SOME SIGNIFICANCE TESTS BASED ON ORDER STATISTICS

By John E. Walsh

Princeton University

1. Summary. In this paper significance tests are developed whose application requires only the determination of one order statistic and the computation of sums of sample values. The simplest case considered is that of testing a new sample value x on the basis of m previous sample values $y_1, \cdots, y_m$, all sample values being assumed from normal populations with the same variance. Two separate tests of whether the mean of the new population from which x was taken exceeds the mean of the population from which $y_1, \cdots, y_m$ were drawn consist in accepting the alternative that the new population mean exceeds the old population mean if

(1) $$x > \left(\frac{\sqrt{m+1}+1}{m}\right)\sum_1^m y_i - \sqrt{m+1}\, y_{(u)}$$

(2) $$x > \left(\frac{\sqrt{m+1}-1}{m}\right)\sum_1^m y_i + \sqrt{m+1}\, y_{(m+1-u)},$$

where $y_{(u)}$ is the uth largest of $y_1, \cdots, y_m$. It can be shown that both of these tests have the same power so that either one might be equally well selected for use. In practical application, however, there may exist reasons for preferring one test to the other. Similarly, the alternative that the new population mean is less than the old population mean will be accepted if

(3) $$x < \left(\frac{\sqrt{m+1}+1}{m}\right)\sum_1^m y_i - \sqrt{m+1}\, y_{(m+1-u)}$$

(4) $$x < \left(\frac{\sqrt{m+1}-1}{m}\right)\sum_1^m y_i + \sqrt{m+1}\, y_{(u)}.$$

All four of these significance tests have the same power, also the same significance level $\alpha(u, m)$. By appropriate choice of u and m the significance level can be made to assume values suitable for significance tests. For example,

$$\alpha(1, 6) = .0156, \qquad \alpha(2, 10) = .0107$$
$$\alpha(3, 13) = .0110, \qquad \alpha(4, 16) = .0107.$$

The above tests are still valid if each of $x, y_1, \cdots, y_m$ equals a sum of r sample values.

These order statistic tests are generalized to the case where x is a sum of r new sample values; $y_1, \cdots, y_m$ each equals a sum of s past sample values and another sum of relatively weighted past sample values is utilized but not as an order statistic. The introduction of this relatively weighted sum allows less reliable past information to be lumped together and weighted according to its relative importance.

In comparing the order statistic tests with the most powerful tests which could be used for these alternatives it is found that the size of the samples used must be increased in order to bring the efficiency of the order statistic test up to that of the corresponding most powerful test. Thus the advisability of using the order statistic test will depend upon whether it is more desirable to take larger samples but have less computation.

2. Introduction. Many statistical problems are concerned with the determination of whether a new sample can be considered as having been drawn from the same population as that from which a previous sample was taken. Frequently this reduces to the question of whether the mean of the population from which the new sample came is greater than the mean of the past sample population. The problem of whether the new population mean is less than that of the old population is also occasionally investigated. If both populations can be considered normal with the same variance, it is well known that the most powerful Studentized test of each of these one-sided alternatives is furnished by use of the appropriate Student t-test. When the number of previous sample values from which the test is determined is large, however, the computation of the numerical value required for the application of the Student t-test becomes lengthy. This calculation difficulty can become very important if the test is to be applied repeatedly as, for example, in quality control work. It is desirable, therefore, to develop other Studentized tests which are easily calculated and whose efficiency with relation to the corresponding Student t-tests is reasonably high. It is the purpose of this paper to develop tests of this type by the use of order statistics.

The class of tests in which a new sample value x is tested on the basis of m previous sample values $y_1, \cdots, y_m$ used as order statistics is developed in detail. The significance tests arising are the ones given in the summary above. For a better intuitive understanding of what takes place rewrite (1) to (4) as

$$(1') \qquad x - \bar{y} > \sqrt{m+1}(\bar{y} - y_{(u)})$$

$$(2') \qquad x + \bar{y} > \sqrt{m+1}(\bar{y} + y_{(m+1-u)})$$

$$(3') \qquad x - \bar{y} < \sqrt{m+1}(\bar{y} - y_{(m+1-u)})$$

$$(4') \qquad x + \bar{y} < \sqrt{m+1}(\bar{y} + y_{(u)}),$$

where $\bar{y}$ is the average of the y_i. The relative efficiencies of these tests with respect to the corresponding Student t-tests are determined and the simplicity of the computation necessary for their application is outlined. The method of attack having been sufficiently indicated by the development of this special class of tests, more general tests based on order statistics are stated but not proved here.

