SOME TOPICS IN SAMPLING THEORY[*]

BY H. L. RIETZ

1. *Introduction.* Many important contributions have been made to "small or exact sampling theory" since 1908, when "Student's" contribution[†] created a new interest in the subject. A general notion of the extent of the literature on this topic may be obtained by an examination of the excellent survey by P. R. Rider[‡] in 1930. Next, J. O. Irwin[§] gave valuable reports on this literature for approximately the years 1930–1934. Then Rider[||] brought the survey on exact sampling theory well up-to-date in a paper of 1935, and again for part of the field in 1936.

The invitation to give a paper at this meeting left me free as to the selection of a subject. This freedom was interpreted to mean that you would probably prefer to have me speak on some topics in which I have a special interest, rather than to attempt a well-balanced discussion of recent progress for which we may well turn to the papers mentioned above as surveying the contributions to exact sampling theory.

On the side of applications, sampling theory is much concerned with judging, by means of one or more tests, whether an observed random sample, taken as a whole, conforms reasonably to samples expected from a specified population. Test criteria may be based on such concepts as the mean, the variance, the standard deviation, the Pearson χ^2, the "Student" ratio, the Fisher z, the correlation ratio, and other statistical estimates of

[*] An address delivered before the Society on November 27, 1936, in Lawrence, by invitation of the Program Committee.

[†] *The probable error of the mean,* Biometrika, vol. 6 (1908–09), pp. 1–25.

[‡] *A survey of the theory of small samples,* Annals of Mathematics, (2), vol. 31 (1930), pp. 577–628.

[§] *Recent advances in mathematical statistics,* Journal of the Royal Statistical Society, vol. 94 (1931), pp. 568–578; vol. 95 (1932), pp. 498–530; vol. 97 (1934), pp. 114–154; vol. 99 (1936), pp. 714–769.

[||] *Recent progress in statistical method,* Journal of the American Statistical Association, vol. 30 (1935), pp. 58–88. *Annual survey of statistical technique developments in the analysis of multivariate data,* Part I, Econometrica, vol. 4 (1936), pp. 264–268. See also C. F. Roos, *Annual survey of statistical techniques: The correlation and analysis of time series,* Part II, Econometrica, vol. 4 (1936), pp. 368–381.

averages or parameters calculated from a sample. In what follows, my apparent freedom will be exercised by choosing to emphasize the "Student" ratio, although commenting briefly on some of the other concepts mentioned above and on certain of their generalizations. This choice of emphasis is perhaps largely one of expediency as it seems simpler to discuss, before a general mathematical audience, the concept and sampling theory of the "Student" ratio than of any one of the other concepts whose exact sampling theory is finding many applications.

It seems convenient to divide the paper into two parts. In Part I it will be assumed that the random samples are drawn from normal parent populations, whereas in Part II we shall deal briefly with random samples drawn from certain specified non-normal populations.

PART I

RANDOM SAMPLES FROM NORMAL PARENT POPULATION

2. *Distribution Function of the Sum of Squares.* While it seems appropriate to cite "Student's" paper* of 1908 as marking the beginning of what is commonly regarded as "small or exact sampling theory" in applied statistics, it also seems to be appropriate and historically correct to direct attention to papers by Helmert† published in 1875–1876 that gave the basis for starting a new small sampling development from his theorems concerning the theoretical distribution function of the sums of squares of true and of apparent errors. In the language of statistics, these theorems may be expressed as follows.

Given a normal parent population of x's with mean 0 *and variance* σ^2 *from which are drawn at random each of* N *independent values,* $x_1, x_2, \cdots, x_N$, *measured from the population mean as the origin, giving as the sample mean* $\bar{x} = (x_1 + x_2 + \cdots + x_N)/N$ *and as the second moment of the sample from the population mean* $s^2 = \bar{\mu} = (x_1^2 + x_2^2 + \cdots + x_N^2)/N$. *Then the probability that the*

* Loc. cit.

† *Ueber die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler und über einige damit im Zusammenhange stehende Fragen*, Zeitschrift für Mathematik und Physik, vol. 21 (1876), pp. 192–218; see also Astronomische Nachrichten, vol. 85 (1875), No. 2039; ibid, vol. 88 (1876), No. 2096–7. See also E. Czuber, *Theorie der Beobachtungsfehler*, 1891, pp. 136–164.

sum of squares of deviations $U = x_1^2 + x_2^2 + \cdots + x_N^2$ will fall into the interval U to $U + dU$ is given by*

$$\frac{1}{2^{N/2}\sigma^N\Gamma\left(\frac{N}{2}\right)} U^{(N-2)/2}e^{-U/(2\sigma^2)}dU. \tag{1}$$

If we let $U = N\bar{\mu}$, we obtain as a corollary of (1) that the probability that a sample value of $\bar{\mu}$ will fall into the interval $d\bar{\mu}$ at $\bar{\mu}$ is given by

$$\left(\frac{N}{2}\right)^{N/2} \frac{1}{\sigma^N\Gamma\left(\frac{N}{2}\right)} \bar{\mu}^{(N-2)/2}e^{-N\bar{\mu}/(2\sigma^2)}d\bar{\mu}. \tag{2}$$

The distribution functions (coefficients of dU and $d\bar{\mu}$) in (1) and (2) were derived by Helmert in a rather elegant but somewhat tedious manner involving mathematical induction. In statistical applications we do not ordinarily know the mean of the population nor the variance, σ^2, but make estimates of their values from a random sample. To deal with this situation, Helmert found that the probability that the sum of squares of residuals

$$U = N\bar{\mu} = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_N - \bar{x})^2 \tag{3}$$

will fall into the interval dU at U is the same as the probability that the sum of squares of $N-1$ discrepancies from the population mean will fall into the same interval, so that the frequency function of the sample variance $s^2 = \bar{\mu}$ given by (3) is equal to

$$\left(\frac{N}{2}\right)^{(N-1)/2} \frac{1}{\sigma^{N-1}\Gamma\left(\frac{N-1}{2}\right)} \bar{\mu}^{(N-3)/2}e^{-N\bar{\mu}/(2\sigma^2)}. \tag{4}$$

