THE POWER OF THE CLASSICAL TESTS ASSOCIATED WITH THE NORMAL DISTRIBUTION

By J. Wolfowitz

Columbia University

Summary. The present paper is concerned with the power function of the classical tests associated with the normal distribution. Proofs of Hsu, Simaika, and Wald are simplified in a general manner applicable to other tests involving the normal distribution. The set theoretic structure of several tests is characterized. A simple proof of the stringency of the classical test of a linear hypothesis is given.

1. Introduction. The present paper is concerned with the optimum properties, from the power function viewpoint, of the classical tests associated with the normal distribution. In 1941 Hsu [2] proved the result stated in Section 2 below, which is concerned with the general linear hypothesis (in this connection his paper [1] of 1938 will be of interest). Also in 1941 Simaika [3] proved similar results for the tests based on the multiple correlation coefficient and Hotelling's generalization of Student's t. In 1942, Wald [4] gave a generalization of Hsu's result.

In the present paper we give short and simple proofs of almost all these results, and a simple proof of the stringency property of the analysis of variance (Section 5). These proofs rest on theorems which characterize the set theoretic structure of the tests. Thus, while the proofs of Hsu, Simaika and Wald are rather elaborate and each problem is essentially attacked *de novo*, the methods of the present paper are in effect applicable to the classical tests based on the normal distribution. For these tests it will not be difficult to demonstrate the analogues of Theorems 1 and 3, and of the results of Hsu, Simaika, and Wald. In the present paper we first treat the general linear hypothesis, because it is the simplest problem, its solution is easiest to describe, and it admits Wald's integration theorem. Multivariate analogues of the latter are rather artificial and not as simple. We then discuss the problem of the multiple correlation coefficient, because it seems to be more difficult than that of Hotelling's T and indeed, to include all the essential multivariate difficulties. Theorems 6 and 7 are the analogues of 1 and 3, respectively, while Theorem 9 describes the essential property of the power function which is of interest to us. In other multivariate problems one will prove the analogues of Theorems 6, 7 and 9. A generally inclusive formulation is no doubt possible. Theorems 5 and 9 are slightly more general than the theorems of Hsu and Simaika.

Many of the statements below may be not valid on exceptional sets of measure zero. Usually this is so stated, but sometimes, for reasons of brevity or to avoid repetition, this qualification may be omitted. The reader will have no difficulty supplying it wherever necessary.

The author is indebted to Erich L. Lehmann of the University of California, who carefully read a first version of this paper. Theorem 4 below was arrived at independently by Professor Lehmann, with a somewhat different proof.

2. The general linear hypothesis. In canonical form the general linear hypothesis may be stated as follows: The chance variables

$$X_1, X_2, \cdots, X_{k+l}$$

have at $x_1, \cdots, x_{k+l}$, the density function

$$(\sqrt{2\pi}\,\sigma)^{-(k+l)} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_1^k (x_i - \eta_i)^2 + \sum_{k+1}^{k+l} x_i^2\right\}\right] = f(\eta, \sigma) \tag{2.1}$$

with $\sigma, \eta_1, \cdots, \eta_k$ all unknown.

Let η be the vector $(\eta_1, \cdots, \eta_k)$. The null hypothesis H_0 states that

$$\eta_1 = \cdots = \eta_k = 0$$

and is to be tested with constant size $\alpha < 1$ (identically in σ).

Let D be any admissible critical region for testing H_0. If A is any event let $P\{A \mid \eta, \sigma\}$ denote the probability of A when η and σ are the parameters of (2.1). We have then

$$P\{D \mid 0, \sigma\} = \alpha$$

identically in σ, where 0 is the vector with k components all of which are zero. We now prove a property which characterizes all D. This theorem is due to Neyman and Pearson [12], and is given here only for completeness.

THEOREM 1. *The fraction of the surface area of the sphere*

$$\sum_1^{k+l} x_i^2 = c^2$$

which lies in D is α for almost all c.

