SYMMETRIC TESTS OF THE HYPOTHESIS THAT THE MEAN OF ONE NORMAL POPULATION EXCEEDS THAT OF ANOTHER

By Herbert A. Simon

Illinois Institute of Technology

1. Introduction. One of the most commonly recurring statistical problems is to determine, on the basis of statistical evidence, which of two samples, drawn from different universes, came from the universe with the larger mean value of a particular variate. Let M_y be the mean value which would be obtained with universe (Y) and M_x be the mean value which would be obtained with universe (X). Then a test may be constructed for the hypothesis[1] $M_y \geq M_x$.

If $x_1, \cdots, x_n$ are the observed values of the variate obtained from universe (X), and $y_1, \cdots, y_n$ are the observed values obtained from universe (Y), then the sample space of the points $E: (x_1, \cdots, x_n; y_1, \cdots, y_n)$ may be divided into three regions ω_0, ω_1, and ω_2. If the sample point falls in the region ω_0, the hypothesis $M_y \geq M_x$ is accepted; if the sample point falls in the region ω_1, the hypothesis $M_y \geq M_x$ is rejected; if the sample point falls in the region ω_2, judgment is withheld on the hypothesis. Regions ω_0, ω_1, and ω_2 are mutually exclusive and, together, fill the entire sample space. Any such set of regions ω_0, ω_1, and ω_2 defines a test for the hypothesis $M_y \geq M_x$.

In those cases, then, where the experimental results fall in the region ω_2, the test leads to the conclusion that there is need for additional data to establish a result beyond reasonable doubt. Under these conditions, the test does not afford any guide to an unavoidable or non-postponable choice. In the application of statistical findings to practical problems it often happens, however, that judgment can not be held in abeyance—that *some* choice must be made, even at a risk of error. For example, when planting time comes, a choice must be made between varieties (X) and (Y) of grain even if neither has been conclusively demonstrated, up to that time, to yield a larger crop than the other. It is the purpose of this paper to propose a criterion which will always permit a choice between two experimental results, that is, a test in which the regions ω_0 and ω_1 fill the entire sample space. In the absence of a region ω_2, any observed result is interpreted as a definite acceptance or rejection of the hypothesis tested.

2. General characteristics of the criterion. Let us designate the hypothesis $M_y \geq M_x$ as H_0 and the hypothesis $M_x > M_y$ as H_1. Then a pair of tests, T_0 and T_1, for H_0 and H_1 respectively must, to suit our needs, have the following properties:

(1) The regions ω_{00} (ω_{00} is the region of acceptance for H_0, ω_{10} the region of rejection for H_0; ω_{01} and ω_{11} the corresponding regions for H_1) and ω_{11} must

[1] This paper presupposes a familiarity with the theory of testing statistical hypotheses as set forth by J. Neyman and E. S. Pearson [1].

coincide; as must the regions ω_{10} and ω_{01}. This correspondence means that when H_0 is accepted, H_1 is rejected, and vice versa. Hence, the tests T_0 and T_1 are identical, and we shall hereafter refer only to the former.

(2) There must be no regions ω_{20} and ω_{21}. This means that judgment is never held in abeyance, no matter what sample is observed.

(3) The regions ω_{00} and ω_{10} must be so bounded that the probability of accepting H_1 when H_0 is true (error of the first kind for T_0) and the probability of accepting H_0 when H_1 is true (error of the second kind for T_0) are, in a certain sense, minimized. Since H_0 and H_1 are composite hypotheses, the probability that a test will accept H_1 when H_0 is true depends upon which of the simple hypotheses that make up H_0 is true.

Neyman and Pearson [2] have proposed that a test, T_α for a hypothesis be termed *uniformly more powerful* than another test, T_β, if the probability for T_α of accepting the hypothesis if it is false, or the probability of rejecting it if it is true, does not exceed the corresponding probability for T_β no matter which of the simple hypotheses is actually true. Since there is no test which is uniformly more powerful than all other possible tests, it is usually required that a test be uniformly most powerful (UMP) among the members of some specified class of tests.

