ABSTRACTS OF PAPERS

(Abstracts of papers to be presented at the Eastern Regional Meeting, Cambridge, Massachusetts, May 6–7, 1963. Additional abstracts will appear in the June, 1963 issue.)

1. A Generalization on Distribution-Free Tolerance Limits. L. Danziger and S. A. Davis, International Business Machines Corporation, Poughkeepsie, New York.

Given an ordered random sample $X_1 \leqq X_2 \leqq \cdots \leqq X_n$ from a population with a continuous probability density function $f(x)$, we wish to make distribution-free inferences about a second, finite, random sample $Y_1, Y_2, \cdots Y_N$ from the same population. *The ONE tolerance-limit problem is*: For any integer r, such that $1 \leqq r \leqq n$, and for any integer N_0, such that $0 \leqq N_0 \leqq N$, find the probability that at least N_0 of the Y_i are greater than or equal to X_r. *The TWO tolerance-limit problem can be similarly stated*: For any pair of integers r_1 and r_2, such that $1 \leqq r_1 < r_2 \leqq n$, and for any integer N_0, such that $0 \leqq N_0 \leqq N$, find the probability that at least N_0 of the Y_i are greater than or equal to X_{r_1}, and less than or equal to X_{r_2}. This paper generalizes certain results given by S. S. Wilks (*Ann. Math. Statist.* **12** (1941) 91–96) by proving that the probability of at least N_0 of the Y_i being greater than X_r is equal to the probability of at least N_0 of the Y_i lying betwen X_{r_1} and X_{r_2}, where $r = n + r_1 - r_2 + 1$. Hence, the distribution-free TWO tolerance-limit problem can be reduced to the ONE tolerance-limit problem.

In conjunction with this proof, an extensive set of distribution-free tolerance-limit tables has been computed for many combinations of r, n, and N.

2. On the Power of Rank Tests on the Equality of Two Distribution Functions. Jean D. Gibbons, University of Cincinnati.

A comparative study is made of power functions of two-sample rank tests of the hypothesis of equal distributions, $H_0: \ H = G$, including the most powerful rank test, Terry Test, one and two-sided Wilcoxon tests, one and two-sided median tests, and runs. Two new tests are proposed, the Gamma test and Psi test, which are locally most powerful against certain nonparametric alternatives. Numerical results are given for the alternative

$$H_1: \ H = 1 - (1 - F)^k, \qquad G = F^k,$$

F unspecified, $k = 2, 3, 4$, $m = n \leqq 6$, $\alpha = .01, .05, .10$, using randomized decision rules. If F is symmetric, H amd G are mirror images. Some results for unequal sample sizes are also given. Comparisons are made with power against normal alternatives having the corresponding standardized differences. The Psi test is found to occupy a position intermediate between the Terry and Mann-Whitney-Wilcoxon tests, but the difference between power functions of these and the most powerful rank test is almost negligible. Asymptotic properties of the Psi test are investigated using the results of Chernoff and Savage (*Ann. Math. Statist.* **29** (1958) 972–994).

3. Dependence in Three Dimensions. H. O. Lancaster, University of Sydney, Sydney, Australia. (By title)

Let a Latin square be described by a triplet of indices $\{x, y, z\}$ each running over the integers $0, 1, 2 \cdots (n - 1)$. Let a three dimensional distribution of $W \equiv \{X, Y, Z\}$ be defined by setting $P\{X = x, Y = y, Z = z\}$ equal to n^{-2} if x, y, z is a triplet occurring in the description of the Latin square and equal to zero otherwise. The variables X, Y and Z are independent in pairs but not in threes. The example of S. N. Bernstein, cited by A. N. Kolmogorov on page 11 of his *Foundations of the Theory of Probability*, Dover (1950), is obtained by setting $n = 2$. Now define an independent collection of the triples, W_i, each having the same distribution as W. Define now sums T, U, and V of the variables X_i, Y_i and Z_i, where $T = \sum_1^\infty n^{-i}X_i$, $U = \sum_1^\infty n^{-i}Y_i$ and $V = \sum_1^\infty n^{-i}Z_i$. Then T, U and V are all rectangular variables and are mutually independent in pairs but not as a triplet. In fact, the joint distribution is not ϕ^2-bounded with respect to the product of the marginal distributions; if the summation is only taken to k terms, $\phi^2 = n^k - 1$, where $\phi^2 + 1$ is

$$\sum_{i,j,k} p_{ijk}^2/(p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}).$$