PROOF. Let a be any positive integer, h a positive parameter, and $\psi(y)$ a measurable function of y defined for $y > 0$ and such that $0 \le \psi(y) \le 1$. In view of the distribution of ΣX_i^2, it will be enough to prove that, if

$$\frac{h^{a+1}}{\Gamma(a+1)} \int_0^\infty \psi(y) y^a e^{-hy}\, dy = \alpha$$

identically for all positive h, that then

$$\psi(y) = \alpha \text{ for almost all } y.$$

Write

$$\frac{1}{\alpha\Gamma(a+1)} \int_0^\infty \psi(y) y^a e^{-hy}\, dy = h^{-(a+1)}. \tag{2.2}$$

Differentiating both members k times with respect to h and then setting $h = 1$

we obtain the following result. The function

$$\frac{1}{\alpha\Gamma(a + 1)}\, \psi(y) y^{a} e^{-y}$$

is a density function with kth moment

$$\mu_k = (a + 1)(a + 2) \cdots (a + k).$$

The moments μ_k are the moments of the density function

$$\frac{1}{\Gamma(a + 1)}\, y^{a} e^{-y}.$$

They satisfy the Carleman criterion [5, p. 19, Th. 1.10], and hence no essentially different distribution can have these moments. This proves the desired result.

THEOREM 2 (Wald). *Among all tests of the general linear hypothesis the analysis of variance test has the property that, for all positive d, the integral of its power on the surface* $\eta^2 = d^2$ *is a maximum.*

PROOF. Let c be any positive number. We have only to show that if we allocate to the critical region D of the test the fraction α of the surface area of the sphere

$$\sum_{1}^{k+l} x_i^2 = c^2 \tag{2.3}$$

for which

$$C = \frac{\sum_{1}^{k} x_i^2}{\sum_{k+1}^{k+l} x_i^2}$$

is as large as possible and that if we do this for all c, the desired maximum of the integral of the power will be achieved. If C is as large as possible so is

$$\frac{\sum_{1}^{k} x_i^2}{\sum_{1}^{k+l} x_i^2} = \frac{\sum_{1}^{k} x_i^2}{c^2}.$$

Let $a_1, \cdots, a_{k+l}$ be any point on the sphere (2.3). Let db be the differential of area on the surface $\eta^2 = d^2$. Then

$$\int_{\eta^2=d^2}\cdots\int f(\eta, \sigma)\, db = (\sqrt{2\pi}\,\sigma)^{-(k+l)} \exp\left\{-\frac{(c^2 + d^2)}{2\sigma^2}\right\} \cdot \int_{\eta^2=d^2}\cdots\int \exp\left\{\frac{(\eta)'z}{\sigma^2}\right\} db, \tag{2.4}$$

where z is the vector $(a_1, \cdots, a_k)$ and $(\eta)'z$ is the scalar product of the two vectors. This last integral is easily seen to depend only upon $|z|$ and to be monotonically increasing in $|z|$. This proves the theorem.

COROLLARY (Hsu). *Among all tests of the general linear hypothesis whose power is a function of η^2 only, the analysis of variance is the most powerful.*

3. The set theoretic structure of tests whose power is a function only of η^2/σ^2. Wald's result (Theorem 2) cannot always be extended, in its simple form, to tests involving the multivariate normal distribution, but this can be done with Hsu's theorem (corollary to Theorem 2). In order to see what is involved we shall investigate the set theoretic structure of tests of the general linear hypothesis whose power is a function only of η^2/σ^2.

Let $q(x_1, \cdots, x_k)$ be the set of points in the region D whose first k coordinates are $x_1, \cdots, x_k$. Let $A(x_1, \cdots, x_k, \sigma)$ be the integral of

$$(2\pi\sigma^2)^{-(l/2)} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_{j=1}^{l} x_{k+j}^2\right\}\right]$$

with respect to $x_{k+1}, \cdots, x_{k+l}$, taken over $q(x_1, \cdots x_k)$. We first prove the following:

LEMMA. *Suppose the power of D is a function only of η^2/σ^2. Then for two points*

$$x_1, \cdots, x_k$$

and

$$x_1', \cdots, x_k'$$

such that

$$\sum_1^k x_i^2 = \sum_1^k x_i'^2 \tag{3.1}$$

we have

$$A(x_1, \cdots, x_k, \sigma) = A(x_1', \cdots, x_k', \sigma) \tag{3.2}$$

identically in σ, with the exception of a set of measure zero.