3. A symmetric test when the two universes have equal standard deviations. Let us consider, first, the hypothesis $M_y \geq M_x$ where the universes from which observations of varieties (X) and (Y), respectively, are drawn are normally distributed universes with equal standard deviations, σ, and means M_x and M_y respectively. Let us suppose a sample drawn of n random observations from the universe of variety (X) and a sample of n independent and random observations from the universe of (Y). The probability distribution of points in the sample space is given by

$$p(x_1, \cdots, x_n; y_1, \cdots, y_n) = (2\pi\sigma^2)^{-n} e^{-\frac{1}{2\sigma^2}\left[\sum_i (x_i - M_x)^2 + \sum_i (y_i - M_y)^2\right]} \tag{1}$$

In testing the hypothesis $M_y \geq M_x$, there is a certain symmetry between the alternatives (X) and (Y). If there is no *a priori* reason for choosing (X) rather than (Y), and if the sample point $E_1: (a_1, \cdots, a_n; b_1, \cdots, b_n)$ falls in the region of acceptance of H_0: then the point $E_2: (b_1, \cdots, b_n; a_1, \cdots, a_n)$ should fall in the region of acceptance of H_1. That is, if E_1 is taken as evidence that $M_y \geq M_x$; then E_2 can with equal plausibility be taken as evidence that $M_x \geq M_y$.

Any test such that $E_1: (a_1, \cdots, a_n; b_1, \cdots, b_n)$ lies in ω_0 whenever E_2: $(b_1, \cdots, b_n; a_1, \cdots, a_n)$ lies in ω_1 and vice versa, will be designated a symmetric test of the hypothesis $M_y \geq M_x$. Let Ω be the class of symmetric tests of H_0. If T_α is a member of Ω, and is uniformly more powerful than every other T_β which is a member of Ω, then *T_α is the uniformly most powerful symmetric test of H_0.*

The hypothesis $M_y \geq M_x$ possesses a UMP symmetric test. This may be shown as follows. From (1), the ratio can be calculated between the probability densities at the sample points $E:(x_1, \cdots, x_n; y_1, \cdots, y_n)$ and E': $(y_1, \cdots, y_n; x_1, \cdots, x_n)$. We get

$$\frac{p(E)}{p(E')} = \exp\left\{\frac{n}{\sigma^2}(\bar{x} - \bar{y})(M_x - M_y)\right\}, \tag{2}$$

where

$$\bar{x} = \frac{1}{n}\sum_i x_i, \qquad \bar{y} = \frac{1}{n}\sum_i y_i.$$

Now the condition $p(E) > p(E')$ is equivalent to $\frac{n}{\sigma^2}(\bar{x} - \bar{y})(M_x - M_y) > 0$. Hence $p(E) > p(E')$ whenever $(\bar{x} - \bar{y})$ has the same sign as $(M_x - M_y)$.

Now for any symmetric test, if E lies in ω_0, E' lies in ω_1, and vice versa. Suppose that, in fact, $M_y > M_x$. Consider a symmetric test, T_α whose region ω_0 contains a sub-region ω_{0U} (of measure greater than zero) such that $\bar{y} < \bar{x}$ for every point in that sub-region. Then for every point E' in ω_{0U}, $p(E') < p(E)$. Hence, a more powerful test, T_β could be constructed which would be identical with T_α, except that ω_{1U}, the sub-region symmetric to ω_{0U}, would be interchanged with ω_{0U} as a portion of the region of acceptance for H_0. Therefore, a test such that ω_0 contained all points for which $\bar{y} > \bar{x}$, and no others, would be a UMP symmetric test. This result is independent of the magnitude of $(M_x - M_y)$ provided only $M_y \geq M_x$. We conclude that *$\bar{y} > \bar{x}$ is a uniformly most powerful symmetric test for the hypothesis $M_y > M_x$.*

The probability of committing an error with the UMP symmetric test is a simple function of the difference $|M_y - M_x|$. The exact value can be found by integrating (1) over the whole region of the sample space for which $\bar{y} < \bar{x}$. There is no need to distinguish errors of the first and second kind, since an error of the first kind with T_0 is an error of the second kind with T_1, and vice versa. The probability of an error is one half when $M_x = M_y$, and in all other cases is less than one half.

4. Relation of UMP symmetric test and test which is UMP of tests absolutely equivalent to it. Neyman and Pearson [2] have shown the test $\bar{y} - \bar{x} > k$ to be UMP among the tests absolutely equivalent to it, for the hypothesis $M_y \geq M_x$. They have defined a class of tests as absolutely equivalent if, for each simple hypothesis in H_0, the probability of an error of the first kind is exactly the same for all the tests which are members of the class. If k be set equal to zero, $\bar{y} > \bar{x}$, and their test reduces to the UMP symmetric test. What is the relation between these two classes of tests?