For a given T and U, V can only have one saltus. If T and U are written in the n-ary scale, $T = 0.x_1x_2x_3 \cdots$, $U = 0.y_1y_2y_3 \cdots$ and if W is a point of increase, (x_1, y_1, z_1), $(x_2, y_2, z_2) \cdots$ must all be triplets occurring in the description of the Latin square. $V \equiv 0.z_1z_2 \cdots$ is thus uniquely determined. Further variations can be made by allowing the Latin square to vary with the index of W_i; moreover, mixtures of the distribution of W_i with the corresponding product distribution of the U_i, V_i, W_i can be formed and linear forms considered as before.

4. **Multivariate Linear Hypothesis With Unknown Covariance Matrix.** ANDRE G. LAURENT, Wayne State University. (By title)

In case of a multivariate regression with n past observations available the results of the preceding abstract provide the distribution of a studentized future observation ξ and tolerance regions for ξ. In case the $p \times k$ matrix $\mathbf{Y}$ is a sample of k column vectors $\mathbf{Y}^j$ with distribution $N(\mathbf{B}^* \mathbf{U}^j, \mathbf{\Sigma})$, where $\mathbf{B}^*$ is $p \times q$ and $\mathbf{U} = (\mathbf{U}^1, \cdots, \mathbf{U}^k)$, let $\bar{\mathbf{Y}}$ and $\bar{\mathbf{U}}$ denote the sample means of the $\mathbf{Y}^j$ and $\mathbf{U}^j$; then $\xi = k^{\frac{1}{2}}[\bar{\mathbf{Y}} - (\mathbf{B}^* - \mathbf{B})\bar{\mathbf{U}}]$ and $\mathbf{Z}_\xi = k^{\frac{1}{2}}\bar{\mathbf{U}}$. The distribution of $\tilde{\mathbf{n}}'\tilde{\mathbf{n}}$ provides a two sample test of H_0: $\mathbf{B}^* =$ known $\mathbf{B}$; the distribution of $\mathbf{n}'\mathbf{n}$ provides a test of H_0: $\mathbf{B}^* - \mathbf{B} = \mathbf{0}$, i.e. of the equality of the regression coefficients of two multivariate regressions.

5. **Generalization of Thompson's Distribution IV.** ANDRE G. LAURENT, Wayne State University.

Let the $p \times n$ matrix $\mathbf{X}_n$ be a sample of n column vectors $\mathbf{X}^i$ with distribution $N(\mathbf{B}\mathbf{Z}^i, \mathbf{\Sigma})$, where $\mathbf{B}$ is $p \times q$ and $\mathbf{Z}_n = (\mathbf{Z}^1, \cdots, \mathbf{Z}^n)$ is $q \times n$ of rank q. Let $\hat{\mathbf{B}}_n$, $\hat{\mathbf{\Sigma}}_n$ be the M.L. estimates of $\mathbf{B}$, $\mathbf{\Sigma}$ and $\tilde{\mathbf{\Sigma}}_n$ that of $\mathbf{\Sigma}$ when $\mathbf{B}$ is known. Let ξ be a vector with distribution $N(\mathbf{B}\mathbf{Z}_\xi, \mathbf{\Sigma})$. The distribution of $\tilde{\mathbf{n}} = n^{-\frac{1}{2}} \tilde{\mathbf{\Sigma}}_n^{-\frac{1}{2}}(\xi - \mathbf{B}\mathbf{Z}_\xi)$ is

$$\pi^{-p/2}\{\Gamma[(n+1)/2]/\Gamma[(n+1-p)/2]\}(1 + \tilde{\mathbf{n}}'\tilde{\mathbf{n}})^{-(n+1)/2} d\,\tilde{\mathbf{n}}$$

and $\tilde{\mathbf{n}}'\tilde{\mathbf{n}}$ is Hotelling distributed with n d.o.f. The distribution of

$$\mathbf{n} = n^{-\frac{1}{2}}(1 - T^2)^{\frac{1}{2}}\hat{\mathbf{\Sigma}}^{-\frac{1}{2}}(\xi - \hat{\mathbf{B}}_n\mathbf{Z}_\xi),$$

where $n \geqq p + q$ and $(1 - T^2) = [1 + \mathbf{Z}_\xi'(\mathbf{Z}_n\mathbf{Z}_n')^{-1}\mathbf{Z}_\xi]^{-1}$, is