3. Statement of the significance tests. Let each of x, $y_1, \ldots, y_m$ be distributed independently of all the others, x according to $N(\nu, \sigma^2)$ and the y_i, $(i = 1, \cdots, m)$, according to $N(\mu, \sigma^2)$, where the notation $N(\xi, \sigma^2)$ signifies the

normal distribution with mean ξ and variance σ^2. As above let $y_{(u)}$ denote the uth largest of $y_1, \cdots, y_m$. The one-sided significance tests are then stated as follows:
If

$$(5) \qquad \begin{aligned} x &> \frac{1}{K_2}\sum_1^m y_i - \frac{K_1}{K_2} y_{(u)} \qquad (K_2 > 0) \\ x &> \frac{1}{K_2}\sum_1^m y_i - \frac{K_1}{K_2} y_{(m+1-u)} \qquad (K_2 < 0) \end{aligned}$$

accept the alternative $\mu < \nu$, otherwise accept the hypothesis tested, namely that $\mu \doteq \nu$.
If

$$(6) \qquad \begin{aligned} x &< \frac{1}{K_2}\sum_1^m y_i - \frac{K_1}{K_2} y_{(m+1-u)} \qquad (K_2 > 0) \\ x &< \frac{1}{K_2}\sum_1^m y_i - \frac{K_1}{K_2} y_{(u)} \qquad (K_2 < 0) \end{aligned}$$

accept $\nu < \mu$, otherwise accept $\nu = \mu$.

The constants K_1 and K_2 are given by

$$(7) \qquad K_1 = m + 1 \pm \sqrt{m+1}, \qquad K_2 = -1 \mp \sqrt{m+1},$$

where all upper signs or all lower signs will be chosen so that to a given value of K_1 there is but one value of K_2. This rule for the choice of signs will hold throughout the paper.

It is to be noted that (5) defines two separate significance tests of the hypothesis $\mu = \nu$ against the alternative $\mu < \nu$ depending upon whether it is decided to use the positive or the negative value given for K_2. A similar statement applies to the two significance tests defined by (6).

Each of these four significance tests can be shown to have the same significance level, which is determined by the values of u and m. Denote this significance level by $\alpha(u, m)$. Then it can be demonstrated that

$$\alpha(1, m) = (\tfrac{1}{2})^m, \qquad \alpha(2, m) = (m+1)(\tfrac{1}{2})^m$$

$$\alpha(3, m) = (m^2 + m + 2)(\tfrac{1}{2})^{m+1}, \qquad \alpha(4, m) = \tfrac{1}{3}(m^3 + 5m + 6)(\tfrac{1}{2})^{m+1}.$$

The general expression for $\alpha(u, m)$ is given by (12).

It is to be observed that the application of these tests is independent of the parameters of the normal populations in question.

4. Analysis. An analysis will be given for the development of the significance test in which the alternative is $\mu < \nu$ and $K_2 > 0$. The developments of the properties of the other three tests are almost identical with that for this case and will not be given here.

Now consider this analysis. Let

$$x' = \frac{x - \nu}{\sigma}, \qquad y_i' = \frac{y_i - \mu}{\sigma}, \qquad (i = 1, \cdots, m).$$

Then x' and the y_i' are independently distributed according to $N(0, 1)$. Define

$$r_u = \frac{1}{K_1}\left(K_1 y_u' - \sum_1^m y_i' + K_2 x'\right), \qquad (u = 1, \cdots, m).$$

It is easily seen that

$$E(r_u) = 0, \qquad E(r_u^2) = \frac{1}{K_1^2}(K_2^2 + K_1^2 - 2K_1 + m),$$

$$E(r_u r_v) = \frac{1}{K_1^2}(K_2^2 - 2K_1 + m), \qquad (u \neq v).$$

Thus the condition which must be fulfilled in order that the r_u be independently distributed according to $N(0, 1)$ is that

$$K_2^2 - 2K_1 + m = 0. \tag{8}$$

To insure that the r_u are independent of μ when $\mu = \nu$ it is evidently necessary that

$$K_1 - m + K_2 = 0. \tag{9}$$

Solving (8) and (9) for K_1 and K_2 one obtains (7).