PROOF. Suppose the statement is false. Then under some orthogonal transformation T of $x_1, \cdots, x_k$ the region D would go over into a region D^* with the following property: Let $A^*(x_1, \cdots, x_k, \sigma)$ have the same definition for the region D^* as $A(x_1, \cdots, x_k, \sigma)$ has for D. Then on a set of positive measure[1] we would have

$$A(x_1, \cdots, x_k, \sigma) \not\equiv A^*(x_1, \cdots, x_k, \sigma). \tag{3.3}$$

We shall now show that (3.3) results in a contradiction. We have

$$P\{D \mid \eta, \sigma\} = P\{D^* \mid T\eta, \sigma\} \tag{3.4}$$

identically in η. By the property of the region D, therefore, we have

$$P\{D \mid \eta, \sigma\} = P\{D \mid T^{-1}\eta, \sigma\}$$

[1] The situation here is similar to that described in footnote 3.

and hence

$$P\{D \mid \eta, \sigma\} = P\{D^* \mid \eta, \sigma\} \tag{3.5}$$

identically in η. Thus we obtain

$$\int (2\pi\sigma^2)^{-(k/2)} A(x_1, \cdots, x_k, \sigma) \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_1^k (x_i - \eta_i)^2\right\}\right] dx_1 \cdots dx_k \tag{3.6}$$
$$\equiv \int (2\pi\sigma^2)^{-(k/2)} A^*(x_1, \cdots, x_k, \sigma) \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_1^k (x_i - \eta_i)^2\right\}\right] dx_1 \cdots dx_k$$

with the integrations taking place over the entire space. Differentiating both members with respect to the components of η and setting $\eta = 0$, we obtain that the two density functions (for fixed σ)

$$(2\pi\sigma^2)^{-(k/2)}\alpha^{-1}A(x_1, \cdots, x_k, \sigma) \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_1^k x_i^2\right\}\right]$$

and

$$(2\pi\sigma^2)^{-(k/2)}\alpha^{-1}A^*(x_1, \cdots, x_k, \sigma) \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_1^k x_i^2\right\}\right]$$

have identical moments. We shall now argue that these moments satisfy the conditions of Cramér and Wold [7, Th. 2], so that the two density functions are essentially the same, in contradiction to (3.3). The Cramér-Wold theorem states the following: *Let $Y_1, \cdots, Y_k$ be k chance variables with a joint distribution function, and write*

$$\lambda_{2n} = \sum_{i=1}^k EY_i^{2n}.$$

Then the divergence of the series

$$\sum_{n=1}^{\infty} \lambda_{2n}^{-(1/2n)}$$

is sufficient to ensure that there exists essentially only one distribution which has these moments. We notice that the factor $1/\alpha$ of course makes no difference. If we set $A(x_1, \cdots, x_k, \sigma)$ and $A^*(x_1, \cdots, x_k, \sigma)$ both identically unity and consider the resulting moments which enter into the λ_{2n}, we see that these moments satisfy the Cramér-Wold condition. Now A and A^* are ≤ 1. Thus, using the true value of A can serve only to increase the value of $\lambda_{2n}^{-(1/2n)}$, so that the series will diverge a fortiori. This proves the lemma.

The following theorem helps to describe the set theoretic structure of tests whose power is a function only of $\lambda = \eta^2/\sigma^2$:

THEOREM 3. *Let D be a test whose power is a function only of λ. Let u be any positive number, and $D(x_1, \cdots, x_k, u)$ be the fraction of the "area" of the sphere $\sum_{j=1}^l x_{k+j}^2 = u^2$ occupied by points which are in D and whose first k coordinates are $x_1, \cdots, x_k$. If*

$$\sum_{1}^{k} x_j^2 = \sum_{1}^{k} x_j'^2 \tag{3.7}$$

then, except on a set of measure zero,

$$D(x_1, \cdots, x_k, u) = D(x_1', \cdots, x_k', u). \tag{3.8}$$

PROOF. We shall show that, if the power of D is a function only of λ, the failure of (3.7) to imply (3.8) would contradict the preceding lemma. Suppose then that (3.8) is not true on a set of positive measure. Under some orthogonal transformation on $x_1, \cdots, x_k$ we obtain[2] a function $D^*(x_1, \cdots, x_k, u)$ which differs from $D(x_1, \cdots, x_k, u)$ on a set of positive measure and such that, for almost every $x_1, \cdots, x_k$,

