SOME RECENT ADVANCES IN MATHEMATICAL STATISTICS, I

By Burton H. Camp[1]

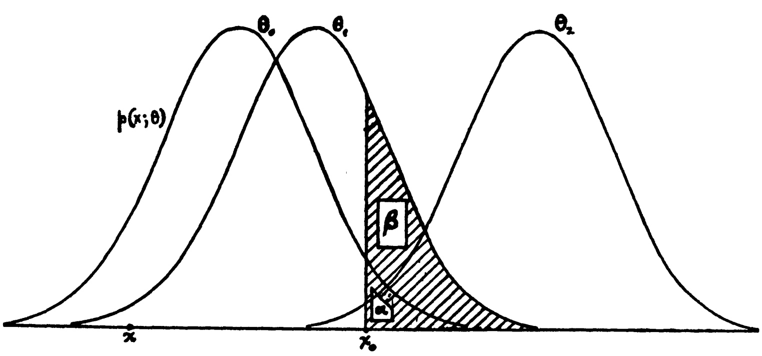
Wesleyan University

The papers considered in this partial review are listed at the end. For the most part they have appeared within the last five years, but in order to explain what has been done within the last five years it has been necessary occasionally to use material that appeared earlier. The subject matter is divided into four parts.

Part I. The Theory of Tests. Since an attempt is being made to present the material of this paper in such a form that it may be read rapidly by those who have not read the underlying literature, the author will endeavor to do little more, in Part I, than to define and illustrate several terms which are being used. Altogether there are nine of these terms. It is fortunate that their meanings can be explained pretty well by reference to an extremely simple picture. Let each of the curves in the figure indicate a probability distribution $p(x; \theta)$, in which there is a single variate x and a single parameter θ.

Example 1. $p(x, \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}$, the normal distribution in which the center is at $x = \theta$, and the standard deviation is unity.

Let a random sample E be drawn from a population indicated by such a curve. In the simplest case $E = x$, a single individual. Shortly, we shall have to suppose that there are N individuals: $E = x_1, \cdots, x_N$. Eventually, the



Univariate Distributions

picture will be generalized much further. The population will be described by a function of n variables, so that, in place of each x of our sample, we shall have

[1] One of two papers read by Cecil C. Craig and by the author at a joint meeting of the Institute of Mathematical Statistics, the Econometric Society and the American Statistical Association, held in New York City on December 30, 1941.

$x^{(1)}, \cdots, x^{(n)}$; moreover there will be, not one parameter, but l parameters $\theta^{(1)}, \cdots, \theta^{(l)}$; so that our probability distribution will be multivariate and will be denoted by

$$p(x^{(1)}, \cdots, x^{(n)}; \theta^{(1)}, \cdots, \theta^{(l)}).$$

A common way of putting this is to say that x and θ are vectors in n and l dimensions, respectively, and to leave the form as originally, $p(x; \theta)$. In the figure the space which the samples ($E = x$) can occupy is of course not more than the x-axis, but in the most general case the sample space will be a part or all of a space of nN dimensions and will be denoted by W. As is well understood, a significance test is an inequality which specifies in W a certain region w as a critical region, and if E is in this w, the hypothesis being tested is rejected. For example, in the figure, one might test the hypothesis H_0 that $\theta = \theta_0$. The rejection region w_0 might be the part of the x-axis where $x > x_0$. In all such cases we shall let α equal the probability that E is in w_0 if $\theta = \theta_0$. This statement will be denoted as follows:

$$\alpha = P(w_0 \mid \theta_0), \tag{1}$$

P standing for probability.

(*i*) *Power of a test.* A good test should satisfy two conditions: (*a*) if our sample is drawn from the population specified by θ_0, the hypothesis H_0 that $\theta = \theta_0$ should be accepted as often as possible, and (*b*) if our sample is drawn from a population specified by some other value of θ, say θ_1, then the hypothesis that $\theta = \theta_1$ should also be accepted as often as possible. Suppose first that there are but these two admissible populations. The probability of (*a*) is $1 - \alpha$. We commonly make the artificial requirement that this shall be some larger fraction such as 0.99. The probability of (*b*) is commonly denoted by β, and in the figure, when $w = w_0$, β is the area under the θ_1 curve which lies to the right of $x = x_0$. Relative to θ_0, θ_1, and α, the quantity β is called the *power* of that test which designates w_0 as the critical region. Also, α and $(1 - \beta)$ are the probabilities of the so-called errors of the first and second kinds, respectively.