Restrict the r_u by conditions (8) and (9) and let $r_{(u)}$ be the uth largest of $r_1, \cdots, r_m$. From (8) $K_1 > 0$; therefore

$$r_{(u)} = \frac{1}{K_1}\left[K_1 y_{(u)}' - \sum_1^m y_i' + K_2 x'\right],$$

where $y_{(u)}'$ is the uth largest of $y_1', \cdots, y_m'$. Then using (9),

$$r_{(u)} = \frac{1}{K_1 \sigma}\left[K_1 y_{(u)} - \sum_1^m y_i + K_2 x + K_2(\mu - \nu)\right].$$

From the definition of the power function and (5) for $K_2 > 0$, it follows that the power function for this test is given by

$$\begin{aligned} \text{Power Function} &= Pr\left[x > \frac{1}{K_2}\sum_1^m y_i - \frac{K_1}{K_2} y_{(u)}\right] \\ &= Pr\left[0 < K_1 y_{(u)} - \sum_1^m y_i + K_2 x < \infty\right] \\ &= Pr\left[\frac{K_2}{K_1\sigma}(\mu - \nu) < \frac{1}{K_1\sigma}\left\{K_1 y_{(u)} - \sum_1^m y_i + K_2 x + K_2(\mu - \nu)\right\} < \infty\right] \\ &= Pr\left[\frac{K_2}{K_1\sigma}(\mu - \nu) < r_{(u)} < \infty\right]. \end{aligned} \tag{10}$$

The distribution function of the order statistic $r_{(u)}$ may be found in [1], from which it follows that

$$\text{(11)}\quad \begin{aligned} \text{Power Function} &= \frac{m!}{(u-1)!(m-u)!} \\ &\cdot \int_{(K_2/K_1\sigma)(\mu-\nu)}^{\infty} \left(\int_{-\infty}^{z} f(y)\,dy \right)^{u-1} \left(\int_{z}^{\infty} f(y)\,dy \right)^{m-u} f(z)\,dz, \end{aligned}$$

where

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}.$$

Consider the value of the power function under the assumption that the hypothesis is true. Then $\mu = \nu$ and from (11) the significance level of the test is given by

$$\text{(12)}\quad \begin{aligned} \alpha(u, m) &= \frac{m!}{(u-1)!(m-u)!} \\ &\cdot \int_{0}^{\infty} \left(\int_{-\infty}^{z} f(y)\,dy \right)^{u-1} \left(\int_{z}^{\infty} f(y)\,dy \right)^{m-u} f(z)\,dz. \end{aligned}$$

The method used to eliminate σ from the quantities required for the application of the significance test, therefore, is to have the limits 0 and ∞ in the probability expression (10) for the power function when the hypothesis is true. Suitable significance levels are obtained by varying the statistical function $r_{(u)}$ by means of the selection of the values of u and m.

5. Comparison with Student t-test. The test considered is that of a single sample value on the basis of m other sample values. Hence, the corresponding Student t-test has $m - 1$ degrees of freedom. The probabilities of Type II errors for the Student t-tests are calculated for values of

$$\delta = \frac{\mu - \nu}{\sigma \sqrt{1 + \frac{1}{m}}}$$

by use of the normal approximation given in [2].

Using this notation

$$\frac{K_2}{K_1 \sigma} (\mu - \nu) = \delta \frac{K_2}{K_1} \sqrt{1 + \frac{1}{m}}$$

and from (11) the power function for the significance test for which the alternative is $\mu < \nu$ and $K_2 > 0$ is found to be

$$\frac{m!}{(u-1)!(m-u)!} \int_{\delta(K_2/K_1)\sqrt{1+(1/m)}}^{\infty} \left(\int_{-\infty}^{z} f(y)\,dy \right)^{u-1} \left(\int_{z}^{\infty} f(y)\,dy \right)^{m-u} f(z)\,dz.$$

The probability of a Type II error for a given value of δ is equal to one minus the value of the power function for this value of δ.

It can be proved that the other three significance tests have the same probabilities of Type II errors as the one considered above.

The numerical comparison of the two types of tests is contained in Table I. In each instance the significance level was chosen to be approximately .01.

The process of increasing the size of each sample by a given percentage has practical meaning if each of $x, y_1, \cdots, y_m$ equals the sum of r sample values. For example, if $x, y_1, \cdots, y_m$ each consist of the sum of ten sample values, increasing the sample size by 30% would amount to letting $x, y_1, \cdots, y_m$ each equal the sum of thirteen sample values. The case where each of $x, y_1, \cdots, y_m$ equals the sum of r sample values will be treated later and will be shown to be a particular case of the one analyzed above.

