NORMAL AND EDGEWORTH APPROXIMATIONS

TO THE DISTRIBUTION OF THE

WILCOXON-MANN-WHITNEY

STATISTIC

By

RALPH EDWIN PORTER JR.

Bachelor of Science

Oklahoma State University

Stillwater, Oklahoma

1972

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
MASTER OF SCIENCE
July, 1974

NORMAL AND EDGEWORTH APPROXIMATIONS

TO THE DISTRIBUTION OF THE

WILCOXON-MANN-WHITNEY

STATISTIC

Thesis Approved:

P. Larry Claypool
Thesis Adviser

J. Leroy Folks

W D Warde

N N Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

The author is grateful to his major adviser, Dr. P. Larry Claypool, for his time, consideration, and constructive criticism during the course of this research.

A sincere thanks is extended to the author's parents, Mr. and Mrs. Ralph Edwin Porter Sr., for their interest, guidance, and encouragement throughout his career.

Deepest appreciation is extended to Mrs. Ralph Edwin Porter Jr., the author's wife, for her undaunted and nonpareil devotion throughout his college career.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Assumptions and Notations

The two sample test introduced by Wilcoxon [12] in 1945 is equivalent to the Mann-Whitney U test introduced in 1947 [6]. The experimental application of the test is as follows. Assume an experiment in which a random sample is taken from each of two independent populations. Also, assume that each population has a continuous distribution of probabilities, and that the scale of measurement is at least ordinal. The objective is to test the hypothesis that the two populations are identically distributed. Let $x_1, x_2, \ldots, x_m$ denote the random sample of size m from the first population and $y_1, y_2, \ldots, y_n$ denote the random sample of size n from the second population. For convenience, it will be assumed that $m \leq n$. (This can always be achieved by denoting the smaller sample by X-values and the larger sample by Y-values.) Also, let the populations have cumulative distribution functions G(x) and H(y), then define the random variable X to be stochastically larger than the random variable Y if $G(t) \geq H(t)$ for all t with strict inequality for at least one t. Thus, we are interested in testing $H_0: G(t) = H(t)$ for all t against an alternative which either specifies a simple negation of H_0 or specifies that one random variable is stochastically larger than the other. The hypotheses are more commonly stated as:

$$H_0: P(X > Y) = P(X < Y) = ½$$

against one of the following:

$$H_1: P(X > Y) \neq ½,$$

$$H_2: P(X > Y) > ½, \text{ or}$$

$$H_3: P(X > Y) < ½.$$

Under the null hypothesis, the data may be viewed as a single random sample of size $N = m + n$ from a common population; hence the N observed values are ranked from 1 to N in ascending algebraic order. That is, the smallest observed value receives rank 1, the second smallest observed value receives rank 2, ..., and the largest observed value receives rank N. Despite the assumption that the samples come from continuous distributions, ties do occur. However, the mid-ranks procedure may be used when the number of ties is moderate. The mid-ranks procedure assigns to the tied observations the average of the ranks that would have been assigned to the observations had no ties occurred.

The test introduced by Wilcoxon [12] is based on either the statistic W_x or the statistic W_y where W_x is the sum of ranks assigned to the m X-values and W_y is the sum of ranks assigned to the n Y-values. Since the sum of all ranks assigned to the data is simply the sum of the first N positive integers it follows that $W_x + W_y = N(N+1)/2$.

The test introduced by Mann and Whitney [6] is based on either the statistic U_x or the statistic U_y where U_x is the number of pairings (x_i, y_j) for which $x_i > y_j$ and U_y is the number of pairings (x_i, y_j) for which $x_i < y_j$ for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. The statistics

U_x and U_y differ from W_x and W_y, respectively, only by a constant; that is,

$$U_x = W_x - m(m+1)/2$$

and

$$U_y = W_y - n(n+1)/2.$$

Since there are mn pairings (x_i, y_j), it follows that $U_x + U_y = mn$.

The observed significance level, denoted by $\hat{\alpha}$, is defined as the smallest significance level α for which H_0 may be rejected. In other words, $\hat{\alpha}$ is the probability of obtaining a result which gives at least as much support to the alternative hypothesis as that obtained given that H_0 is true. Let w denote the observed value of W_x; then $u = w - m(m+1)/2$ is the observed value of U_x. Letting $\hat{\alpha}_i$ denote the observed significance level for the alternative hypothesis $H_i (i = 1, 2, 3)$, then

$$\hat{\alpha}_1 = 2 \min[P(W_x \leq w | H_0 \text{ true}),\ P(W_x \geq w | H_0 \text{ true})]$$

$$= 2 \min[P(U_x \leq u | H_0 \text{ true}),\ P(U_x \geq u | H_0 \text{ true})],$$

$$\hat{\alpha}_2 = P(W_x \geq w | H_0 \text{ true}) = P(U_x \geq u | H_0 \text{ true}), \text{ and}$$

$$\hat{\alpha}_3 = P(W_x \leq w | H_0 \text{ true}) = P(U_x \leq u | H_0 \text{ true}).$$

The distribution properties of the statistics discussed above will be treated in Chapter II.