(*ii*) *Unbiased test.* As stated, we would like to have β large. In any case we would like to have $\beta \geqq \alpha$. If $\beta \geqq \alpha$, the test and the corresponding region w_0 are "unbiased" (relative to the preassigned quantities θ_0, θ_1, and α). The region w_0 appears to be unbiased in our figure. This definition can obviously be extended to the case where, in addition to θ_1, there is an infinity of admissible values of θ; then the test is unbiased relative to the whole family of admissible values of θ if, for every one of these θ's, $\beta \geqq \alpha$.

(*iii*) *UMP test and CBC region.* If, with respect to a family of admissible θ's, a critical region w_0 exists such that, for each of these θ's ($\neq \theta_0$), β is greater than it would be for any other critical region satisfying (1), then this w_0 is said to be the common best critical (CBC) region and the corresponding test is the uniformly most powerful (UMP) test.

(*iv*) *UMPU test and CBCU region.* If there is no CBC region, still it may happen that, if one restricts one's view to only unbiased regions, there may be among them a CBC region. Such a region is said to be a common best critical unbiased (CBCU) region, and the corresponding test is the uniformly most powerful unbiased (UMPU) test.

In the following examples, and elsewhere, we shall now use H_0 to indicate the hypothesis being tested, H^* to indicate all admissible alternatives.

Example 2: $p(x, \theta)$ normal as in Example 1, $E = x$, $H_0 : \theta = \theta_0$, $H^*: \theta > \theta_0$. The CBC region is where $x > k$ if

$$\int_k^\infty p(x; \theta_0)\, dx = \alpha.$$

This region is the interval indicated by w_0 in the figure.

Example 3: Same as the preceding example except that now we have as $H^*: \theta \neq \theta_0$. There is no CBC region, but the CBCU region consists of two tail intervals, where $|x| > k$ if

$$\int_k^\infty p(x, \theta_0)\, dx = \tfrac{1}{2}\alpha.$$

A little reflection will convince the reader that the statements in these two examples are at least apparently true. It is geometrically evident, for example, that the last mentioned region (two tail intervals) is not as powerful with respect to the alternatives of Example 1 ($\theta > \theta_0$) as is the single tail region w_0 in the figure.

(*v*) *Type A regions.* It is often difficult to find even a CBCU region, or such a region may not exist, but it may be that there is a region which has the required properties if one admits only values of θ near to the value θ_0 being tested. Type A regions have this property. More precisely, they have the property that the power of w_0 is a minimum at θ_0 with respect to small changes in θ, and that this is a sharper minimum at θ_0 than is the power of any other w_0 which satisfies equation (1). Here the words "small changes" are used as in the calculus. The full definition [4] of an unbiased region of type A is that it shall satisfy (1) and also the following conditions:

(2) θ shall be a single parameter (not a vector),

(3) $\frac{d}{d\theta} P(w_0 \mid \theta) = 0$ if $\theta = \theta_0$,

(4) $\frac{d^2}{d\theta^2} P(w_0 \mid \theta) \geqq \frac{d^2}{d\theta^2} P(w \mid \theta)$ when $\theta = \theta_0$ for all regions w which satisfy the preceding conditions imposed on w_0. There are also other types of regions designated by A_1, B, C, and D, which resemble Type A [9]. The following example illustrates Type A; it is a familiar problem with an unfamiliar solution [4].

Example 4. $p(x; \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$; $E = x_1, \cdots, x_N$; $H_0 : \sigma = \sigma_0$;

H^*: $\sigma \neq \sigma_0$. The CBCU region of type A is determined by two tail areas (but they are not equal tail areas) of the distribution of Σx_i^2.

(*vi*) *Test unbiased in the limit* [11]. (*vii*) *Asymptotically MP test* [15]. (*viii*) *Asymptotically MPU test* [15]. In these cases the complete definitions are too lengthy to be repeated here, and they cannot be recapitulated briefly. The general idea is that, if none of the regions of the preceding types exist, still it may be true that there are regions which do have approximately the desired properties if $E = x_1, \cdots, x_N$, and N is large. The following example [11] illustrates (*vi*).