TABLE I

Test	Degrees of Freedom	m	% Increase in Sample Size	Significance Level	Probability of Type II Error			
					$\delta = -1$	$\delta = -2$	$\delta = -3$	$\delta = -4$
t	5		0	.0156	.919	.750	.477	.215
O.S.		6	0	.0156	.919	.752	.506	.276
O.S.		6	5	.0156	.916	.742	.486	.256
O.S.		6	10	.0156	.914	.732	.469	.239
t	9		0	.0107	.930	.735	.413	.142
O.S.		10	0	.0107	.936	.782	.527	.270
O.S.		10	20	.0107	.927	.738	.448	.191
O.S.		10	30	.0107	.921	.715	.411	.161
t	12		0	.0110	.920	.699	.358	.106
O.S.		13	0	.0110	.933	.771	.492	.245
O.S.		13	30	.0110	.919	.717	.378	.139
O.S.		13	40	.0110	.913	.679	.353	.119
t	15		0	.0107	.919	.688	.337	.092
O.S.		16	0	.0107	.938	.765	.488	.234
O.S.		16	40	.0107	.917	.687	.351	.111
O.S.		16	50	.0107	.912	.664	.310	.090

In Table I the order statistic tests (O.S.) are calculated for cases where the size of each sample is increased by the same percentage. This amounts to saying that the amount of information used for the test has been increased by this percentage. This method furnishes a quantitative estimate of the relative efficiency of the order statistic test as compared with the corresponding Student t-test. For example, if 30% more information is required for the order statistic test to have the same probabilities of Type II errors as the corresponding Student

t-test, then the order statistic test will be said to have a relative efficiency of $\frac{1}{1.3} = 77\%$.

Examination of Table I shows that the order statistic tests have the approximate relative efficiencies listed in Table II. These relative efficiencies can be shown to be approximately the same as those for other significance levels.

6. Computation required. Since application of the order statistic test requires only the determination of one order statistic, the calculation of one sum, the multiplication of each of these quantities by given constants and the subtraction of the resulting values, the amount of computation required for application of the order statistic test is obviously much less than is necessary for the application of the corresponding Student t-test.

If the test is applied continuously from one sample to the next, as in quality control work, the value of $\sum_1^m y_i$ can be calculated by a continuous process. For let the sample values be taken in the order $y_1, \cdots, y_m, x$, where x is the new sample value which is to be tested on the basis of the previous m sample values $y_1, \cdots, y_m$. Then x for the present test becomes y_m for the next test; y_m becomes y_{m-1}; $\cdots$; y_2 becomes y_1, and y_1 for the present test is no longer used. The value of x will be furnished by the next sample value drawn. Thus, $\sum_1^m y_i$ for the next test is calculated by adding $x - y_1$ for the present test to $\sum_1^m y_i$ for the present test. The order statistic can be easily determined from a plot of the sample values which is also applied continuously from one sample to the next.

TABLE II

m	Significance Level	% Increase in Sample Size	Relative Efficiency
6	.0156	5	95%
10	.0107	25	80%
13	.0110	35	74%
16	.0107	43	70%

7. Generalization of results. The derivations given above are immediately applicable to the case where x represents the sum of r sample values from a population with distribution $N(\nu', \sigma'^2)$, and each y_i, $(i = 1, \cdots, m)$, equals the sum of r sample values from a population with distribution $N(\mu', \sigma'^2)$. Then x would be distributed according to $N(r\nu', r\sigma'^2)$ and the y_i would be distributed according to $N(r\mu', r\sigma'^2)$. These distributions are of the form $N(\nu, \sigma^2)$ and $N(\mu, \sigma^2)$, where $\mu = r\mu'$, $\nu = r\nu'$ and $\sigma^2 = r\sigma'^2$.

If x equals the sum of r sample values from a population with distribution $N(\nu, \sigma^2)$ and each y_i, $(i = 1, \cdots, m)$, equals the sum of s sample values from a population with distribution $N(\mu, \sigma^2)$, the significance tests are derived in a similar manner and can still be stated in the forms (5) and (6), but the values of K_1 and K_2 become

$$K_1 = m + \frac{r}{s} \pm \sqrt{\frac{r}{s}\left(m + \frac{r}{s}\right)}, \qquad K_2 = -\sqrt{\frac{r}{s}} \mp \sqrt{m + \frac{r}{s}}.$$

The power function for the test in which the alternative is $\mu < \nu$ and $K_2 > 0$ is found by replacing $\frac{K_2}{K_1 \sigma}(\mu - \nu)$ by $\frac{K_2\sqrt{r}}{K_1 \sigma}(\mu - \nu)$ in (11). The significance level of each of the four tests is again furnished by (12) and it can be shown that each test has the same power.

