NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Report No. 32-755

Tests of Hypotheses and Estimation of the Correlation Coefficient Using Quantiles, II

Isidore Eisenberger

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

September 15, 1965

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Report No. 32-755

Tests of Hypotheses and Estimation of the Correlation Coefficient Using Quantiles, II

Isidore Eisenberger

Richard Goldstein

Richard Goldstein, Manager
Communications Systems Research Section

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

September 15, 1965

Prepared Under Contract No. NAS 7-100
National Aeronautics & Space Administration

CONTENTS

TABLES

TABLES (Cont'd)

FIGURES

N65-34902

ABSTRACT

Further results are presented of the investigation into the use of quantiles in data compression of space telemetry. Tests of hypotheses are given using one, two, and four optimum sample quantiles. In Test $\bar{A}$, one decides whether the mean of a normal population has a value of μ_1 or a value of μ_2 when the variance is unknown. Test $\bar{D}$ decides whether the unknown means of two normal populations are identical when the common variance is unknown. Test $\bar{E}$ decides whether the unknown variances of two normal populations are identical when the common mean and variance are unknown. Test $\bar{F}$ decides whether or not two normal populations are independent when their common mean and variance are unknown. In addition, estimators of the correlation coefficient are constructed. Sub-optimum test statistics and estimators using the same four quantiles are also given. In all cases, the sample sizes are assumed to be large.

Author

I. INTRODUCTION

This Report presents further results of the continuing investigation into the use of sample quantiles for data compression of space telemetry. The first results of this investigation are given in Ref. 1, where estimators of the mean of a normal distribution using $1, 2, 3, 4, 6, \cdots, 20$ optimal quantiles and estimators of the standard deviation using $2, 4, 6, \cdots, 20$ optimal quantiles are constructed under various constraints. In addition, two goodness-of-fit tests are devised, one designed for high power against bimodal distributions. The next results are given in Ref. 2, where six tests of simple hypotheses using quantiles are described and estimators of the correlation between two normal populations are constructed.

Although entirely self-contained, this Report in effect continues the discussion of four of the six tests of hypotheses considered in Ref. 2, those designated there as Tests A, D, E, and F. Corresponding tests (designated for the purpose of comparison as Tests $\bar{A}$, $\bar{D}$, $\bar{E}$, and $\bar{F}$) will be given for which, in general, less knowledge of the parameters of the population distributions will be assumed.

In Test A, we were given a set of n independent observations from a normal population with known variance σ^2; the test was designed to decide whether the mean μ had a value of μ_1 or a value of μ_2. In Test $\bar{A}$, the assumption that σ^2 is known is not used.

In Tests D and E, it was assumed that we are given sets of independent samples taken from two independent normal populations; the following two problems were considered:

1a. If $\sigma = \sigma_1 = \sigma_2$ is known and μ_1 not known, is $\mu_2 = \mu_1$ or is $\mu_2 = \mu_1 + \theta$, $\theta \neq 0$?

2a. If μ_1 and μ_2 are known and σ_1 is not known, is $\sigma_2 = \sigma_1$ or is $\sigma_2 = \theta\sigma_1$, $\theta > 0$?

For Tests $\bar{D}$ and $\bar{E}$, the corresponding problems that will be considered are:

1b. If both $\sigma = \sigma_1 = \sigma_2$ and μ_1 are not known, is $\mu_2 = \mu_1$ or is $\mu_2 = \mu_1 + \theta$, $\theta \neq 0$?

2b. If both $\mu = \mu_1 = \mu_2$ and σ_1 are not known, is $\sigma_2 = \sigma_1$ or is $\sigma_2 = \theta\sigma_1$, $\theta > 0$?

In Test F, we assumed n independent pairs of observations taken from two normal populations with known means and variances and tested for independence. In addition, estimators of the correlation coefficient ρ were constructed. In Test $\bar{F}$, the assumption will be that $\mu_1 = \mu_2 = \mu$ and $\sigma_1 = \sigma_2 = \sigma$ and, with one exception, that both μ and σ are unknown. Estimators of ρ will be constructed assuming that σ_1 and σ_2 are known but that $\mu = \mu_1 = \mu_2$ is unknown. All tests and estimators will be based on sample quantiles. Test $\bar{A}_i$ will denote Test $\bar{A}$ using i quantiles, Test $\bar{D}_i$ will denote Test $\bar{D}$ using i quantiles, and so on. In all cases the sample sizes are assumed to be large ($\gtrsim 200$).

The power functions P_0 of the quantile tests and the power functions P'_0 of the best tests using all the sample values are derived. The efficiencies of the quantile tests, defined as P_0/P'_0 are determined. The efficiencies, $\mathrm{Var}(\hat{\rho})/\mathrm{Var}(r)$, of $\hat{\rho}$, the estimators of ρ using quantiles, are also determined for the special case $\rho = 0$, where r is the sample correlation coefficient.

Table 1 describes Tests $\bar{A}$, $\bar{D}$, $\bar{E}$, and $\bar{F}$ and lists the assumptions that are made in connection with each test and with the estimation of ρ.

When applied to data resulting from space experiments, the transmission to Earth of a comparatively small number of sample quantiles instead of all the sample values for the purpose of statistical estimation and prediction can result in a significant amount of data compression, particularly when the sample size is large. The advantage to be gained by this procedure, however, depends upon two factors. The first consideration is the amount of information that is lost in using sample quantiles instead of all the sample values. The second consideration is the amount and complexity of the equipment aboard the spacecraft necessary to select the required quantiles, as compared with the equipment which could perform the same statistical analyses aboard the spacecraft using all the sample values as would be performed on Earth using quantiles. If the loss in information (defined according to reasonable criteria) is large, the price that one would have to pay for data compression in terms of loss in precision may be higher than one could afford to pay. Moreover, even if the loss in information were not excessive, it is evident that if the mechanization of a quantile system aboard the spacecraft were not relatively simple, there would be little justification for using quantiles for data compression of space telemetry.

Table 1. Hypotheses and assumptions relating to Tests $\bar{A}$, $\bar{D}$, $\bar{E}$, and $\bar{F}$, and assumptions relating to estimating ρ

Test	Null hypothesis	Alternative hypothesis	Assumptions
$\bar{A}$	$g(x) = N(\mu_1, \sigma)$	$g(x) = N(\mu_2, \sigma)$	σ unknown
$\bar{D}$	$g_1(x) = N(\mu, \sigma)$ $g_2(y) = N(\mu, \sigma)$	$g_1(x) = N(\mu_1, \sigma_1)$ $g_2(y) = N(\mu + \theta, \sigma)$ $\theta \neq 0$	x and y independent; μ and σ unknown
$\bar{E}$	$g_1(x) = N(\mu, \sigma)$ $g_2(y) = N(\mu, \sigma)$	$g_1(x) = N(\mu, \sigma)$ $g_2(y) = N(\mu, \theta\sigma)$ $\theta > 0$	x and y independent; μ and σ unknown
$\bar{F}$	$g_1(x) = N(\mu_1, \sigma_1)$ $g_2(y) = N(\mu_2, \sigma_2)$ $\rho = 0$	$g_1(x) = N(\mu_1, \sigma_1)$ $g_2(y) = N(\mu_2, \sigma_2)$ $\rho \neq 0$	For one pair of quantiles, μ_1 and μ_2 known; $\sigma = \sigma_1 = \sigma_2$ unknown. For two and four pairs of quantiles, $\mu = \mu_1 = \mu_2$ and $\sigma = \sigma_1 = \sigma_2$ unknown
	Estimating ρ		$\mu = \mu_1 = \mu_2$ unknown; σ_1 and σ_2 known

In a detailed discussion relative to the quantile estimators given in Ref. 1, Edward C. Posner, in Ref. 3, shows that a significant amount of data compression accompanied by high efficiencies can be achieved using quantiles. He also suggests several possible applications of their use in civilian technology. In addition, a design of a quantile system, called a Quantiler, which uses no arithmetic operations is described in Ref. 4, and has since been built. Thus, it has been shown that even if quantiles were to be used only for estimating population parameters, the advantage to be gained in terms of data compression is not offset by the possible disadvantages given previously.

Statistical analysis, however, is not confined to parameter estimation. The results obtained in Ref. 2 and in this Report show that quantile test statistics can be constructed which are as efficient as quantile estimators. This fact indicates that the possible uses of sample quantiles in statistical analysis have not been exhausted. Since the advantage gained in terms of data compression is augmented with each new statistical use that is found for quantiles, every area of statistic should be investigated to determine how useful the substitution of sample quantiles for the entire set of samples will prove to be.

II. REVIEW OF QUANTILES

To define a quantile, consider a sample of n independent values, $x_1, x_2, \cdots, x_n$ taken from a distribution of a continuous type with distribution function $G(x)$ and density function $g(x)$. The pth quantile, or the quantile of order p of the distribution or population, denoted by ζ_p, is defined as the root of the equation $G(\zeta_p) = p$; that is

$$p = \int_{-\infty}^{\zeta_p} dG(x) = \int_{-\infty}^{\zeta_p} g(x)\,dx$$

The corresponding *sample* quantile z_p is defined as follows: If the sample values are arranged in nondecreasing order of magnitude

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$

then $x_{(i)}$ is called the ith order statistic and

$$z_p = x_{([np]+1)}$$

where $[np]$ is the greatest integer $\leq np$.

If $g(x)$ is differentiable in some neighborhood of each quantile value considered, it has been shown (Ref. 5) that the joint distribution of any number of quantiles is asymptotically normal as $n \to \infty$ and that, asymptotically,

$$E(z_p) = \zeta_p$$

$$\operatorname{Var}(z_p) = \frac{p(1-p)}{ng^2(\zeta_p)}$$

$$\rho_{12} = \left[\frac{p_1(1-p_2)}{p_2(1-p_1)}\right]^{1/2}$$

where ρ_{12} is the correlation between z_{p_1} and z_{p_2} and $p_1 < p_2$.

Throughout this Report, we will denote by $F(x)$ and $f(x) = F'(x)$ the distribution function and density function, respectively, of the standard normal distribution; that is

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

where

$$f(x) = \frac{1}{(2\pi)^{1/2}} e^{-1/2 x^2}$$

The statement "$g(x) = N(\mu, \sigma)$" will mean that the random variable under consideration is normally distributed with mean μ and variance σ^2 and has the density function $g(x)$ associated with it. For simplification, when only one quantile is being considered, the sample quantile of order p will be denoted by z, the corresponding population quantile by ζ^*, and the corresponding population quantile of the standard normal distribution by ζ. Thus, one has

$$p = \int_{-\infty}^{\zeta^*} g(x)\,dx = \frac{1}{\sigma(2\pi)^{1/2}} \int_{-\infty}^{\zeta^*} e^{-1/2[(x-\mu)^2/\sigma^2]}\,dx$$

$$= \int_{-\infty}^{(\zeta^*-\mu)/\sigma} f(x)\,dx = \int_{-\infty}^{\zeta} f(x)\,dx$$

Hence, one sees that, asymptotically,

$$E(z) = \zeta^* = \sigma\zeta + \mu$$

and, since $g(\zeta^*) = (1/\sigma) f(\zeta)$,

$$\operatorname{Var}(z) = \frac{\sigma^2 F(\zeta)[1 - F(\zeta)]}{nf^2(\zeta)}$$

so that the moments of the sample quantiles of normal distributions are expressable in terms of the standard normal distribution. When m quantiles are being considered, the sample quantiles will be denoted by z_i of order p_i, $i = 1, 2, \cdots, m$, and $p_i < p_j$ for $i < j$. ζ_i will denote the corresponding population quantiles of the standard normal. Since n is assumed to be large, the statistical analyses to be given in the sequel will be based on the asymptotic normal distribution of the sample quantiles.

III. TEST $\bar{A}$. TESTING THE MEAN OF A NORMAL DISTRIBUTION USING QUANTILES WHEN σ IS UNKNOWN

A. Test $\bar{A}_1$

Suppose one is given a set of n sample values taken from a normally distributed population with density function $g(x)$, and one wishes to test the null hypothesis:

$$H_0: \quad g(x) = g_1(x) = N(\mu_1, \sigma), \qquad \sigma \text{ unknown}$$

against the alternative hypothesis:

$$H_1: \quad g(x) = g_1(x) = N(\mu_2, \sigma), \qquad \sigma \text{ unknown}$$

where $\mu_2 > \mu_1$ ($\mu_2 < \mu_1$). If $\mu_1 = 0$, this would correspond to the problem of detecting a DC signal of known amplitude in the presence of additive stationary Gaussian noise with unknown noise power σ. The tests given here will be based on one, two, and four sample quantiles.

Beginning with one quantile, let z denote the sample quantile of order p. Then one has

Under H_0:

$$E(z) = \sigma\zeta + \mu_1, \qquad \text{Var}(z) = \sigma^2 a^2$$

where

$$a^2 = \frac{F(\zeta)[1 - F(\zeta)]}{nf^2(\zeta)}$$

Under H_1:

$$E(z) = \sigma\zeta + \mu_2, \qquad \text{Var}(z) = \sigma^2 a^2$$

If σ were known, the best critical (or rejection) region determined by the likelihood ratio inequality (Ref. 6, p. 166) is given by

$$\begin{aligned} z &> \mu_1 + \sigma(ab + \zeta), \qquad \mu_2 > \mu_1 \\ z &< \mu_1 - \sigma(ab + \zeta), \qquad \mu_2 < \mu_1 \end{aligned} \tag{1}$$

where $F(-b) = \varepsilon$, the significance level of the test (that is, ε is the probability of rejecting H_0 when H_0 is true). The order of z can be chosen arbitrarily but, in the absence of compelling reasons to the contrary, should be chosen to maximize P_0, the power of the test (that is, the probability of rejecting H_0 when H_0 is false). It is obvious, however, that a rejection region of this form is useless unless σ is known. What is needed is a rejection criterion for which the probability of occurrence when H_0 is true can be calculated when σ is *unknown*. A rejection region which can meet this condition is given by

$$\begin{aligned} z &> \mu_1, \qquad \mu_2 > \mu_1 \\ z &< \mu_1, \qquad \mu_2 < \mu_1 \end{aligned} \tag{2}$$

Now the order of z is no longer arbitrary, but must be chosen such that the probability of Ineq. (2) occurring is equal to ε. This is accomplished as follows:

Under H_0, for $\mu_2 > \mu_1$:

$$\Pr(z < \mu_1) = F\left(\frac{\mu_1 - \mu_1 - \sigma\zeta}{\sigma a}\right) = F\left(\frac{-\zeta}{a}\right)$$
$$= F(b) = 1 - \varepsilon$$

Thus, if p can be chosen so that $-\zeta/a = b$, the probability of Ineq. (2) occurring will be equal to ε, as required. Now for fixed n, $u = u(p) = -\zeta/a$ is a function only of p with the following properties for $0 < p < 1$. $u(p) = -u(1 - p)$, $u(p) \to 0$ as $p \to$ zero and one, and $u(p)$ attains a maximum of $0.78\, n^{1/2}$ at $p = 0.0576$ and a minimum of $-0.78\, n^{1/2}$ at $p = 0.9424$. Thus, for $n \gtrless 200$, since $u(p)$ is continuous on $0 < p < 1$, $u(p)$ takes on all values between $-0.78\, n^{1/2} \leqq -11.03$ and $0.78\, n^{1/2} \geqq 11.03$. Since $F(-11.03) \cong 0$ and $F(11.03) \cong 1$, it is readily seen that, for all realistic values of ε, values of p can indeed be determined such that $\Pr(z > \mu_1) = \varepsilon$ for $\mu_2 > \mu_1$ and such that $\Pr(z < \mu_1) = \varepsilon$ for $\mu_2 < \mu_1$. The rejection criteria given by Ineqs. (1) and (2) differ essentially in that, in the former case, the order of the statistic z can be fixed for all ε and n, and then the right-hand side depends on ε and n. In the latter case, the right-hand side is fixed for all ε and n, while the order of z depends on both n and ε. It will be seen, however, that this restriction on the order of z will not occur in Tests $\bar{A}_2$ and $\bar{A}_4$, although a different type of restriction will be placed on the statistic used for each of these tests.

The power of the test, which of course depends on the unknown σ, is determined as follows:

Under H_1, for $\mu_2 > \mu_1$:

$$\Pr(z < \mu_1) = F\left(\frac{\mu_1 - \mu_2 - \sigma\zeta}{\zeta a}\right) = F\left(b - \frac{\mu_2 - \mu_1}{\sigma a}\right)$$
$$= 1 - P_0$$

The expression for P_0 is identical with that in Test A where σ is known. In Test A, we were at liberty to choose the median, $z(0.5)$, as this is the sample quantile which maximizes P_0. In the present case, however, the order of z, and hence the value of a^{-1}, is determined by the restriction $-\zeta/a = b$. In order to compare the power functions of Tests A_1 and $\bar{A}_1$, we take, as an example, $n = 200$ and $\varepsilon = 0.01$. Then for Test A, using the median for $\mu_2 > \mu_1$, one has

$$P_0\,(\text{Test } A_1) = 1 - F\left(2.326 - 11.285\,\frac{\mu_2 - \mu_1}{\sigma}\right) \tag{3}$$

For Test $\bar{A}_1$, one uses $z(0.4177)$ since $p = 0.4177$, $\zeta =$ -0.2078, $a^{-1} = 11.1955$, and $-\zeta/a = 2.326$, resulting in

$$P_0\,(\text{Test } \bar{A}_1) = 1 - F\left(2.326 - 11.196\,\frac{\mu_2 - \mu_1}{\sigma}\right) \tag{4}$$

If σ is known, the optimum test statistic using all the sample values is based on their sum, and the power function P_0' of this test is easily determined to be

$$P_0' = 1 - F\left[b - \frac{n^{1/2}(\mu_2 - \mu_1)}{\sigma}\right], \qquad \mu_2 > \mu_1 \tag{5}$$

Although Student's t-test is generally used if σ is unknown, since n is large very little error will be introduced if one uses the expression given by Eq. (5) as the optimum P_0' for Test $\bar{A}_1$, and the results obtained with respect to the efficiency of the test will be conservative. We will therefore adopt this procedure in this and all subsequent tests.

A comparison between the coefficient of $(\mu_2 - \mu_1)/\sigma$ in Eq. (3) and the same coefficient in Eq. (4) shows that very little power is lost by dropping the assumption that σ is unknown. It is true that as ε decreases the value of a increases, thereby decreasing P_0. However, the effect is negligible unless ε is taken to be exceedingly small. For example, for $n = 200$, if $\varepsilon = 0.001$ then $p = 0.3906$, and a^{-1}, the coefficient of $(\mu_2 - \mu_1)/\sigma$ in P_0, is 11.1265, whereas if $\varepsilon = 0.05$, $p = 0.4414$ and $a^{-1} = 11.240$.