Example 5. $p(x;\theta) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}$; $E = x_1, \cdots, x_N$; $H_0 : \theta = 0$; H^*: $\theta \neq 0$. Regions of Type A unbiased in the limit are defined by the inequality,

$$4\sum_i \frac{x_i^2}{(1+x_i^2)^2} - 2\sum_i \frac{1}{1+x_i^2} + 4\left(\sum \frac{x_i}{1+x_i^2}\right)^2 \geq M\sqrt{\frac{5N}{3}} - \frac{N}{2}.$$

Here M is a quantity that has to be approximated and tabulated. The inequality is not simple, but it furnishes a definite answer to the problem.

(*ix*) *Regions similar to sample space.* All the preceding definitions apply to the case where x is a vector in n space, but not all to the case where θ is a vector in l space. Suppose now that this is the case, or, as we have said before, that there are l different parameters $\theta^{(1)}, \cdots, \theta^{(l)}$, each being capable of taking on a variety of values. Suppose we fix our attention on $\theta^{(1)}$ and wish to test the hypothesis that $\theta^{(1)} = \theta_0^{(1)}$. First of all we wish to find a critical region w_0 for which an equation like (1) will be true, independently of what the values of the other parameters may be. Such a region is said to be "similar" to sample space; the "similarity" consists in the fact that the equation like (1) would be true independently of the other parameters, if w_0 were replaced by all of sample space W, and if $\alpha = 1$. Feller [10] has shown that there are simple cases in which there is no region similar to sample space. He and others have investigated the conditions under which such regions do exist. "Generally speaking it seems that for most of the probability laws $p(x, \theta^{(1)}, \cdots, \theta^{(l)})$ in which the composite probability law for sample space is made up by multiplication,

$$\prod_{i=1}^{N} (p(x_i) \mid \theta^{(1)}, \cdots, \theta^{(l)}), \tag{2}$$

there do exist such similar regions, at least if $N > l$."

Part II. Estimation. (*i*) *Estimation by interval.* So far we have been considering possible answers to the question: Shall specified values of $\theta^{(1)}, \cdots, \theta^{(l)}$ be accepted? The totality of values of the θ's which are so acceptable might be called the acceptable point set in parameter (l-dimensional) space. This point set is determined by the sample or experiment (E), and usually different point sets are determined by different E's. Frequently this set of points constitutes a simple closed region, or, in the case of only one parameter, it may be a single interval. Such an interval is called a fiducial or confidence interval. The fundamental property of such a point set or interval is well known, but has to be stated with some care: If $\alpha = 0.01$, and if one is about to take a sample from a population in which the true values of the parameters $\theta^{(1)}, \cdots, \theta^{(l)}$ are $\theta_0^{(1)}, \cdots, \theta_0^{(l)}$, then the probability is 0.99 that the sample will be such that the point set determined by it will contain this true parameter point $\theta_0^{(1)}, \cdots, \theta_0^{(l)}$. It does not matter whether or not one knows what these true values of the parameters are. If there is more than one parameter, the fiducial interval for one of these parameters often does not exist; that is, there is often no such interval which is independent of the values of the other parameters. The question whether there is such an interval is obviously connected with the question whether there are regions similar to sample space. But if one fiducial interval does exist, then usually there are an infinity of them, and our problem is to choose the best one. This problem is called "estimation by interval." One answer is to choose the shortest interval. More precisely one should say, the shortest system of intervals. One gets a system of intervals by fixing α but not E. What is desired is a formula which will give the shortest interval for every E, but it may well happen that one formula (system) will supply the shortest intervals for some E's, and another will supply the shortest intervals for other E's. The choice between the two systems will then depend on the relative frequency with which the shortest intervals will be supplied by one system or by the other.

Example 6: $p(x; \xi, \sigma)$ is normal, ξ indicating the mean and σ the standard deviation. Given $E = x_1, \cdots, x_N$; to estimate ξ. The shortest system of confidence intervals does not exist (independently of σ).

Example 7. Same as Example 6, except that now one seeks only an upper limit to the confidence interval which the parameter must not exceed. Then the shortest system (best one-sided estimate) is: $\xi \leqq \bar{x} + ts$, where Fisher's t and s are meant; t corresponds to a preassigned α, and $\bar{x}$ is the mean of the sample.

In cases like Example 6, where the shortest system does not exist, Neyman [7] defines a "short unbiased system."

Example 8. The short unbiased system for Example 6 is: $\bar{x} - ts \leqq \xi \leqq \bar{x} + ts$, $(t, s, \bar{x})$ as in Example 7.