To this point all significance tests considered have consisted of testing a new sample on the basis of m previous samples used as order statistics. In some cases, however, it may be desirable to utilize additional samples in the test but not as order statistics. These sample values can be gathered together in a summation term in which values from different samples are given relative numerical weighting. This procedure can be used to emphasize those sample values which appear to be more important from practical consideration with relation to those which seem to have less importance. The determination of what relative weighting scheme to use is to be decided by the person applying the test and is not considered as a problem of this paper. The significance tests with this property can be stated as follows:

Let each of x_a, y_{ib}, z_{jc}, $(a = 1, \cdots, r; b = 1, \cdots, s; c = 1, \cdots, n_j; i = 1, \cdots, m; j = 1, \cdots, n)$, be distributed independently of all the others, the x_a according to $N(\nu, \sigma^2)$ and the y_{ib} and z_{jc} according to $N(\mu, \sigma^2)$. Define $y_u = \sum_{b=1}^{s} y_{ub}$, $(u = 1, \cdots, m)$, and let $y_{(u)}$ be the uth largest of $y_1, \cdots, y_m$. The one-sided significance tests are then given by

If

$$\sum_1^r x_a > \frac{-\sqrt{r}}{K_2} V_{m+1-u} \qquad (K_2 > 0)$$

$$\sum_1^r x_a > \frac{-\sqrt{r}}{K_2} V_u \qquad (K_2 < 0)$$

accept the alternative $\mu < \nu$, otherwise accept $\mu = \nu$.

If

$$\sum_1^r x_a < \frac{-\sqrt{r}}{K_2} V_u \qquad (K_2 > 0)$$

$$\sum_1^r x_a < \frac{-\sqrt{r}}{K_2} V_{m+1-u} \qquad (K_2 < 0),$$

accept $\mu > \nu$, otherwise accept $\mu = \nu$.

The quantity V_u is given by

$$V_u = \frac{1}{\sqrt{s}} \sum_{i=1}^{m} \sum_{b=1}^{s} y_{ib} - \frac{K_1}{\sqrt{s}} y_{(u)} + \sum_{j=1}^{n} \frac{C_j}{\sqrt{n_j}} \left(\sum_{c=1}^{n_j} z_{jc} \right),$$

where the constants C_j, $(j = 1, \cdots, n)$, are defined by $C_j = w_j \eta$, the w_j being given positive weights. The values of η, K_1 and K_2 are

$$\eta = \frac{\frac{A}{B} \sqrt{\frac{r}{s}} K_2}{m + A^2/B}, \qquad K_1 = \frac{m}{m + A^2/B} \left(m + \frac{A^2}{B} + \sqrt{\frac{r}{s}} K_2 \right),$$

$$K_2 = \frac{m + A^2/B}{\left[B(m + A^2/B)^2 + A^2 \left(\frac{r}{s} \right) \right]} \cdot \left\{ Bm \sqrt{\frac{r}{s}} \pm \sqrt{B^2 m^2 \left(\frac{r}{s} \right) + Bm \left[B(m + A^2/B)^2 + A^2 \left(\frac{r}{s} \right) \right]} \right\},$$

where

$$A = \frac{1}{\sqrt{s}} \sum_{1}^{n} w_j \sqrt{n_j}, \qquad B = \sum_{1}^{n} w_j^2.$$

The quantity η in the expressions for the C_j is not considered given but is determined in the derivation of the tests. The two equations corresponding to (8) and (9) then contain three undetermined quantities η, K_1 and K_2. Thus there are infinitely many possible selections of these quantities, each selection resulting in a valid significance test. The values of η, K_1 and K_2 given above, however, are the ones which result in the maximum power function and consequently the smallest probabilities of Type II errors. The power function for the test in which the alternative is $\mu < \nu$ and $K_2 > 0$ is that given in (11) with $\frac{K_2}{K_1 \sigma} \cdot (\mu - \nu)$ replaced by $\frac{K_2 \sqrt{r}}{K_1 \sigma} (\mu - \nu)$. The significance level of each of the four tests given above is still that of (12). It can also be shown that each of the tests has the same probabilities of Type II errors.

REFERENCES

[1] S. S. Wilks, *Mathematical Statistics*, Princeton Univ. Press, 1943, p. 90.

[2] N. L. Johnson and B. L. Welch, "Applications of the non-central t-distribution", *Biometrika*, Vol. 31 (1940), p. 376.