original data are percentages or fractions whose experimental errors are derived from the binomial distribution.

In practice the type of analysis outlined above is unlikely to supplant the simple use of transformations, because it can seldom be assumed that the experimental variance is entirely of the Poisson or binomial type. The more exact analysis may, however, be useful (*i*) for cases in which the plot yields are very small integers or the ratios of very small integers (*ii*) in showing how to give proper weight to an occasional zero plot yield.

The Standard Errors of Geometric and Harmonic Types of Index Numbers. By NILAN NORRIS, Hunter College.

Various statisticians have made empirical studies of the sampling errors of certain types of index numbers used in the United States and England. None of these writers has taken advantage of the tools afforded by the modern theory of estimation, including fiducial inference, as a means of arriving at direct and general expressions for estimating the standard deviations of the sampling errors of geometric and harmonic types of index numbers.

A known expression for the first approximation to the variance of a function, as given by the relation between the variance of the function and the variance of the argument, is valid for that general class of distributions of which the variance and a higher moment are finite. With the aid of this relation, there appear simple and useful forms for estimating the standard errors of geometric and harmonic types of indexes. For sufficiently large samples, these forms are valid for all of the types of distributions of price relatives, production relatives, and similar observations ordinarily encountered, provided that there are satisfied the necessary conditions for drawing sound inferences on the basis of sampling without reference to the value of the variate.

Necessary conditions for using tests of significance soundly in connection with index number problems are those of realistic and intimate acquaintance with observations, and careful attention to certain broad theoretical considerations which determine whether or not the index is suited for the purpose for which it is used.

A Study of R. A. Fisher's z Distribution and the Related F Distribution. L. A. AROIAN, Hunter College.

The following results for the z distribution and related F distribution are investigated:

(1) Geometric properties.
(2) Exact values of the seminvariants and moments of z. Exact values of the first four central moments of F.
(3) The approach to normality of both distributions as n_1 and n_2 become large in any manner whatever.
(4) The Pearson types of approximating curves, the logarithmic normal approximation, the Gram-Charlier approximation, and the uses of these in finding any level of significance of z and of F.

A Note on the Analysis of Variance with Unequal Class Frequencies. ABRAHAM WALD, Columbia University.

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

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

Confidence limits for ρ have been derived only in case of equal class frequencies, i.e. $m_1 = m_2 = \cdots = m_p$. We give here the confidence limits for ρ in case of unequal class frequencies. Since ρ is a monotonic function of $\frac{\sigma'^2}{\sigma^2}$, it is sufficient to derive confidence limits for $\frac{\sigma'^2}{\sigma^2}$. Denote $\frac{\sigma'^2}{\sigma^2}$ by λ^2 and the arithmetic mean of the j-th group by $\bar{x}_j$. Let

$$w_j = \frac{m_j}{1 + m_j\lambda^2},$$

and denote by F_1 and F_2 the lower and upper confidence limits respectively of F, where F has the analysis of variance distribution with $p - 1$ and $N - p = m_1 + \cdots + m_p - p$ degrees of freedom. Then the lower confidence limit λ_1^2 of λ^2 is given by the root of the equation in λ^2:

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

and the upper confidence limit λ_2^2 of λ^2 is given by the root of

$$f(\lambda^2) = F_1. \tag{2}$$

For calculating the roots of (1) and (2), we can make use of the fact that $f(\lambda^2)$ is monotonically decreasing with increasing λ^2.

An Approach to Problems Involving Disproportionate Frequencies. BURTON D. SEELEY, Washington, D. C.

Applied mechanics offers an analysis of variance solution to problems of multiple classification involving disproportionate sub-class numbers. The quality of orthogonality may be attained in such problems by measuring the variability between classes of any one classification after centering the others. This approach, which is not limited by the number of classes or the number of classifications, treats the problem involving equal sub-class numbers as a special phase of the general analysis of variance.