If T_α be the UMP symmetric test, then it is clear from Section 2 that there is no other symmetric test, T_β, which is absolutely equivalent to T_α. Hence Ω, the class of symmetric tests, and Λ, the class of tests aboslutely equivalent to T_α, have only one member in common—the test T_α itself. Neyman and Pearson have shown T_α to be the UMP test of Λ, while the results of Section 4 show T_α to be the UMP test of Ω.

5. Justification for employing a symmetric test. In introducing Section 3, a heuristic argument was advanced for the use of a symmetric, rather than an asymmetric test for the hypothesis $M_y \geq M_x$. This argument will now be given a precise interpretation in terms of probabilities.

Assume, not a single experiment for testing the hypothesis $M_y \geq M_x$, but a series of similar experiments. Suppose a judgment to be formed independently on the basis of each experiment as to the correctness of the hypothesis. Is there any test which, if applied to the evidence in each case, will maximize the probability of a correct judgment in that experiment? Such a test can be shown to exist, providing one further assumption is made: that if any criterion be applied *prior* to the experiment to test the hypothesis $M_y \geq M_x$, the probability of a correct decision will be one half. That is, it must be assumed that there is no evidence which, prior to the experiment, will permit the variety with the greater yield to be selected with greater-than-chance frequency.

Consider now any asymmetric test for the hypothesis H_0—that is, any test which is not symmetric. The criterion $\bar{y} - \bar{x} > k$, where $k > 0$, is an example of such a test. Unlike a symmetric test, an asymmetric test may give a different result if applied as a test of the hypothesis H_0 than if applied as a test of the hypothesis H_1. For instance, a sample point such that $\bar{y} - \bar{x} = \epsilon$, where $k > \epsilon > 0$, would be considered a rejection of H_0 and acceptance of H_1 if the above test were applied to H_0; but would be considered a rejection of H_1 and an acceptance of H_0 if the test were applied to H_1. Hence, before an asymmetric test can be applied to a problem of dichotomous choice—a problem where H_0 *or* H_1 must be determinately selected—a decision must be reached as to whether the test is to be applied to H_0 or to H_1. This decision cannot be based upon the evidence of the sample to be tested—for in this case, the complete test, which would of course include this preliminary decision, would be symmetric by definition.

Let H_c be the correct hypothesis (H_0 or H_1, as the case may be) and let H_* be the hypothesis to which the asymmetric test is applied. Since by assumption there is no prior evidence for deciding whether H_c is H_0 or H_1, we may employ any random process for deciding whether H_* is to be identified with H_0 or H_1. If such a random selection is made, it follows that the probability that H_c and H_* are identical is one half.

We designate as the region of asymmetry of a test the region of points E_1: $(a_1, \cdots, a_n; b_1, \cdots, b_n)$ and E_2: $(b_1, \cdots, b_n; a_1, \cdots, a_n)$ of aggregate measure greater than zero such that E_1 and E_2 both fall in ω_0 or both fall in ω_1. Suppose ω_{0a} and ω_{0b} are a particular symmetrically disposed pair of subregions of the region of asymmetry, which fall in ω_0 of a test T_0. Suppose that, for every point, E_1, in ω_{0a}, $\bar{b} > \bar{a}$, and that ω_{0a} and ω_{0b} are of measure greater than zero. The sum of the probabilities that the sample point will fall in ω_{0a} or ω_{0b} is exactly the same whether H_c and H_* are the same hypothesis or are contradictory hypotheses. In the first case H_c will be accepted, in the second case H_c will be rejected. These two cases are of equal probability, hence there is a probability of one half of accepting or rejecting H_c if the sample point falls in the region of asymmetry of T_0. But from equation (2) of Section 2 above, we see that if the subregions ω_{0a} and ω_{0b} had been in a region of symmetry, and if ω_{0a} had been in ω_0, the probability of accepting H_c would have been greater than the probability of rejecting H_c.

Hence, if it is determined by random selection to which of a pair of hypotheses an asymmetric test is going to be applied, the probability of a correct judgment with the asymmetric test will be less than if there were substituted for it the UMP symmetric test. It may be concluded that the UMP symmetric test is to be preferred unless there is prior evidence which permits a tentative selection of the correct hypothesis with greater-than-chance frequency.

6. Symmetric test when standard deviations of universes are unequal. Thus far, we have restricted ourselves to the case where $\sigma_x = \sigma_y$. Let us now relax this condition and see whether a UMP symmetric test for $M_y \geq M_x$ exists in this more general case.