(*ii*) *Single estimators.* Suppose that, as before, we have a sample (E) and wish to choose the best single value for one of the parameters, not as before its best fiducial interval. It is well known that there often exists a fiducial function $g(\theta)$ which, like a probability function, is everywhere positive or zero and has an integral,

$$\int_{-\infty}^{\infty} g(\theta)\, d\theta = 1,$$

and is further useful in determining confidence intervals. In particular, if θ is a location parameter and if the composite probability function is as in (2), with only one parameter θ: $g(\theta) = kp(x_1 - \theta) \cdots p(x_N - \theta)$, k being a constant. An estimate commonly thought of as best is the maximum likelihood estimate: this is the mode of $g(\theta)$. Other estimates that have interesting properties are the mean and the median of $g(\theta)$. Pitman [14] defines a new "best" estimate θ_B. This has the property that, for every $h > 0$, θ_B is within h of the true value θ more often than is any other estimate. More precisely, if

$$P(|\theta_B - \theta| \leqq h) \geqq P(|\theta_1 - \theta| \leqq h)$$

for all positive values of h, and if the *inequality* sign between the P's holds for some positive value of h, θ_1 being every other estimate, then θ_B is the "best" estimate. As before P stands for probability.

Example 9. If $p(x; \xi, \sigma)$ is normal and the sample $E = x_1, \cdots, x_N$, the "best" estimate of σ^2 is $\frac{\Sigma x_i^2}{N - \frac{2}{3}}$, instead of the usual estimates: $\frac{\Sigma x_i^2}{N - 1}, \frac{\Sigma x_i^2}{N}$.

(*iii*) *Weight function.* Wald [13] defines a weight function $V(\theta, \theta_E)$ which depends on the seriousness of the error committed when the estimate θ_E is used in place of the true value of the parameter θ. The sample $E = x_1, \cdots, x_N$; and θ may be a vector. Thence he defines a risk function,

$$r(\theta) = \int_W V \cdot p(x_1, \cdots, x_N \mid \theta)\, dW,$$

and the "best" θ_E as that value of θ which minimizes the total risk,

$$\int Vp\, df(\theta),$$

this integral being taken over all of the parameter space, and $f(\theta)$ being the a priori distribution of θ. It is undesirable to introduce $f(\theta)$, but it can be shown that, subject to slight restrictions on the nature of f, one can obtain a best estimate by finding a value θ_E which for all θ's makes r equal to a constant and also satisfies other general conditions; this equation and these conditions do not contain $f(\theta)$. In a symmetrical but otherwise fairly general case θ_E is the maximum likelihood solution.

Part III. Likelihood Tests. This part has to do mostly with special cases of likelihood tests. As is well known, this test consists in selecting a critical rejection region w in sample space where

(*a*) $P(w \mid H_0) = \alpha$,

(*b*) the relative likelihood of H_0 is small; more precisely, where $\lambda <$ constant, and

$$\lambda = \frac{\max_\omega P(E \mid \omega)}{\max_\Omega P(E \mid \Omega)},$$

ω being the region in parameter space specified by the hypothesis tested H_0, and Ω being the region in parameter space specified by all admissible hypotheses. (In special cases *max* is replaced by *least upper bound.*) If H_0 is simple (ω being a point) and if the CBC region w exists, then w is bounded by the contour,

λ = constant [19]. Otherwise this λ test does not necessarily yield the same critical regions as do any of the preceding tests. But it is generally much easier to apply, and, in many of the cases that follow, these λ tests are good ones as judged by the preceding theory. They are powerful even if they are not the most powerful of all tests, and often this power can be found and tabulated. In fact Wilks [28] has shown that the appropriate distribution of λ (omitting terms of order $1/\sqrt{N}$) can be found[2] if the distribution of E is

$$\prod_{i=1}^{N} p(x_i, \theta^{(1)}, \cdots, \theta^{(l)}), \qquad (N \text{ large})$$

and[3] if the optimum estimates $\hat{\theta}^{(1)}, \cdots, \hat{\theta}^{(l)}$ exist and are distributed (except for certain terms of order $1/\sqrt{N}$) normally. This theorem has now been generalized by Wald, in a paper presented to the American Mathematical Society in December, 1941.

There are many of these tests, made to fit all sorts of hypotheses. The author will try to summarize a considerable group of them; all members of this group might be called generalizations of the Student-Fisher t-test. They fall naturally into two classes, according as to whether the individuals of the sample are taken from a univariate or from a multivariate universe. Unless otherwise stated all universes shall be normal. H_0 shall stand for the hypothesis being tested, and H^* for all admissible alternatives to H_0.