The frequency functions (1), (2), and (4) are Pearson Type III functions. Although Czuber in his *Beobachtungsfehler* (1891) gave, in an improved notation, a good account of Helmert's contributions, neither the theoretical nor the practical import of the discovery was recognized in practical statistics until long

* It is to be understood throughout this paper that "is given by," used before an element for integration, means "is given, to within infinitesimals of higher order, by."

after Pearson in 1900 had arrived at the distribution of his χ^2 and "Student" had in 1908 inferred the distribution of the variances, s^2, from the relations he had found among the expected values of the first four moments of variance.

"Student" was first to recognize the fundamental importance, for the theory of small samples, of taking account of the simultaneous sampling variations of $\bar{x}$ and s since the ratio, $\bar{x}/s$, is used whenever we enter a normal probability table with s as an estimate of σ. On the basis of finding a linear correlation equal to zero between $\bar{x}^2$ and s^2, "Student" correctly inferred, by means of a remarkable intuition, the independence of $\bar{x}$ and s^2 in the probability sense of independence. This property of independence of $\bar{x}$ and s^2 was established explicitly by R. A. Fisher* by showing that the simultaneous distribution function of $\bar{x}$ and s^2 is the product of two functions, one of which contains $\bar{x}$ but not s and the other contains s but not $\bar{x}$, and it was implicitly contained in earlier work by Karl Pearson† and others.‡

By considering the ratio $z=\bar{x}/s$ and taking into account the sampling variations of s as well as those of $\bar{x}$, "Student" found the probability that a random sampling value of z will fall into an assigned dz to be

$$\frac{\Gamma(N/2)dz}{\pi^{1/2}\Gamma\left(\frac{N-1}{2}\right)(1+z^2)^{N/2}}. \tag{5}$$

For values of $N>30$, it is frequently satisfactory to employ the normal function

$$\left(\frac{N-3}{2\pi}\right)^{1/2} e^{-(N-3)z^2/2}$$

for (5) in applications.

A small probability table of the integral of (5) for $N=4$ to $N=10$ is given by "Student"§ in his original paper and later in a somewhat more extensive table|| for $N=2$ to $N=30$ in 1915.

* R. A. Fisher, *Applications of "Student's" distribution*, Metron, vol. 5 (1925), No. 3, pp. 90–93.

† Karl Pearson, *On the distribution of the standard deviations of small samples*, Biometrika, vol. 10 (1915), pp. 522–529.

‡ U. Romanowsky, *On the moments of standard deviations and of correlation in samples from a normal population*, Metron, vol. 5 (1925), No. 4, pp. 8–12.

§ Loc. cit., p. 19.

|| Biometrika, vol. 11, pp. 414–417.

These tables give the probability (to four places of decimals) that a sample value of z shall fall between $-\infty$ and an assigned z.

The revolutionary character of the idea introduced by "Student" comes forcibly to light by making applications that involve drawing probable inferences from small samples, say from a sample with $N=10$, and perhaps even more forcibly in following the generalizations of the "Student" idea.

3. *The Fisher Modifications and Generalizations.* For purposes of generalization, it seems that, in tests of significance, there are some advantages in following R. A. Fisher's modification of the "Student" ratio by using the ratio of $\bar{x}$ to its own standard deviation, $s/(N-1)^{1/2}$, as estimated from a sample. Then with

$$t = \frac{\bar{x}}{s}(N-1)^{1/2} \qquad \text{or} \qquad z = \frac{\bar{x}}{s} = \frac{t}{(N-1)^{1/2}},$$

the probability that t, from a sample of N, will fall into dt is given by

$$\frac{\Gamma\left(\frac{N}{2}\right)dt}{(N-1)^{1/2}\pi^{1/2}\Gamma\left(\frac{N-1}{2}\right)\left(1+\frac{t^2}{N-1}\right)^{N/2}}. \tag{6}$$

The extensions of the "Student" ratio by R. A. Fisher may be described as schemes for making the "Student" idea applicable to a wide range of data largely by emphasis on building a variable t in the nature of a fraction whose numerator is a variate normally distributed about zero and whose denominator is an unbiassed estimate of the standard deviation of the numerator. Fisher applied this principle in the building of appropriate variables to obtain new tests of the significance of the difference between two means, of a linear regression coefficient, of nonlinear regression coefficients of orthogonal functions, and of the coefficients in a multiple regression surface.

4. *Generalization of the "Student" Ratio.*[*] Both applications to situations involving more than one variable and the natural

[*] Harold Hotelling, *The generalization of "Student's" ratio,* Annals of Mathematical Statistics, vol. 2 (1931), pp. 360–378.

tendency to generalize important ideas probably suggested the generalization of the "Student" ratio to multivariate situations. Hotelling's generalization sets up a function T of n variables $x_1, x_2, \cdots, x_n$, each of which is measured for N individuals.