$$A(x_1, \cdots, x_k, \sigma) = K \int_0^\infty D(x_1, \cdots, x_k, u)\sigma^{-l} u^{l-1} e^{(-u^2)/2\sigma^2}\, du$$

$$= K \int_0^\infty D^*(x_1, \cdots, x_k, u)\sigma^{-l} u^{l-1} e^{(-u^2)/2\sigma^2}\, du$$

identically in σ, where K is a suitable constant of no interest to us. Multiplying by σ^l, differentiating repeatedly under the integral sign with respect to σ, and setting $\sigma = 1$, we obtain the result that the two density functions in u,

$$\frac{KD(x_1, \cdots, x_k, u)}{A(x_1, \cdots, x_k, 1)} u^{l-1} e^{(-u^2)/2}$$

and

$$\frac{KD^*(x_1, \cdots, x_k, u)}{A(x_1, \cdots, x_k, 1)} u^{l-1} e^{(-u^2)/2}$$

are identical except perhaps on a set of measure zero. This contradiction proves the theorem.

THEOREM 4. *A necessary and sufficient condition that the power of D be a function of λ only, is that, with the usual exception of a set of measure zero, $D(x_1, \cdots, x_k, u)$ be a function only of*

$$\frac{\sum_{1}^{k} x_i^2}{u^2}.$$

The proof of this theorem is not essentially different from that of the preceding theorem, and we shall therefore sketch it only briefly. Let Z be a transformation on $(x_1, \cdots, x_k, u) = (x, u)$ which consists of a rotation of the vector x, followed by a multiplication of u and the components of x by a positive constant c. If $D(x, u)$ is not a function of $\Sigma_1^k x_i^2/u^2$ alone, then, just as before[3], we can use some

[2] See footnote 1.

[3] This statement implies that a function of $x_1, \cdots, x_k, u$, which is invariant to within sets of measure zero under all transformations Z (the exceptional set may depend on the

transformation Z to give us a function $D^*(x, u)$ such that

$$D(x, u) \not\doteq D^*(x, u)$$

on a set of positive measure, while

$$ED(x, u) = ED^*(x, u)$$

identically in η, σ. This yields a contradiction in the usual manner and proves the necessity of the condition.

To prove sufficiency, write $D(x, u) = \nu(\Sigma x_i^2/u^2) = \nu(v)$. Let $\gamma(v, \eta, \sigma)$ be the density function of v. Then

$$P\{D \mid \eta, \sigma\} = \int_0^\infty \nu(v)\gamma(v, \eta, \sigma)\, dv.$$

By hypothesis, $\nu(v)$ is a function only of v. We know [9, p. 140, eq. 101] that $\gamma(v, \eta, \sigma)$ is a function only of v and λ. Hence $P\{D \mid \eta, \sigma\}$ is a function only of λ. This completes the proof of the theorem.

THEOREM 5. *Among all tests of the general linear hypothesis which have the properties described in the conclusions of Theorems* 1 *and* 3, *the classical analysis of variance test is the most powerful.*

We shall omit the proof of this theorem, which is very similar to that of the more difficult Theorem 9 below.

Theorem 4 above shows that there exist regions D which satisfy the conclusions of Theorems 1 and 3 and such that $P\{D \mid \eta, \sigma\}$ is not a function of λ alone. It follows that the content of Theorem 5 is greater than that of Hsu's theorem (Corollary to Theorem 2).

It is instructive to note that Hsu's theorem follows almost immediately from Theorem 4 and the form of $\gamma(v, \lambda)$. For let λ be fixed but arbitrary. One verifies immediately from the form of $\gamma(v, \lambda)$ that

$$\frac{\gamma(v, \lambda)}{\gamma(v, 0)}$$

is, for fixed λ, a monotonically increasing function of v. This, by Neyman's lemma, immediately proves Hsu's result.