(*i*) *Univariate case.* The sample consists of N elements, as before, $x_1, \cdots, x_N$, chosen independently from N normal populations indicated by their parameters $(\xi_1, \sigma_1), \cdots, (\xi_N, \sigma_N)$. About these populations we may ask a variety of questions resulting in a variety of problems and tests.

Problem a: If the populations are all identical (ξ, σ), does $\xi = \xi_0$ (specified in advance)? This results in the well-known t-test. The hypothesis tested H_0 is that $\xi = \xi_0$, and the alternative hypothesis H^* is that $\xi \neq \xi_0$; it being assumed at the outset that all the populations are identical. The t-test has been shown to be an UMPU test relative to H^*.

Problems b, c, d: Let these same samples be arranged in k groups or "columns"

$$\begin{matrix} x_1^{(1)} & \cdots & x_1^{(k)} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ x_{n_1}^{(1)} & \cdots & x_{n_k}^{(k)} \end{matrix}$$

where the n_i are not necessarily all equal. Let it be assumed that the populations (ξ, σ) do not change within the columns. Problems (b), (c) and (d), with their corresponding tests, may be indicated as follows:

(b) Are (ξ, σ) constant from column to column? (The $\lambda_H = L$ test.)

[2] Distribution of $(-2 \log \lambda)$ is like that of χ^2 except for terms of order $1/\sqrt{N}$.

[3] See Doob's conditions, Transactions of American Mathematical Society, vol. 36 (1934), pp. 759–775.

(*c*) Is σ constant from column to column regardless of what values the ξ's may have? (The $\lambda_{H_1} = L_1$ test.)

(*d*) Is ξ constant from column to column assuming the σ's constant? (The $\lambda_{H_2} = L_2$ test.)

In Problem (*b*), H_0 is that (ξ, σ) are constant, H^* that they are not constant. In Problem (*c*), H_0 is that σ is constant, H^* that it is not constant. In Problem (*d*), H_0 is that ξ is constant, H^* that it is not constant. The test of Problem (*c*) has recently been shown to be unbiased only if the numbers in all the columns are the same ($n_1 = \cdots = n_k$). It is, however, unbiased in the limit. Power tables were published in 1937 [23]. Bartlett's (1937) μ is another test for this problem, and Pitman's [36] L test is another, but it has been shown that these two tests are equivalent. Both are unbiased; they are not likelihood tests. This problem is frequently called the problem of the "homogeneity" of a set of variances.

All these tests are, of course, functions of the observations, and the details are readily available in the papers listed. For example, Pitman's

$$L = \tfrac{1}{2}N \log \frac{\Sigma S_i}{N/2} - \Sigma \left(n_i \log \frac{S_i}{n_i/2} \right),$$

where S_i is what he calls the "squariance" for the ith column, and a large value of L is significant. The squariance is what the physicists had called and what statisticians ought therefore to have called the second moment, *viz.*: $N\mu_2$; μ_2 is really the unit second moment.

(*e*) *Linear Hypothesis.* Problems like the above, and many others, can be included in a general theorem by Kolodziejczyk, who showed how to write out quite simply the likelihood test if each ξ is a linear function of l parameters ($l < N$) and if the hypothesis H_0 specifies the values of r different linear functions of the θ's ($r \leqq l$). Furthermore, the power of this test (with numerous applications) was discussed and tabulated by Tang in an important paper [39].

Problem (*f*). This method (*e*) has been used by Neyman [43] to test the homogeneity of a set of variances, the problem already studied by a number of authors. It has been stated that some of their tests were unbiased with respect to the alternative hypothesis that the σ's were not all equal. Neyman gives reasons for supposing, in the industrial problem he is considering, that it would be more realistic to consider another alternative hypothesis, namely, H^* that the σ's are not all equal and that their distribution can be approximately described by saying that $1/\sigma^2$ has a χ^2 distribution. No UMP test exists but there does exist a critical region whose power, with respect to a sub-family of H^* is independent of the means, and the corresponding test is the most powerful test for this sub-family of alternatives. Tables of its power are furnished. More applications are promised.

