ON THE NON-EXISTENCE OF TESTS OF "STUDENT'S" HYPOTHESIS HAVING POWER FUNCTIONS INDEPENDENT OF σ

By George B. Dantzig

1. Introduction. Consider a system of n random variables $x_1, x_2, \cdots, x_n$ where each is known to be normally distributed about the same but unknown mean, ξ, and with the same, but also unknown standard deviation σ. The assumption, H_0, that ξ has some specified value, ξ_0, e.g. $\xi_0 = 0$, while nothing is assumed about σ, is known as the "Student" Hypothesis. Two aspects of the hypothesis H_0 have been already studied extensively. If the alternatives with respect to which it is desired to test H_0 assume specifically that $\xi > \xi_0$, (or $\xi < 0$), then we have the so-called asymmetric case of "Student's Hypothesis" and it is known, [1], that there exists a uniformly most powerful test of H_0. This consists in the rule, originally suggested by "Student," of rejecting H_0 whenever

$$t = \frac{\bar{x} - \xi_0}{S}\sqrt{n-1} > t_\alpha, \tag{1}$$

where $\bar{x}$ and S denote the mean and the standard deviation of the observed x_i's and t_α is taken, for example, from Fisher's Tables [2] with his $P = 2\alpha$. In other words t_α is such that

$$P\{t > t_\alpha \mid H_0\} = \alpha, \tag{2}$$

where α is the chosen level of significance. In accordance with the definition of the uniformly most powerful test, whenever any other rule, R, offered to test the same hypothesis H_0 has the same probability α of H_0 being rejected when it is true, the power of this alternative test cannot exceed that of "Student's" Test. In other words, if it happens that the true value of ξ is not equal to ξ_0 but is greater, then the probability of this circumstance being detected by "Student's" test is at least equal to that corresponding to the rule R.

If the set of alternative hypotheses is not limited to those specifying the value of ξ either greater or smaller than ξ_0, but includes both those categories, then it is known, [1], that there is no uniformly most powerful test of the hypothesis, H_0. However in this case there exists a slightly different test, also based on "Student's" criterion t, possessing the remarkable property of being unbiased of type B_1, [3]. The test, in common use for a long time, consists in rejecting H_0 when

$$|t| > t_\alpha, \tag{3}$$

with t_α being taken again from Fisher's tables, this time corresponding to his $P = \alpha$, where α is the chosen level of significance.

In order to describe the optimum property of this test we must use the concept of the power function of a test, [3]. Denote by $\beta(\xi, \sigma)$ the probability of the hypothesis H_0 being rejected when ξ and σ are the true mean and the true standard error of the observable x_i's. The function $\beta(\xi, \sigma)$ is just what is called the power function of the test. If we substitute $\xi = \xi_0$, then we shall have $\beta(\xi_0, \sigma) = \alpha$ irrespective of the value of σ. Now the optimum property of "Student's" test mentioned above consists in that (1) its power function has a minimum at $\xi = \xi_0$ and this is true whatever be the value of σ, (2) whatever be any other test of the same hypothesis which has the same level of significance α and has property (1), its power function $\beta'(\xi, \sigma)$ cannot exceed that of "Student's" test.

These two properties, demonstrating the excellence of the criterion suggested by "Student," fully justify the general confidence in the test as described above, or in its extended form where it is applied to two or more samples. However, it is known that "Student's" test in both its forms, $t > t_\alpha$, and $| t | > t_\alpha$, has one very undesirable property which causes great difficulties in various problems of rational planning of experiments.

One of the most important questions to have in mind when planning an experiment is: What is the probability that the experiment and the subsequent statistical test will detect a difference or effect when it actually exists? If we perform an experiment and then apply some statistical analysis to test "Student's" hypothesis that $\xi = \xi_0$, we do hope that, if the actual value of ξ is different from ξ_0, the test will discover this circumstance. But apart from mere hope, it is desirable to take precautions so that when the *difference*, $\xi - \xi_0 = \Delta$, has some appreciable value, the chance of the hypothesis H_0 being rejected will be reasonably large. This may be done by calculating the value of the power function $\beta(\xi, \sigma)$ corresponding to the value $\xi = \xi_0 + \Delta$. And here we come to the unfortunate property of "Student's" test.