Literature Review

The two sample test based on ranks was first proposed by Wilcoxon [12] in 1945. His motivation for the test was "to obtain a rapid approximate idea of the significance of the differences in [unpaired] experiments." He assumed that the data resulted from two unpaired samples of equal sizes for the purpose of testing the equality of means of the populations, H_0:E(X) = E(Y), against one of the following alternatives:

$$H_1: E(X) \neq E(Y),$$

$$H_2: E(X) > E(Y), \text{ or}$$

$$H_3: E(X) < E(Y).$$

In order to be able to state the hypotheses as above, the additional assumption must be made that should there be any difference in the two populations, the difference is in the location of the distribution of each population. (The assumption implies equal variances for the two populations. Should the experimenter be reluctant to make such an assumption, it is worth noting that there is a popular distribution-free test known as the Siegel & Tukey rank sum test for testing equality of variances.)

In 1947, Mann and Whitney [6] proposed a test procedure which is equivalent to that proposed by Wilcoxon [12]. They expanded the scope of application by relaxing the assumption implied above and including the case in which the sample sizes are unequal. Based on their assumptions (as stated in the previous section), they stated the null hypothesis H_0 in terms of the two populations having identical cumulative distribution functions against the less restrictive alternative that specified a simple negation of H_0 or one of the random variables to be stochastically larger than the other. They transformed the test statistics W_x and W_y to U_x and U_y, respectively, and derived a recurrence relation involving m and n to calculate the distribution of probabilities for U_x (or U_y), assuming there exist no ties. Using this recurrence relation they constructed tables of the probabilities of U for $\max(m, n) \leq 8$. Even at the comparatively small sample sizes $m = n = 8$, they found the distribution of U to be "almost normal." Mann and Whitney [6] then proved that under H_0 the U statistic is distributed asymptotically normally for large values of m and n.

Most applications of this test procedure do not include the restrictive assumption that if the distributions differ, they differ only in location parameters. Regardless of whether or not this assumption is included, the test procedure is referred as the Wilcoxon Rank Sum Test when W_x or W_y is used as the test statistic and as the Mann-Whitney U test when U_x or U_y is used. More generally, the proceudre may be referred to as the Wilcoxon-Mann-Whitney test.

Most subsequent papers dealing with this test procedure are concerned with the problem of either tabulating the exact distribution of probabilities or approximating the distribution of either W_x or U_x, under H_0. All are based on the case of no ties. Virtually all tabulations give lower-tail critical values of one of the statistics for certain ones of the more commonly used significance levels. An α-level lower-tail critical value t_α may be defined for any integer valued test statistic T as the largest interger t such that $P(T \leq t) \leq \alpha$.

Wilcoxon [13] presented tables of the .05, .02, and .01 level critical values of W_x for the case $m = n = 5(1)20$ in 1947. In the same year, Mann and Whitney [6] constructed tables of probabilities of U for $\max(m, n) \leq 8$ using their recurrence relation. These tables were expanded by Owen [8] to $\max(m, n) \leq 10$. The tabulation by Wilcoxon, Katti, and Wilcox [14] is the most extensive of these. For $\alpha = .005$,

.01, .025, .05, and max(m, n) $\leq$ 50, they tabulate the pair (w_α, $T - w_\alpha$) along with $P(W_x \leq w_\alpha) = P(W_x \geq T - w_\alpha)$ to four decimal places and the pair ($w_\alpha + 1$, $T - w_\alpha - 1$) along with $P(W_x \leq w_\alpha + 1) = P(W_x \geq T - w_\alpha - 1)$ to four decimal places, where w_α is the lower-tail critical value of W_x at level α and $T = m(N + 1)$. Milton [7] tabulates lower-tail critical values of U at the significance levels α = .0005, .001, .0025, .005, .01, .025, .05, and .10 for m $\leq$ 20, n $\leq$ 40. Other tabulations are referenced in the bibliography by Jacobson [5]. A non-recurrence procedure for constructing these distributions is given by van der Vaart [9]. However, this procedure has gained limited acceptance due to the tedium involved in its application.