(*ii*) *Multivariate case.* The sample consists of N elements, exactly as before, except that now each x is a vector in n space and comes from a multivariate normal universe whose means may be represented again by ξ if we think of ξ as being a vector in n space. The other parameters of this universe are the variances and covariances α_{ij}. So, with these changes, we may repeat the statement at the beginning of (*i*) that the sample is $x_1, \cdots, x_N$, and that the populations are $(\xi_1, \alpha_{ij1}), \cdots, (\xi_N, \alpha_{ijN})$. The questions to be asked about these populations correspond exactly to those asked in the simpler case.

Problem (*a*): If the populations are all identical (ξ, α_{ij}), does $\xi = \xi_0$ (specified in advance)? The answer is given by Hotelling's T test. The hypothesis tested is H_0 that the vector $\xi = \xi_0$, and the alternative hypothesis H^* is that these two vectors are not identical. P. Hsu [28] has shown that this test is the most powerful in a special sense, and has given a new demonstration of it by the use of the Laplace transform. Incidentally he has shown that the Laplace transform of an elementary probability law determines the law uniquely except perhaps at a null set of points.

Problems (*b*), (*c*), (*d*): Now let the same sample be arranged in k groups or columns, as in (*i*) *b*, *c*, *d*; and let it be assumed that the populations (ξ, α_{ij}) do not change within the columns. Problems (*b*), (*c*), and (*d*), with their corresponding tests, may be indicated as follows:

(*b*) Are (ξ, α_{ij}) constant from column to column? (The $\lambda_{H(n)}$ test).

(*c*) Are α_{ij} constant from column to column regardless of what values the ξ's may have? (The $\lambda_{H(n')}$ test).

(*d*) Is the vector ξ constant from column to column assuming the α_{ij} constant from column to column? (The λ_H test).

Unfortunately, in the customary notation, the λ's for this case (*ii*) do not follow the pattern adopted in (*i*). It would be better to put (n) after each of the λ's (or L's) in (*i*) to signify the corresponding tests in (*ii*). But, even if this were agreed upon, there would still be a confused notation because there are many other "λ" and "L" tests besides those listed here. Apparently[4] the power functions of these last three multivariate tests have not been found yet.

(*e*) The linear hypothesis theory was shown to be applicable to the multivariate case in a special instance by P. Hsu in 1940 [38]. Since then he has generalized it further [45].

(*iii*) *Bivariate case.* This important special case of (*ii*) has now been pretty thoroughly solved. A general summary of various tests which have been devised by Finney, Pitman, Morgan, Wilks, and E. S. Pearson was given by C. Hsu in 1940 [42], with some slight additions and with tables of power functions with respect to certain alternatives. Altogether there are seven of these tests corresponding to seven different problems, including the four just referred to as Problems *a*, *b*, *c*, and *d*.

Part IV. The Method of Randomization. This part concerns randomization of the individuals within a sample to obtain a method of testing hypotheses without making use of any characteristic of the population from which the sample was drawn. It does not deal with randomization in field experi-

[4] So far as the author is aware; but he does not pretend to have made a careful search.

ments to off-set the effects of variable fertility. Also, in this discussion, the hypothesis being tested is not that the sample was a random sample. It is *assumed* that the given sample is random. We begin with an example from Pitman [46]. Two samples, $(x_1, \cdots, x_N)$ and $(y_1, \cdots, y_N)$, have been drawn at random from two populations. The means of the samples are $\bar{x}$ and $\bar{y}$, respectively. Let $|\bar{x} - \bar{y}|$ be called the spread of these samples. Now rearrange these same x's and y's with each other in all possible ways to obtain all possible spreads. The larger the observed spread, among all these possible spreads, the more significant it is supposed to be as a test of the (null) hypothesis that the two populations were identical. Similarly, tests have been devised for correlations, variances, etc.

E. S. Pearson [51] in 1938 published a criticism of this general theory which in substance seems to be that the reason why one calls the largest spreads significant, rather than the smallest ones, in the illustration just used, is that one is assuming tacitly that the admissible populations are such that large spreads would be more likely on some other than the null hypothesis; that if one does not make some such implicit assumption, then one might quite as well call the smallest spreads significant; and that therefore, barring such implicit assumptions, one can control only errors of the first kind by this method.

It seems to the author that Pearson's criticism is sound, and that, if indeed one is unwilling to make any assumption whatever about the populations considered, then this device is of no[5] value in testing the null hypothesis. For, if all that one pretends to do is to control errors of the first kind, one can do that by consulting a table of random numbers of two digits. Thus one can control errors of the first kind without performing the experiment at all, let alone making the long computations usually required by the method of randomization. Or, better, one can reduce that error to zero simply by making up one's mind that one will never reject the hypothesis being tested: certainly one will never reject it improperly if one never rejects it at all.