To obtain a general notion of the nature of T, consider first the deviations of sample values from a hypothetical set of mean values $m_1, m_2, \cdots, m_n$. We may calculate the means $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n$, of the samples, and put $\xi_i = (\bar{x}_i - m_i)/N^{1/2}$. Assuming the individual items to be taken independently from an infinite population, the expected value of ξ_i will be zero. To outline a procedure for estimating the values of the variance and covariances, we may write $x_{ik} = X_{ik} - \bar{x}_i$, where X_{ik} is the value of x_i for the kth individual. Then take

$$(7) \qquad a_{ij} = a_{ji} = \frac{1}{N-1} \sum_{k=1}^{N} x_{ik} x_{jk},$$

and

$$A_{ij} = A_{ji} = (\text{cofactor of } a_{ij} \text{ in } |a_{ij}|)/|a_{ij}|,$$

where $|a_{ij}|$ denotes the nth order determinant of elements a_{ij}. Then the measure of simultaneous deviations employed in the generalization is the quadratic form

$$(8) \qquad T^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} \xi_i \xi_j.$$

In problems of examining the deviation of a single variable from zero, T reduces to the Fisher modification of the "Student" ratio. For problems of examining the deviations from zero of two variables, say x and y, T^2 reduces to

$$T^2 = \frac{N}{1-r^2}\left(\frac{\bar{x}^2}{s_x^2} - \frac{2r\bar{x}\bar{y}}{s_x s_y} + \frac{\bar{y}^2}{s_y^2}\right)$$

in a familiar notation involving sample means, variances, and the coefficient of correlation.

Illustrations of the situations to which the T^2 may be applied include the comparison of means of two samples of n variables, and comparison of regression coefficients for functions of more than one variable. Deviations, x_{ik}, of the observations from means, or from regression functions, or other such functions, are used to estimate the variances and covariances, a_{ij}. The distribution function of T is found to be given by

$$(9) \qquad \frac{2\Gamma\left(\frac{N}{2}\right)T^{n-1}}{\Gamma\left(\frac{N-n}{2}\right)\Gamma\left(\frac{n}{2}\right)(N-1)^{n/2}\left(1+\frac{T^2}{N-1}\right)^{N/2}}.$$

5. *Generalization of Variance.* In 1932, Wilks* gave generalizations to n-variate populations of the concept of variance and found the distributions of several important functions of the sample observations whose distribution laws had been established when the sampling was limited to univariate normal populations.†

More specifically, the extensions in question are concerned with the generalization of the concept of variance itself, and of such other concepts as the Fisher z, the correlation ratio, γ, the related $1-\gamma^2$, the "Student" ratio previously generalized by Hotelling, the λ-criteria of Pearson and Neyman, so as to give each of these concepts a meaning for multivariate populations and for samples from such populations. In his article of 1934 giving "recent advances," Irwin states very appropriately that this paper by Wilks gives generalizations of the greatest interest.

In a notation differing only slightly from that used above, Wilks takes

$$(10) \qquad a_{ij} = \frac{1}{N}(X_{ik}-x_i)(X_{jk}-x_j), \qquad (i, j = 1, 2, \cdots, n).$$

The determinant $|a_{ij}|$ was adopted as the generalized sample variance. Wilks points out that $|a_{ij}|$ for n-variate samples and the ordinary variance for univariate samples are similar in the way they arise in maximizing likelihood functions.

6. *Distribution of Generalized Variance.* Wilks found the distribution function of $|a_{ij}|$ by the method of moment generating

* S. S. Wilks, *Certain generalizations in the analysis of variance*, Biometrika, vol. 24 (1932), pp. 471–494; *Moment-generating operators for determinant of product moments in samples from a normal system*, Annals of Mathematics, (2), vol. 35 (1934), pp. 312–339.

† See J. Wishart, *The generalized product moment distribution in samples from a normal multivariate population*, Biometrika, vol. 20A (1928), pp. 32–52.

functions. His success seems to depend largely on the fact that he obtained the solutions of two integral equations that have a wide application. The distribution function of $\xi = |a_{ij}|$ took the form of a multiple integral which could be integrated and expressed explicitly for special values of the parameters and for $n = 1$ and $n = 2$. Thus, for $n = 1$, the distribution function reduced to the well known distribution of ordinary variance given in (4). For $n = 2$, the distribution function for generalized variance is

$$D_2(\xi) = \frac{2^{N-3}A_2{}^{(N-2)/2}\xi^{(N-4)/2}e^{-2(A_2\xi)^{1/2}}}{\Gamma(N-2)}, \tag{11}$$

where

$$A_2 = \frac{N^2}{4\sigma_1^2\sigma_2^2(1-\rho_{12}^2)}.$$

In 1934, Kullback* obtained the distribution function of ξ in explicit form for any positive integral value of n, by the use of characteristic functions.

7. *Distribution of the Ratio of Generalized Variances.* Fisher's z-distribution in which $z = \frac{1}{2}\log s_2^2/s_1^2$, where s_1^2, and s_2^2 are two independent estimates of variance, is much used in the analysis of variance for univariate populations. Put $u = s_2^2/s_1^2 = e^{2z}$. It is this u which Wilks has generalized into $\psi = \xi/\eta$, where ξ and η are generalized variances for two samples from populations of n variables for which the generalized variances are given. He then developed the distribution function of ψ, and, in particular, the distribution function of ψ for the bivariate population as a special case.