Test $\bar{A}_1$ can now be described: For given values of n and ε, choose the sample quantile z of order p such that $-\zeta/a = b$. Then, for $\mu_2 > \mu_1$, if $z < \mu_1$, accept H_0; otherwise, reject H_0. For $\mu_2 < \mu_1$, if $z > \mu_1$, accept H_0; otherwise, reject H_0. It should be noted that for $\mu_2 > \mu_1$, $p < 0.5$, $\zeta < 0$; and for $\mu_2 < \mu_1$, $p \geq 0.5$, $\zeta > 0$. Fig. 1 gives the power and efficiency of Test $\bar{A}_1$ and the efficiency of Test A_1 for $n = 200$, $\varepsilon = 0.01$.

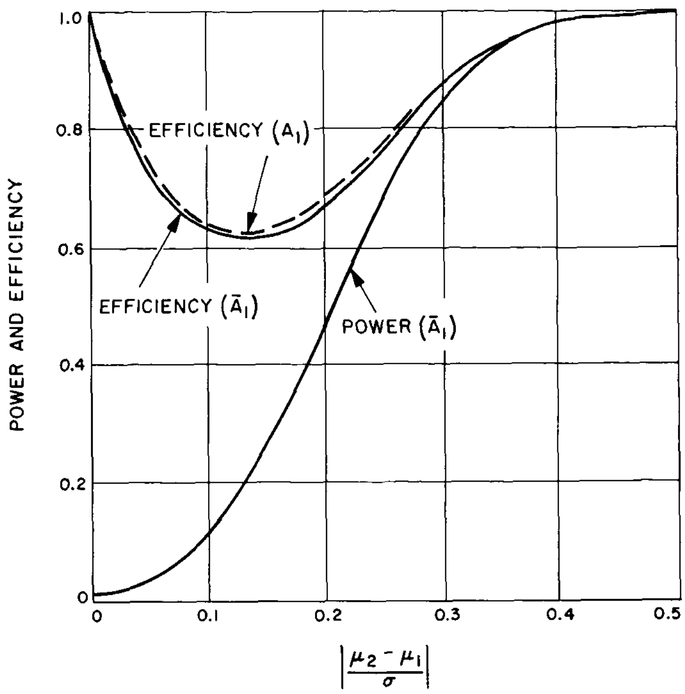

Fig. 1. Power and efficiency of Test $\bar{A}_1$ and efficiency of Test A_1, for $n = 200$, $\varepsilon = 0.01$

B. Test $\bar{A}_2$

Let z_1 and z_2 denote the sample quantiles of orders p_1 and $p_2 = 1 - p_1$. Then one has:

Under H_0:

$$E(z_1) = -\sigma\zeta_2 + \mu_1 \qquad \mathrm{Var}(z_1) = \mathrm{Var}(z_2) = \sigma^2 a^2$$

$$E(z_2) = \sigma\zeta_2 + \mu_1,$$

where

$$a^2 = \frac{F(\zeta_2)\,[1 - F(\zeta_2)]}{nf^2(\zeta_2)}$$

Under H_1:

$$E(z_1) = -\sigma\zeta_2 + \mu_2, \qquad \mathrm{Var}(z_1) = \mathrm{Var}(z_2) = \sigma^2 a^2$$

$$E(z_2) = \sigma\zeta_2 + \mu_2$$

When σ is known, the likelihood ratio inequality gives rejection regions of the form

$$\begin{aligned} z_1 + z_2 &> 2\mu_1 + d_1\sigma, \qquad \mu_2 > \mu_1 \\ z_1 + z_2 &< 2\mu_1 - d_1\sigma, \qquad \mu_2 < \mu_1 \end{aligned} \tag{6}$$

Since the rejection regions again depend upon σ, they cannot be used in the present test. However, since an *estimate* of σ can be obtained from two quantiles of the form $\hat{\sigma} = c(z_2 - z_1)$, we substitute $\hat{\sigma}$ for σ in Ineq. (6), which results, for $\mu_2 > \mu_1$, in a rejection region of the form

$$y = (1 + \alpha) z_1 + (1 - \alpha) z_2 > 2\mu_1 \tag{7}$$

provided that the probability of the occurrence of Eq. (7) can be shown to be equal to ε when σ is unknown. Now

Under H_0:

$$E(y) = (1 + \alpha)(\mu_1 - \sigma\zeta_2) + (1 - \alpha)(\mu_1 + \sigma\zeta_2)$$
$$= 2(\mu_1 - \alpha\sigma\zeta_2)$$

$$\mathrm{Var}(y) = \sigma^2 a^2 [(1 + \alpha)^2 + (1 - \alpha)^2 + 2(1 - \alpha)(1 + \alpha)\rho]$$
$$= 2\sigma^2 a^2 [1 + \rho + \alpha^2(1 - \rho)]$$

where ρ denotes the correlation between z_1 and z_2.

$$\Pr(y < 2\mu_1) = F\left\{\frac{2\mu_1 - 2\mu_1 + 2\alpha\sigma\zeta_2}{2^{1/2}\sigma a[1 + \rho + \alpha^2(1 - \rho)]^{1/2}}\right\}$$
$$= F\left\{\frac{2^{1/2}\alpha\zeta_2}{a[1 + \rho + \alpha^2(1 - \rho)]^{1/2}}\right\}$$
$$= F(b) = 1 - \varepsilon$$

It is thus seen that the probability of Eq. (7) occurring is independent of σ, and the orders of the symmetric quantiles z_1 and z_2 are subject only to the restriction that

$$\alpha^2 = \frac{a^2 b^2 (1 + \rho)}{2\zeta_2^2 - b^2 a^2 (1 - \rho)} > 0 \tag{8}$$

To determine P_0, one has

Under H_1:

$$E(y) = 2(\mu_2 - \alpha\sigma\zeta_2)$$

$$\mathrm{Var}(y) = 2\sigma^2 a^2 [1 + \rho + \alpha^2(1 - \rho)]$$

so that

$$\Pr(y < 2\mu_1) = F\left\{\frac{2\mu_1 - 2\mu_2 + 2\alpha\sigma\zeta_2}{2^{1/2}\sigma a[1 + \rho + \alpha^2(1 - \rho)]^{1/2}}\right\}$$
$$= F\left\{b - \frac{2^{1/2}(\mu_2 - \mu_1)}{\sigma a[1 + \rho + \alpha^2(1 - \rho)]^{1/2}}\right\}$$
$$= 1 - P_0 \tag{9}$$

Substituting in Eq. (9) the value of α^2 given by Eq. (8) and putting $a^2 = \bar{a}^2/n$ results in

$$P_0(\text{Test } \bar{\mathrm{A}}_2) = 1 - F\left\{b - \frac{(2n)^{1/2}\left(\frac{\mu_2 - \mu_1}{\sigma}\right)}{\bar{a}\left[(1 + \rho)\left(1 + \frac{1}{\frac{2\zeta_2^2 n}{\bar{a}^2 b^2 (1 - \rho)} - 1}\right)\right]^{1/2}}\right\}$$

Comparing the above expression with

$$P_0(\text{Test A}_2) = 1 - F\left[b - \frac{(2n)^{1/2}\left(\frac{\mu_2 - \mu_1}{\sigma}\right)}{\bar{a}(1 + \rho)^{1/2}}\right]$$

one sees that P_0 (Test $\bar{\mathrm{A}}_2$) $\to P_0$ (Test A_2) asymptotically as $n \to \infty$. One also sees that for fixed n, the coefficient of $(\mu_2 - \mu_1)/\sigma$ in P_0 (Test A_2) depends only upon the orders of the quantiles chosen for the test, while the coefficient in P_0 (Test $\bar{\mathrm{A}}_2$) depends also upon b (which depends on ε). However, for large n the dependence upon b of the coefficient in P_0 (Test $\bar{\mathrm{A}}_2$) is not very sensitive. For example, for $n = 200$, if one uses the quantiles which maximize P_0 (Test A_2) those of orders $p_1 = 0.2703$, $p_2 = 0.7297$, the coefficient of $(\mu_2 - \mu_1)/\sigma$ in P_0 (Test $\bar{\mathrm{A}}_2$) is equal to 12.462 for $\varepsilon = 0.01$, 12.595 for $\varepsilon = 0.05$, and 12.257 for $\varepsilon = 0.001$. In P_0 (Test A_2), this coefficient equals 12.728.

Although the use in Test $\bar{\mathrm{A}}_2$ of the quantiles which maximize P_0 (Test A_2) is not strictly optimum, the loss in power by doing so is negligible and the advantage to be gained by standardizing the test is obvious. Accordingly, for $p_1 = 0.2703$, $p_2 = 0.7207$, one has

$$\alpha^2 = \frac{2.4699b^2}{0.7491n - 1.1347b^2} \tag{10}$$

For $\mu_2 > \mu_1$,

$$F\left\{\frac{2^{1/2}\alpha\zeta_2}{a[1 + \rho + \alpha^2(1 - \rho)]^{1/2}}\right\} = F(b) = 1 - \varepsilon$$

and hence for the usual small values of ε, $b > 0$ and, since $\zeta_2 > 0$, α must be positive, so one must use the positive root of α^2. For $\mu_2 < \mu_1$,

$$F\left\{\frac{2^{1/2}\alpha\zeta_2}{a[1 + \rho + \alpha^2(1 - \rho)]^{1/2}}\right\} = F(-b) = \varepsilon$$

Therefore, in this case one must use the negative root of α^2. Thus, the test using two symmetric quantiles can be expressed as follows. For $\mu_2 > \mu_1$, if $[(1 + \alpha)/2]\, z(0.2703) + [(1 - \alpha)/2]\, z(0.7297) < \mu_1$, accept H_0; otherwise, reject H_0. For $\mu_2 < \mu_1$, if $[(1 - \alpha)/2]\, z(0.2703) + [(1 + \alpha)/2]\, z(0.7297) > \mu_1$, accept H_0; otherwise, reject

H_0. α^2 is given by Eq. (10) and it is readily seen that for all realistic values of ε, $\alpha^2 > 0$ even for moderate values of n. In order to illustrate the small loss in power in using the above values of p_1 and p_2 instead of the optimum values, for $n = 200$, $\varepsilon = 0.01$, the optimum quantiles for Test $\bar{A}_2$ are those of orders $p_1 = 0.2498$, $p_2 = 0.7502$. Using these values, the coefficient of $(\mu_2 - \mu_1)/\sigma$ in P_0 (Test $\bar{A}_2$) equals 12.475 as compared to 12.462 for $p_1 = 0.2703$, $p_2 = 0.7297$. Nevertheless, the optimum quantiles *can* be determined with a small amount of effort for fixed values of b and n, and then the value of α^2 is given by Eq. (8).

For $n = 200$, $\varepsilon = 0.01$, $\alpha = 0.3050$ and the *acceptance* regions are given by

$$0.6525\ z(0.2703) + 0.3475\ z(0.7297) < \mu_1, \qquad \mu_2 > \mu_1$$

$$0.3475\ z(0.2703) + 0.6525\ z(0.7297) > \mu_1, \qquad \mu_2 < \mu_1$$

Fig. 2 gives the power and efficiency of Test $\bar{A}_2$ using the optimum quantiles and the efficiency of Test A_2 for $n = 200$, $\varepsilon = 0.01$.

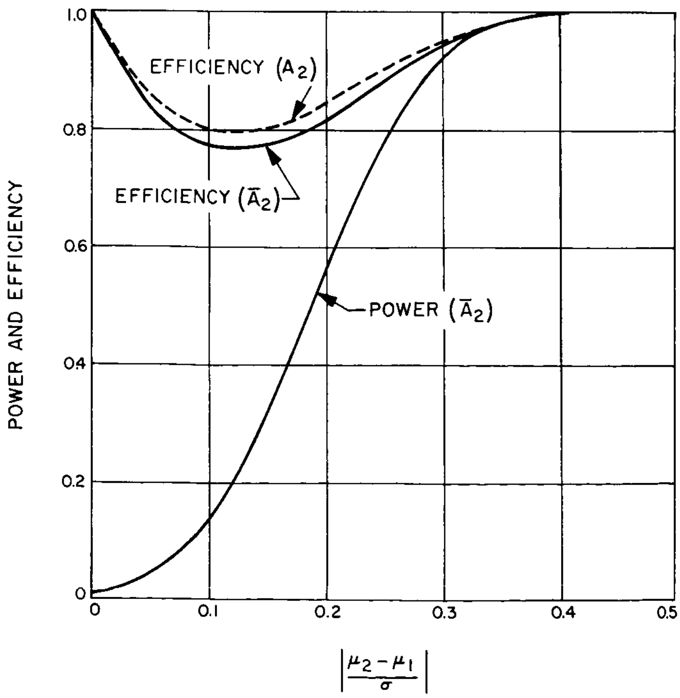

Fig. 2. Power and efficiency of Test $\bar{A}_2$ and efficiency of Test A_2, for $n = 200$, $\varepsilon = 0.01$

C. Test $\bar{A}_4$

Let z_i, $i = 1, 2, 3, 4$ denote four sample quantiles such that $p_1 + p_4 = p_2 + p_3 = 1$. The best statistic in Test A_4 was found to be a linear combination of the z_i, and the rejection regions, determined by the likelihood ratio inequality, were of the form

$$0.192\,(z_1 + z_4) + 0.308\,(z_2 + z_3) > \mu_1 + d_2\sigma, \qquad \mu_2 > \mu_1$$

$$0.192\,(z_1 + z_4) + 0.308\,(z_2 + z_3) < \mu_1 - d_2\sigma, \qquad \mu_2 < \mu_1 \tag{11}$$

Using the same technique here as in Test $\bar{A}_2$, we substitute for the unknown σ in Ineq. (11) the estimate $\hat{\sigma} = c\,(z_4 - z_1) + c\,(z_3 - z_1)$, and this results; for $\mu_2 > \mu_1$, in the rejection region given by

$$y = (0.192 + \alpha)\,z_1 + (0.192 - \alpha)\,z_4 + (0.308 + \alpha)\,z_2 + (0.308 - \alpha)\,z_3 > \mu_1 \tag{12}$$

where α must be determined such that the probability of Eq. (12) occurring is equal to ε when σ is unknown.

Under H_0:

$$E(z_1) = \mu_1 - \sigma\zeta_4\,, \qquad \mathrm{Var}(z_1) = \mathrm{Var}(z_4) = \sigma^2 a_4^2$$

$$E(z_2) = \mu_1 - \sigma\zeta_3\,, \qquad \mathrm{Var}(z_2) = \mathrm{Var}(z_3) = \sigma^2 a_3^2$$

$$E(z_3) = \mu_1 + \sigma\zeta_3$$

$$E(z_4) = \mu_1 + \sigma\zeta_4$$

where

$$a_i^2 = \frac{F(\zeta_i)\,[1 - F(\zeta_i)]}{nf^2(\zeta_i)}, \qquad i = 3, 4$$

One also has, under H_0,

$$\begin{aligned} E(y) &= (0.192 + \alpha)(\mu_1 - \sigma\zeta_4) + (0.192 - \alpha)(\mu_1 + \sigma\zeta_4) \\ &\quad + (0.308 + \alpha)(\mu_1 - \sigma\zeta_3) + (0.308 - \alpha)(\mu_1 + \sigma\zeta_3) \\ &= \mu_1 - 2\alpha\sigma(\zeta_3 + \zeta_4) \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(y) &= 2\sigma^2\{a_4^2\,[0.03686\,(1 + \rho_{14}) + \alpha^2(1 - \rho_{14})] \\ &\quad + a_3^2\,[0.09486\,(1 + \rho_{23}) + \alpha^2(1 - \rho_{23})] \\ &\quad + 2a_3a_4\,[(0.05914 + \alpha^2)\,\rho_{12} + (0.05914 - \alpha^2)\,\rho_{13}]\} \\ &= 2\sigma^2\gamma^2 \end{aligned}$$

where ρ_{ij} denotes the correlation between z_i and z_j.

$$\Pr(y < \mu_1) = F\left[\frac{2^{1/2}\alpha(\zeta_3 + \zeta_4)}{\gamma}\right] = F(b) = 1 - \varepsilon$$

so that

$$\alpha^2 = \frac{b^2\,[0.03686a_4^2(1+\rho_{14}) + 0.09486a_3^2(1+\rho_{23}) + 0.11827a_3a_4(\rho_{12}+\rho_{13})]}{2(\zeta_3+\zeta_4)^2 - b^2\,[a_4^2(1-\rho_{14}) + a_3^2(1-\rho_{23}) + 2a_3a_4(\rho_{12}-\rho_{13})]} \tag{13}$$

To determine P_0, one has

Under H_1:

$$E(y) = \mu_2 - 2\alpha\sigma(\zeta_3+\zeta_4)$$

$$\mathrm{Var}(y) = 2\sigma^2\gamma^2$$

$$\Pr(y<\mu_1) = F\left[\frac{\mu_1-\mu_2+2\alpha\sigma(\zeta_3+\zeta_4)}{\sigma\gamma\,(2)^{1/2}}\right]$$

$$= F\left[b - \frac{\mu_2-\mu_1}{\sigma\gamma\,(2)^{1/2}}\right] = 1-P_0 \tag{14}$$

The optimum quantiles for Test A_4, which we will also use in Test $\bar{A}_4$, are those of orders

$$p_1 = 0.1068$$

$$p_2 = 0.3512$$

$$p_3 = 0.6488$$

$$p_4 = 0.8932$$

Using these quantiles, one obtains

$$\alpha^2 = \frac{0.5434b^2}{5.2864n - 4.1696b^2} \tag{15}$$

For $n = 200$, $\epsilon = 0.01$, the coefficient of $(\mu_2-\mu_1)/\sigma$ in Eq. (14) is equal to 13.420 compared to 13.562 for the corresponding coefficient in P_0 (Test A_4), so that the power loss in assuming σ is unknown and using in Test $\bar{A}_4$ the optimum quantiles for Test A_4 is again small. For $\mu_2 > \mu_1$, the positive root of α^2 is used in the test statistic; for $\mu_2 < \mu_1$, the negative root is used. The test can thus be given as follows. For $\mu_2 > \mu_1$, if $(0.192+\alpha)\,z(0.1068) + (0.192-\alpha)\,z(0.8932) + (0.308+\alpha)\,z(0.3512) + (0.308-\alpha)\,z(0.6488) < \mu_1$, accept H_0; otherwise, reject H_0. For $\mu_2 < \mu_1$, if $(0.192-\alpha)\,z(0.1068) + (0.192+\alpha)\,z(0.8932) + (0.308-\alpha)\,z(0.3512) + (0.308+\alpha)\,z(0.6488) > \mu_1$, accept H_0; otherwise, reject H_0.