However, if one is willing to make in the illustration used the very mild assumption that the populations considered are such that unusually large spreads would more probably be obtained from some admissible hypothesis other than the null hypothesis, then it would seem to the author that the method would be useful. Similar remarks apply to the tests for correlations, variances, etc.

REFERENCES

PART I. THEORY OF TESTS AND PART II. ESTIMATION

[1] J. NEYMAN and E. S. PEARSON, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. R. Soc.*, A., Vol. 231 (1933), pp. 289–337.

[2] J. NEYMAN and E. S. PEARSON, "The testing of statistical hypotheses in relation to probabilities a priori," *Camb. Phil. Soc. Proc.*, Vol. 29 (1933), pp. 492–510.

[3] S. S. WILKS, "Test criteria for statistical hypotheses involving several variables," *Jour. Am. Stat. Asso.*, Vol. 30 (1935), pp. 549–560.

[4] J. NEYMAN and E. S. PEARSON, "Contributions to the theory of testing statistical

[5] Pearson's language is not so strong as this. He says "perhaps it should be described as a valuable device rather than a fundamental principle."

hypotheses. I. Unbiased critical regions of type A and type A_1," *Stat. Res. Mem.*, Univ. of London, Vol. 1 (1936), pp. 1–37.
[5] J. Neyman and E. S. Pearson, "Sufficient statistics and uniformly most powerful tests of statistical hypotheses," *Stat. Res. Mem.*, Vol. 1 (1936), pp. 113–137.
[6] E. J. G. Pitman, "The closest estimates of statistical parameters," *Camb. Phil. Soc. Proc.*, Vol. 33 (1937), pp. 212–222.
[7] J. Neyman, "Outline of a theory of estimation based on the classical theory of probability," *Phil. Trans. Roy. Soc.*, A, Vol. 236 (1937), pp. 333–381.
[8] J. Neyman, "On statistics the distribution of which are independent of the parameters involved in the probability law of the original variables," *Stat. Res. Mem.*, Vol. 2 (1938), pp. 58–59.
[9] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypotheses," *Stat. Res. Mem.*, Vol. 2 (1938), pp. 25–57.
[10] W. Feller, "Note on regions similar to the sample space," *Stat. Res. Mem.*, Vol. 2 (1938), pp. 117–125.
[11] J. Neyman, "Tests of statistical hypotheses which are unbiased in the limit," *Annals of Math. Stat.*, Vol. 9 (1938), 69–86.
[12] S. S. Wilks, "Fiducial distributions in fiducial inference," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 272–280.
[13] A. Wald, "Contributions to the theory of estimation and testing hypotheses," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 299–326.
[14] E. J. G. Pitman, "The estimation of the location and scale parameters of a continuous population of any given form," *Biometrika*, Vol. 30 (1939), pp. 391–421.
[15] A. Wald, "Asymptotically most powerful tests of statistical hypotheses," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 1–19.

Part III. Likelihood Tests—Special Cases

[16] H. Hotelling, "Generalized t-test," *Annals of Math. Stat.*, Vol. 2 (1931), pp. 360–378.
[17] S. S. Wilks, "Certain generalizations in analysis of variance," *Biometrika*, Vol. 24 (1932), pp. 471–494.
[18] E. S. Pearson and S. S. Wilks, "Methods of statistical analysis appropriate for k samples of two variates," *Biometrika*, Vol. 25 (1933), pp. 353–376.
[19] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy. Soc.*, A., Vol. 231 (1933), pp. 289–337.
[20] B. L. Welch, "Some problems in analysis of regression among k samples of two variables," *Biometrika*, Vol. 27 (1935), pp. 145–160.
[21] S. S. Wilks, "Test criteria for statistical hypotheses involving several variables," *Jour. Am. Stat. Asso.*, Vol. 30 (1935), pp. 549–560.
[22] P. P. N. Nayer, "An investigation into the application of Neyman and Pearson's L_1 test, with tables of percentage limits," *Stat. Res. Mem.*, Vol. 1 (1936), University College, London, pp. 38–56.
[23] S. S. Wilks and C. M. Thompson, "The sampling distribution of the criterion H when the hypothesis tested is not true," *Biometrika*, Vol. 29 (1937), pp. 124–132.
[24] D. J. Finney, "The distribution of the ratio of estimates of the two variances in a sample from a normal bivariate population," *Biometrika*, vol. 30 (1938), pp. 190–192.
[25] D. N. Lawley, "A generalization of Fisher's z-test," *Biometrika*, Vol. 30 (1938), pp. 180–187.
[26] P. L. Hsu, "Contribution to the theory of 'Student's' t-test as applied to the problem of two samples," *Stat. Res. Mem.*, Vol. 2 (1938), pp. 1–24.
[27] P. L. Hsu, "On the best unbiased quadratic estimate of the variance," *Stat. Res. Mem.*, Vol. 2 (1938), pp. 91–104.