We now have for the ratio of $p(E)$ to $p(E')$:

$$(3)\quad \frac{p(E)}{p(E')} = \exp\left\{-\frac{n}{2\sigma_x^2\sigma_y^2}\left[(\sigma_y^2 - \sigma_x^2)(\mu_x - \mu_y) - 2(\sigma_y^2 M_x - \sigma_x^2 M_y)(\bar{x} - \bar{y})\right]\right\},$$

where

$$\mu_x = \sum_i x_i^2/n, \qquad \mu_y = \sum_i y_i^2/n.$$

Even if σ_y and σ_x are known, which is not usually the case, there is no UMP symmetric test for the hypothesis $M_y \geq M_x$. From (3), the symmetric critical region which has the lowest probability of errors of the first kind for the hypothesis ($M_y = k_1$; $M_x = k_2$; $k_1 > k_2$) is the set of points E such that:

$$(4)\qquad (\sigma_y^2 - \sigma_x^2)(\mu_x - \mu_y) - 2(\sigma_y^2 k_2 - \sigma_x^2 k_1)(\bar{x} - \bar{y}) > 0.$$

Since this region is not the same for all values of k_1 and k_2 such that $k_1 > k_2$, there is no UMP symmetric region for the composite hypothesis $M_y \geq M_x$. This result holds, *a fortiori* when σ_y and σ_x are not known.

If there is no UMP symmetric test for $M_y \geq M_x$ when $\sigma_y \neq \sigma_x$, we must be satisfied with a test which is UMP among some class of tests more restricted than the class of symmetric tests. Let us continue to restrict outselves to the case where there are an equal number of observations, in our sample, of (X) and of (Y). Let us pair the observations x_i, y_i, and consider the differences $u_i = x_i - y_i$. Is there a UMP test among the tests which are symmetric with respect to the u_i's for the hypothesis that $M_y - M_x = -U \geq 0$? By a symmetric test in this case we mean a test such that whenever the point $(u_1, \cdots, u_n)$ falls into region ω_0, the point $(-u_1, \cdots, -u_n)$ falls into region ω_1.

If x_i and y_i are distributed normally about M_x and M_y with standard deviations σ_x and σ_y respectively, then u_i will be normally distributed about $U =$

$M_x - M_y$ with standard deviation $\sigma_u = \sqrt{\sigma_x^2 + \sigma_y^2}$. The ratio of probabilities for the sample points $E_v: (u_1, \cdots, u_n)$ and $E_v': (-u_1, \cdots, -u_n)$ is given by:

$$\frac{p(E_v)}{p(E_v')} = \exp\left\{\frac{-2n}{\sigma_u^2} \bar{u} U\right\}, \tag{5}$$

where

$$\bar{u} = \frac{1}{n} \sum_i u_i.$$

Hence, $p(E_v) > p(E_v')$ whenever $\bar{u}$ has the same sign as U. Therefore, by the same process of reasoning as in Section 2, above, we may show that $\bar{u} \leq 0$ is a UMP test among tests symmetric in the sample space of the u's for the hypothesis $U \leq 0$.

It should be emphasized that Ω_{su}, the class of symmetric regions in the space of $E_v: (u_1 \cdots u_n)$, is far more restricted than Ω_s, the class of symmetric regions in the sample space of $E: (x_1 \cdots x_n ; y_1 \cdots y_n)$. In the latter class are included all regions such that:

(A) $E: (a_1, \cdots, a_n ; b_1, \cdots, b_n)$ falls in ω_0 whenever $E: (b_1, \cdots, b_n ; a_1, \cdots, a_n)$ falls in ω_1. Members of class Ω_{su} satisfy this condition together with the further condition:

(B) For all possible sets of n constants $k_1, \cdots, k_n$, $E: (x_1 + k_1, \cdots, x_n + k_n ; y_1 + k_1, \cdots, y_n + k_n)$ falls in ω_0 whenever $E: (x_1, \cdots, x_n ; y_1, \cdots, y_n)$ falls in ω_0. When $\sigma_y \neq \sigma_x$, a UMP test for $M_y \geq M_x$ with respect to the symmetric class Ω_{su} exists, but a UMP test with respect to the symmetric class Ω_s does not exist.

REFERENCES

[1] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy. Soc.*, Series A, 702, Vol. 231 (1933), pp. 289–337.

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