For $n = 200$, $\epsilon = 0.01$, $\alpha = 0.05331$ and the *acceptance* regions are given by

$$0.2445z(0.1068) + 0.1391z(0.8932) + 0.3609z(0.3512) + 0.2555z(0.6488) < \mu_1, \qquad \mu_2 > \mu_1$$

$$0.3191z(0.1068) + 0.2445z(0.8932) + 0.2555z(0.3512) + 0.3609z(0.6488) > \mu_1, \qquad \mu_2 < \mu_1$$

For other values of n and ϵ, α^2 is determined by Eq. (15) if the optimum quantiles for Test A_4 are used in Test $\bar{A}_4$. If the true optimum quantiles for Test $\bar{A}_4$ are determined, then α^2 is given by Eq. (13). However, since P_0 (Test $\bar{A}_4$) is already very nearly equal to P_0 (Test A_4) when the optimum quantiles for Test A_4 are used in both tests and, moreover, P_0 (Test $\bar{A}_4$) cannot exceed the maximum P_0 (Test A_4) under the best of circumstances, from a practical point of view it seems hardly worthwhile determining the optimum quantiles for Test $\bar{A}_4$. Fig. 3 gives the power and efficiency of Test $\bar{A}_4$ and the efficiency of Test A_4, for $n = 200$, $\epsilon = 0.01$.

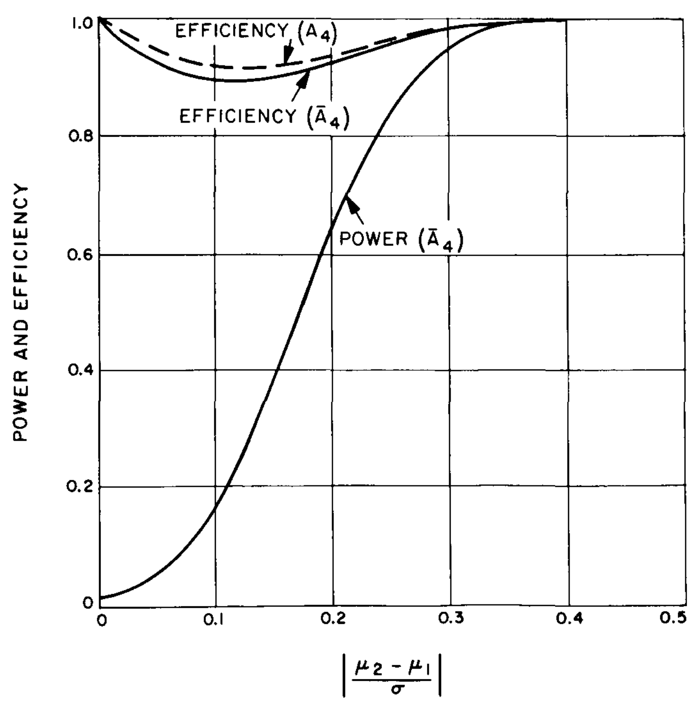

Fig. 3. Power and efficiency of Test $\bar{A}_4$ and efficiency of Test A_4, for $n = 200$, $\epsilon = 0.01$

IV. TESTS $\bar{D}$ AND $\bar{E}$. TWO SAMPLE TESTS

A. Statement of the Problem

In this Section, it is assumed that we are given sets of independent samples taken from two independent normally distributed populations with density functions $g_1(x)$ and $g_2(y)$ and consider the following two tests:

Test $\bar{D}$

$$
\begin{aligned}
H_0: \quad & g_1(x) = N(\mu, \sigma) \\
& g_2(y) = N(\mu, \sigma) \\
H_1: \quad & g_1(x) = N(\mu, \sigma) \\
& g_2(y) = N(\mu + \theta, \sigma), \qquad \theta \neq 0
\end{aligned}
$$

where *both* μ and σ are unknown.

Test $\bar{E}$

$$
\begin{aligned}
H_0: \quad & g_1(x) = N(\mu, \sigma) \\
& g_2(y) = N(\mu, \sigma) \\
H_1: \quad & g_1(x) = N(\mu, \sigma) \\
& g_2(y) = N(\mu, \theta\sigma), \qquad \theta > 0
\end{aligned}
$$

where *both* μ and σ are unknown.

Tests $\bar{D}$ and $\bar{E}$ differ from Tests D and E considered in Ref. 2 in that in Test D σ is assumed known, and in Test E μ_1 and μ_2 are assumed known but not necessarily equal. It should be noted that in Test $\bar{E}$ we have not achieved complete generality, since we are assuming that, although μ_1 and μ_2 are unknown, they are nevertheless equal. For simplicity we will assume identical sample sizes for both sets of samples. Modification of the tests for sample sizes n_1 and $n_2 \neq n_1$ will be obvious. The tests will be based on one, two, and four pairs of sample quantiles, one of each pair from the first set of samples and the other from the second set.

B. Test $\bar{D}_1$. One Pair of Quantiles

Let z denote the sample quantile of order p of the samples taken from the first population, that with density $g_1(x)$, and let z' be the sample quantile of order $1 - p$ of the samples taken from the second population. The test statistic will be $w = z - z'$, and the rejection regions are given by

$$
\begin{aligned}
w &= z - z' < 0, \qquad \text{for } \theta > 0 \\
w &= z - z' > 0, \qquad \text{for } \theta < 0
\end{aligned} \tag{16}
$$

As in Test $\bar{A}_1$, the orders of the quantiles will be chosen so that the probability of Eq. (16) occurring will be equal to ε when μ and σ are unknown.

Under H_0:

$$
\begin{aligned}
E(z) &= \mu + \sigma\zeta, \qquad \mathrm{Var}(z) = \mathrm{Var}(z') = \sigma^2 a^2 \\
E(z') &= \mu - \sigma\zeta, \qquad \mathrm{Var}(w) = 2\sigma^2 a^2 \\
E(w) &= 2\sigma\zeta
\end{aligned}
$$

where

$$
a^2 = \frac{F(\zeta)[1 - F(\zeta)]}{nf^2(\zeta)}
$$

For $\theta > 0$,

$$
\Pr(w < 0) = F\left(\frac{-2\sigma\zeta}{2^{1/2}\sigma a}\right) = F\left(\frac{-2^{1/2}\zeta}{a}\right) = F(b) = \varepsilon \tag{17}
$$

From Eq. (17) it is seen that the order of z must be chosen such that

$$
\frac{-\zeta}{a} = \frac{b}{2^{1/2}} \tag{18}
$$

and that, since $b < 0$, $\zeta > 0$ and $p > 0.5$. To determine P_0, one has

Under H_1:

$$
\begin{aligned}
E(z) &= \mu + \sigma\zeta, \qquad \mathrm{Var}(z) = \mathrm{Var}(z') = \sigma^2 a^2 \\
E(z') &= \mu + \theta - \sigma\zeta, \qquad \mathrm{Var}(w) = 2\sigma^2 a^2 \\
E(w) &= 2\sigma\zeta - \theta
\end{aligned}
$$

and, for $\theta > 0$,

$$
\Pr(w < 0) = F\left(\frac{-2\sigma\zeta + \theta}{2^{1/2}\sigma a}\right) = F\left(b + \frac{\theta}{\sigma a 2^{1/2}}\right) = P_0
$$

The power function of the best test using all the samples when σ is known is given by

$$
P_0' = F\left[b + \frac{\theta}{\sigma}\left(\frac{n}{2}\right)^{1/2}\right]
$$

As in Test $\bar{A}_1$ the condition given by Eq. (18) can be met for all realistic values of ϵ. For $n = 200$ and $\epsilon = 0.01$, $p = 0.5582$. Under these conditions, one has

$$P_0(\text{Test } \bar{D}_1) = F\left(-2.326 + \frac{7.978\theta}{\sigma}\right)$$

$$P_0(\text{Test } D_1) = F\left(-2.326 + \frac{7.948\theta}{\sigma}\right)$$

$$P_0' = F\left(-2.326 + \frac{10\theta}{\sigma}\right)$$

and it is seen again that the power of the test when σ is unknown differs very little from that when σ is known. Test $\bar{D}_1$ can now be described as follows: Choose those sample quantiles z and z' of orders p and $1-p$, respectively, which satisfy Eq. (18) and, for $\theta > 0$, if $z - z' > 0$, accept H_0; otherwise, reject H_0. For $\theta < 0$, $b > 0$ and hence $\zeta < 0$, $p < 0.5$. For this case, if $z - z' < 0$, accept H_0; otherwise, reject H_0. For $n = 200$, $\epsilon = 0.01$, the *acceptance* regions are given by

$$z(0.5582) - z'(0.4418) > 0, \quad \text{for } \theta > 0$$

$$z(0.4418) - z'(0.5582) < 0, \quad \text{for } \theta < 0$$

Table 2 shows the power and efficiency of Test $\bar{D}_1$, for $n = 200$, $\epsilon = 0.01$.

Table 2. Power and efficiency of Test $\bar{D}_1$, and efficiency of the same test with σ known, for $n = 200$, $\epsilon = 0.01$

$\left\lvert\frac{\theta}{\sigma}\right\rvert$	P_0	Efficiency	Efficiency (σ known)
0.01	0.0123	0.9461	0.9538
0.05	0.0269	0.7912	0.7941
0.10	0.0629	0.6807	0.6851
0.15	0.1284	0.6282	0.6336
0.20	0.2307	0.6198	0.6252
0.25	0.3676	0.6479	0.6507
0.30	0.5233	0.6979	0.6990
0.40	0.8032	0.8429	0.8467
0.50	0.9503	0.9538	0.9553

C. Test $\bar{D}_2$. Two Pairs of Quantiles

Let z_1 and z_2 be the sample quantiles of the first population of orders p_1 and $p_2 = 1 - p_1$, and let z_1' and z_2' be the corresponding sample quantiles of the second population. To eliminate dependence on μ and σ, the rejection regions of the test will be taken as

$$y = (1+\alpha)(z_2 - z_1') + (1-\alpha)(z_1 - z_2') \begin{cases} < 0, & \text{for } \theta > 0 \\ > 0, & \text{for } \theta < 0 \end{cases} \tag{19}$$

α will be determined such that the probability of Eq. (19) occurring is equal to ϵ when μ and σ are unknown.

Under H_0:

$$E(z_1) = E(z_1') = \mu - \sigma\zeta_2$$

$$E(z_2) = E(z_2') = \mu + \sigma\zeta_2$$

$$\mathrm{Var}(z_1) = \mathrm{Var}(z_2) = \mathrm{Var}(z_1') = \mathrm{Var}(z_2') = \sigma^2 a^2$$

where

$$a^2 = \frac{F(\zeta_2)[1 - F(\zeta_2)]}{nf^2(\zeta_2)}$$

and

$$E(y) = 2\sigma\zeta_2[(1+\alpha) - (1-\alpha)] = 4\alpha\sigma\zeta_2$$

$$\mathrm{Var}(y) = 2\sigma^2 a^2[(1+\alpha)^2 + (1-\alpha)^2 + \alpha(1-\alpha)(1+\alpha)\rho]$$

$$= 4\sigma^2 a^2[1 + \rho + \alpha^2(1-\rho)]$$

where ρ denotes the correlation between z_1 and z_2 and also that between z_1' and z_2'. For $\theta > 0$

$$\Pr(y < 0) = F\left\{\frac{-4\alpha\sigma\zeta_2}{2\sigma a[1 + \rho + \alpha^2(1-\rho)]^{1/2}}\right\}$$

$$= F\left\{\frac{-2\alpha\zeta_2}{a[1 + \rho + \alpha^2(1-\rho)]^{1/2}}\right\} = F(b) = \epsilon$$

and hence

$$\alpha^2 = \frac{a^2 b^2(1+\rho)}{4\zeta_2^2 - a^2 b^2(1-\rho)} \tag{20}$$

Under H_1:

$$E(y) = 4\alpha\sigma\zeta_2 - 2\theta$$

$$\mathrm{Var}(y) = 4\sigma^2 a^2([1 + \rho + \alpha^2(1-\rho)]$$

and, for $\theta > 0$,

$$\Pr(y < 0) = F\left\{\frac{2\theta - 4\alpha\sigma\zeta_2}{2\sigma a[1 + \rho + \alpha^2(1-\rho)]^{1/2}}\right\}$$

$$= F\left\{b + \frac{\theta}{\sigma a[1 + \rho + \alpha^2(1-\rho)]^{1/2}}\right\} = P_0$$

The orders of the quantiles which maximize

$$P_0\,(\text{Test D}_2) = F\left[b + \frac{\theta}{\sigma a\,(1+\rho)^{1/2}}\right]$$

are

$$p_1 = 0.2703, \qquad p_2 = 0.7297 \tag{21}$$

which are identical with the orders of the quantiles which maximize P_0 (Test A_2). Using these values, one obtains from Eq. (20)

$$\alpha^2 = \frac{2.4699b^2}{1.4982n - 1.1347b^2} \tag{22}$$

For $n = 200$ and $\epsilon = 0.01$, $\alpha = 0.21338$, so that for these conditions

$$P_0\,(\text{Test }\bar{\text{D}}_2) = F\left[-2.326 + \frac{8.9058\theta}{\sigma}\right]$$

$$P_0\,(\text{Test D}_2) = F\left[-2.326 + \frac{9.000\theta}{\sigma}\right]$$

The quantiles which actually maximize P_0 (Test $\bar{\text{D}}_2$) for $n = 200$, $\epsilon = 0.01$, are those of orders $p_1 = 0.2644$ and $p_2 = 0.7356$, and for those values

$$P_0\,(\text{Test }\bar{\text{D}}_2) = F\left[-2.326 + \frac{8.90898\theta}{\sigma}\right]$$

which again indicates that only a slight advantage can be gained by determining the optimum quantiles.

Thus, for given values of n and ϵ, one either uses the quantiles of orders given by Eq. (21) and determines α from Eq. (22), or one finds the optimum quantiles and then determines α from Eq. (20). For $\theta > 0$, the positive root of α^2 is used in the test statistic and for $\theta < 0$, the negative root is used. This results in the following tests using two pairs of symmetric quantiles. For $\theta > 0$, if

$$(1+\alpha)\,[z_2 - z_1'] + (1-\alpha)\,[z_1 - z_2'] > 0$$

accept H_0; otherwise, reject H_0. For $\theta < 0$, if

$$(1-\alpha)\,[z_2 - z_1'] + (1+\alpha)\,[z_1 - z_2'] < 0$$

accept H_0; otherwise, reject H_0. For $n = 200$, $\epsilon = 0.01$, and using the optimum quantiles of Test D_2, the *acceptance* regions are given by

$$1.5425\,[z\,(0.7297) - z'\,(0.2703)] + [z\,(0.2703) - z'\,(0.7297)] > 0, \qquad \text{for } \theta > 0$$

$$z\,(0.7297) - z'\,(0.2703) + 1.5425\,[z\,(0.2703 - z'(0.7297)] < 0, \qquad \text{for } \theta < 0$$

Table 3 gives the power and efficiency of Test $\bar{\text{D}}_2$ and the efficiency of Test D_2.

Table 3. Power and efficiency of Test $\bar{\text{D}}_2$, and efficiency of the same test with σ known, for $n = 200$, $\epsilon = 0.01$

$\left\lvert\frac{\theta}{\sigma}\right\rvert$	P_0	Efficiency	Efficiency (σ known)
0.01	0.0126	0.9692	0.9769
0.05	0.0300	0.8824	0.8912
0.10	0.0756	0.8182	0.8323
0.15	0.1611	0.7882	0.8048
0.20	0.2931	0.7875	0.8047
0.25	0.4606	0.8093	0.8253
0.30	0.6355	0.8476	0.8613
0.40	0.8920	0.9361	0.9431
0.50	0.9833	0.9870	0.9888

D. Test $\bar{\text{D}}_4$. Four Pairs of Quantiles

Let z_i, $i = 1, 2, 3, 4$, denote the sample quantiles of the first population of orders p_i such that $p_1 + p_4 = p_2 + p_3 = 1$, and let z_i' be the corresponding sample quantiles of the second population. The rejection regions of the test will be taken as

$$y = (0.192+\alpha)\,(z_4 - z_1') + (0.192-\alpha)\,(z_1 - z_4') + (0.308+\alpha)\,(z_3 - z_2') + (0.308-\alpha)\,(z_2 - z_3') \begin{cases} < 0, & \text{for } \theta > 0 \\ > 0, & \text{for } \theta < 0 \end{cases} \tag{23}$$

α will be determined such that the probability of Eq. (23) occurring will be equal to ϵ when μ and σ are unknown.

Omitting much of the tedious details, one has

$$\alpha^2 = \frac{b^2\,[0.03686a_4^2\,(1+\rho_{14}) + 0.09486a_3^2\,(1+\rho_{23}) + 0.1183a_3a_4\,(\rho_{12}-\rho_{13})]}{4\,(\zeta_3+\zeta_4)^2 - b^2\,[a_4^2\,(1-\rho_{14}) + \alpha_3^2\,(1-\rho_{23}) + 2a_3a_4\,(\rho_{12}-\rho_{13})]} \tag{24}$$

where

$$a_i^2 = \frac{F(\zeta_i)\,[1-F(\zeta_i)]}{nf^2(\zeta_i)}, \qquad i = 3, 4$$

$$F(b) = \varepsilon$$

ρ_{ij} denotes the correlation between z_i and z_j, and also between z_i' and z_j'.

For $\theta > 0$, the power function of the test is given by

$$P_0 = F\left(b + \frac{\theta}{2\sigma\gamma}\right)$$

where

$$\gamma^2 = a_4^2\,[0.03686\,(1+\rho_{14}) + \alpha^2\,(1-\rho_{14})] + a_3^2\,[0.09486\,(1+\rho_{23}) + \alpha^2\,(1-\rho_{23})] + 2a_3a_4\,[0.05914 + \alpha^2)\,\rho_{12} + (0.05914-\alpha^2)\,\rho_{13}]$$

The orders of the quantiles which maximize P_0 (Test D_4) will be used in Test $\bar{D}_4$, and are given by

$$\rho_1 = 0.1068$$

$$\rho_2 = 0.3512$$

$$\rho_3 = 0.6488$$

$$\rho_4 = 0.8932$$

Using these values, Eq. (24) becomes

$$\alpha^2 = \frac{0.5435b^2}{10.5729n - 4.1696b^2} \tag{25}$$

and for $n = 200$, $\varepsilon = 0.01$

$$P_0\,(\text{Test } \bar{D}_4) = F\left(-2.326 + \frac{9.541\theta}{\sigma}\right)$$

$$P_0\,(\text{Test } D_4) = F\left(-2.326 + \frac{9.586\theta}{\sigma}\right)$$

For $\theta > 0$, the positive root of α^2 is used; for $\theta < 0$, the negative root is used. Thus, the test can be expressed as follows. For $\theta > 0$, if

$$(0.192+\alpha)\,(z_4 - z') + (0.192-\alpha)\,(z_1 - z_4') + (0.308+\alpha)\,(z_3 - z_2') + (0.308-\alpha)\,(z_2 - z_3') > 0$$

accept H_0; otherwise, reject H_0. For $\theta < 0$, if

$$(0.192-\alpha)_4(z_4 - z_1') + (0.192+\alpha)\,(z_1 - z_4') + (0.308-\alpha)\,(z_3 - z_2') + (0.308+\alpha)\,(z_2 - z_3') < 0$$

accept H_0; otherwise, reject H_0. For $n = 200$, $\varepsilon = 0.01$, the *acceptance* regions are given by

$$0.2293\,[z\,(0.8932) - z'\,(0.1068)] + 0.1543\,[z\,(0.1068) - z'\,(0.8932)] + 0.3457\,[z\,(0.6488) - z'\,(0.3512)] + 0.2707\,[z\,(0.3512) - z'\,(0.6488)] > 0, \qquad \text{for } \theta > 0$$

$$0.1543\,[z\,(0.8932) - z'\,(0.1068)] + 0.2293\,[z\,(0.1068) - z'\,(0.8932)] + 0.2707\,[z\,(0.6488) - z'\,(0.3512)] + 0.3457\,[z\,(0.3512) - z'\,(0.6488)] < 0, \qquad \text{for } \theta < 0$$

Table 4 gives the power and efficiency of Test $\bar{D}_4$ and the efficiency of Test D_4, for $n = 200$, $\varepsilon = 0.01$.