8. *Generalized Correlation Ratio.* Consider p samples ω_β, $(\beta = 1, 2, \cdots, p)$, of N_β items respectively drawn from a normal population of one variable. Let $\bar{x}_\beta$ and s_β^2 be the mean and variance of ω_β. Let Ω be the sample formed by pooling the ω's, and let its mean and variance be denoted by $\overline{X}$ and S^2, respectively. Then the correlation ratio, η, for the p categories, as defined by R. A. Fisher, is given by

* Solomon Kullback, *An application of characteristic functions to the distribution problem of statistics,* Annals of Mathematical Statistics, vol. 5 (1934), pp. 263–307.

$$(12)\qquad \eta^2 = \frac{\sum_{\beta=1}^{p} N_\beta(\bar{x}_\beta - \bar{x})^2}{NS^2}, \qquad \left(\sum_{\beta=1}^{p} N_\beta = N\right),$$

this being a weighted variance of the means of p sub-samples (arrays) divided by the variance of Ω. To generalize (12), consider the p samples ω_β', $(\beta = 1, 2, \cdots, p)$, of n_β items drawn from an n-variate normal population. Let the sample formed by pooling the drawings be Ω' with $\sum n_\beta = N$ items. Then the generalized η^2 may be written

$$(13)\qquad U = \frac{|b_{ji}|}{|a_{ij}|},$$

where

$$b_{ij} = b_{ji} = \frac{1}{N}\sum_{\beta=1}^{p} n_\beta(\bar{x}_{i\beta} - \overline{X}_i)(\bar{x}_{j\beta} - \overline{X}_j),$$

and

$$a_{ij} = a_{ji} = \frac{1}{N}\sum_{\beta=1}^{p}\sum_{k=1}^{n_\beta} (x_{i\beta k} - \overline{X}_i)(x_{j\beta k} - \overline{X}_j),$$

where $\overline{X}_{i\beta}$ is the mean of the ith variate in the βth sample and $x_{i\beta k}$ is the value of the kth individual of the ith variate in the βth sample, ω_β'.

The distribution function of U is found. The integrations are carried out for $n = 1$ and yield the well known results of R. A. Fisher and of Hotelling for the distribution of η^2. We may write $a_{ij} = b_{ij} + c_{ij}$, where

$$c_{ij} = \frac{1}{N}\sum_{\beta=1}^{p}\sum_{k=1}^{n_\beta} (x_{i\beta k} - \overline{X}_{i\beta})(x_{j\beta k} - \overline{X}_{j\beta}).$$

It is shown that $W = |c_{ij}| / |a_{ij}|$ may be regarded as a generalization of $1 - \eta^2$, and the distribution function of W is found. For $n = 1$, the integrations are carried out completely. Moreover, it is shown that W serves as a maximum likelihood criterion, λ_H, of the type used by Neyman and E. S. Pearson. It turns out that a simple relation $\lambda_H = W^{n/2}$ exists, thus showing a simple connection among these fundamental statistical concepts.

For samples from multivariate populations, Wilks* has devised criteria of the Neyman-Pearson likelihood type for testing the following classes of hypotheses:

(1) That a sample is from a normal population with a specified set of means.

(2) That two or more samples are from populations having a common system of: (a) means; (b) variances and covariances; (c) means, variances, and covariances.

(3) That several sets of variates are mutually independent.

Wilks directs attention to the fact that all the criteria he has considered may be called "Studentized" functions by which he means that the criteria and their probability functions are completely expressible in terms of observations.

Part II

Random Samples from Some Non-Normal Parent Populations

9. *Non-Normal Parent Populations.* It is fairly obvious that many of the samples used in the application of statistical coefficients or ratios are drawn from non-normal parent populations that differ very much from normal populations. Moreover, a small sample is almost sure to be inadequate to give information essential to pass a reasonable judgment about the type of parent population. This situation suggests the importance of learning, if possible, how far departures from normality in a parent population influence the distribution of statistical estimates such as the mean, the variance, and the "Student" ratio used to decide whether a sample belongs to a given parent population.

We shall consider first a few non-normal parent populations for which the exact form of the distribution function of the mean, standard deviation, or "Student" ratio has been found. Incidentally, in some cases, we shall comment also very briefly on the exact distributions of some other averages such as the median, harmonic mean, geometric mean, center, or variance, if time and space permit. Later, we shall discuss briefly the progress that has been made in characterizing the distributions of $\bar{x}$, s^2, and $\bar{x}/s$ for samples from various types of parent populations by means of theoretical moments and by experimental sampling.

* S. S. Wilks, *Test criteria for statistical hypotheses involving several variables,* Journal of the American Statistical Association, vol. 30 (1935), pp. 549–560.

10. *On Exact Distributions of Some Averages for Samples from Certain Non-Normal Parent Populations with Special Reference to the Mean* $\bar{x}$. For a continuous rectangular parent population given by $y=1/a$ from $x=0$ to $x=a$, $(a>0)$, the distribution function of $\bar{x}$ for samples of N items consists of N polynomials, each being of degree $N-1$, and each being applicable to a subinterval of length a/N. The curve is bell-shaped and resembles the normal curve when $N\geqq 3$. For $N=2$, the curve degenerates into the two equal sides of an isosceles triangle.

It seems nearly certain that Laplace* knew the distribution of means of samples of N items, each drawn from a continuous rectangular universe; for he knew the exact frequency function, $f(u)$, of the sum $u=x_1+x_2+\cdots+x_N$ of N elements, each x_i being a real number taken at random from a given interval 0 to a, $(a>0)$, of a rectangular universe, and it is a small step from this distribution of u to the distribution of the corresponding mean $\bar{x}$, where $\bar{x}=Nu$. Laplace applied the distribution function of u to the solution of the historic problem of the probability that the inclinations of the orbits of the ten planets besides the earth known at the beginning of the year 1801 do not constitute a random distribution. Perhaps it would not be without some interest to remark that the distribution formula for the sum, u, of N elements drawn at random from a rectangular universe as described above, has again appeared in 1936 in an up-to-date problem† in a paper by Condon and Breit, on the energy distribution of neutrons. The distribution of means of samples drawn from a continuous rectangular universe was given in explicit form by Irwin‡ and by Hall§ in 1927.

In 1929, Rider|| developed the exact distribution function of

* See H. L. Rietz, *On a certain law of probability of Laplace*, Proceedings, International Congress of Mathematics, Toronto, vol. 2 (1924), pp. 795–799.