As previously mentioned, the normal approximation was first proposed by Mann and Whitney [6] in 1947. White [11] used the normal approximation to approximate α-level critical values of the statistic W_x for α = .001, .01, and .05. White found the normal approximation to be excellent at the α = .05 level and good at the α = .01 level provided m was of sufficient size. As an example, he found that for m + n = 30 the normally approximated value was in consonance with the exact value in 11 of the 14 cases, the remaining 3 cases differing only by unity. However, Fix and Hodges [4] found that the percent relative errors incurred in using the normal approximation for approximating cumulative probabilities of U_x are quite large when m $\leq$ 12 and u $\leq$ 100. For example, when m = n = 12 and u = 55, they found the normally approximated value to be 0.17039 and the exact value to be 0.17368.

Another area of confusion arising from the several discussions on the normal approximation is the use of a correction for continuity. Verdooren [10] first suggested its use in 1963. However, no mention has been made as to the advantages or disadvantages of such a correction. The normal approximation is discussed even further by Verdooren [10], Jacobson [5], and Buckle, Kraft, and van Eeden [1].

The use of the Edgeworth approximation was first suggested by Fix and Hodges [4] in 1955. They found that a more efficient approximation was needed due to the limitations of the tables of exact distributions and due to the large percent relative errors incurred by the normal approximation. As a result, they derived an approximation using the Edgeworth series to terms of order $1/m^2$. They found the Edgeworth approximation to be accurate to about four decimal places for $m = n = 12$, the accuracy of the approximation increasing for increasing values of m. In 1963, Verdooren [10] discussed the use of the Edgeworth approximation to terms of order $1/m$ and compared it to the normal approximation for selected values of m and n, α = .001, .005, .010, .025, .050, and .10. The Edgeworth approximation to terms of order $1/m$ shows a considerable improvement over the normal approximation and is accurate to about four decimal places for $n = 25$, $m \geq 5$. It is interesting to note that the statement of accuracy made for the Edgeworth approximation to terms of order $1/m^2$ (as discussed by Fix and Hodges [4]) extends to the Edgeworth approximation to terms of order $1/m$ (as discussed by Verdooren [10]). Furthermore, as in the case of the normal approximation, no mention is made as to the effect of the correction for continuity on the Edgeworth approximation. Finally, it may be of interest to compare the corrected Edgeworth approximation to terms of order $1/m$ to the uncorrected Edgeworth approximation to terms of order $1/m^2$.

Statement of the Problem

Concerning the Edgeworth approximation, it remains unclear as to whether or not there is any advantage to using the approximation to terms of order $1/m^2$ over the approximation to terms of order $1/m$. It is apparent from Fix and Hodges' [4] comparison of the two versions of the Edgeworth approximation that greater accuracy is gained from the approximation to terms of order $1/m^2$. However, the comparisons are made at only a few values of m, n, and u, and it remains unknown as to whether or not the additional computations of terms merit the increase in accuracy.

Thus, there remain many unanswered questions concerning the normal and the Edgeworth approximations. The objectives of this study are: To describe the general behavior of each approximation over the complete distribution of W_x or U_x, to determine the magnitude of the percentage errors for the normal approximation, to determine the significance levels for which it is advantageous to use the correction for continuity, to determine the magnitude of the percentage errors for the Edgeworth approximation, to compare the accuracy of the Edgeworth approximation to terms of order $1/m^2$ to the accuracy of the approximation to terms of order $1/m$, to determine any advantages or disadvantages of using the correction for continuity in the Edgeworth approximations, and to compare the uncorrected Edgeworth approximation to terms of order $1/m^2$ to the corrected Edgeworth approximation to terms of order $1/m$.

CHAPTER II

DISTRIBUTION PROPERTIES

The Null Distribution

The number of ways in which the integers 1, 2, ..., N may be partitioned into a set of m integers to be assigned as ranks to the X-values is $\binom{N}{m}$. For each of these, there remains a second set of n integers to be assigned as ranks to the Y-values; thus, there are $\binom{N}{n} = \binom{N}{m}$ ways in which the integers 1, 2, ..., N may be partitioned into a set of n integers. Under the null hypothesis that the two populations sampled are identically distributed, each of these partitions is equally likely to occur. Thus, one may obtain the null distribution of each of the statistics W_x, W_y, U_x, and U_y in the following manner. First, list each of the possible partitions of the integers 1, 2, ..., N. Next, calculate the value of the desired statistic for each partition, then tabulate the frequency of occurrence for the distinct values of the statistic. Finally, divide each frequency by $\binom{N}{m}$ to obtain the relative frequency of occurrence for the range of values for the statistic. For example, $P_{n,m}(U_x = u) = f_u/\binom{N}{m}$ for all integer values of u where f_u is the number of partitions giving $U_x = u$. A similar statement may be made for each of the other statistics.