Table 4. Power and efficiency of Test $\bar{D}_4$, and efficiency of the same test with σ known, for $n = 200$, $\varepsilon = 0.01$

$\left\lvert\frac{\theta}{\sigma}\right\rvert$	P_0	Efficiency	Efficiency (σ known)
0.01	0.0128	0.9846	0.9923
0.05	0.0322	0.9471	0.9559
0.10	0.0850	0.9197	0.9286
0.15	0.1854	0.9070	0.9173
0.20	0.3380	0.9081	0.9177
0.25	0.5236	0.9200	0.9290
0.30	0.7041	0.9391	0.9463
0.40	0.9319	0.9780	0.9808
0.50	0.9928	0.9965	0.9969

The following modifications should be made if the sample sizes are n_1 and $n_2 \neq n_1$, where n_1 and n_2 are both large. Define

$$\bar{a}_i^2 = \frac{F(\zeta_i)\,[1-F(\zeta_i)]}{f^2(\zeta_i)}$$

Then the condition $-\zeta/a = b/2$ in Eq. (18) in Test $\bar{\text{D}}_1$ is replaced by the condition

$$\frac{-\zeta}{\bar{a}\left(\frac{n_1+n_2}{n_1 n_2}\right)^{1/2}} = \frac{b}{2}$$

In Test $\bar{\text{D}}_2$, the values of α^2 in Eqs. (20) and (22) are replaced by

$$\alpha^2 = \frac{\bar{a}^2 b^2 \left(\frac{n_1+n_2}{n_1 n_2}\right)(1+\rho)}{8\zeta_2^2 - \bar{a}^2 b^2\left(\frac{n_1+n_2}{n_1 n_2}\right)(1-\rho)}$$

and

$$\alpha^2 = \frac{2.4699 b^2 \left(\frac{n_1+n_2}{n_1 n_2}\right)}{2.9964 - 1.1347 b^2 \left(\frac{n_1+n_2}{n_1 n_2}\right)}$$

In Test $\bar{\text{D}}_4$, the values of α^2 in Eqs. (24) and (25) are replaced by

$$\alpha^2 = \frac{b^2\left(\frac{n_1+n_2}{n_1 n_2}\right)[0.03686\bar{a}_4(1+\rho_{14}) + 0.09846\bar{a}_3^2(1+\rho_{23}) + 0.1183\bar{a}_3\bar{a}_4(\rho_{12}-\rho_{13})]}{8(\zeta_3+\zeta_4)^2 - b^2\left(\frac{n_1+n_2}{n_1 n_2}\right)[\bar{a}_4^2(1-\rho_{14}) + \bar{a}_3^2(1-\rho_{23}) + 2\bar{a}_3\bar{a}_4(\rho_{12}-\rho_{13})]}$$

and

$$\alpha^2 = \frac{0.5435 b^2 \left(\frac{n_1+n_2}{n_1 n_2}\right)}{21.1458 - 4.1696 b^2 \left(\frac{n_1+n_2}{n_1 n_2}\right)}$$

E. Test $\bar{E}_1$. One Pair of Quantiles

In Test $\bar{\text{E}}$, it is assumed that both $\mu = \mu_1 = \mu_2$ and $\sigma = \sigma_1 = \sigma_2$ are unknown. In Test E, by virtue of the assumption of known means, we were permitted to assume without loss of generality that $\mu_1 = \mu_2 = 0$. In effect, then, Test E is a special case of Test $\bar{\text{E}}$ for which the means are equal to zero, but known. Hence, any test that is applicable to the hypotheses of Test $\bar{\text{E}}$ is, *a fortiori*, applicable to the hypotheses of Test E. Now for Test E, in order to eliminate dependence on σ, ratios of linear combinations of sample quantiles of the first population to the same linear combination of sample quantiles from the second population were taken as test statistics. The distribution of the ratios was determined, and rejection regions were calculated only for the special case $n = 200$ and $\varepsilon = 0.01$, using one, two, and four pairs of quantiles. For Test $\bar{\text{E}}$, however, this type of test statistic cannot be used because μ is unknown. A different type of test statistic is needed which will eliminate dependence on *both* μ and σ. For this purpose, linear combinations of one, two, and four pairs of sample quantiles will be used and, as we shall see, the efficiencies of these tests compare very favorably with the efficiencies of the tests of Test E. Moreover, the present tests have the additional advantage that the test statistics are no more difficult to determine, for given values of n and ε, than those of Test $\bar{\text{D}}$.

Beginning with one pair of quantiles, let z be the sample quantile of order p of the first population and z' the sample quantile of order p' of the second population. The test statistic that will be used is

$$y = z - z'$$

Under H_0:

$$E(y) = \sigma(\zeta - \zeta'), \qquad \text{Var}(y) = \sigma^2[a^2 + (a')^2]$$

where

$$a^2 = \frac{F(\zeta)[1-F(\zeta)]}{nf^2(\zeta)}, \qquad (a')^2 = \frac{F(\zeta')[1-F(\zeta')]}{nf^2(\zeta')}$$

The rejection region for $\theta > 1$ will be taken as

$$y < 0$$

so that

$$\Pr(y<0) = F\left\{\frac{-(\zeta-\zeta')}{[a^2+(a')^2]^{1/2}}\right\} = F(b) = \varepsilon$$

and hence the orders of the quantiles must satisfy the condition

$$\frac{-(\zeta-\zeta')}{[a^2+(a')^2]^{1/2}} = b \tag{26}$$

Under H_1:

$$E(y) = \sigma(\zeta - \theta\zeta'), \qquad \text{Var}(y) = \sigma^2[a^2 + \theta^2(a')^2]$$

and

$$\Pr(y < 0) = F\left\{\frac{-(\zeta - \theta\zeta')}{[a^2 + (a')^2]^{1/2}}\right\} = P_0 \qquad (27)$$

If, as in Test $\bar{D}_1$, we now put $p' = 1 - p\,(\zeta' = -\zeta)$, thereby eliminating one degree of freedom in Eq. (26), the following unsatisfactory result would ensue:

$$\frac{-\zeta\,(2)^{1/2}}{a} = b$$

and

$$F\left[\frac{-\zeta\,(1+\theta)}{a\,(1+\theta^2)^{1/2}}\right] = P_0$$

Since $b < 0$, ζ is constrained to be positive so that P_0 would vary from $P_0 = \epsilon$ as $\theta \to 1$ to a maximum of $F(b/2^{1/2})$ as $\theta \to \infty$, which, for $\epsilon = 0.01$, for example, would be $P_0 = 0.05$. A similar situation would arise if one takes $y > 0$ as the rejection region. For this case, since $F(b) = 1 - \epsilon$, b is positive and ζ would have to be negative. As a result, $F\,[-\zeta\,(1+\theta)/(1+\theta^2)^{1/2}] = 1 - P_0$ and $-\zeta\,(1+\theta)/(1+\theta^2)^{1/2}$ would never be negative for $\theta > 1$. P_0 would again vary from $1 - F(b) = \epsilon$ to $1 - F(b/2^{1/2}) = 0.05$ for $\epsilon = 0.01$. The difficulty lies in the fact that for $\zeta > 0$ and $\zeta' < 0$, in the first case, $-(\zeta - \theta\zeta')/[a^2 + (a')^2]^{1/2}$ is never positive for $\theta > 1$. In order to maximize P_0, the condition $\zeta > \zeta' > 0$ must be imposed so that for sufficiently large θ, $-(\zeta - \theta\zeta')/[a^2 + (a')^2]^{1/2} > 0$ and $F\{-(\zeta - \theta\zeta')/[a^2 + (a')^2]^{1/2}\} > 0.5$. A few sample calculations show that $p = 0.9424$ ($\zeta = 1.575$) should be used and then p' determined so as to satisfy Eq. (26). For $\theta < 1$, $y > 0$ is the rejection region, $p' = 0.9424$ and p is determined so as to satisfy Eq. (26) where $F(b) = 1 - \epsilon$. It is easy to show that, for given values of n and ϵ, if $p = 0.9424$ and p' maximizes P_0 for $\theta > 1$, then $\bar{p} = p'$ and $\bar{p}' = 0.9424$ will maximize P_0 for θ^{-1} and

$$P_0(\theta) = P_0(\theta^{-1})$$

Test $\bar{E}_1$ can now be described as follows: If

$$y = z\,(0.9424) - z'\,(p') > 0, \qquad \text{for } \theta > 1$$

$$y = z\,(p) - z'\,(0.9424) < 0, \qquad \text{for } \theta < 1$$

accept H_0; otherwise, reject H_0. p' for $\theta > 1$ and p for $\theta < 1$ are determined so as to satisfy Eq. (26) where $F(b) = \epsilon$ for $\theta > 1$ and $F(b) = 1 - \epsilon$ for $\theta < 1$. For $n = 200$, $\epsilon = 0.01$ the *acceptance* regions are

$$z\,(0.9424) - z'\,(0.8751) > 0, \qquad \text{for } \theta > 1$$

$$z\,(0.8751) - z'\,(0.9424) < 0, \qquad \text{for } \theta < 1$$

Table 5 gives the power and efficiency of Test $\bar{E}_1$ and the efficiency of Test E_1 for $n = 200$, $\epsilon = 0.01$.

Table 5. Power and efficiency of Test $\bar{E}_1$, and efficiency of Test E_1, for $n = 200$, $\epsilon = 0.01$

θ	P_0	Efficiency (Test $\bar{E}_1$)	Efficiency (Test E_1)
1.025	0.0160	0.6667	0.6696
1.050	0.0245	0.4813	0.4872
1.100	0.0527	0.3213	0.3262
1.150	0.0970	0.2669	0.2795
1.200	0.1630	0.2719	0.2876
1.250	0.2494	0.3127	0.3321
1.300	0.3508	0.3826	0.4072
1.350	0.4598	0.4729	0.5024
1.400	0.5676	0.5719	0.6050
1.500	0.7522	0.7524	0.7858
1.600	0.8755	0.8755	0.9013
1.700	0.9440	0.9440	0.9600

F. Test $\bar{E}_2$. Two Pairs of Quantiles

Let z_1 and z_2 be sample quantiles of orders p_1 and $p_2 = 1 - p_1$ of the first population and z_1' and z_2' be the corresponding sample quantiles of the second population. The test statistic to be used is given by

$$y = (1+\alpha)(z_2 - z_1) + (1-\alpha)(z_2' - z_1')$$

and the rejection region will be taken as

$$y < 0$$

Under H_0:

$$E(y) = 2\sigma\zeta_2(1+\alpha) + 2\sigma\zeta_2(1-\alpha) = 4\sigma\zeta_2$$

$$\begin{aligned}\mathrm{Var}(y) &= 2\sigma^2 a^2 (1-\rho_{12})\,[(1+\alpha)^2 + (1-\alpha)^2] \\ &= 4\sigma^2 a^2 (1+\alpha^2)(1-\rho_{12})\end{aligned}$$

where

$$a^2 = \frac{F(\zeta_2)\,[1 - F(\zeta_2)]}{nf^2(\zeta_2)}$$

and ρ_{12} denotes the correlation between z_1 and z_2 as well as that between z_1' and z_2'.

$$\Pr(y<0) = F\left\{\frac{-2\zeta_2}{a[(1-\rho_{12})(1+\alpha^2)]^{1/2}}\right\} = F(b) = \epsilon$$

and

$$\alpha^2 = \frac{4\zeta_2^2}{a^2b^2(1-\rho_{12})} - 1 \qquad (28)$$

Under H_1:

$$E(y) = 2\sigma\zeta_2(1+\alpha) + 2\sigma\theta\zeta_2(1-\alpha)$$

$$= 2\sigma\zeta_2[1+\alpha+\theta(1-\alpha)]$$

$$\mathrm{Var}(y) = 2\sigma^2a^2(1-\rho)(1+\alpha)^2$$

$$+ 2\sigma^2a^2\theta^2(1-\rho)(1-\alpha)^2$$

$$= 2\sigma^2a^2(1-\rho)[(1+\alpha)^2+\theta^2(1-\alpha)^2]$$

$$\Pr(y<0) = F\left\{\frac{-2^{1/2}\zeta_2[1+\alpha+\theta(1-\alpha)]}{a(1-\rho)^{1/2}[(1+\alpha)^2+\theta^2(1-\alpha)^2]^{1/2}}\right\}$$

$$= F\left\{\frac{b(1+\alpha^2)^{1/2}[1+\alpha+\theta(1-\alpha)]}{2^{1/2}[(1+\alpha)^2+\theta^2(1-\alpha)^2]^{1/2}}\right\} = P_0$$

As in Test E_1, we will use the quantiles which minimize the variance of the estimate of σ from a single set of samples. The orders of these quantiles are

$$p_1 = 0.0690, \qquad p_2 = 0.9310 \qquad (29)$$

Using these values, Eq. (28) becomes

$$\alpha^2 = \frac{2.609n}{b^2} - 1 \qquad (30)$$

Now, since $b<0$, for $\theta>1$, P_0 will increase as θ increases from $\theta=1$ if $\alpha>1$, and for $\theta<1$, P_0 will increase as θ decreases from $\theta=1$ if $\alpha<-1$. From Eq. (30) one sees that for sufficiently large n and realistic values of ϵ, $\alpha^2>1$. For example, for $n=200$ and $\epsilon=0.01$, $b=2.326$ and $\alpha^2=95.445$. Thus, Test $\bar{E}_2$ can now be described as follows. For $\theta>1$, if $y=(1+\alpha)(z_2-z_1)+(1-\alpha)(z_2'-z_1')<0$ reject H_0; otherwise, accept H_0. For $\theta<1$, if $y=(1-\alpha)(z_2-z_1)+(1+\alpha)(z_2'-z_1')<0$ reject H_0; otherwise, accept H_0. For $n=200$, $\epsilon=0.01$, and using the quantiles with orders given by Eq. (29), the *acceptance* regions are given by

$$1.228[z(0.9310)-z(0.0690)] - [z'(0.9310)-z'(0.0690)]>0, \qquad \theta>1$$

$$-[z(0.9310)-z(0.0690)] + 1.228[z'(0.9310)-z'(0.0690)]>0, \qquad \theta<1$$

α^2 is determined from Eq. (28) if the values in Eq. (29) are not used, or from Eq. (30) for general values of n and ϵ if the values in Eq. (29) are used.

Table 6 gives the power and efficiency of Test $\bar{E}_2$ and the efficiency of Test E_2 for $n=200$, $\epsilon=0.01$.

Table 6. Power and efficiency of Test $\bar{E}_2$, and efficiency of Test E_2, for $n=200$, $\epsilon=0.01$

θ	P_0	Efficiency (Test $\bar{E}_2$)	Efficiency (Test E_2)
1.025	0.0202	0.8417	0.8333
1.050	0.0375	0.7376	0.7308
1.100	0.1049	0.6396	0.6348
1.150	0.2268	0.6239	0.6206
1.200	0.3958	0.6602	0.6577
1.250	0.5801	0.7273	0.7253
1.300	0.7421	0.8094	0.8079
1.350	0.8597	0.8841	0.8832
1.400	0.9320	0.9390	0.9384
1.500	0.9883	0.9886	0.9884

G. Test $\bar{E}_4$. Four Pairs of Quantiles

Let z_i, $i=1,2,3,4$, be four sample quantiles of the first population such that $p_1+p_4=p_2+p_3=1$, and let z_i' be the corresponding sample quantiles of the second population. The statistic that will be used in the test is given by

$$y = (1+\alpha)[0.116(z_4-z_1)+0.236(z_3-z_2)]$$

$$+ (1-\alpha)[0.116(z_4'-z_1')+0.236(z_3'-z_2')]$$

and the rejection region will be taken as

$$y<0$$

Under H_0:

$$E(y) = (1+\alpha)[0.232\sigma\zeta_4 + 0.472\sigma\zeta_3] + (1-\alpha)[0.232\sigma\zeta_4 + 0.472\sigma\zeta_3] = \sigma[0.464\zeta_4 + 0.944\zeta_3]$$

$$\mathrm{Var}(y) = 4(1+\alpha^2)\sigma^2[0.013456a_4^2(1-\rho_{14}) + 0.055696a_3^2(1-\rho_{23}) + 0.0547752a_3a_4(\rho_{34}-\rho_{24})] = 4(1+\alpha^2)\sigma^2\gamma^2$$

where

$$a_i^2 = \frac{F(\zeta_i)[1-F(\zeta_i)]}{nf^2(\zeta_i)}, \qquad i = 3, 4$$

and ρ_{ij} denotes the correlation between z_i and z_j, as well as that between z'_i and z'_j.

$$\Pr(y<0) = F\left[\frac{-(0.464\zeta_4 + 0.944\zeta_3)}{2\gamma(1+\alpha^2)^{1/2}}\right] = F(b) = \varepsilon$$

and

$$\alpha^2 = \frac{(0.232\zeta_4 + 0.472\zeta_3)^2}{b^2\gamma^2} - 1 \qquad (31)$$

Under H_1:

$$E(y) = \sigma(0.234\zeta_4 + 0.472\zeta_3)[1+\alpha+\theta(1-\alpha)]$$

$$\mathrm{Var}(y) = 2\sigma^2\gamma^2[(1+\alpha)^2 + \theta^2(1-\alpha)^2]$$

$$\Pr(y<0) = F\left\{\frac{-(0.234\zeta_4 + 0.472\zeta_3)[1+\alpha+\theta(1-\alpha)]}{\gamma(2)^{1/2}[(1+\alpha)^2+\theta^2(1-\alpha)^2]^{1/2}}\right\} = F\left\{\frac{b(1+\alpha^2)^{1/2}[1+\alpha+\theta(1-\alpha)]}{2^{1/2}[(1+\alpha)^2+\theta^2(1-\alpha)]^{1/2}}\right\} = P_0$$

By using the optimum quantiles in estimating σ from a single set of samples, those of orders

$$p_1 = 0.0230$$
$$p_2 = 0.1271$$
$$p_3 = 0.8729$$
$$p_4 = 0.9770 \qquad (32)$$

Eq. (31) becomes

$$\alpha^2 = \frac{3.2977n}{b^2} - 1 \qquad (33)$$

Since the expression for P_0 in this test is identical with that in Test $\bar{E}_2$, the positive root of α^2 must be used for $\theta > 1$ and the negative root must be used for $\theta < 1$, in order to maximize P_0. Test $\bar{E}_4$ can now be described as follows. For $\theta > 1$, if

$$(1+\alpha)[0.116(z_4 - z_1) + 0.236(z_3 - z_2)] + (1-\alpha)[0.116(z'_4 - z'_3) + 0.236(z'_3 - z'_2)] < 0$$

reject H_0; otherwise, accept H_0. For $\theta < 1$, if

$$(1-\alpha)[0.116(z_4 - z_1) + 0.236(z_3 - z_2)] + (1+\alpha)[0.116(z'_4 - z'_3) + 0.236(z'_3 - z'_2)] < 0$$

reject H_0; otherwise, accept H_0. For $n = 200$, $\varepsilon = 0.01$, and using the quantiles with orders given in Eq. (32), $\alpha^2 = 120.904$ and the *acceptance* regions are

$$0.1392[z(0.9770) - z(0.0230)] + 0.2832[z(0.8729) - z(0.1271)] - 0.116[z'(0.9770) - z'(0.0230)] - 0.236[z'(0.8729) - z'(0.1271)] > 0, \qquad \theta > 1$$

$$-0.116[z(0.9770) - z(0.0230)] - 0.236[z(0.8729) - z(0.1271)] + 0.1392[z'(0.9770) - z'(0.0230)] + 0.2832[z'(0.8729) - z'(0.1271)] > 0, \qquad \theta < 1$$

As in Test $\bar{E}_2$, α^2 is determined by Eq. (31) if the quantiles used are not those with orders given by Eq. (32) and by Eq. (33) if the quantiles are of the orders given by (32).