† E. U. Condon and G. Breit, *The energy distribution of neutrons slowed by elastic impacts*, Physical Review, (2), vol. 49 (1936), No. 3, p. 230.

‡ J. O. Irwin, *On the frequency distribution of the means of samples from a population having any law of frequency with finite moments, with special reference to Pearson's Type* II, Biometrika, vol. 19 (1927), pp. 225–239.

§ Phillip Hall, *The distribution of means of samples of size N drawn from a population in which the variate takes values between* 0 *and* 1, *all such values being equally probable*, Biometrika, vol. 19 (1927), pp. 240–244.

|| P. R. Rider, *On the distribution of ratio of mean to standard deviation in small samples from non-normal universes*, Biometrika, vol. 21 (1929), pp. 124–143.

means of samples of four items drawn from a discrete five class rectangular parent population with x taking values from -2.5 to $+2.5$. He found that third differences of the resulting probabilities vanish except at $x=0$, ± 0.25. He fitted the probabilities by means of cubic curves, and also exhibited the results in tabular form. Similar results were obtained by using curves of degree one and two for samples of two and three items, respectively.

The paper to which I have just now referred presents some exact distributions of medians and of some other averages for both discrete or continuous rectangular parent populations. However, it is concerned largely with exact distributions of "Student" ratios to be discussed later in this paper. E. L. Dodd* had given a formula in 1922, in terms of integrals, for the distribution of medians of samples from a rather general parent population, but Rider gave the results in explicit integrated form for the rectangular parent population.

In 1920, Karl Mayr† developed a theory for the determination of the distribution of the sum of N items. He then applied the theory to the parent population given by $\phi(x)=e^{|-x|}/2$ for all real values of x.

The determination of the distribution functions of sample means for the Pearson types as parent populations has been the subject of considerable investigation by several authors. A. E. R. Church‡ seems to have been first to find the exact distribution of the mean, $\bar{x}$, of samples from the Type III population. He carried out the integrations. The resulting distribution function is a Type III curve. J. O. Irwin§ arrived a little later at the same result by a method that involves the use of integral equations and complex variables. He found also, in the form of an integral, the distribution of the means of samples from a Type II population, and evaluated the integrals in some special

* E. L. Dodd, *Functions of measurements under general laws of error*, Skandinavisk Aktuarietidskrift, vol. 5 (1922), pp. 132–158.

† Karl Mayr, *Wahrscheinlichkeitsfunctionen und ihre Anwendungen*, Monatshefte für Mathematik und Physik, vol. 30 (1920), pp. 17–43.

‡ A. E. R. Church, *Means and squared standard deviations of small samples from any population*, Biometrika, vol. 18 (1926), pp. 321–394.

§ Loc. cit., pp. 225–239.

cases. In 1929, C. C. Craig* obtained the distribution of sample means from a Type III parent population by the use of semi-invariants of Thiele. In 1930, G. A. Baker† found the distribution function of means, $\bar{x}$, of samples of N drawn from a parent population defined by the first $m-1$ terms of a Type A Gram-Charlier series

$$f(x) = a_0\phi_0(x) + a_3\phi_3(x) + \cdots + a_m\phi_m(x),$$

where

$$\phi_i(x) = \frac{d^i(e^{-x^2/2})}{dx^i}.$$

In 1931, C. C. Craig‡ arrived at Baker's results in a very few steps, by use of the semi-invariants of Thiele. In 1931, Rider§ gave the distributions, in tabular form, for means of samples drawn from triangular and from U-shaped populations.

In 1932, A. T. Craig‖ classified probability functions, $f(x)$, of the parent populations from which samples are drawn, into three classes according as x is allowed the range $(-\infty, \infty)$, $(0, \infty)$, or $(0, A)$, $A > 0$. Craig then established general theorems by which the problems of finding the distributions of arithmetic means, harmonic means, geometric means, medians, quartiles, and deciles are reduced to problems of integration. The illustrative parent universes for which integrations were carried out for the arithmetic mean, for some or all values of N, are given

* C. C. Craig, *Sampling when the parent population is of Pearson's Type* III, Biometrika, vol. 20 (1929), pp. 287–293.

† G. A. Baker, *Distribution of the means of samples of n drawn at random from a population represented by a Gram-Charlier series*, Annals of Mathematical Statistics, vol. 1 (1930), pp. 199–204. See also B. H. Camp, *Problems in sampling*, Journal of the American Statistical Association, vol. 18 (1923), pp. 964–977.

‡ C. C. Craig, *Note on the distribution of means of samples of N drawn from a Type A population*, The Annals of Mathematical Statistics, vol. 2 (1931), pp. 99–101.

§ P. R. Rider, *On small samples from certain non-normal universes*, Annals of Mathematical Statistics, Vol. 2 (1931), pp. 48–65.

‖ A. T. Craig, *On the distribution of certain statistics*, American Journal of Mathematics, vol. 54 (1932), pp. 353–366.

by assigning to the parent population, $f(x)$, the following special forms:

$$(1) \qquad f(x) = \frac{1}{a}, \qquad (0 \leqq x \leqq a),$$

$$(2) \qquad f(x) = \frac{k}{\sigma} e^{-x/\sigma}, \qquad (0 \leqq x < \infty),$$

$$(3) \qquad f(x) = kx^{-1/2}e^{-x/2}, \qquad (0 \leqq x < \infty),$$

$$(4) \qquad f(x) = \frac{k}{\sigma} e^{-|x|/2}, \qquad (-\infty < x < \infty),$$

$$(5) \qquad f(x) = \frac{4x}{a^2}, \qquad (0 \leqq x \leqq a/2),$$

$$= \frac{4}{a^2}(a - x), \qquad (a/2 \leqq x \leqq a).$$

The integration for (4) is carried out for $N=2$, 3, and 4 only, and that for (5) is carried out for $N=2$ only. For the distributions of harmonic means, geometric means, medians, and ranges, a somewhat similar set of illustrations is given for which integrations are carried out.