Although the form of the power function of "Student's" test is known and tabled [4], [5], [6], [7], there are occasionally considerable difficulties in applying these tables, because it appears that the values n and Δ are not all its arguments, for it also depends on σ. Consequently in order to have an idea of the probability that the test will detect the falsehood of the hypothesis H_0 that $\xi = \xi_0$ when actually $\xi = \xi_0 + \Delta$ we need not only the knowledge of n but also a likely value of σ. The latter is known accurately only in exceptional cases and then in those cases one would apply a test which is different from "Student's" test. Usually we have only a vague notion of the magnitude of σ and accordingly the tables of $\beta(\xi, \sigma)$ may be used to obtain a rough idea as to whether the arrangement of the experiment planned is satisfactory or not. Frequently we have no idea of what may be the values of σ.

To Dr. P. L. Hsu is due the idea of looking for tests, the power of which is independent of the parameters unspecified by the hypothesis tested. In an unpublished paper, he proved among other things that the λ test of the general *linear hypothesis* is the most powerful of all those, the power function of which depends on the same argument as that of the λ test and not on other parameters. The above circumstances suggest the following problem: to see whether it is possible to devise a test of "Student's" hypothesis such that its power function would be independent of σ. If such a test could be devised and proved to be reasonably powerful then the tables of its power function could be used for the purpose of planning experiments.

The purpose of the present paper is to show that no such test exists and, consequently, this negative result implies in still another way that it is impossible to improve on the test originally suggested by "Student."

2. Statement of the Problem. The problem of finding a test whose power function is independent of σ is equivalent to finding a critical region w such that the value of the power function

$$\beta(\xi, \sigma) = P\{E \epsilon w \mid \xi, \sigma\} \tag{4}$$

for any fixed ξ is independent of the value of σ, where E denotes the sample point $(x_1, x_2, \cdots x_n)$. We shall show specifically that if this is the case, then the power function is also independent of ξ; so that the test will reject the hypothesis tested with the same frequency independently of whether it be correct or wrong.

3. Theorem. *If there exists a region w such that, whatever be the value of σ,*

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \int \cdots \int_w e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\xi_0)^2}\, dx_1\, dx_2 \cdots dx_n \equiv \alpha \tag{5}$$

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \int \cdots \int_w e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\xi_1)^2}\, dx_1\, dx_2 \cdots dx_n \equiv \beta, \tag{6}$$

where $\xi_0 \neq \xi_1$, α, β are constants, then

$$\alpha = \beta. \tag{7}$$

A region w is called *similar* [1] to the whole sample space, W, of size α, with respect to a set of elementary probability laws $p(E \mid \theta)$ given in terms of a parameter θ, if $P\{E \epsilon w \mid \theta\} = \alpha$, whatever be the value of θ. Essentially, then, the region, w, above is a similar region with respect to two different sets of elementary laws each being given parametrically in terms of the parameter σ. Denote by w_r the portion of the surface of the hypersphere, $\sum_{i=1}^{n}(x_i - \xi_0)^2 = r^2$, which is common to w, and let the total surface be denoted by W_r. Neyman and Pearson have shown [1], that a necessary and sufficient condition that w be a similar region, in the above case, is that, whatever be r, the probability that the sample point E will fall on the subsurface w_r, when it is known that the sample point lies on the surface W_r is α, i.e.

$$P\{E \epsilon w_r \mid (E \epsilon W_r)(\xi = \xi_0)\} = \alpha \tag{8}$$

for all r.

In a similar manner let w_ρ denote the portion of the surface of the hypersphere $\sum_{i=1}^{n} (x_i - \xi_1)^2 = \rho^2$ common to w, and let the total surface be denoted by W_ρ. Since w is similar to the set of probability laws indicated in (6), we have also

$$P\{E \epsilon w_\rho \mid (E \epsilon W_\rho)(\xi = \xi_1)\} = \beta \tag{9}$$

for all ρ.