4. The multiple correlation coefficient. We shall now apply our methods to a multivariate test. For typographic ease we shall conduct the discussion for the

---

transformation), is a function of $\frac{\Sigma x_i^2}{u^2}$, except on a set of measure zero. This statement would be completely trivial were it not for the exceptional sets; in any case it must be well known to set theorists. The author constructed an unnecessarily long proof of it, and believes that a more expeditious proof can be constructed using the ideas of [11, page 91, Theorem 11.1, and page 318, p. 7]. Professor C. M. Stein of the University of California has informed the author that this result is a special case of one established by himself and G. H. Hunt in a forthcoming paper. For these reasons the proof is omitted. (See also [13, page 27, Lemma 9.1].)

case of three variates, but the reader will observe that the procedure is really perfectly general.

The chance variables $\{Y_{ij}\}$, $i = 1, 2, 3$, $j = 1, \cdots, n$, have the density function

$$g(B) = (2\pi)^{(-3n)/2}(|B|)^{n/2} \exp\left\{-\tfrac{1}{2}\sum_{j=1}^{n}\sum_{i,l=1}^{3} b_{il}\, y_{ij}\, y_{lj}\right\} \tag{4.1}$$

where 1) $B = \{b_{il}\}$ is a positive definite (symmetric) 3×3 matrix, 2) y_{ij} is the value assumed by Y_{ij}. The null hypothesis H_0 asserts that a given multiple correlation coefficient is zero, say that of Y_1 with Y_2 and Y_3, i.e.,

$$b_{12} = b_{21} = b_{13} = b_{31} = 0. \tag{4.2}$$

The test is to be made on the level of significance α, i.e., if B_0 is any matrix which satisfies (4.2), and if G is a critical region for testing H_0, then

$$P\{G \mid B_0\} = \alpha \tag{4.3}$$

where the symbol in the left member means the probability of G according to $g(B_0)$.

Write

$$ns_{ij} = \sum_{k=1}^{n} y_{ik}\, y_{jk}$$

$$S = \begin{Bmatrix} s_{22} & s_{23} \\ s_{32} & s_{33} \end{Bmatrix}.$$

Let $M(c_{11}, C)$ be the manifold in the $3n$-space of

$$y_{11}, \cdots, y_{ik}, \cdots, y_{3n}$$

where $s_{11} = c_{11}$, $S = C$. First we prove the following:

THEOREM 6. *Any region G which satisfies* (4.3) *must have the property that the fraction of the area of $M(c_{11}, C)$ which lies in G is α, for any positive c_{11} and any positive definite* 2×2 *matrix* $C = \{c_{ij}\}$. (We remind the reader that exceptional sets of measure zero are not precluded).

PROOF. Let $\psi(c_{11}, C)$ be the fraction of the area of $M(c_{11}, C)$ in G. Recall equation (4.3) and the fact that s_{11}, s_{22}, s_{23}, s_{33} are sufficient statistics for the elements of B_0. On the manifold $M(c_{11}, C)$ the conditional density is uniform. Employing Wishart's distribution [6] we conclude that

$$K' \int \psi(s_{11}, S) \mid B_0 \mid N \mid S \mid^{(n-3)/2} s_{11}^{(n-2)/2} \cdot \exp\left[-\frac{n}{2}\{b_{11} s_{11} + b_{22} s_{22} + 2b_{23} s_{23} + b_{33} s_{33}\}\right] ds_{11}\, ds_{22}\, ds_{23}\, ds_{33} \equiv \alpha \tag{4.4}$$

where K' is a suitable constant which need not concern us. Here the symbol

"$\equiv$" means identically in b_{11}, b_{22}, b_{23}, b_{33}, provided only that $b_{11} > 0$, $b_{22} > 0$ $b_{22}b_{33} - b_{23}^2 > 0$. Of course s_{11} is distributed independently of s_{22}, s_{23}, s_{33}. Proceeding as in section 2, we can, by differentiation with respect to the b's, obtain all the moments of the s_{ij}'s. Now let the b's take any admissible constant values. The moments of the s_{ij}'s are then seen to satisfy the criterion of Cramér and Wold [7, Th. 2], and consequently essentially uniquely determine the distribution of the s_{ij}. The desired conclusion follows as before.

The six parameters which uniquely determine the trivariate normal distribution (of Y_1, Y_2, Y_3) with zero means may be taken to be the following:

1) The covariance matrix $\{\sigma_{ij}\}$, $i, j = 2, 3$, of Y_2 and Y_3.