To illustrate, for the median, the integrations were carried out for the following besides the rectangular distribution:

$$f(x) = \frac{k}{\sigma} e^{-|x|/\sigma}, \qquad (-\infty < x < \infty),$$

$$f(x) = \frac{2x}{a^2}, \qquad (0 \leqq x \leqq a),$$

$$f(x) = e^{-x}, \qquad (0 \leqq x < \infty).$$

In 1934, Baten* found in explicit form, the distribution function for the sum of N independent items of a sample from a parent population defined by $(1/(2h))$ sech $(\pi x/(2h))$.

In a paper of 1935,† Weida developed the distribution func-

* W. D. Baten, *The probability law for the sum of n independent variables, each subject to the law* $(1/(2h))$ sech $(\pi x/(2h))$, this Bulletin, vol. 40 (1934), pp. 284–290.

† F. M. Weida, *On certain distribution functions when the law of the universe is Poissons' First Law of Error*, Annals of Mathematical Statistics, vol. 6 (1935), pp. 102–110.

tion for the samples of N items for the parent population

$$f(x) = \frac{k}{\sigma} e^{-|x|/\sigma}, \qquad (-\infty < x < \infty),$$

in explicit form, for any value of N. The development was carried out by the use of characteristic functions.

11. *On Some Exact Distributions of Standard Deviations.* With respect to exact distributions of sample standard deviations, s, from non-normal parent populations, Rider* found, for samples of 2 items, the distribution function of s to be $f(s) = 4(1-2s)$, when the drawings are from the rectangular parent universe given by $y=1$ from $x=0$ to $x=1$.

Rietz† found the exact distribution function of the standard deviation of samples of 3 drawn from a rectangular parent population given by $y=1/a$ from $x=0$ to $x=a$ to be

$$f(s) = \frac{6\cdot 3^{1/2}}{a^3}(\pi a - 3s6^{1/2})s \qquad \text{when} \qquad 0 \leqq s \leqq \frac{a}{6}6^{1/2},$$

$$f(s) = \frac{18}{a^3}\left[-a3^{1/2}\text{arc cos}\left(\frac{a^2}{3s^2}-1\right) + 2(18s^2 - 3a^2)^{1/2} + \frac{\pi}{3}a3^{1/2} - 3s2^{1/2}\right] \qquad \text{when} \qquad \frac{a}{6}6^{1/2} \leqq s \leqq \frac{a}{3}2^{1/2},$$

where s is the standard deviation of samples.

It is easily shown that there is continuity both in the two curves and in their slopes at the junction at $s = a6^{1/2}/6$. We may note that the distribution curve for the interval 0 to $a6^{1/2}/6$ on s is a parabola, but the distribution curve for the interval $a6^{1/2}/6$ to $a2^{1/2}/3$ on s is a transcendental curve with a rather complicated equation. When $a=4$, we find 98.86 per cent of the area under the curve is under the parabola to the left of the ordinate $s=a6^{1/2}/6$ at the junction with the more complicated curve.

Rider‡ directed attention to the fact that several of the es-

* Loc. cit., p. 141.

† H. L. Rietz, *Note on the distribution of the standard deviation of sets of three varieties drawn at random from a rectangular population*, Biometrika, vol. 23 (1931), pp. 424–426.

‡ Loc. cit. (1929), p. 139.

timates of parameters (mean, median, range, and extreme average) in samples from a rectangular universe are distributed in accord with polynomials, which are, apparently, of degree one less than the number in the sample. This rule still holds for the distribution of the standard deviation from two items and for the major portion of the range for the case of three items. It seems natural to surmise that the distribution of s from N items may be a polynomial of degree $N-1$ for a part of the range on s starting at zero. My attempts to prove or disprove this conjecture have been unsuccessful. However, I shall present a bit of experimental evidence that the distribution of s for samples of four items drawn from a rectangular population is a polynomial of degree 3 for an interval at the left end of the range. Using Tippett's Random Numbers, we* drew 400 samples of 4 items each from a rectangular distribution with 19 class intervals, and obtained the following distribution of standard deviation s:

s	Frequency	s	Frequency
0.000-0.249	1	2.500-2.749	38
0.250-0.499	5	2.750-2.999	42
0.500-0.749	5	3.000-3.249	38
0.750-0.999	15	3.250-3.499	20
1.000-1.249	24	3.500-3.749	11
1.250-1.499	24	3.750-3.999	3
1.500-1.749	28	4.000-4.249	2
1.750-1.999	44	total	400
2.000-2.249	49		
2.250-2.499	51		

Without taking the space to discuss our scheme of testing this distribution, I will merely state that it turns out, when degrees of freedom are taken into account, that a third degree polynomial will fit the first 12 class frequencies of the 17 shown above better than a fourth or fifth degree polynomial. This conforms to Rider's observation.

* A. C. Olshen made the drawings.

In 1935, G. A. Baker[*] found the distribution function of the standard deviation for samples of 2 drawn from a parent population given by the first three terms of the Gram-Charlier Type A, and the distribution function for the standard deviation of samples of 3 drawn from a parent distribution given by the first two terms of the Type A series.

In 1932, A. T. Craig[†] found in the form of integrals the simultaneous frequency functions of $\bar{x}$ and s for drawings from a general type of non-normal parent populations when $N=2$, 3, or 4. The integrations were carried out in explicit form for several special functions when $N=2$, and $N=3$.