Since on the surface W_r, the elementary probability law,

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\xi_0)^2} = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{r^2}{2\sigma^2}}, \tag{10}$$

is constant, we see that an equivalent statement of (8) is that *the hyper-area of w_r is a constant proportion, α, of the total hyper-area W_r.* Similarly, from (9), we have that *the hyper-area of w_ρ is a constant proportion, β, of the area of the hypersurface W_ρ, whatever be the values of r and ρ.*

Consider the transformation which expresses $x_1, x_2, \cdots x_n$ in terms of generalized polar coordinates with pole at the point $(\xi_0, \xi_0, \cdots, \xi_0)$, i.e.

$$\begin{aligned}
x_1 - \xi_0 &= r \cos\theta_2 \cos\theta_3 \cdots \cos\theta_{n-2} \cos\theta_{n-1} \cos\theta_n \\
x_2 - \xi_0 &= r \cos\theta_2 \cos\theta_3 \cdots \cos\theta_{n-2} \cos\theta_{n-1} \sin\theta_n \\
x_3 - \xi_0 &= r \cos\theta_2 \cos\theta_3 \cdots \cos\theta_{n-2} \sin\theta_{n-1} \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
x_{n-1} - \xi_0 &= r \cos\theta_2 \sin\theta_3 \\
x_n - \xi_0 &= r \sin\theta_2
\end{aligned} \tag{11}$$

Let Δ be the Jacobian of the transformation:

$$|\Delta| = r^{n-1}\left|\prod_{i=2}^{n} \cos^i \theta_{n+2-i}\right| = r^{n-1} T(\theta_i). \tag{12}$$

Consider also a transformation which expresses $(x_1, x_2, \cdots x_n)$ in terms of polar coordinates, the point $(\xi_1, \xi_1, \cdots, \xi_1)$ being pole. It may be obtained by replacing in (11), ξ_0 by ξ_1, r by ρ, and θ_i by $\bar\theta_i$. The Jacobian of this transformation is given by $|\bar\Delta| = \rho^{n-1} T(\bar\theta_i)$.

We are now able to express the *hyper-area* of W_r:

$$\iint_{W_r} |\Delta|\, d\theta_2\, d\theta_3 \cdots d\theta_n = r^{n-1} \iint_{W_r} T(\theta_i)\, d\theta_2\, d\theta_3 \cdots d\theta_n = K r^{n-1}, \tag{13}$$

where the integral $K > 0$ is a constant independent of r. Similarly the hyper-area of W_ρ is $K\rho^{n-1}$, where K is the same as in (13). According to (8) and (9) we have, now

$$\iint_{w_r} |\Delta| \, d\theta_2 \, d\theta_3 \cdots d\theta_n = \alpha \cdot K \cdot r^{n-1}, \tag{14}$$

$$\iint_{w_\rho} |\bar{\Delta}| \, d\bar{\theta}_2 \, d\bar{\theta}_3 \cdots d\bar{\theta}_n = \beta \cdot K \cdot \rho^{n-1}. \tag{15}$$

Let us consider the distances between the three points: $(x_1, x_2, \cdots, x_n)$, $(\xi_0, \xi_0, \cdots, \xi_0)$, and $(\xi_1, \xi_1, \cdots, \xi_1)$. The distances of the first point to the second point and to the third point we have already denoted by r and ρ. Let the distance between last two be L, then, since the sum of two sides is at least equal to the third side of a triangle, we have

$$r \leqq \rho + L, \qquad \rho \leqq r + L, \quad \text{where} \quad L = \sqrt{N} \, | \xi_0 - \xi_1 |. \tag{16}$$

Let $\varphi(t) \geqq 0$ be an *arbitrary* monotonic nonincreasing function of t, such that the product $t^{n-1}\varphi(t)$ is integrable from 0 to $+\infty$. Since $\varphi(t)$ is a decreasing function it follows from (16) that

$$\varphi(r) \geqq \varphi(\rho + L) \quad \text{and} \quad \varphi(\rho) \geqq \varphi(r + L). \tag{17}$$

Consider the integral I:

$$I = \iint_w \varphi(r) \, dx_1 \, dx_2 \cdots dx_n. \tag{18}$$

We shall express it in terms of the variables $r, \theta_2, \cdots, \theta_n$ and also in terms of $\rho, \bar{\theta}_2, \cdots \bar{\theta}_n$ and compare the results. Thus

$$\begin{aligned} I &= \iint_w |\Delta| \, \varphi(r) \, dr \, d\theta_2 \cdots d\theta_n \\ &= \int_0^\infty \varphi(r) \, dr \iint_{w_r} |\Delta| \, d\theta_2 \cdots d\theta_n \\ &= \alpha \cdot K \cdot \int_0^\infty r^{n-1} \varphi(r) \, dr. \end{aligned} \tag{19}$$