2) The partial regression coefficients β_2, β_3, of Y_1 on Y_2 and Y_3. These are defined as follows: Let $E(Y_1 \mid Y_2 = y_2, Y_3 = y_3)$ denote the conditional expected value of Y_1, given $Y_2 = y_2$, $Y_3 = y_3$. Then

$$E(Y_1 \mid Y_2 = y_2, Y_3 = y_3) = \beta_2 y_2 + \beta_3 y_3 .$$

3) The conditional variance ω^2 of Y_1, given $Y_2 = y_2$, $Y_3 = y_3$.

The population multiple correlation coefficient $\bar{R}$ of Y_1 with Y_2 and Y_3 is then defined by

$$\frac{\bar{R}^2\omega^2}{(1 - \bar{R}^2)} = \beta_2^2 \sigma_{22} + 2\beta_2\beta_3\sigma_{23} + \beta_3^2\sigma_{33}.$$

The six parameters above may be chosen arbitrarily, provided only that $\{\sigma_{ij}\}$ is positive definite. $\bar{R}$ and ω are, by definition, non-negative.

Let y_i be the column vector $y_{i1}, \cdots, y_{in}$; let y_i' be its transpose, and let y denote the point $y_{11}, y_{12}, \cdots, y_{1n}, y_{21}, \cdots, y_{3n}$ in $3n$-space. Let $z(y) = z(y_1, y_2, y_3)$ be the component of y_1 in the plane of y_2 and y_3; let $r = | z(y) |$ and θ the angle between z and y_2, measured positively say in the direction of y_3. Finally let h be the absolute value of the vector $y_1 - z(y_1, y_2, y_3)$.

We intend now to investigate the set theoretic structure of tests whose power is a function only of $\bar{R}$, and for this purpose prove the following:

THEOREM 7. *Let H be a region whose power is a function only of $\bar{R}$. Let $V(h, r, \theta, s_{22}, s_{23}, s_{33})$ be the fraction of the "volume" of the manifold on which $h, r, \theta, s_{22}, s_{23}, s_{33}$ are fixed which is contained in H. With the usual exception of a set of measure zero, for fixed $h, r, s_{22}, s_{23}, s_{33}$, the quantity V above is constant for all θ.*

Later, after this theorem is proved, we shall write V without exhibiting θ. This procedure is justified by Theorem 7.

PROOF. Suppose the theorem false, and proceed as in Theorem 3. A suitable[4] rotation of the radius vector $z(y)$ implies an orthogonal transformation T on the generic point y which leaves h, r, s_{22}, s_{23}, and s_{33} unaltered, and takes the region H into a region H^* such that H and H^* differ on a set of positive measure. T leaves R invariant, hence leaves invariant $\bar{R}$ which uniquely determines the distribution

[4] See footnote 1.

of R. Hence an argument almost the same as that which led us to (3.5) yields the conclusion that the power of H and the power of H^* are equal, identically in B. Proceeding as in Theorem 3, we obtain two essentially different density functions in h, r, θ, s_{22}, s_{23}, s_{33}, whose integrals over the entire space are identical in the elements of B. From these functions we obtain two different density functions in $s_{ij}(i, j = 1, 2, 3)$, with identical moments (obtained by differentiation with respect to the elements of B). The rest of the proof is essentially no different from that of Theorem 3.

THEOREM 8. *In order that the power of H be a function of $\bar{R}$ alone, it is necessary and sufficient that, with the usual exception of a set of measure zero, $V(h, r, s_{22}, s_{23}, s_{33})$ be a function only of h/r* (i.e., *of* R).

The proof of this theorem is essentially the same as the proof of Theorem 4. The place of the transformation Z is taken by one which consists of any linear transformation on the vectors y_2 and y_3, the addition of a constant angle to θ (rotation of $z(y)$), and multiplication of the vector y_1 by a positive scalar c. This transformation leaves $\bar{R}$ invariant. In the proof of sufficiency we use the distribution of R (see, for example, [10, p. 384, equation (15.55)]). The remainder of the proof is essentially the same as that of Theorem 4.