Table 7 gives the power and efficiency of Test $\bar{E}_4$ and the efficiency of Test E_4 for $n = 200$, $\varepsilon = 0.01$.

Table 7. Power and efficiency of Test $\bar{E}_4$, and efficiency of Test E_4, for $n = 200$, $\varepsilon = 0.01$

θ	P_0	Efficiency (Test $\bar{E}_4$)	Efficiency (Test E_4)
1.025	0.0225	0.9375	0.9292
1.050	0.0438	0.8605	0.8723
1.100	0.1322	0.8061	0.8177
1.150	0.2921	0.8036	0.8132
1.200	0.4996	0.8334	0.8409
1.250	0.6975	0.8745	0.8823
1.300	0.8478	0.9247	0.9277
1.350	0.9341	0.9606	0.9625
1.400	0.9755	0.9829	0.9837

As in Test $\bar{E}_1$, in Tests $\bar{E}_2$ and $\bar{E}_4$, $P_0(\theta) = P_0(\theta^{-1})$. This can be seen by considering the factor

$$\frac{1 + \theta + \alpha(1-\theta)}{[(1+\alpha)^2 + \theta^2(1-\alpha)^2]^{1/2}}$$

in the expression for P_0. For $\theta' = \theta^{-1}$, $\alpha' = -\alpha$ and the above factor becomes

$$\frac{1 + \theta' + \alpha'(1-\theta')}{[(1+\alpha')^2 + (\theta')^2(1-\alpha')^2]^{1/2}} = \frac{1 + \theta^{-1} - \alpha(1-\theta^{-1})}{\{(1-\alpha)^2 + [(\theta')^{-1}]^2(1+\alpha)^2\}^{1/2}} = \frac{1 + \theta + \alpha(1-\theta)}{[(1+\alpha)^2 + \theta^2(1-\alpha)^2]^{1/2}}$$

V. TESTS OF INDEPENDENCE AND ESTIMATION OF THE CORRELATION COEFFICIENT ρ USING QUANTILES

A. Statement of the Problem

Given a set of n independent pairs of observations $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, taken from two normally distributed populations, one is naturally interested in whether the set of observations $x = \{x_1, x_2, \cdots, x_n\}$ is independent of the set of observations $y = \{y_1, y_2, \cdots, y_n\}$, and in the correlation between them, if any. With respect to the question of independence, the problem of testing the null hypothesis

$$H_0: \quad g_1(x) = N(\mu_1, \sigma_1), \qquad g_2(y) = N(\mu_2, \sigma_2), \qquad \rho = 0$$

against the alternative hypothesis

$$H_1: \quad g_1(x) = N(\mu_1, \sigma_1), \qquad g_2(y) = N(\mu_2, \sigma_2), \qquad \rho \neq 0$$

will be considered. One, two, and four pairs of quantiles are used for the tests. For the case of one pair of quantiles, it will be assumed that μ_1 and μ_2 are known and hence can, without loss of generality, be put equal to zero, and that $\sigma = \sigma_1 = \sigma_2$ is unknown. When two and four pairs of quantiles are used, it will be assumed that *both* $\mu = \mu_1 = \mu_2$ and $\sigma = \sigma_1 = \sigma_2$ are unknown.

Unbiased estimators of ρ are constructed, first using one quantile and then using one, two, and four pairs of quantiles. The assumption here will be that $\mu = \mu_1 = \mu_2$ is unknown and that σ_1 and σ_2 are known and hence can be put equal to one.

The power functions of the tests are derived and the efficiencies are determined relative to the best test using the entire sample when the four parameters are *known*. The efficiencies of the estimators are also determined relative to the sample correlation coefficient for the case $\rho = 0$.

B. Test $\bar{F}_1$. One Pair of Quantiles

It is necessary at this point to form two new sets of values $\{u_i\}$ and $\{v_i\}$ from the sample values $\{x_i\}$ and $\{y_i\}$ by means of the linear transformations

$$u_i = \frac{2^{1/2}}{2}(x_i + y_i)$$

$$v_i = \frac{2^{1/2}}{2}(-x_i + y_i)$$

With $\mu_1 = \mu_2 = \mu$ and $\sigma_1 = \sigma_2 = \sigma$, it is easily verified that:

Under H_0:

$$E(u_i) = u(2)^{1/2}$$

$$E(v_i) = 0$$

$$\mathrm{Var}(u_i) = \mathrm{Var}(v_i) = \sigma^2$$

$$E(u_i v_i) = 0$$

Under H_1:

$$E(u_i) = u(2)^{1/2}$$

$$E(v_i) = 0$$

$$\mathrm{Var}(u_i) = \sigma^2(1+\rho)$$

$$\mathrm{Var}(v_i) = \sigma^2(1-\rho)$$

$$E(u_i v_i) = 0$$

so that the set of values $\{u_i\}$ is independent of the set of values $\{v_i\}$ *under both hypotheses.* All the tests and estimators will be based on the quantiles of the transformed sets of variables $\{u_i\}$ and $\{v_i\}$, which are all normally distributed.

Now let z denote the sample quantile of order p of the u_i and z' the sample quantile of order p' of the $\{v_i\}$. The test statistic that will be used is

$$y = z - z'$$

and the rejection region for $\rho < 0$ will be taken as $y < 0$. For $\mu = 0$, one has:

Under H_0:

$$E(y) = \sigma(\zeta - \zeta'), \qquad \mathrm{Var}(y) = \sigma^2[a^2 + (a')^2]$$

where

$$a^2 = \frac{F(\zeta)[1-F(\zeta)]}{nf^2(\zeta)}$$

$$(a')^2 = \frac{F(\zeta')[1-F(\zeta')]}{nf^2(\zeta')}$$

$$\Pr(y<0) = F\left\{\frac{-(\zeta-\zeta')}{[a^2+(a')^2]^{1/2}}\right\} = F(b) = \varepsilon$$

and hence the orders of the quantiles must satisfy the condition

$$\frac{-(\zeta-\zeta')}{[a^2+(a')^2]^{1/2}} = b \qquad (34)$$

Under H_1:

$$E(y) = \sigma[\zeta(1+\rho)^{1/2} - \zeta'(1-\rho)^{1/2}]$$

$$\mathrm{Var}(y) = \sigma^2[a^2(1+\rho) + (a')^2(1-\rho)]$$

$$\Pr(y<0) = F\left\{\frac{-[\zeta(1+\rho)^{1/2} - \zeta'(1-\rho)^{1/2}]}{[a^2(1+\rho)+(a')^2(1-\rho)]^{1/2}}\right\} = P_0$$

Since the condition specified in Eq. (34) is identical with that given in Eq. (26) for Test $\bar{\mathrm{E}}_1$, one can use the results derived for that test. Thus, let $p = 0.9424$ and determine $p' < p$ to satisfy Eq. (34). For $\rho > 0$, the rejection region is $y > 0$. For this case, $p' = 0.9424$ and p is determined to satisfy Eq. (34) where now $F(b) = 1 - \varepsilon$. It is easy to show that $P_0(\rho) = P_0(-\rho)$.

Test $\bar{\mathrm{F}}_1$ can now be described as follows. For $\rho < 0$, if

$$z(0.9424) - z'(p') < 0$$

reject H_0; otherwise, accept H_0. For $\rho > 0$, if

$$z(p) - z'(0.9424) > 0$$

reject H_0; otherwise, accept H_0. For $\rho < 0$, determine p' to satisfy Eq. (34), where $F(b) = \varepsilon$. For $\rho > 0$, determine p to satisfy Eq. (34) where $F(b) = 1 - \varepsilon$.

Table 8 gives the power and efficiency of Test $\bar{\mathrm{F}}_1$ and the efficiency of Test F_1, for $n = 200$, $\varepsilon = 0.01$.

Table 8. Power and efficiency of Test $\bar{\mathrm{F}}_1$, and efficiency of Test F_1, for $n = 200$, $\varepsilon = 0.01$

ρ	P_0	Efficiency (Test $\bar{\mathrm{F}}_1$)	Efficiency (Test F_1)
0.025	0.0161	0.6680	0.6846
0.050	0.0250	0.4808	0.5077
0.075	0.0378	0.3765	0.4074
0.100	0.0558	0.3181	0.3489
0.150	0.1124	0.2764	0.3061
0.200	0.2035	0.2981	0.3284
0.250	0.3317	0.3738	0.4048
0.300	0.4878	0.4993	0.5287
0.350	0.6501	0.6516	0.6745
0.400	0.7925	0.7925	0.8055
0.450	0.8962	0.8962	0.8999
0.500	0.9572	0.9572	0.9565

C. Test $\bar{F}_2$. Two Pairs of Quantiles

Let z_1 and z_2 be two sample quantiles of orders p_1 and $p_2 = 1 - p_1$ of the $\{u_i\}$, and z_1' and z_2' the corresponding sample quantiles of the $\{v_i\}$. The statistic that will be used in the test is

$$y = (1+\alpha)(z_2 - z_1) + (1-\alpha)(z_2' - z_1')$$

and the rejection region will be taken as $y < 0$.

Under H_0:

$$E(y) = 4\sigma\zeta_2$$
$$\mathrm{Var}(y) = 4\sigma^2 a^2 (1 - \rho_{12})(1 + \alpha^2)$$

where

$$a^2 = \frac{F(\zeta_2)[1 - F(\zeta_2)]}{nf^2(\zeta_2)}$$

and ρ_{12} denotes the correlation between z_1 and z_2 as well as that between z'_1 and z'_2.

$$\Pr(y < 0) = F\left[\frac{-2\zeta_2}{a(1 + \alpha^2)^{1/2}(1 - \rho_{12})^{1/2}}\right] = F(b) = \varepsilon$$

and

$$\alpha^2 = \frac{4\zeta_2^2}{a^2 b^2 (1 - \rho_{12})} - 1 \qquad (35)$$

Under H_1:

$$E(y) = 2\sigma\zeta_2[(1 + \alpha)(1 + \rho)^{1/2} + (1 - \alpha)(1 - \rho)^{1/2}]$$

$$\mathrm{Var}(y) = 2\sigma^2 a^2 (1 - \rho_{12})[(1 + \alpha)^2(1 + \rho) + (1 - \alpha)^2(1 - \rho)] = 4\sigma^2 a^2 (1 - \rho_{12})(1 + 2\alpha\rho + \alpha^2)$$

$$\Pr(y < 0) = F\left\{\frac{-\zeta_2[(1 + \alpha)(1 + \rho)^{1/2} + 1 - \alpha)(1 - \rho)^{1/2}]}{a(1 - \rho_{12})^{1/2}(1 + 2\alpha\rho + \alpha^2)^{1/2}}\right\}$$
$$= F\left\{\frac{b(1 + \alpha^2)^{1/2}[(1 + \alpha)(1 + \rho)^{1/2} + (1 - \alpha)(1 - \rho)^{1/2}]}{2(1 + 2\alpha\rho + \alpha^2)^{1/2}}\right\}$$
$$= P_0 \qquad (36)$$

Using the same values of p_1 and p_2 as in Test $\overline{\mathrm{E}}_2$, given by

$$p_1 = 0.0690, \qquad p_2 = 0.9310 \qquad (37)$$

Eq. (35) becomes

$$\alpha^2 = \frac{2.609n}{b^2} - 1 \qquad (38)$$

and is identical with Eq. (30). Now, using the same argument as in Test $\overline{\mathrm{E}}_2$, we conclude that one should use the positive root of α^2 when $\rho < 0$ and the negative root when $\rho > 0$. Thus, Test $\overline{\mathrm{F}}_2$ can be described as follows. For $\rho > 0$, if

$$y = (1 - \alpha)(z_2 - z_1) + (1 + \alpha)(z'_2 - z'_1) < 0$$

reject H_0; otherwise, accept H_0. For $\rho < 0$, if

$$y = (1 + \alpha)(z_2 - z_1) + (1 - \alpha)(z'_2 - z'_1) < 0$$

reject H_0; otherwise, accept H_0. For $n = 200$, $\varepsilon = 0.01$ and using the quantiles of orders given by Eq. (37), $\alpha^2 = 95.445$, and the *acceptance* regions are given by

$$1.228[z(0.9310) - z(0.0690)] - [z'(0.9310) - z'(0.0690)] > 0, \qquad \rho < 0$$

$$-[z(0.9310) - z(0.0690)] + 1.228[z'(0.9310) - z'(0.0690)] > 0, \qquad \rho > 0$$

Table 9 gives the power and efficiency of Test $\overline{\mathrm{F}}_2$ and the efficiency of Test F_2.

Table 9. Power and efficiency of Test $\overline{\mathrm{F}}_2$, and efficiency of Test F_2, for $n = 200$, $\varepsilon = 0.01$

ρ	P_0	Efficiency (Test $\overline{\mathrm{F}}_2$)	Efficiency (Test F_2)
0.025	0.0203	0.8423	0.8589
0.050	0.0387	0.7442	0.7635
0.075	0.0690	0.6873	0.7062
0.100	0.1155	0.6585	0.6756
0.150	0.2677	0.6582	0.6678
0.200	0.4877	0.7144	0.7147
0.250	0.7158	0.8067	0.7997
0.300	0.8822	0.9031	0.8935
0.350	0.9655	0.9677	0.9609

D. Test $\overline{F}_4$. Four Pairs of Quantiles

Let z_i, $i = 1, 2, 3, 4$, be four sample quantiles of the $\{u_i\}$ such that $p_1 + p_4 = p_2 + p_3 = 1$, and let z'_i be the corresponding sample quantiles of the $\{v_i\}$. The statistic that will be used in the test is given by

$$y = (1 + \alpha)[0.116(z_4 - z_1) + 0.236(z_3 - z_2)] + (1 - \alpha)[0.116(z'_4 - z'_1) + 0.236(z'_3 - z'_2)]$$

and the rejection region will be taken as

$$y < 0$$

Under H_0:

$$E(y) = \sigma(0.464\zeta_4 + 0.944\zeta_3)$$

$$\mathrm{Var}(y) = 4(1 + \alpha^2)\sigma^2[0.013456a_4^2(1 - \rho_{14}) + 0.055696a_3^2(1 - \rho_{23}) + 0.054775a_3a_4(\rho_{34} - \rho_{24})]$$
$$= 4(1 + \alpha^2)\sigma^2\gamma^2$$

where

$$a_i^2 = \frac{F(\zeta_i)\,[1 - F(\zeta_i)]}{nf^2(\zeta_i)}, \qquad i = 3, 4$$

and ρ_{ij} denotes the correlation between z_i and z_j, as well as that between z'_i and z'_j.

$$\Pr(y < 0) = F\left[\frac{-(0.464\zeta_4 + 0.944\zeta_3)}{\gamma(1 + \alpha^2)^{1/2}}\right] = F(b) = \varepsilon$$

and

$$\alpha^2 = \frac{(0.232\zeta_4 + 0.472\zeta_3)^2}{b^2\gamma^2} - 1 \qquad (39)$$

Under H_1:

$$E(y) = \sigma(0.232\zeta_4 + 0.472\zeta_3)\,[(1 + \alpha)(1 + \rho)^{1/2} + (1 - \alpha)(1 - \rho)^{1/2}]$$

$$\mathrm{Var}(y) = 4\sigma^2\gamma^2(1 + 2\alpha\rho + \alpha^2)$$

$$\Pr(y < 0) = F\left\{\frac{-(0.116\zeta_4 + 0.236\zeta_3)\,[1 + \alpha)(1 + \rho)^{1/2} + (1 - \alpha)(1 - \rho)^{1/2}]}{\gamma(1 + 2\alpha\rho + \alpha^2)^{1/2}}\right\}$$

$$= F\left\{\frac{b(1 + \alpha^2)^{1/2}[(1 + \alpha)(1 + \rho)^{1/2} + (1 - \alpha)(1 - \rho)^{1/2}]}{2(1 + 2\alpha\rho + \alpha^2)^{1/2}}\right\} = P_0$$

Using the same quantiles as in Test $\overline{\mathrm{E}}_4$, those of orders

$$p_1 = 0.0230$$
$$p_2 = 0.1271$$
$$p_3 = 0.8729$$
$$p_4 = 0.9770$$

Eq. (39) becomes

$$\alpha^2 = \frac{3.2977n}{b^2} - 1$$

The negative root of α^2 must be used for $\rho > 0$ and the positive root for $\rho < 0$, in order to maximize P_0. Thus, Test $\overline{\mathrm{F}}_4$ can be described as follows. For $\rho > 0$, if

$$(1 - \alpha)\,[0.116(z_4 - z_1) + 0.236(z_3 - z_2)] + (1 + \alpha)\,[0.116(z'_4 - z'_1) + 0.236(z'_3 - z'_2)] < 0$$

reject H_0; otherwise, accept H_0. For $\rho < 0$, if

$$(1 + \alpha)\,[0.116(z_4 - z_3) + 0.236(z_3 - z_2)] + (1 - \alpha)\,[0.116(z'_4 - z'_3) + 0.236(z'_3 - z'_2)] < 0$$

reject H_0; otherwise, accept H_0. For $n = 200$, $\varepsilon = 0.01$, and using the quantiles with orders given by Eq. (40), $\alpha^2 = 120.904$ and the *acceptance* regions are given by

$$-0.116\,[z(0.9770) - z(0.0230)] - 0.236\,[z(0.8729) - z(0.1271)] + 0.1392\,[z'(0.9770) - z'(0.0230)] + 0.2832\,[z'(0.8729) - z'(0.1271)] > 0, \qquad \rho > 0$$

$$0.1392\,[z(0.9770) - z(0.0230)] + 0.2832\,[z(0.8729) - z(0.1279)] - 0.116\,[z'(0.9770) - z'(0.0230)] - 0.236\,[z'(0.8729) - z'(0.1271)] > 0, \qquad \rho < 0$$

Table 10. Power and efficiency of Test $\overline{\mathrm{F}}_4$, and efficiency of Test F_4, for $n = 200$, $\varepsilon = 0.01$

ρ	P_0	Efficiency (Test $\overline{\mathrm{F}}_4$)	Efficiency (Test F_4)
0.025	0.0222	0.9212	0.9336
0.050	0.0452	0.8692	0.8865
0.075	0.0847	0.8436	0.8616
0.100	0.1464	0.8347	0.8501
0.150	0.3441	0.8461	0.8522
0.200	0.6030	0.8833	0.8799
0.250	0.8255	0.9304	0.9226
0.300	0.9481	0.9705	0.9642
0.350	0.9902	0.9925	0.9898

Table 10 gives the power and efficiency of Test $\overline{\mathrm{F}}_4$ and the efficiency of Test F_4.