$$\pi^{-p/2}\{\Gamma[(n-q+1)/2]/\Gamma[(n-q-p+1)/2]\}(1 + \mathbf{n}'\mathbf{n})^{-(n-q+1)/2}\, d\mathbf{n},$$

and $\mathbf{n}'\mathbf{n}$ is Hotelling distributed with $n - q$ d.o.f. The results above follow from "Generalization of Thompson's Distribution III", W. S. U. Memo., April 1960, by means of the change of variables $\boldsymbol{\zeta} = (\mathbf{I} + \mathbf{n}\mathbf{n}')^{-\frac{1}{2}}\mathbf{n}$ ($\mathbf{n}$ with or without a tilde), where

$$\tilde{\boldsymbol{\zeta}} = (n+1)^{-\frac{1}{2}}\, \tilde{\mathbf{\Sigma}}^{-\frac{1}{2}}(\xi - \mathbf{B}\mathbf{Z}_\xi),$$

$\boldsymbol{\zeta} = (n+1)^{-\frac{1}{2}}(1 - T^2)^{-\frac{1}{2}}\hat{\mathbf{\Sigma}}^{-\frac{1}{2}}(\xi - \hat{\mathbf{B}}\mathbf{Z}_\xi)$, with $X = (X_n, \xi)$, $\mathbf{Z} = (\mathbf{Z}_n, \mathbf{Z}_\xi)$ of rank q; $\hat{\mathbf{B}}$, $\hat{\mathbf{\Sigma}}$, the M.L. estimates of $\mathbf{B}$, $\mathbf{\Sigma}$; and $\tilde{\mathbf{\Sigma}}$ that of $\mathbf{\Sigma}$ when $\mathbf{B}$ is known.

(Abstract of paper to be presented at the Central Regional Meeting, Madison, Wisconsin, June 14–15, 1963. Additional abstracts will appear in the June, 1963 issue.)

1. **A Note on Interpenetrating Subsampling.** P. K. PATHAK, Michigan State University. (Introduced by Leo Katz)

It is proved that for estimating the population mean from interpenetrating subsamples when the subsamples are drawn by simple random sampling (without replacement) a better estimator than the usual overall average of subsample means is given by the average of distinct units selected. This result is also true when the subsample sizes are unequal and also when the subsamples are drawn by systematic sampling. A similar result in simple random sampling (with replacement) by Basu (1958) and by Des Raj and Khamis (1958) follows as a special case of the above result. The variance and the variance estimator of the average of distinct units are derived. A better estimator of the population variance is given.

(Abstract of paper to be presented at the Annual Meeting of the Institute, Ottawa, August 27–29, 1963. Additional abstracts will appear in the June, 1963 issue.)

1. **On Characterizing Nonadditivity.** MARY D. LUM, Aeronautical Research Laboratories, Wright-Patterson Air Force Base, Ohio.

Since the validity of analysis of variance tests is strongly dependent on assumptions concerning additivity of effects, the question of how to characterize departures from additivity of main effects becomes an important consideration. If this can be done in a straightforward manner, it could conceivably indicate how the effects in a linear model may be made "more additive" (that is, having fewer or no interactions) by a suitable transformation to a different scale. In this paper a multiple regression is proposed as a general basis for characterizing this nonadditivity. This multiple regression of nonadditivity is one in which the dependent variable represents deviations of observed cell means from the fitted linear model (the grand mean plus the sum of the observed main effects) and in which the independent variables are chosen using the set of all possible products among observed main effects and all possible products (with the exception of the highest order interaction) of observed interactions with observed main effects. This procedure includes the Tukey and Harter-Lum techniques as special cases, and numerical comparisons are made with them.

(Abstracts not connected with any meeting of the Institute)

1. **On the Equiprobability of Two Rank Orders.** JEAN D. GIBBONS and H. A. DAVID, University of Cincinnati and Virginia Polytechnic Institute.