THEOREM 9. *Among all tests H which have the properties described in the conclusions of Theorems* 6 *and* 7, *the classical test based on R is the most powerful.*

As a corollary to this theorem we have the following result due to Simaika [3]: Of all tests H whose power is a function of $\bar{R}$ only, the classical test based on R is the most powerful.

Simaika's result also follows easily from Theorem 8 and the density function of R in the same manner that Hsu's result followed from Theorem 4 and the density function of v.

In the course of the proof of Theorem 9, the various symbols W, with or without subscripts, will denote suitable functions of the variables exhibited, and the various symbols k, with or without subscripts, will denote suitable constants.

We have that

$$
\begin{aligned}
P\{H \mid B\} &= \int_H (2\pi)^{(-3n)/2} \mid B \mid^{n/2} \exp\left\{-\tfrac{1}{2} \sum_{j=1}^{n} y_j' B y_j\right\} dy_{11} \cdots dy_{3n} \\
&= \int_H (2\pi\omega^2)^{(-n)/2} \exp\left[-\frac{1}{2\omega^2} \{y_1 - (\beta_2 y_2 + \beta_3 y_3)\}^2\right] \cdot \\
&\quad \cdot W_0(s_{22}, s_{23}, s_{33}, \{\sigma_{ij}\})\, dy_{11} \cdots dy_{3n} = (2\pi\omega^2)^{(-n)/2} \int_H \exp\left\{\frac{1}{\omega^2} (\beta_2 y_2 + \beta_3 y_3)' z\right\} \cdot \\
&\quad \exp\left[-\frac{1}{2\omega^2} \{y_1^2 + \beta_2^2 s_{22} + 2\beta_2\beta_3 s_{23} + \beta_3^2 s_{33}\}\right] \cdot \\
&\quad \cdot W_0(s_{22}, s_{23}, s_{33}, \{\sigma_{ij}\})\, dy_{11} \cdots dy_{3n}.
\end{aligned} \tag{4.5}
$$

Now $(\beta_2 y_2 + \beta_3 y_3)'z$ is a function only of β_2, β_3, s_{22}, s_{23}, s_{33}, r, and θ. Also

$h^2 + r^2 = s_{11} = y_1^2$. Thus

$$P\{H \mid B\} = \int V(h, r, s_{22}, s_{23}, s_{33}) W_1(h, r, s_{22}, s_{23}, s_{33}, \{B\})$$

$$\cdot \exp\left\{\frac{1}{\omega^2}(\beta_2 y_2 + \beta_3 y_3)'z\right\} d\theta\, dh\, dr\, ds_{22}\, ds_{23}\, ds_{33} = \int V(h, r\, s_{22}, s_{23}, s_{33})$$

$$(4.6) \qquad \cdot W_1(h, r, s_{22}, s_{23}, s_{33}, \{B\})(4hr)^{-1} \exp\left\{\frac{1}{\omega^2}(\beta_2 y_2 + \beta_3 y_3)'z\right\}$$

$$\cdot d\theta\, dh^2\, dr^2\, ds_{22}\, ds_{23}\, ds_{33} = \int V\left(\sqrt{y_1^2 - r^2}, r, s_{22}, s_{23}, s_{33}\right)$$

$$\cdot W_2(\sqrt{y_1^2 - r^2}, r, s_{22}, s_{23}, s_{33}, \{B\}) \exp\left\{\frac{1}{\omega^2}(\beta_2 y_2 + \beta_3 y_3)'z\right\}$$

$$\cdot d\theta\, dr^2\, dy_1^2\, ds_{22}\, ds_{23}\, ds_{33}.$$

Integrating with respect to θ and designating

$$W_2 \int \exp\left\{\frac{1}{\omega^2}(\beta_2 y_2 + \beta_3 y_3)'z\right\} d\theta$$

by $W(\sqrt{y_1^2 - r^2}, r, s_{22}, s_{23}, s_{33}, \{B\})$ we observe that just as in (2.4), W is monotonically increasing in r (all other variables fixed). Thus we have

$$(4.7) \qquad P\{H \mid B\} = \int VW\, dr^2\, dy_1^2\, ds_{22}\, ds_{23}\, ds_{33}.$$

In constructing H only the function V is at our disposal, and this subject to the limitations imposed by the conclusions of Theorems 6 and 7 and the fact that $h^2 + r^2 = y_1^2 = s_{11}$. The function W is not within our control at all. With y_1^2, s_{22}, s_{23}, s_{33} fixed, W is monotonically increasing with r. To maximize the power it is therefore best to distribute the "mass" so that V is as large as possible for large values of r and hence of R. This implies the classical test and proves the theorem.