[28] P. L. Hsu, "Notes on Hotelling's generalized T," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 231-244.

[29] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 60-63.

[30] F. N. David and J. Neyman, "Extension of the Markoff theorem on least squares," *Stat. Res. Mem.*, Vol. 2 (1938), pp. 105-116.

[31] D. J. Bishop and U. S. Nair, "A note on certain methods of testing for the homogeneity of a set of estimated variances," *Jour. Roy. Stat. Soc., Supp.*, Vol. 6 (1939), pp. 89-99.

[32] G. W. Brown, "On the power of the L_1 test for the equality of several variances," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 119-128.

[33] E. J. G. Pitman, "A note on normal correlation," *Biometrika*, Vol. 31 (1939), pp. 9-12.

[34] W. A. Morgan, "A test for the significance of the difference between the two variances in a sample from a normal bivariate population," *Biometrika*, Vol. 31 (1939), pp. 13-19.

[35] D. J. Bishop, "On a comprehensive test for the homogeneity of variances and covariances in multivariate problems," *Biometrika*, Vol. 31 (1939), pp. 31-55.

[36] E. J. G. Pitman, "Tests of hypotheses concerning location and scale parameters," *Biometrika*, Vol. 31 (1939), pp. 200-215.

[37] P. O. Johnson and J. Neyman, "Tests of certain linear hypotheses and their application to some educational problems," *Stat. Res. Mem.*, Vol. 1 (1939), pp. 57-93.

[38] P. L. Hsu, "On generalized analysis of variance, I," *Biometrika*, Vol. 31 (1940), pp. 221-237.

[39] P. C. Tang, "The power function of analysis of variance tests with tables and illustrations of their use," *Stat. Res. Mem.*, Vol. 2 (1938), pp. 126-149.

[40] H. O. Hartley, "Testing the homogeneity of a set of variances," *Biometrika*, Vol. 31 (1940), pp. 249-255.

[41] J. F. Daly, "On the unbiased character of likelihood ratio tests for independence in normal systems," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 1-32.

[42] C. T. Hsu, "On samples from a normal bivariate population," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 410-426.

[43] J. Neyman, "A statistical problem arising in routine analysis and in sampling inspections of mass production," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 46-76.

[44] A. Wald and R. J. Brookner, "On the distribution of Wilks' statistic for testing the independence of several groups of variates," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 137-152.

[45] P. L. Hsu, "Canonical reduction of the general regression problem," *Annals of Eugenics*, Vol. 11 (1941), pp. 42-46.

Part IV. Randomization Tests

[46] E. J. G. Pitman, "Significance tests which may be applied to samples from any population, (I)," *Jour. Roy. Stat. Soc. Supp.*, Vol. 4 (1937), pp. 119-130.

[47] E. J. G. Pitman, "Significance tests which may be applied to samples from any population, (II), The correlation coefficient test," *Jour. Roy. Stat. Soc. Supp.*, Vol. 4 (1937), pp. 225-232.

[48] E. J. G. Pitman, "Significance tests which may be applied to samples from any population, (III), The analysis of variance test," *Biometrika*, Vol. 29 (1938), pp. 322-335.

[49] B. L. Welch, "On the z-test in randomized blocks and Latin squares," *Biometrika*, Vol. 29 (1937), pp. 26-52.

[50] B. L. Welch, "On tests for homogeneity," *Biometrika*, Vol. 30 (1938), pp. 149-158.

[51] E. S. Pearson, "Some aspects of the problem of randomization," *Biometrika*, Vol. 29 (1938), pp. 53-64.

Note: None of the 1941 *Biometrika* was received until after this paper had been read and prepared for publication.