E. An Unbiased Estimator of ρ Using One Quantile

We are now assuming that $\mu_1 = \mu_2 = \mu$ is unknown and that $\sigma_1 = \sigma_2 = 1$. Under these assumptions, one has

$$E(u_i) = u(2)^{1/2}, \qquad \mathrm{Var}(u_i) = 1 + \rho$$

$$E(v_i) = 0, \qquad \mathrm{Var}(v_i) = 1 - \rho$$

Unbiased estimators of ρ will be constructed using a single quantile and a pair of quantiles of the $\{v_i\}$ and two, and four pairs of quantiles, one of each pair from the $\{u_i\}$ and the other from the $\{v_i\}$. The efficiency of the estimators will be determined relative to the sample correlation coefficient r, the minimum-variance unbiased estimator of ρ given by

$$r = \frac{\sum_{i=1}^{n} (u_i - \bar{u})(v_i - \bar{v})}{\left\{ \sum_{i=1}^{n} (u_i - \bar{u})^2 \sum_{i=1}^{n} (v_i - \bar{v})^2 \right\}^{1/2}}$$

where

$$\bar{u} = n^{-1} \sum_{i=1}^{n} u_i, \qquad \bar{v} = n^{-1} \sum_{i=1}^{n} v_i$$

for the special case $\rho = 0$. Since the asymptotic variance, $\mathrm{Var}\,(r \mid \rho = 0)$ is $(n-1)^{-1}$ (Ref. 7), the efficiency will be defined as

$$\mathrm{Eff}\,(\hat{\rho}) = \frac{\mathrm{Var}\,(r \mid \rho = 0)}{\mathrm{Var}\,(\hat{\rho} \mid \rho = 0)} = \frac{1}{(n-1)\,\mathrm{Var}\,(\hat{\rho} \mid \rho = 0)}$$

Hence, let z denote the quantile of order p of the $\{v_i\}$. Then an unbiased estimator of ρ is given by

$$\hat{\rho} = 1 - \frac{z^2}{a^2 + \zeta^2}$$

where

$$a^2 = \frac{F(\zeta)\,[1 - F(\zeta)]}{nf^2(\zeta)}$$

Since $E(z) = \zeta(1-\rho)^{1/2}$ and $\mathrm{Var}\,(z) = a^2(1-\rho)$, one has

$$E(\hat{\rho}) = 1 - \frac{a^2(1-\rho) + \zeta^2(1-\rho)}{a^2 + \zeta^2} = 1 - (1-\rho) = \rho$$

so that $\hat{\rho}$ is seen to be unbiased. The variance of $\hat{\rho}$ is given by

$$\mathrm{Var}\,(\hat{\rho}) = \frac{1}{(a^2+\zeta^2)^2}\,[2a^4(1-\rho)^2 + 4\zeta^2 a^2 (1-\rho)^2]$$

$$= \frac{2(a^4 + 2\zeta^2 a^2)(1-\rho)^2}{(a^2+\zeta^2)^2} \tag{41}$$

For $\rho = 0$, Eq. (41) can be written as

$$\mathrm{Var}\,(\hat{\rho} \mid \rho = 0) = \frac{2a_4 + 4\zeta^2 a^2}{a^4 + 2\zeta^2 a^2 + \zeta^4} \simeq \frac{4\zeta^2 a^2}{2\zeta^2 a^2 + \zeta^4} = \frac{2}{1 + \dfrac{\zeta^2}{2a^2}} \tag{42}$$

if one neglects the a^4 term in the numerator and denominator of Eq. (42). (This term is small compared with $\zeta^2 a^2$ and ζ^4 for large n.) Now the approximate value of $\mathrm{Var}\,(\hat{\rho} \mid \rho = 0)$ in Eq. (42) is minimized if one chooses the order of z to maximize ζ^2/a^2. It is known that $p = 0.9424$, the order of the quantile which minimizes the variance of the estimate of σ, will maximize ζ/a and will, of course, also maximize ζ^2/a^2. [Since $\zeta(0.0576) = -\zeta(0.9424)$ and $a(0.0576) = a(0.9424)$, $p = 0.0576$ can also be used.] Thus, using either $p = 0.9424$ or $p = 0.0576$, one obtains

$$\hat{\rho} = 1 - \frac{z^2}{\dfrac{4.0771}{n} + 2.4806}$$

$$\mathrm{Var}\,(\hat{\rho} \mid \rho = 0) = \frac{2\left[\dfrac{16.627}{n^2} + \dfrac{20.230}{n}\right]}{\dfrac{16.627}{n^2} + \dfrac{20.230}{n} + 6.1535} \simeq \frac{2}{1 + 0.3042n}$$

For $n = 200$

$$\hat{\rho} = 1 - 0.3998z^2$$

$$\mathrm{Var}\,(\hat{\rho} \mid \rho = 0) = 0.03247$$

$$\frac{2}{1 + 0.3042n} = 0.03234$$

$$\mathrm{Eff}\,(\hat{\rho}) = 0.1548$$

F. An Unbiased Estimator of ρ Using One Pair of Quantiles

Let z_1 and z_2 be two sample quantiles of orders p_1 and $p_2 = 1 - p_1$ of the $\{v_i\}$. Then an unbiased estimator of ρ using one pair of symmetric quantiles is given by

$$\hat{\rho} = 1 - \frac{z_1^2 + z_2^2}{2(\zeta_2^2 + a^2)}$$

where

$$a^2 = \frac{F(\zeta_2)\,[1 - F(\zeta_2)]}{nf^2(\zeta_2)}$$

Since

$$E(z_1) = -\zeta_2(1-\rho)^{1/2}$$

$$\mathrm{Var}(z_1) = \mathrm{Var}(z_2) = a^2(1-\rho)$$

$$E(z_2) = \zeta_2(1-\rho)^{1/2}$$

$$E(\hat{\rho}) = 1 - \frac{2\zeta_2^2(1-\rho) + 2a^2(1-\rho)}{2(\zeta_2^2 + a^2)} = 1 - (1-\rho) = \rho$$

and

$$\mathrm{Var}(\hat{\rho}) = \frac{1}{4(\zeta_2^2 + a^2)^2}[2\,\mathrm{Var}(z_2^2) + 2\,\mathrm{Cov}(z_1^2, z_2^2)]$$

$$= \frac{1}{(\zeta_2^2 + a^2)^2}[a^4(1+\rho_{12}^2) + 2a^2\zeta_2^2(1-\rho_{12})](1-\rho)^2 \tag{43}$$

where ρ_{12} denotes the correlation between z_1 and z_2. For $\rho = 0$, Eq. (43) can be written as

$$\mathrm{Var}(\hat{\rho}\,|\,\rho = 0) = \frac{a^4(1+\rho_{12}^2) + 2a^2\zeta_2^2(1-\rho_{12})}{a_2^4 + 2a^2\zeta_2^2 + \zeta_4^2}$$

$$\simeq \left[1 + \frac{\zeta_2^2}{2a^2(1-\rho_{12})}\right]^{-1} \tag{44}$$

if one neglects the a^4 term in the numerator and denominator of Eq. (44). The orders of the quantiles which minimize the variance of the estimate of σ will maximize $\zeta_2^2/[2a^2(1-\rho_{12})]$, and hence minimize the approximate expression of $\mathrm{Var}(\hat{\rho}\,|\,\rho = 0)$ in Eq. (44). These are $p_1 = 0.0690$, $p_2 = 0.9310$. Using these values, one obtains

$$\hat{\rho} = 1 - \frac{z^2(0.0690) + z^2(0.9310)}{\dfrac{7.2839}{n} + 4.3986}$$

$$\mathrm{Var}(\hat{\rho}\,|\,\rho = 0) = \frac{\dfrac{13.337}{n^2} + \dfrac{14.831}{n}}{\dfrac{13.264}{n^2} + \dfrac{16.019}{n} + 4.8369} \simeq \frac{1}{1 + 0.3261n}$$

For $n = 200$

$$\hat{\rho} = 1 - 0.2255\,[z^2(0.0690) + z^2(0.9310)]$$

$$\mathrm{Var}(\hat{\rho}\,|\,\rho = 0) = 0.01515$$

$$\frac{1}{1 + 0.3261n} = 0.01510$$

$$\mathrm{Eff}(\hat{\rho}) = 0.3317$$

G. An Unbiased Estimator of ρ Using Two Pairs of Quantiles

Let z_1 and z_2 be two sample quantiles of orders p_1 and $p_2 = 1 - p_1$ of the $\{u_i\}$ and z_1' and z_2' the corresponding sample quantiles of the $\{v_i\}$ and let

$$y = z_2 - z_1$$

$$y' = z_2' - z_1'$$

Then one has

$$E(y) = 2\zeta_2(1+\rho)^{1/2}$$

$$\mathrm{Var}(y) = 2a^2(1+\rho)(1-\rho_{12})$$

$$E(y') = 2\zeta_2(1-\rho)^{1/2}$$

$$\mathrm{Var}(y') = 2a^2(1-\rho)(1-\rho_{12})$$

where

$$a^2 = \frac{F(\zeta_2)[1 - F(\zeta_2)]}{nf^2(\zeta_2)}$$

and ρ_{12} denotes the correlation between z_1 and z_2 as well as that between z_1' and z_2'. An unbiased estimator of ρ using two pairs of symmetric quantiles is given by

$$\hat{\rho} = \frac{y^2 - (y')^2}{4[a^2(1-\rho_{12}) + 2\zeta_2^2]}$$

and

$$E(\hat{\rho}) = \frac{1}{4[a^2(1-\rho_{12}) + {}^2\zeta_2^2]}[2a^2(1+\rho)(1-\rho_{12}) + 4\zeta_2^2(1+\rho) - 2a^2(1-\rho)(1-\rho_{12}) - 4\zeta_2^2(1-\rho)] = \rho$$

$$\mathrm{Var}(\hat{\rho}) = \frac{1}{16[a^2(1-\rho_{12}) + 2\zeta_2^2]^2}[\mathrm{Var}(y^2) + \mathrm{Var}(y')^2]$$

$$= \frac{8a^4(1+\rho)^2(1-\rho_{12})^2 + 8a^4(1-\rho)^2(1-\rho_{12})^2 + 32\zeta_2^2a^2(1+\rho)^2(1-\rho_{12}) + 32\zeta_2^2a^2(1-\rho)^2(1-\rho_{12})}{16[a^2(1-\rho_{12}) + 2\zeta_2^2]^2}$$

$$= \frac{[a^4(1-\rho_{12})^2 + 4a^2\zeta_2^2(1-\rho_{12})](1+\rho^2)}{[a^2(1-\rho_{12}) + 2\zeta_2^2]^2} \tag{45}$$

For $\rho = 0$, Eq. (45) can be written as

$$\mathrm{Var}(\hat{\rho}\,|\,\rho = 0) = \frac{a^4(1-\rho_{12})^2 + 4a^2\zeta_2^2(1-\rho_{12})}{a^4(1-\rho_{12})^2 + 4a^2\zeta^2(1-\rho_{12}) + 4\zeta_2^2}$$

$$\cong \left[1 + \frac{\zeta_2^2}{a^2(1-\rho_{12})}\right]^{-1} \qquad (46)$$

if one neglects the a^4 term in the numerator and denominator of Eq. (46). Since the same values of p_1 and p_2 which maximize $\zeta_2^2/[2a^2(1-\rho_{12})]$ will also maximize $\zeta_2^2/[a^2(1-\rho_{12})]$, we again use $p_1 = 0.0690$ and $p_2 = 0.9310$, which results in

$$\hat{\rho} = \frac{[z(0.9310) - z(0.0690)]^2 - [z'(0.9310) - z'(0.0690)]^2}{\dfrac{13.4875}{n} + 17.5959}$$

$$\mathrm{Var}(\hat{\rho}\,|\,\rho = 0) = \frac{\dfrac{11.369}{n^2} + \dfrac{29.663}{n}}{\dfrac{11.369}{n^2} + \dfrac{29.663}{n} + 19.347} \cong \frac{1}{1 + 0.6522n}$$

For $n = 200$,

$$\hat{\rho} = 0.05661\{[z(0.9310) - z(0.0690)]^2 - [z'(0.9310) - z'(0.0690)]^2\}$$

$$\mathrm{Var}(\hat{\rho}\,|\,\rho = 0) = 0.007622$$

$$(1 + 0.6522n)^{-1} = 0.007607$$

$$\mathrm{Eff}(\hat{\rho}) = 0.6593$$

H. An Unbiased Estimator of ρ Using Four Pairs of Quantiles

Let z_i, $i = 1, 2, 3, 4$ be four sample quantiles of the $\{u_i\}$ such that $p_1 + p_4 = p_2 + p_3 = 1$, and let z_i' be the corresponding sample quantiles of the $\{v_i\}$. Furthermore, let

$$y_1 = z_4 - z_1$$
$$y_2 = z_3 - z_2$$
$$y_1' = z_4' - z_1'$$
$$y_2' = z_3' - z_2'$$

An unbiased estimator of ρ using four pairs of symmetric quantiles is given by

$$\hat{\rho} = \frac{\alpha[y_1^2 - (y_1')^2] + \beta[y_2^2 - (y_2')^2]}{4\{\alpha[a_4^2(1-\rho_{14}) + 2\zeta_4^2] + \beta[a_3^2(1-\rho_{23}) + 2\zeta_3^2]\}}$$

where

$$a_i = \frac{F(\zeta_i)[1 - F(\zeta_i)]}{nf^2(\zeta_i)}, \qquad i = 3, 4$$

and ρ_{ij} denotes the correlation between z_i and z_j, as well as the correlation between z_i' and z_j'.

It was seen previously that the orders of the quantiles which are optimum with respect to estimating σ are very nearly optimum with respect to estimating ρ, and were therefore used in the estimators. We will adopt the same procedure in this case and, in addition, will use the optimum values of the weighting factors, α and β. Accordingly, the following values will be used

$$\begin{aligned} \alpha &= 0.116 \\ \beta &= 0.236 \\ p_1 &= 0.0230 \\ p_2 &= 0.1269 \\ p_3 &= 0.8731 \\ p_4 &= 0.9770 \end{aligned} \qquad (47)$$

Omitting the details, one has the following:

$$\mathrm{Var}(\hat{\rho}) = \frac{1}{\{\alpha[a_4^2(1-\rho_{14}) + 2\zeta_4^2] + \beta[a_3^2(1-\rho_{23}) + 2\zeta_3^2]\}^2}\{\alpha^2[a_4^4(1-\rho_{14})^2 + 4a_4^2\zeta_4^2(1-\rho_{14})]$$
$$+ \beta^2[a_3^4(1-\rho_{23}^2) + 4a_3^2\zeta_3^2(1-\rho_{23})] + 2\alpha\beta[a_3^2a_4^2(\rho_{12}-\rho_{13})^2$$
$$+ 4a_3a_4\zeta_3\zeta_4(\rho_{12}-\rho_{13})]\}(1+\rho^2)$$

and, using the values given in Eqs. (47),

$$\hat{\rho} = \frac{0.116\{[z(0.9770) - z(0.0230)]^2 - [z'(0.9770) - z'(0.0230)]^2\}}{\dfrac{5.4892}{n} + 6.1471}$$
$$+ \frac{0.236\{[z(0.8731) - z(0.1269)]^2 - [z'(0.8731) - z'(0.1269)]^2\}}{\dfrac{5.4892}{n} + 6.1471}$$

$$\mathrm{Var}(\hat{\rho}\,|\,\rho = 0) = \frac{\dfrac{1.1247}{n^2} + \dfrac{2.9673}{n}}{\dfrac{1.8832}{n^2} + \dfrac{4.2178}{n} + 2.3617}$$

For $n = 200$,

$$\hat{\rho} = 0.01879\{[z(0.9770) - z(0.0230)]^2 - [z'(0.9770) - z'(0.0230)]^2\}$$
$$+ 0.03822\{[z(0.8731) - z(0.1269)]^2 - [z'(0.8731) - z'(0.1269)]^2\}$$

$$\mathrm{Var}(\hat{\rho}\,|\,\rho = 0) = 0.006238$$
$$\mathrm{Eff}(\hat{\rho}) = 0.8056$$

VI. APPLYING THE TESTS

Two sets of samples, each containing 200 sample values were drawn from a table of random numbers (Ref. 8), in which the entries were distributed $N(0, 1)$. Hence, the sets of sample values can be considered as samples of two independent normal random variables x and y, with means $\mu_x = \mu_y = 0$ and variances $\sigma_x^2 = \sigma_y^2 = 1$. The sample quantiles (denoted by $z(p)$ and $z'(p)$, respectively) necessary to perform Tests $\bar{\mathrm{A}}$, $\bar{\mathrm{D}}$, $\bar{\mathrm{E}}$, and $\bar{\mathrm{F}}$, as well as those for the estimation of ρ, were determined. All the tests were performed at a significance level of 0.01. From the samples of x, the following values were obtained:

$z(0.5) = 0.006$	$z(0.0576) = -1.601$	$z(0.127) = -1.226$
$z(0.2703) = -0.681$	$z(0.9424) = 1.355$	$z(0.873) = 1.014$
$z(0.7297) = 0.526$	$z(0.0690) = -1.562$	$z(0.4177) = -0.227$
$z(0.1068) = -1.349$	$z(0.9310) = 1.303$	$z(0.5823) = 0.241$
$z(0.8932) = 1.119$	$z(0.023) = -2.067$	$z(0.4418) = -0.144$
$z(0.3512) = -0.396$	$z(0.977) = 1.939$	$z(0.5582) = 0.169$
$z(0.6488) = 0.356$		$z(0.8751) = 1.044$

From the samples of y, the following values were also obtained:

$z'(0.5) = 0.027$	$z'(0.0576) = -1.554$	$z'(0.127) = -1.095$
$z'(0.2703) = -0.697$	$z'(0.9424) = 1.663$	$z'(0.873) = 1.156$
$z'(0.7297) = 0.635$	$z'(0.0690) = -1.475$	$z'(0.4177) = -0.224$
$z'(0.1068) = -1.231$	$z'(0.9310) = 1.624$	$z'(0.5823) = 0.254$
$z'(0.8932) = 1.316$	$z'(0.023) = -2.068$	$z'(0.4418) = -0.171$
$z'(0.3512) = -0.359$	$z'(0.977) = 2.169$	$z'(0.5582) = 0.188$
$z'(0.6488) = 0.393$		$z'(0.8751) = 1.179$

The sample means, $\bar{x}$ and $\bar{y}$, and the sample standard deviations, s_x and s_y, as well the corresponding estimates using four optimal quantiles were computed and found to be:

$\bar{x} = -0.0710$	$s_x = 1.000$	$\hat{\mu}_x = -0.056$	$\hat{\sigma}_x = 0.993$
$\bar{y} = 0.0109$	$s_y = 1.039$	$\hat{\mu}_y = 0.027$	$\hat{\sigma}_y = 1.023$

The estimators of $\rho = 0$ using one quantile and one, two, and four pairs of quantiles, denoted by $\hat{\rho}_0$, $\hat{\rho}_1$, $\hat{\rho}_2$, and $\hat{\rho}_4$, as well as the sample correlation r, were also computed and found to be:

$\hat{\rho}_0 = -0.1057$	$\hat{\rho}_2 = 0.0790$	$r = 0.0434$
$\hat{\rho}_1 = -0.0853$	$\hat{\rho}_4 = 0.0377$	

Test $\bar{\mathrm{A}}$ using one, two, and four sample quantiles was performed independently on both sets of samples with H_0 being true. In all six tests, H_0 was accepted. For Tests $\bar{\mathrm{D}}$ and $\bar{\mathrm{E}}$, which require sample quantiles from both sets of samples for each test, H_0 was accepted in all six tests when H_0 was true. For Test $\bar{\mathrm{F}}$, it was assumed that the given sets of sample values were actually transformed values $\{u_i\}$ and $\{v_i\}$ obtained from sets $\{x_i\}$ and $\{y_i\}$ taken from two normal distributions with $\rho = 0$. With H_0 being true, in each of the three tests of Test $\bar{F}$, H_0 was accepted.