5. Stringency of the classical tests. Wald [8] calls a test T_1 "most stringent" if the following is true: Let $\{T\}$ be the totality of tests. Let θ be the generic point in the parameter space, and $P\{T \mid \theta\}$ be the power of T at the point θ. Let T_2 be any test other than T_1. Then

$$\sup_{\theta} [\sup_{\{T\}} P\{T \mid \theta\} - P\{T_1 \mid \theta\}] \leq \sup_{\theta} [\sup_{\{T\}} P\{T \mid \theta\} - P\{T_2 \mid \theta\}].$$

Of course, we have omitted to specify the totality $\{T\}$. One can admit all tests whose size $\leq \alpha$, a given constant between 0 and 1, or restrict one's self to tests whose size is exactly α. We shall do the latter.

Under these circumstances we shall prove that the classical test of a linear hypothesis is most stringent. Our proof will occupy but a few lines, and is an easy consequence of the structure of the classical tests as described in the lemma of section 2. The result itself is a special case of an unpublished theorem due to G. H. Hunt and C. M. Stein, and all priority on this result is theirs.

Return then to the notation of section 2. Let σ be fixed at any arbitrary positive value, and the surface

$$\eta^2 = c_0^2$$

be that one on which

$$\omega_1(\eta) = \sup_{\{T\}} P\{T \mid \eta\} - P\{L_1 \mid \eta\}$$

is a maximum, where L_1 is the classical test of the linear hypothesis. It is clear that this maximum is actually achieved, and that $\omega_1(\eta)$ is a constant on the surface $\eta^2 = c_0^2$. Let L_2 be any other test (of size α), and $\omega_2(\eta)$ be the corresponding function for L_2. We have only to show that on the surface $\eta^2 = c_0^2$ we cannot have everywhere $\omega_2(\eta) < \omega_1(\eta)$, and our proof is complete. If everywhere on the surface $\eta^2 = c_0^2$ we had $\omega_2(\eta) < \omega_1(\eta)$, we would have, also on the same surface, $P\{L_2 \mid \eta\} > P\{L_1 \mid \eta\}$. This would, however, violate Wald's Theorem 2 (section 2) and proves the desired result.

REFERENCES

[1] P. L. Hsu, "Notes on Hotelling's generalized T," *Annals of Math. Stat.*, Vol. 9 (1938) p. 231.
[2] P. L. Hsu, "Analysis of variance from the power function standpoint," *Biometrika*, Vol. 32 (1941), p. 62.
[3] J. B. Simaika, "On an optimum property of two important statistical tests," *Biometrika*, Vol. 32 (1941), p. 70.
[4] A. Wald, "On the power function of the analysis of variance test," *Annals of Math. Stat.*, Vol. 13 (1942), p. 434.
[5] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, The American Mathematical Society, New York, 1943.
[6] John Wishart, "The generalized product moment distribution, etc.," *Biometrika*, Vol. 20A (1928), p. 32.
[7] H. Cramér and H. Wold, "Some theorems on distribution functions," *Lond. Math. Soc. Jour.*, Vol. 11 (1936).
[8] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Am. Math. Soc. Trans.*, Vol. 54 (1943), p. 426.
[9] P. C. Tang, "The power function of the analysis of variance etc.," *Stat. Res. Memoirs*, Vol. 2 (1938) (University of London), p. 126.
[10] M. G. Kendall, *The Advanced Theory of Statistics*, Vol. 1, Charles Griffin and Company, London, 1945.
[11] S. Saks, *Theory of the Integral*, (Second Edition), G. E. Stechert and Company, New York, 1937.
[12] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Roy. Soc. London Phil. Trans.*, Ser. A, Vol. 231 (1933), pp. 289–337.
[13] Eberhard Hopf, *Ergodentheorie*, Chelsea, New York, 1948.