12. *On Some Exact Distributions of the "Student" Ratio.* In his 1929 paper,[‡] Rider gave, in tabular form, the distributions of "Student" ratios for samples of 2, 3, and 4 items drawn from discrete rectangular parent populations of five classes, and also of ten classes for samples of 4. Some features of the distributions will be discussed later in this paper. Rider[§] gave also the distribution function of the "Student" ratio for samples of two items drawn from a continuous rectangular parent population. In 1931, he published[||] the results of his study of samples from both triangular and U-shaped distributions. He drew the inferences that the general characteristics of the z-distribution from the U-shaped parent population resemble those for the rectangular population, that the negative skewness in the triangular population tended to produce skewness of the opposite type[¶] in the distribution of the "Student" z, and that the cumulative probability of $|z|$, for the triangular population, tends to follow the results from a normal universe fairly well.

[*] G. A. Baker, *Note on the distributions of the standard deviations and second moments of samples from a Gram-Charlier population*, Annals of Mathematical Statistics, vol. 6 (1935), pp. 127–130.

[†] A. T. Craig, *The simultaneous distribution of mean and standard deviation in small samples*, Annals of Mathematical Statistics, vol. 3 (1932), pp. 126–140.

[‡] Loc. cit., pp. 124–143.

[§] An error in sign has been corrected by Victor Perlo. Reference cited later.

[||] P. R. Rider, *On small samples from certain non-normal universes*, Annals of Mathematical Statistics, vol. 2 (1931), pp. 48–65.

[¶] See Neyman and E. S. Pearson, loc. cit., Biometrika, vol. 20A (1928), p. 198. See also "Sophister", loc. cit., Biometrika, vol. 20A (1928), p. 408.

In 1933, Victor Perlo* found the distribution functions of the "Student ratio" as modified by R. A. Fisher, for samples of 2 items, drawn from a rectangular parent population, to be

$$\frac{1}{2(1+|t|)^2};$$

for samples of 3, to be

$$\text{(i)}\quad \frac{-9}{4(t+1)(t^2-4)}\left(\frac{1}{t+1}+\frac{3t}{t^2-4}\right) + \frac{3^{3/2}(t^2+2)}{(t^2-4)^{5/2}}\tan^{-1}\frac{(t^2-4)^{1/2}}{3^{1/2}(t+2)}, \qquad (\infty \geqq t \geqq 2),$$

$$\text{(ii)}\quad \frac{-9}{4(t+1)(t^2-4)}\left(\frac{1}{t+1}+\frac{3t}{t^2-4}\right) + \frac{3^{3/2}(t^2+2)}{(4-t^2)^{5/2}}\tanh^{-1}\frac{(4-t^2)^{1/2}}{3^{1/2}(t+2)}, \qquad (2 \geqq t \geqq \tfrac{1}{2}),$$

$$\text{(iii)}\quad \frac{3^{1/2}}{2(4-t^2)(1-t^2)^{1/2}}\left(1-\frac{9t^2}{4-t^2}\right) + \frac{3^{3/2}(t^2+2)}{(4-t^2)^{5/2}}\tanh^{-1}\left(\frac{1-t^2}{4-t^2}\right)^{1/2}, \qquad (\tfrac{1}{2} \geqq t \geqq 0).$$

The function (i), (ii), (iii), is continuous with continuous derivatives except at $t = \pm 1/2$. For these values of t, the derivative is discontinuous. When this distribution function is compared with "Student's" exact distribution for samples of 3 from a normal parent population, the frequency of values of the "Student" ratio is greater at the ends and the middle, and less elsewhere, when the parent distribution is rectangular than when it is normal.

In bringing to a conclusion our comments on exact distribution functions, perhaps it should be said that it is not lack of interest, but only limitations of time and space that cause me to make my comments so brief on the contributions concerned with the distribution of averages other than the mean and standard deviation. No comments have been made on some

* Victor Perlo, *On the distribution of Student's ratio for samples of three drawn from a rectangular distribution*, Biometrika, vol. 25 (1933), pp. 203–204.

very interesting cases. Among these are the exact explicit distribution functions which Neyman* and E. S. Pearson found, from a rectangular parent population, for the distribution of range, center, and ratio of deviation of sample center from population center to the semirange.

13. *Investigations of the Distributions of $\bar{x}$ and s^2 by Moments and by Experimental Sampling.* In 1928, Church† contributed some experimental sampling investigations that are especially important when viewed as an attempt to make use of theoretical moments of mean and variance to obtain Pearson frequency curves to serve as distribution functions of the mean and variance. He found the distribution of the means of samples of 10 drawn from two infinite skew populations and from one finite population. In his interpretation of the results, he emphasized the strong tendency of the distribution of the means, $\bar{x}$, to assume an approximately normal form, and gave a method of predicting rather rapidly the Pearson type to which the distribution of $\bar{x}$ is likely to approximate in the case of samples drawn from an infinite supply. Later, in a study of means of samples from a U-shaped parent population, Holzinger‡ and Church found that the distribution of means, $\bar{x}$, in samples of N obtained by sampling from a U-shaped population is quite unsatisfactorily represented by a simple continuous curve until N is at least of the order of 50. It is further inferred that this effect is due mainly to the fact that when N is quite small, the distribution of $\bar{x}$ is likely to be composite in form. Returning to the part of Church's work on sample variance, s^2, we observe that it was inferred that the distribution of s^2 from the samples of 10, from the infinite supply, for each of the two populations may be described by Pearson curves in a manner useful for applied statistics.

"Sophister"§ extended the experimental work of Church by

* Neyman and E. S. Pearson, *On the use and interpretation of certain test criteria for purposes of statistical inference*, Biometrika, vol. 20A (1928), pp. 175–240.

† Loc. cit., pp. 321–394.

‡ K. J. Holzinger and A. E. R. Church, *On the means of samples from a U-shaped population*, Biometrika, vol. 20A (1928), pp. 361–388.