Now, if x is distributed $N(\mu, \sigma)$, then $x' = ax + b$, $a > 0$, is distributed $N(a\mu + b, a\sigma)$. If the above transformation were applied to all the sample values taken from a population distributed $N(\mu, \sigma)$, one sees that not only would the new sample values be distributed $N(a\mu + b, a\sigma)$, but the *order* of the samples would remain unchanged, that is, if $x_i < x_j$, then $x'_i < x'_j$. Hence, if $z(p)$ were the quantile of order p of the $\{x_i\}$, then $az(p) + b$ would be the quantile of order p of the $\{x'_i\}$. This fact permits us to perform Tests $\bar{A}$, $\bar{D}$, $\bar{E}$, and $\bar{F}$ when H_0 is *not true* by simply performing a linear transformation on the sample quantiles of the x_i and y_i. These tests will be given in detail. The best tests using all the sample values will also be given.

In Test $\bar{A}$, by adding 0.25 to each quantile $z(p)$ and $z'(p)$, one can assume in each case that $\mu_1 = 0$, $\mu_2 = 0.25$, $\sigma = 1$, and H_1 is true. The results of each test and the decision are as follows. [$\tilde{z}(p)$ and $\tilde{z}'(p)$ will denote the sample quantiles *after* the transformation.]

$$\tilde{z}(0.4177) = 0.023 > \mu_1, \qquad \text{reject } H_0$$

$$\tilde{z}'(0.4177) = 0.026 > \mu_1, \qquad \text{reject } H_0$$

$$0.6525\tilde{z}(0.2703) + 0.3475\tilde{z}(0.7297) = -0.0120 < 0, \qquad \text{accept } H_0$$

$$0.6525\tilde{z}'(0.2703) + 0.3475\tilde{z}'(0.7297) = 0.016 > 0, \qquad \text{reject } H_0$$

$$0.2445\tilde{z}(0.1068) + 0.1391\tilde{z}(0.8932) + 0.3609\tilde{z}(0.3512) + 0.2555\tilde{z}(0.6488) = 0.024 > 0, \qquad \text{reject } H_0$$

$$0.2445\tilde{z}'(0.1068) + 0.1391\tilde{z}'(0.8932) + 0.3609\tilde{z}'(0.3512) + 0.2555\tilde{z}'(0.6488) = 0.103 > 0, \qquad \text{reject } H_0$$

Adding 0.25 to each sample value and applying Test A to all the sample values results in

$$n^{-1} \sum_{i=1}^{n} x_i = 0.179 > 0.1645, \qquad \text{reject } H_0$$

$$n^{-1} \sum y_i = 0.261 > 0.1645, \qquad \text{reject } H_0$$

In Test $\bar{D}$, by putting $\theta = 0.25$ and hence adding 0.25 to each $z'(p)$ and leaving $z(p)$ unchanged, one can assume that $\mu_2 = \mu_1 + 0.25$, $\sigma_1 = \sigma_2 = 1$ and H_1 is true. Then one has:

$$z(0.5582) - \tilde{z}'(0.4418) = 0.090 > 0, \qquad \text{accept } H_0$$

$$1.5425[z(0.7297) - \tilde{z}'(0.2703)] + z(0.2703) - \tilde{z}'(0.7297) = -0.065 < 0, \qquad \text{reject } H_0$$

$$0.2293[z(0.8932) - \tilde{z}'(0.1068)] + 0.1543[z(0.1068) - \tilde{z}'(0.8932)] + 0.3457[z(0.6488) - \tilde{z}'(0.3512)] + 0.2707[z(0.3512) - \tilde{z}'(0.6488)] = -0.089 < 0, \qquad \text{reject } H_0$$

Adding 0.25 to each y_i and leaving each x_i unchanged, and then applying Test D to all the sample values results in:

$$n^{-1} \sum_{i=1}^{n} x_i^2 = n^{-1} \sum_{i=1}^{n} y_i^2 = -0.332 < 0.233, \qquad \text{reject } H_0$$

In Test $\bar{E}$, by multiplying each $z'(p)$ by 1.15 and leaving each $z(p)$ unchanged, one can assume that $\theta = 1.15$, $\mu_1 = \mu_2 = 0$, $\sigma_2 = 1.15\sigma_1$, and H_1 is true. Then one has

$$z(0.9424) - \tilde{z}'(0.8751) = -0.001 < 0, \qquad \text{reject } H_0$$

$$1.228[z(0.9310) - z(0.0690)] - [\tilde{z}'(0.9310) - \tilde{z}'(0.0690)] = -0.046 < 0, \qquad \text{reject } H_0$$

$$0.1392[z(0.9770) - z(0.0230)] + 0.2832[z(0.8729) - z(0.1271)] - 0.116[\tilde{z}'(0.9770) - \tilde{z}'(0.0230)] - 0.236[\tilde{z}'(0.8729) - \tilde{z}'(0.1271)] = 0.016 > 0, \qquad \text{accept } H_0$$

Multiplying each y_i by 1.15 and leaving each x_i unchanged, and then applying Test E to all the sample values results in:

$$\frac{1}{2} \ln \left(\frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} y_i^2} \right) = -0.1752 < 0.1645, \qquad \text{reject } H_0$$

In Test $\bar{F}$, by multiplying each $z(p)$ by $(1.25)^{1/2}$ and each $z'(p)$ by $(0.75)^{1/2}$, it can be assumed that each $\tilde{z}(p)$ is the quantile of order p of a transformed set of variables $\{u_i\}$ distributed $N[0, (1 + \rho)^{1/2}]$, and each resulting $\tilde{z}'(p)$

is the quantile of order p of a transformed set $\{v_i\}$, distributed $N[0,(1-\rho)^{1/2}]$, and the transformations were applied to correlated sets $\{x_i\}$ and $\{y_i\}$, each distributed $N(0,1)$ with $\rho = 0.25$. Hence, for Test $\overline{\text{F}}$, one has $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$, $\rho = 0.25$ and H_1 is true. The results of the tests are

$$\tilde{z}(0.8751) - \tilde{z}'(0.9424) = -0.273 < 0, \qquad \text{accept } H_0$$

$$-[\tilde{z}(0.9310) - \tilde{z}(0.0690)] + 1.228[\tilde{z}'(0.9310) - \tilde{z}'(0.0690)] = 0.093 > 0, \qquad \text{accept } H_0$$

$$-0.116[\tilde{z}(0.9770) - \tilde{z}(0.0230)] - 0.236[\tilde{z}(0.8729) - \tilde{z}(0.1271)] + 0.1392[\tilde{z}'(0.9770) - \tilde{z}'(0.0230)] + 0.2832[\tilde{z}'(0.8729) - z'(0.1271)] = -0.048 < 0, \qquad \text{reject } H_0$$

Multiplying each x_i by $(1.25)^{1/2}$ and each y_i by $(0.75)^{1/2}$ and applying Test F to all the sample values results in:

$$\sum_{i=1}^{n} (\rho - 1) u_i^2 + \sum_{i=1}^{n} (\rho + 1) v_i^2 = 13.78 < 32.19, \qquad \text{reject } H_0$$

VII. SUB-OPTIMUM TEST STATISTICS

Tables 11 through 14 give the test statistics and acceptance regions of the tests as functions of n and ε. When more than one quantile is used in Test $\overline{\text{A}}$ and more than one pair of quantiles is used in Tests $\overline{\text{D}}$, $\overline{\text{E}}$, and $\overline{\text{F}}$, the choice of which pairs of symmetric quantiles to use is at our disposal. For these cases, we have chosen the same quantiles as those used in the corresponding tests discussed in Ref. 2. The fact that this choice is near-optimum is evident from the data given in Figs. 1 through 3 and in Tables 2 through 10.

Table 11. Test statistics and acceptance regions of Test $\overline{\text{A}}$, using near-optimum quantiles for Tests $\overline{\text{A}}_2$ and $\overline{\text{A}}_4$

H_0: $g(x) = g_1(x) = N(\mu_1, \sigma)$

H_1: $g(x) = g_2(x) = N(\mu_2, \sigma)$, σ unknown

Conditions	Acceptance regions	Restraints
Test $\overline{\text{A}}_1$ ($\mu_2 > \mu_1$)	$z(p) < \mu_1$	$\frac{-\zeta}{a} = b$; $F(b) = 1 - \varepsilon$
Test $\overline{\text{A}}_1$ ($\mu_2 < \mu_1$)	$z(p) > \mu_1$	$\frac{-\zeta}{a} = b$; $F(b) = \varepsilon$
Test $\overline{\text{A}}_2$ ($\mu_2 > \mu_1$)	$\left(\frac{1+\alpha}{2}\right) z(0.2703) + \left(\frac{1-\alpha}{2}\right) z(0.7297) < \mu_1$	$\alpha^2 = \frac{2.1767b^2}{0.6602n - b^2}$; $F(b) = \varepsilon$
Test $\overline{\text{A}}_2$ ($\mu_2 < \mu_1$)	$\left(\frac{1-\alpha}{2}\right) z(0.2703) + \left(\frac{1+\alpha}{2}\right) z(0.7297) > \mu_1$	$\alpha^2 = \frac{2.1767b^2}{0.6602n - b^2}$; $F(b) = \varepsilon$
Test $\overline{\text{A}}_4$ ($\mu_2 > \mu_1$)	$(0.192 + \alpha) z(0.1068) + (0.192 - \alpha) z(0.8932) + (0.308 + \alpha) z(0.3512) + (0.308 - \alpha) z(0.6488) < \mu_1$	$\alpha^2 = \frac{0.1303b^2}{1.2678n - b^2}$; $F(b) = \varepsilon$
Test $\overline{\text{A}}_4$ ($\mu_2 < \mu_1$)	$(0.192 - \alpha) z(0.1068) + (0.192 + \alpha) z(0.8932) + (0.308 - \alpha) z(0.3512) + (0.308 + \alpha) z(0.6488) > \mu_1$	$\alpha^2 = \frac{0.1303b^2}{1.2678n - b^2}$; $F(b) = \varepsilon$

Table 12. Test statistics and acceptance regions for Test $\bar{D}$, using near-optimum quantiles for Tests $\bar{D}_2$ and $\bar{D}_4$

H_0: $g_1(x) = N(\mu, \sigma)$, $g_2(y) = N(\mu, \sigma)$, μ and σ unknown

H_1: $g_1(x) = N(\mu, \sigma)$, $g_2(y) = N(\mu + \theta, \sigma)$, $\theta \neq 0$

Conditions	Acceptance regions	Restraints
Test $\bar{D}_1$ ($\theta > 0$)	$z(p) - z'(1-p) > 0$	$p > 0.5$ $\frac{-\zeta}{a} = \frac{b}{2^{1/2}}$ $F(b) = \epsilon$
Test $\bar{D}_1$ ($\theta < 0$)	$z(p) - z'(1-p) < 0$	$p < 0.5$ $\frac{-\zeta}{a} = \frac{b}{2^{1/2}}$ $F(b) = 1 - \epsilon$
Test $\bar{D}_2$ ($\theta > 0$)	$(1+\alpha)[z(0.7297) - z'(0.2703)] + (1-\alpha)[z(0.2703) - z'(0.7297)] > 0$	$\alpha^2 = \frac{2.1767b^2}{1.3204n - b^2}$ $F(b) = \epsilon$
Test $\bar{D}_2$ ($\theta < 0$)	$(1-\alpha)[z(0.7297) - z'(0.2703)] + (1+\alpha)[z(0.2703) - z'(0.7297)] < 0$	$\alpha^2 = \frac{2.1767b^2}{1.3204n - b^2}$ $F(b) = \epsilon$
Test $\bar{D}_4$ ($\theta > 0$)	$(0.192+\alpha)[z(0.8932) - z'(0.1068)] + (0.192-\alpha)[z(0.1068) - z'(0.8932)]$ $+ (0.308+\alpha)[z(0.6488) - z'(0.3512)] + (0.308-\alpha)[z(0.3512) - z'(0.6488)] > 0$	$\alpha^2 = \frac{0.1303b^2}{2.5356n - b^2}$ $F(b) = \epsilon$
Test $\bar{D}_4$ ($\theta < 0$)	$(0.192-\alpha)[z(0.8932) - z'(0.1068)] + (0.192+\alpha)[z(0.1068) - z'(0.8932)]$ $+ (0.308-\alpha)[z(0.6488) - z'(0.3512)] + (0.308+\alpha)[z(0.3512) - z'(0.6488)] < 0$	$\alpha^2 = \frac{0.1303b^2}{2.5356n - b^2}$ $F(b) = \epsilon$

Table 13. Test statistics and acceptance regions for Test $\bar{E}$, using near-optimum quantiles for Tests $\bar{E}_2$ and $\bar{E}_4$

H_0: $g_1(x) = N(\mu, \sigma)$, $g_2(y) = N(\mu, \sigma)$, μ and σ unknown

H_1: $g_1(x) = N(\mu, \sigma)$, $g_2(y) = N(\mu, \theta\sigma)$, $\theta > 0$

Conditions	Acceptance regions	Constraints
Test $\bar{E}_1$ ($\theta > 1$)	$z(0.9424) - z'(p') > 0$	$0 < p' < 0.9424$ $\frac{-(1.575 - \zeta')}{\left[\frac{4.07714}{n} + (a')^2\right]^{1/2}} = b$ $F(b) = \epsilon$
Test $\bar{E}_1$ ($\theta < 1$)	$z(p) - z'(0.9424) < 0$	$0 < p < 0.9424$ $\frac{-(\zeta - 1.575)}{\left(a^2 + \frac{4.07714}{n}\right)^{1/2}} = b$ $F(b) = 1 - \epsilon$
Test $\bar{E}_2$ ($\theta > 1$)	$(1+\alpha)[z(0.9310) - z(0.0690)] + (1-\alpha)[z'(0.9310) - z'(0.0690)] > 0$	$\alpha^2 = \frac{2.609n}{b^2} - 1$ $F(b) = \epsilon$
Test $\bar{E}_2$ ($\theta < 1$)	$(1-\alpha)[z(0.9310) - z(0.0690)] + (1+\alpha)[z'(0.9310) - z'(0.0690)] > 0$	$\alpha^2 = \frac{2.609n}{b^2} - 1$ $F(b) = \epsilon$
Test $\bar{E}_4$ ($\theta > 1$)	$(1+\alpha)\{0.116[z(0.9770) - z(0.0230)] + 0.236[z(0.8729) - z(0.1271)]\}$ $+ (1-\alpha)\{0.116[z'(0.9770) - z'(0.0230)] + 0.236[z'(0.8729) - z'(0.1271)]\} > 0$	$\alpha^2 = \frac{3.2977n}{b^2} - 1$ $F(b) = \epsilon$
Test $\bar{E}_4$ ($\theta < 1$)	$(1-\alpha)\{0.116[z(0.9770) - z(0.0230)] + 0.236[z(0.8729) - z(0.1271)]\}$ $+ (1+\alpha)\{0.116[z'(0.9770) - z'(0.0230)] + 0.236[z'(0.8729) - z'(0.1271)]\} > 0$	$\alpha^2 = \frac{3.2977n}{b^2} - 1$ $F(b) = \epsilon$

Table 14. Test statistics and acceptance regions for Test $\bar{F}$, using near-optimum quantiles for Tests $\bar{F}_2$ and $\bar{F}_4$

$$H_0: \quad g_1(x) = N(\mu, \sigma), \quad g_2(y) = N(\mu, \sigma), \quad \rho = 0$$

$$H_1: \quad g_1(x) = N(\mu, \sigma), \quad g_2(y) = N(\mu, \sigma), \quad \rho \neq 0$$

Conditions	Acceptance regions	Constraints
Test $\bar{F}_1$ ($\mu = 0$) σ unknown ($\rho < 0$)	$z(0.9424) - z'(p') > 0$	$0 < p' < 0.9424$ $\dfrac{-(1.575 - \zeta')}{\left[\dfrac{4.07714}{n} + (a')^2\right]^{1/2}} = b$ $F(b) = \varepsilon$
Test $\bar{F}_1$ ($\mu = 0$) σ unknown ($\rho > 0$)	$z(p) - z'(0.9424) < 0$	$0 < p < 0.9424$ $\dfrac{-(\zeta - 1.575)}{\left(a^2 + \dfrac{4.07714}{n}\right)} = b$ $F(b) = 1 - \varepsilon$
Test $\bar{F}_2$ μ and σ unknown ($\rho < 0$)	$(1 + \alpha)[z(0.9310) - z(0.0690)] + (1 - \alpha)[z'(0.9310) - z'(0.0690)] > 0$	$\alpha^2 = \dfrac{2.609n}{b^2} - 1$ $F(b) = \varepsilon$
Test $\bar{F}_2$ μ and σ unknown ($\rho > 0$)	$(1 - \alpha)[z(0.9310) - z(0.0690)] + (1 + \alpha)[z'(0.9310) - z'(0.0690)] > 0$	$\alpha^2 = \dfrac{2.609n}{b^2} - 1$ $F(b) = \varepsilon$
Test $\bar{F}_4$ μ and σ unknown ($\rho < 0$)	$(1 + \alpha)\{0.116[z(0.9770) - z(0.0230)] + 0.236[z(0.8729) - z(0.1271)]\}$ $+ (1 - \alpha)\{0.116[z'(0.9770) - z'(0.0230)] + 0.236[z'(0.8729) - z'(0.1271)]\} > 0$	$\alpha^2 = \dfrac{3.2977n}{b^2} - 1$ $F(b) = \varepsilon$
Test $\bar{F}_4$ μ and σ unknown ($\rho > 0$)	$(1 - \alpha)\{0.116[z(0.9770) - z(0.0230)] + 0.236[z(0.8729) - z(0.1271)]\}$ $+ (1 + \alpha)\{0.116[z'(0.9770) - z'(0.0230)] + 0.236[z'(0.8729) - z'(0.1271)]\} > 0$	$\alpha^2 = \dfrac{3.2977n}{b^2} - 1$ $F(b) = \varepsilon$

Now, in order to apply the results developed here to statistical experiments performed aboard a spacecraft, it is necessary to specify the orders of the quantiles in advance. For maximum data compression, only one set of k quantiles should be so specified for a k quantile test or estimator, regardless of which test or estimator is required. The problem then, of course, is to decide on which set of k quantiles to use. Since a set of quantiles which is optimum for one test is not necessarily optimum with respect to another, it is obvious that a compromise is required based on some reasonable criterion. This problem is not a new one. It was encountered in our previous investigations into the use of quantiles for data compression; hence, a proposed solution, which will be restricted to the four-quantile case, is at hand.