Let X_i and Y_j $(i = 1, 2, \cdots, m; j = 1, 2, \cdots, n)$ be $N = m + n$ mutually independent random variables drawn from the absolutely continuous cdf's $G(x)$ and $H(x)$, respectively. In order to calculate the power function of a rank test of the hypothesis $G = H$ against some chosen alternative, it is important to know the probability of rank orders of the type $\tilde{z} = (z_1, z_2, \cdots, z_N)$, where $z_k = 1$ or 0 $(k = 1, 2, \cdots, N)$ according as the kth smallest variable in the combined sample of N variables is a X or a Y. Let $\tilde{z}'$ be the rank order $(1 - z_N, 1 - z_{N-1}, \cdots, 1 - z_1)$ of nX's and mY's. *Theorem.* The rank orders $\tilde{z}$ and $\tilde{z}'$ have the same probability if G and H are mutually symmetric; i.e., if there exists a constant c such that $H(x - c) = 1 - G(c - x)$. *Corollaries* (i) If $\varphi(x)$ is a monotonic function of x, the theorem continues to hold for cdf's of the form $H[\varphi(x)]$ and $G[\varphi(x)]$. (ii) If $F(x)$ is any absolutely continuous cdf, the theorem holds if G and H are expressible as $G = \varphi(F)$, $H = 1 - \varphi(F)$. The special case $\varphi(F) = F^r (r > 1)$ corresponds to a two-sided version of Lehmann's (1953) nonparametric alternative.

2. Some Applications of a Result of Herz's to Noncentral Multivariate Distribution Problems in Statistics. D. G. KABE, Wayne State University. (Introduced by A. T. Bharucha-Reid)

In previous papers (Kabe, D. G., Some results on the distribution of two random matrices used in classification procedures, *Ann. Math. Statist.* (to appear)), and (Kabe, D. G., On the distribution of the latent roots of the Wishart matrices, (submitted for publication)), the author shows that some noncentral distribution problems in Multivariate Statistics depend on the evaluation of the moments of the noncentral generalized variance. In this paper the author uses a result of Herz's (Herz, Carl S., Bessel functions of matrix argument. *Ann. of Math.* **61** (1955) 474–523) to give formal solutions of the noncentral distributions of the roots of a single Wishart matrix, the noncentral distributions of the roots of two Wishart matrices, and the noncentral multivariate beta distribution.

3. Multivariate Normal Linear Hypothesis, Studentized Observations. ANDRE G. LAURENT and D. G. KABE, Wayne State University.

Let $\mathbf{Y} = (\mathbf{Y}^1, \cdots, \mathbf{Y}^k)$ be a $p \times k$ matrix and $\mathbf{Y}^*$ the pk dimensional vector made of $\mathbf{Y}$'s elements put in lexicographic order; let $\mathbf{Y}^*$ be normally distributed $N(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{B})$, where $\mathbf{\Sigma}$ is $p \times p$, $\mathbf{B}$ is $k \times k$, and both are positive definite; let $\mathbf{V}_\nu$ be a $p \times p$ matrix which is Wishart $W(\nu, \mathbf{\Sigma})$ distributed. Then the studentized $\mathbf{Y}$, namely $\mathbf{n} = \mathbf{V}_\nu^{-\frac{1}{2}}\mathbf{Y}\mathbf{B}^{-\frac{1}{2}}$ has distribution $C \mid \mathbf{I} + \mathbf{n}\mathbf{n}' \mid^{-(\nu+k)/2} d\mathbf{n}$; the distribution of $\mathbf{n}'\mathbf{n}$ when $k \leqq p$ and $\mathbf{n}\mathbf{n}'$ when $k \geqq p$ follow from this result. The cases $\mathbf{Y} = \boldsymbol{\xi} - \hat{\mathbf{B}}_n\mathbf{Z}$ and $\mathbf{Y} = \boldsymbol{\xi} - \mathbf{B}\mathbf{Z}$, where $\boldsymbol{\xi} = (\boldsymbol{\xi}^1, \cdots, \boldsymbol{\xi}^k)$, $\mathbf{Z} = (\mathbf{Z}^1, \cdots, \mathbf{Z}^k)$, the $\boldsymbol{\xi}^i$ are independently $N(\mathbf{B}\mathbf{Z}^i, \mathbf{\Sigma})$ distributed, and $\hat{\mathbf{B}}_n$ is the maximum likelihood estimate of $\mathbf{B}$ based on an independent sample, lead to interesting applications to tolerance regions, tests of hypotheses, "errors of predictions", when the multivariate normal linear hypothesis is valid and the covariance matrix is unknown. The case where $\mathbf{Y}^*$ is $N(\boldsymbol{\mu}, \mathbf{\Sigma} \otimes \mathbf{B})$ and $\boldsymbol{\mu}$ is of rank 1 or 2 has also been considered.