§ "Sophister", *Discussion of small samples drawn from an infinite skew population*, Biometrika, vol. 20A (1928), pp. 389–423.

giving the distribution of variance for samples of $N=5$ and $N=20$ items drawn from a Pearson Type III population differing much from a normal population. He inferred that the distribution of variances was adequately described by a Type VI curve. Attention may be called to the fact that Karl Pearson devised an equation for the distribution function of variances by assuming a Type VI curve which starts at zero and has its origin at zero, thereby using the constants of the population sampled only up to β_4.

In 1931 Le Roux* gave a numerical and graphic analysis of the formulas for the moments of variance with special reference to the relations among the β's with a view to obtaining suitable Pearson curves to represent the sampling distributions of statistics in the case of parent populations which can themselves be represented by Pearson curves. By this analysis light was thrown on the manner in which the variance distribution changes as the character of the population and sample size change. The methods were tested on twenty-one experimental distributions of s^2, among which are included "Sophister's" samples of 5 to 10 items, and Church's samples of 10 for one population. The inference is drawn that Karl Pearson's fixed start method of fitting the distribution of s^2 is satisfactory, giving for "goodness of fit" tests an average of $P=0.49$, whereas the four moment method of fitting may be quite unsatisfactory or even impossible in cases of small samples.

14. *More About the "Student" Ratio.* An investigation of the "Student" z from samples drawn from the skew populations used by "Sophister",† showed that the distributions of z were markedly skew. Nevertheless, he concluded that the value of "Student's" Table in practice is still indicated, even when the parent population is definitely skew.

It has been clearly shown by Shewhart that the theory of small samples is useful in applications to the quality control of manufactured articles. Shewhart‡ and Winters found, largely

* J. M. Le Roux, *A study of the distribution of variance in small samples*, Biometrika, vol. 23 (1931), pp. 134–190.

† Loc. cit., p. 408.

‡ W. A. Shewhart and F. W. Winters, *Small samples—new experimental results*, Journal of the American Statistical Association, vol. 23 (1928), pp. 144–153.

on the basis of experimental sampling, that the "Student" theory gives a marked improvement, for small samples, over the classical theory, but they indicated further that the "Student" procedure fails, in certain practical cases, to give a set of errors in means consistent with the actual errors they obtained from small samples. For the results in cases of sampling from both rectangular and triangular populations, part of the effect is traced to the correlation between $\bar{x}$ and s^2, and part to the departure of the distribution of s^2 from what it would be if the parent distribution were normal. Briefly stated, the probability that a value of z from observation will fall outside a prescribed interval from $-z_1$ to $+z_1$ is likely to be larger than the estimate from "Student's" probability tables, if the numerical value of z_1 is of moderate size, say in the neighborhood of 2 or more.

In 1929, Rider* gave further explanation for the failure of the "Student" probability tables to yield probabilities that would check well with the Shewhart and Winters experiments. After studying several types of parent populations, but especially the rectangular type, he also attributed the failure largely to the correlation between $\bar{x}$ and s^2 in the cases of non-normal parent distributions. In general, his results showed not only a greater number of numerical values of z outside an assigned interval $-z_1$ to z_1, as in the experiments of Shewhart and Winters, but also a greater clustering of numerical values of z about the population mean† than in the case of a normal parent population.

In 1929, E. S. Pearson‡ threw further light on the sensitiveness of values of the "Student" z so far as they concern the fundamental tests dealing with the means of samples, and the differences between means of two samples. He investigated the question as to how well the observed distributions of z, from a variety of non-normal populations, follow the "Student" theory based on a normal parent population. For a set of non-normal parent populations, the following classes of data were selected with $0 \leqq \beta_1 \leqq .5$:

* Loc. cit., pp. 124–143.

† See Victor Perlo, loc. cit.

‡ E. S. Pearson, *The distribution of frequency constants in small samples from non-normal symmetrical and skew populations*, Biometrika, vol. 21 (1929), pp. 259–286.

1000 samples of 2 from 7 populations with $1.8 \leqq \beta_2 \leqq 7.07$,
500 samples of 5 from 6 populations with $2.5 \leqq \beta_2 \leqq 7.07$,
500 samples of 10 from 7 populations with $2.5 \leqq \beta_2 \leqq 7.07$,
500 samples of 20 from 6 populations with $2.5 \leqq \beta_2 \leqq 7.07$.

The inference is drawn that a completely satisfactory analysis of the position of the "Student" z-test will only be possible when the theoretical distribution of z in samples from the non-normal parent population in question is found. It seems to the writer that the probable date of finding such exact distributions, for rather general skew parent distributions, is far in the future. However, the experimental results reported in E. S. Pearson's excellent paper and elsewhere suggest some inferences as to the nature of departures of the distribution of z from that of the normal theory with sampling from a fairly wide range of populations. The least satisfactory agreement seems to occur in the cases of extremely leptokurtic parent distributions.

In concluding these remarks on the distribution of certain averages and ratios, for samples from non-normal parent populations, it seems fairly obvious that it is easy to propose a simple sounding problem by merely asking for the distribution function of some statistic for samples drawn from a simple parent population, but it is usually very difficult to solve the problem. On this account, we know relatively little at present of all we wish to know about the exact distributions of statistics for samples drawn from non-normal parent populations.

Although the prospects of obtaining the exact distribution functions of such statistics as the standard deviation, s, or the "Student" ratio, z, for samples from a considerable variety of non-normal populations, do not seem promising, nevertheless, by the use of moments of moments, and experimental sampling, along with the exact determination of some distribution functions, significant contributions are being made towards an understanding of the probable nature of certain important features of the distributions in question.

The University of Iowa