It has no doubt been noted that only two sets of k quantiles have been used for the tests and for estimating ρ, for $k = 2, 4$. The sets used in Tests $\bar{A}$ and $\bar{D}$ are those which provide the asymptotically unbiased estimators of the mean of a single normal population with minimum variance; the sets used in Tests $\bar{E}$ and $\bar{F}$ and for estimating ρ are those which provide the asymptotically unbiased estimators of the standard deviation with minimum variance. In the four-quantile cases, the weighting factors are also identical with those used in the estimators of μ and σ. Thus, we are faced with the problem of effecting a compromise between two sets of quantiles, one which minimizes $\mathrm{Var}(\hat{\mu})$ and another which minimizes $\mathrm{Var}(\hat{\sigma})$. The compromise we now propose is one which was adopted previously for estimating μ and σ and for Tests

A_4, D_4, E_4, and F_4. Determine the orders of the set of two pairs of symmetric quantiles and weights α_1, β_1, α_2, β_2 such that unbiased estimators of μ and σ are given by

$$\hat{\mu} = \alpha_1 (z_1 + z_4) + \beta_1 (z_2 + z_3)$$

$$\hat{\sigma} = \alpha_2 (z_4 - z_1) + \beta_2 (z_3 - z_2)$$

and for which the linear combination $\mathrm{Var}(\hat{\mu}) + c\,\mathrm{Var}(\hat{\sigma})$ is a minimum, $c = 1, 2, \cdots$

The same sets of quantiles are to be used in all tests and for estimating ρ. Weights α_1 and β_1 are to be used in the test statistics of Tests $\bar{A}_4$ and $\bar{D}_4$, while α_2 and β_2 are to be used in the test statistics of Tests $\bar{E}_4$ and $\bar{F}_4$ as well as for estimating ρ.

In Ref. 1, sets of four quantiles are given which meet the previously given conditions for $c = 1, 2, 3$. The orders of the quantiles and the weights are as follows. For $c = 1$:

$\alpha_1 = 0.141$	$\beta_1 = 0.359$
$\alpha_2 = 0.258$	$\beta_2 = 0.205$
$p_1 = 0.0668$	$p_2 = 0.2912$
$p_4 = 0.9332$	$p_3 = 0.7088$

For $c = 2$:

$\alpha_1 = 0.106$	$\beta_1 = 0.394$
$\alpha_2 = 0.196$	$\beta_2 = 0.232$
$p_1 = 0.0434$	$p_2 = 0.2381$
$p_4 = 0.9566$	$p_3 = 0.7619$

For $c = 3$:

$\alpha_1 = 0.097$	$\beta_1 = 0.403$
$\alpha_2 = 0.179$	$\beta_2 = 0.235$
$p_1 = 0.0389$	$p_2 = 0.2160$
$p_4 = 0.9611$	$p_3 = 0.7840$

Table 15. Sub-optimum test statistics and acceptance regions for $k = 4$, $c = 1$

$p_1 = 0.0668$; $\rho_2 = 0.2912$; $\rho_3 = 0.7088$; $p_4 = 0.9332$

Conditions	Acceptance regions	Restraints
Test $\bar{A}_4$ ($\mu_2 > \mu_1$)	$(0.141 + \alpha) z_1 + (0.141 - \alpha) z_4 + (0.359 + \alpha) z_2 + (0.359 - \alpha) z_3 < \mu_1$	$\alpha^2 = \dfrac{0.09591 b^2}{1.4642n - b^2}$ $F(b) = \varepsilon$
Test $\bar{A}_4$ ($\mu_2 < \mu_1$)	$(0.141 - \alpha) z_1 + (0.141 + \alpha) z_4 + (0.359 - \alpha) z_2 + (0.359 + \alpha) z_3 > \mu_1$	$\alpha^2 = \dfrac{0.09591 b^2}{1.4642n - b^2}$ $F(b) = \varepsilon$
Test $\bar{D}_4$ ($\theta > 0$)	$(0.141 + \alpha)(z_4 - z_1') + (0.141 - \alpha)(z_1 - z_4') + (0.359 + \alpha)(z_3 - z_2')$ $+ (0.359 - \alpha)(z_2 - z_3') > 0$	$\alpha^2 = \dfrac{0.09591 b^2}{2.9284n - b^2}$ $F(b) = \varepsilon$
Test $\bar{D}_4$ ($\theta < 0$)	$(0.141 - \alpha)(z_4 - z_1') + (0.141 + \alpha)(z_1 - z_4') + (0.359 - \alpha)(z_3 - z_2')$ $+ (0.359 + \alpha)(z_2 - z_3') < 0$	$\alpha^2 = \dfrac{0.09591 b^2}{2.9284n - b^2}$ $F(b) = \varepsilon$
Test $\bar{E}_4$ ($\theta > 1$)	$(1 + \alpha)[0.258(z_4 - z_1) + 0.205(z_3 - z_2)]$ $+ (1 - \alpha)[0.258(z_4' - z_1') + (0.205(z_3' - z_2')] > 0$	$\alpha^2 = \dfrac{2.9458n}{b^2} - 1$ $F(b) = \varepsilon$
Test $\bar{E}_4$ ($\theta < 1$)	$(1 - \alpha)[0.258(z_4 - z_1) + (0.205(z_3 - z_2)]$ $+ (1 + \alpha)[0.258(z_4' - z_1') + (0.205(z_3' - z_2')] > 0$	$\alpha^2 = \dfrac{2.9458n}{b^2} - 1$ $F(b) = \varepsilon$
Test $\bar{F}_4$ ($\rho < 0$)	$(1 + \alpha)[0.258(z_4 - z_1) + 0.205(z_3 - z_2)]$ $+ (1 - \alpha)[0.258(z_4' - z_1') + (0.205(z_3' - z_2')] > 0$	$\alpha^2 = \dfrac{2.9458n}{b^2} - 1$ $F(b) = \varepsilon$
Test $\bar{F}_4$ ($\rho > 0$)	$(1 - \alpha)[0.258(z_4 - z_1) + (0.205(z_3 - z_2)]$ $+ (1 + \alpha)[0.258(z_4' - z_1') + 0.205(z_3' - z_2')] > 0$	$\alpha^2 = \dfrac{2.9458n}{b^2} - 1$ $F(b) = \varepsilon$

New test statistics using the sub-optimum sets of quantiles were computed and are given in Tables 15 through 17. The power of each test (and hence the efficiency) will, of course, decrease. Table 18 gives the near-optimum and sub-optimum minimum efficiencies, and it can be seen that the decrease in efficiency is never critical. It should be noticed that as c increases, the efficiencies of Test $\bar{A}_4$ and $\bar{D}_4$, which are concerned with μ and use coefficients α_1 and β_1, decrease. The efficiencies of Tests $\bar{E}_4$ and $\bar{F}_4$, which use coefficients α_2 and β_2, increase as c increases. This is consistent with the fact that as c increases, greater weight is given to $\mathrm{Var}(\hat{\sigma})$ than to $\mathrm{Var}(\hat{\mu})$. This suggests that the choice of c should depend on the relative importance of the tests performed.

Table 19 lists the near-optimum and sub-optimum estimators of ρ. The efficiencies of the sub-optimum estimators increase with increasing values of c, which is again consistent with the use of coefficients α_2 and β_2. The loss in efficiency in going from near-optimum to sub-optimum conditions does not appear to be excessive.

Table 16. Sub-optimum test statistics and acceptance regions for $k = 4$, $c = 2$

$p_1 = 0.0434$; $p_2 = 0.2381$; $p_3 = 0.7619$; $p_4 = 0.9566$

Conditions	Acceptance regions	Restraints
Test $\bar{A}_4$ $(\mu_2 > \mu_1)$	$(0.106 + \alpha) z_1 + (0.106 - \alpha) z_4 + (0.394 + \alpha) z_2 + (0.394 - \alpha) z_3 < \mu_1$	$\alpha^2 = \dfrac{0.08248b^2}{1.7288n - b^2}$ $F(b) = \varepsilon$
Test $\bar{A}_4$ $(\mu_2 < \mu_1)$	$(0.106 - \alpha) z_1 + (0.106 + \alpha) z_4 + (0.394 - \alpha) z_2 + (0.394 + \alpha) z_3 > \mu_1$	$\alpha^2 = \dfrac{0.08248b^2}{1.7288n - b^2}$ $F(b) = \varepsilon$
Test $\bar{D}_4$ $(\theta > 0)$	$(0.106 + \alpha)(z_1 - z_1') + (0.106 - \alpha)(z_4 - z_4') + (0.394 + \alpha)(z_3 - z_2') + (0.394 + \alpha)(z_2 - z_3') > 0$	$\alpha^2 = \dfrac{0.08248b^2}{3.4576n - b^2}$ $F(b) = \varepsilon$
Test $\bar{D}_4$ $(\theta < 0)$	$(0.106 - \alpha)(z_1 - z_1') + (0.106 + \alpha)(z_4 - z_4') + (0.394 - \alpha)(z_3 - z_2') + (0.394 - \alpha)(z_2 - z_3') < 0$	$\alpha^2 = \dfrac{0.08248b^2}{3.4576n - b^2}$ $F(b) = \varepsilon$
Test $\bar{E}_4$ $(\theta > 1)$	$(1 + \alpha)[0.196(z_4 - z_1) + 0.232(z_3 - z_2)] + (1 - \alpha)[0.196(z_4' - z_2') + 0.232(z_3' - z_2')] > 0$	$\alpha^2 = \dfrac{3.0984n}{b^2} - 1$ $F(b) = \varepsilon$
Test $\bar{E}_4$ $(\theta < 1)$	$(1 - \alpha)[0.196(z_4 - z_1) + 0.232(z_3 - z_2)] + (1 + \alpha)[0.196(z_4' - z_1') + 0.232(z_3' - z_2')] > 0$	$\alpha^2 = \dfrac{3.0984n}{b^2} - 1$ $F(b) = \varepsilon$
Test $\bar{F}_4$ $(\rho < 0)$	$(1 + \alpha)[0.196(z_4 - z_1) + (0.232(z_3 - z_2)] + (1 - \alpha)[0.196(z_4' - z_3') + 0.232(z_3' - z_2')] > 0$	$\alpha^2 = \dfrac{3.0984n}{b^2} - 1$ $F(b) = \varepsilon$
Test $\bar{F}_4$ $(\rho > 0)$	$(1 - \alpha)[0.196(z_4 - z_1) + (0.232(z_3 - z_2)] + (1 + \alpha)[0.196(z_4' - z_1') + 0.232(z_3' - z_2')] > 0$	$\alpha^2 = \dfrac{3.0984n}{b^2} - 1$ $F(b) = \varepsilon$

Table 17. Sub-optimum test statistics and acceptance regions for $k = 4$, $c = 3$

$p_1 = 0.0389$; $\rho_2 = 0.2160$; $\rho_3 = 0.7840$; $p_4 = 0.9611$

Conditions	Acceptance regions	Restraints
Test $\bar{A}_4$ $(\mu_2 > \mu_1)$	$(0.097 + \alpha) z_1 + (0.097 - \alpha) z_4 + (0.403 + \alpha) z_2 + (0.403 - \alpha) z_3 < \mu_1$	$\alpha^2 = \dfrac{0.07875b^2}{1.5709n - b^2}$ $F(b) = \varepsilon$
Test $\bar{A}_4$ $(\mu_2 < \mu_1)$	$(0.097 - \alpha) z_1 + (0.097 + \alpha) z_4 + (0.403 - \alpha) z_2 + (0.403 + \alpha) z_3 > \mu_1$	$\alpha^2 = \dfrac{0.07875b^2}{1.5709n - b^2}$ $F(b) = \varepsilon$
Test $\bar{D}_4$ $(\theta > 0)$	$(0.097 + \alpha)(z_4 - z_1') + (0.097 - \alpha)(z_1 - z_4') + (0.403 + \alpha)(z_3 - z_2') + (0.403 - \alpha)(z_2 - z_3') > 0$	$\alpha^2 = \dfrac{0.07875b^2}{3.1418n - b^2}$ $F(b) = \varepsilon$
Test $\bar{D}_4$ $(\theta < 0)$	$(0.097 - \alpha)(z_4 - z_1') + (0.097 + \alpha)(z_1 - z_4') + (0.403 - \alpha)(z_3 - z_2') + (0.403 + \alpha)(z_2 - z_3') < 0$	$\alpha^2 = \dfrac{0.07875b^2}{3.1418n - b^2}$ $F(b) = \varepsilon$
Test $\bar{\bar{E}}_4$ $(\theta > 1)$	$(1 + \alpha)[0.179(z_4 - z_1) + 0.235(z_3 - z_2)] + (1 - \alpha)[0.179(z_4' - z_1') + 0.235(z_3' - z_2')] > 0$	$\alpha^2 = \dfrac{3.1623n}{b^2} - 1$ $F(b) = \varepsilon$
Test $\bar{\bar{E}}_4$ $(\theta < 1)$	$(1 - \alpha)[0.179(z_4 - z_1) + 0.235(z_3 - z_2)] + (1 + \alpha)[0.179(z_4' - z_1') + 0.235(z_3' - z_2')] > 0$	$\alpha^2 = \dfrac{3.1623n}{b^2} - 1$ $F(b) = \varepsilon$
Test $\bar{F}_4$ $(\rho < 0)$	$(1 + \alpha)[0.179(z_4 - z_1) + 0.235(z_3 - z_2)] + (1 - \alpha)[0.179(z_4' - z_1') + 0.235(z_3' - z_2')] > 0$	$\alpha^2 = \dfrac{3.1623n}{b^2} - 1$ $F(b) = \varepsilon$
Test $\bar{F}_4$ $(\rho > 0)$	$(1 - \alpha)[0.179(z_4 - z_1) + 0.235(z_3 - z_2)] + (1 + \alpha)[0.179(z_4' - z_1') + 0.235(z_3' - z_2')] > 0$	$\alpha^2 = \dfrac{3.1623n}{b^2} - 1$ $F(b) = \varepsilon$

Table 18. Minimum efficiency under near-optimum and sub-optimum conditions for $k = 4$, $n = 200$, $\varepsilon = 0.01$

Test	Minimum efficiency			
	Near-optimum	$c = 1$	$c = 2$	$c = 3$
$\bar{A}_4$	0.9002	0.8887	0.8768	0.8661
$\bar{D}_4$	0.9070	0.8967	0.8835	0.8705
$\bar{\bar{E}}_4$	0.8036	0.7117	0.7516	0.7687
$\bar{F}_4$	0.8347	0.7423	0.7834	0.7999

Table 19. Estimators of ρ under near-optimum and sub-optimum conditions for $k = 4$

Conditions	Estimators of ρ	Efficiency ($n = 200$)
Near-optimum	$\hat{\rho} = \dfrac{0.116\,[(z_4 - z_1)^2 - (z_4' - z_1')^2] + 0.236\,[(z_3 - z_2)^2 - (z_3' - z_2')^2]}{\dfrac{5.4892}{n} + 6.1471}$	0.8056
$c = 1$	$\hat{\rho} = \dfrac{0.258\,[(z_4 - z_1)^2 - (z_4' - z_1')^2] + 0.205\,[(z_3 - z_2)^2 - (z_3' - z_2')^2]}{\dfrac{4.4083}{n} + 5.1399}$	0.7135
$c = 2$	$\hat{\rho} = \dfrac{0.196\,[(z_4 - z_1)^2 - (z_4' - z_1')^2] + 0.232\,[(z_3 - z_2)^2 - (z_3' - z_2')^2]}{\dfrac{4.8827}{n} + 5.5420}$	0.7445
$c = 3$	$\hat{\rho} = \dfrac{0.179\,[(z_4 - z_1)^2 - (z_4' - z_1')^2] + 0.235\,[(z_3 - z_2)^2 - (z_3' - z_2')^2]}{\dfrac{4.9631}{n} + 5.6148}$	0.7590

REFERENCES

1. Eisenberger, I. and Posner, E. C., "Systematic Statistics Used for Data Compression in Space Telemetry," *Journal of the American Statistical Association*, Vol. 60, March 1965, pp. 97–133 and Technical Report No. 32–510, October 1, 1963.

2. Eisenberger, I., *Tests of Hypotheses and Estimation of the Correlation Coefficient Using Quantiles, I*, Technical Report No. 32–718, Jet Propulsion Laboratory, Pasadena, California, June 1, 1965.

3. Posner, E. C., "The Use of Quantiles for Space Telemetry Data Compression," *Proceedings of the National Telemetry Conference*, Sect. 1–3, pp. 1–6, 1964.

4. "Design of a Quantile System for Data Compression of Space Telemetry," *Space Programs Summary No. 37-32*, Vol. III, pp. 103–112, Jet Propulsion Laboratory, Pasadena, California, May 31, 1964.

5. Cramér, H., *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N. J., 1946.

6. Kendall, M. G. and Stuart, A., *The Advanced Theory of Statistics*, Vol. 2, Hafner Publishing Co., New York, 1958.

7. Pearson, E. S. and Hartley, H. O., *Biometrica Tables for Statisticians*, Vol. 2, Cambridge University Press, 1962.

8. The Rand Corporation, *A Million Random Digits With 100,000 Normal Deviates*, The Free Press, Glencoe, Ill., 1955.