Also we have by (16)

$$I = \iint_w |\bar{\Delta}| \, \varphi(r) \, d\rho \, d\bar{\theta}_2 \cdots d\bar{\theta}_n$$

$$\geqq \iint_{w} |\bar{\Delta}| \varphi(\rho + L)\, d\rho\, d\bar{\theta}_2 \cdots d\bar{\theta}_n \tag{20}$$

$$\geqq \int_0^\infty \varphi(\rho + L)\, d\rho \iint_{w_\rho} |\bar{\Delta}|\, d\bar{\theta}_2 \cdots d\bar{\theta}_n$$

and consequently

$$I \geqq \beta \cdot K \int_0^\infty \rho^{n-1} \varphi(\rho + L)\, d\rho. \tag{21}$$

Since $K > 0$, we have from (19) and (21)

$$\alpha/\beta \geqq \int_0^\infty t^{n-1} \varphi(t + L)\, dt \Big/ \int_0^\infty t^{n-1} \varphi(t)\, dt. \tag{22}$$

By interchanging ρ and r in (18), (19), (20), and (21) we have also

$$\beta/\alpha \geqq \int_0^\infty t^{n-1} \varphi(t + L)\, dt \Big/ \int_0^\infty t^{n-1} \varphi(t)\, dt. \tag{23}$$

Let us set in (22) and (23), $\varphi(t) = e^{-pt}$ and $\varphi(t + L) = e^{-pL}e^{-pt}$ where $p > 0$ is arbitrary. Then

$$\alpha/\beta \geqq e^{-pL} \quad \text{and} \quad \beta/\alpha \geqq e^{-pL}. \tag{24}$$

Since (24) holds for all $p > 0$, let p approach zero. Then $\text{Lim}\, e^{-pL} = 1$, and the above inequalities can hold only if

$$\alpha = \beta, \qquad \text{Q.E.D.} \tag{25}$$

It is of interest to note that there do exist regions such that the power function is independent of both ξ and σ. For example, let S_n be the standard deviation of the observed values $(x_1, x_2, \cdots, x_n)$ and let S_{n-1} be the standard deviation of the values $(x_1, x_2, \cdots, x_{n-1})$, then the region w given by all points $(x_1, x_2, \cdots x_n)$ which satisfy the inequality $(S_{n-1}/S_n) \geqq C$ is such a region, i.e.

$$P\{(S_{n-1}/S_n) \geqq C \mid \xi, \sigma\} \tag{26}$$

is constant, whatever be the values of ξ and σ. Such regions are, however, unsuitable for testing "Student's" hypothesis $\xi = \xi_0$, because they will reject this hypothesis when it is wrong and when it is correct with equal frequency.

The author is indebted to Professor J. Neyman for assistance in preparing the present paper.

REFERENCES

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

[2] R. A. Fisher, *Statistical methods for research workers*. Oliver & Boyd, 7th edition, London, 1938.

[3] J. Neyman, "Sur la vérification des hypothèses statistiques composées," *Bull. Soc. Math. de France*, T. 63 (1935), pp. 246–266.

[4] S. Kolodziejczyk, "Sur l'erreur de la deuxième catégorie dans le problème de 'Student,' " *Compte Rendus*. T. 197(1933), p. 814.

[5] J. Neyman with co-operation of K. Iwaszkiewicz and S. Kolodziejczyk, "Statistical problems in agricultural experimentation," *Suppl. Jour. Roy. Stat. Soc.*, Vol. II(1935), pp. 107–80.

[6] J. Neyman and B. Tokarska, "Errors of the second kind in testing 'Student's' hypothesis," *Jour. Am. Stat. Ass.*, Vol. 31(1936), pp. 318–26.

[7] P. C. Tang, "The power function of the analysis of variance tests with tables and illustrations of their use," *Stat. Res. Memoirs*, Vol. II(1938), pp. 126–57.

University of California,
Berkeley, California.