The distribution of W_x (or W_y) under the null hypothesis was first given by Wilcoxon [12] in 1945. In terms of computing moments, however, it is easier to use the Mann-Whitney U statistic. Mann and Whitney [6] derived a recurrence relation for calculating the probability of a particular value of u for fixed m and n (under H_0) to be as follows:

$$P_{n,m}(U_x = u) = \frac{n}{n+m} P_{n-1,m}(U_x = u - m) + \frac{m}{n+m} P_{n,m-1}(U_x = u)$$

where $P_{i,o}(U_x = 0) = P_{o,j}(U_x = 0) = 1$, $P_{i,j}(U_x = u) = 0$ for $u < 0$, and $P_{i,o}(U_x = u) = P_{o,j}(U_x = u) = 0$ for $u \neq 0$. Thus, the cumulative distribution function of U_x under H_0 is given by:

$$P_{n,m}(U_x \leq u) = \sum_{z=o}^{u} P_{n,m}(U_x = z).$$

The above may be expressed in terms of cumulative frequencies and will be given explicitly in the next chapter. (See formula (3.1).)

It may be shown that the statistics U_x and U_y are identically distributed under H_0 and that the statistics W_x and W_y are each identical to them except for a shift in scale values. That is:

$$\begin{aligned} P(U_x = u) &= P(U_y = u) \\ &= P(W_x = u + m(m+1)/2) \\ &= P(W_y = u + n(n+1)/2). \end{aligned} \tag{2.1}$$

The common probability in (2.1) is positive for $0 \leq u \leq mn$ and has value zero otherwise. Thus, it follows that

$$\begin{aligned} P(U_x \leq u) &= P(U_y \leq u) \\ &= P(W_x \leq u + m(m+1)/2) \\ &= P(W_y \leq u + n(n+1)/2) \end{aligned} \tag{2.2}$$

for all real numbers u. Due to the fact that each statistic is symmetric about its mean, it can be shown that

$$P(U_x \leq u) = P(U_x \geq mn - u)$$
$$= P(U_y \geq mn - u)$$
$$= P(W_x \geq m(N+n+1)/2 - u)$$
$$= P(W_y \geq n(N+m+1)/2 - u) \qquad (2.3)$$

for all real numbers u.

Moments

Define z_i as follows:

$$z_i = 1 \text{ if the rank } i \text{ is assigned to an X-value}$$
$$= 0 \text{ otherwise.}$$

Then, the test introduced by Wilcoxon [12] is based on W_x or W_y where $W_x = \sum_{i=1}^{N} i \cdot z_i$ is the sum of the ranks assigned to X-values and $W_y = \sum_{i=1}^{N} i(1 - z_i)$ is the sum of the ranks assigned to Y-values. As it was assumed that the samples are random samples drawn from two independent populations, the z_i are mutually independent and, under the null hypothesis, are identically distributed. It follows that $P(z_i = 1) = m/N$ and $P(z_i = 0) = n/N$ which implies

$$E(z_i) = m/N \text{ for all } i, \text{ and}$$

$$Var(z_i) = \frac{mn}{N^2} \text{ for all } i.$$

Therefore,

$$E(W_x) = E(\sum_{i=1}^{N} i \cdot z_i) = m(N+1)/2,$$

$$E(W_y) = E[\sum_{i=1}^{N} i(1 - z_i)] = n(N+1)/2,$$

$$E(U_x) = E[W_x - m(m+1)/2] = mn/2, \text{ and}$$

$$E(U_y) = E[W_y - n(n+1)/2] = mn/2.$$

Due to the fact that the statistics W_x, W_y, U_x, and U_y are identically distributed except for a shift in the scale value, the central moments of order r (denoted μ_r) are equal for all of the statistics, $r = 2, 3, \cdots$. Due to the symmetry of each of the statistics about its mean, all odd moments will be zero; that is, $\mu_{2r+1} = 0$ for $r = 1, 2, \cdots$. Thus, each statistic has the common variance given by:

$$\sigma^2 = \sum_{i=1}^{N} i^2 Var(z_i) = mn(N+1)/12.$$

Mann and Whitney [6] derived the even moments to be as follows:

$$\mu_{2r} = \frac{1}{n+m} \sum_{\alpha=0}^{r} \binom{2r}{2\alpha} \frac{1}{4^\alpha} [nm^{2\alpha} E_{n-1,m}(U - mn/2)^{2r-2\alpha}$$

$$+ mn^{2\alpha} E_{n,m-1}(U - mn/2)^{2r-2\alpha}]$$

for $r = 1, 2, \cdots$.

The Normal Approximation

The limiting distribution of U where U denotes either U_x or U_y was proved by Mann and Whitney [6] to be asymptotically normal as m and n become infinitely large. In general, their method of proof consists of showing by the method of undetermined coefficients that the central moment μ_{2r} contains a polynomial in m and n of degree 3r which is divisible by mn(N+1), considering the ratio of μ_{2r} to μ_2^r, and taking the limit of this ratio as m and n approach infinity in any arbitrary manner.

Due to the results of Mann and Whitney [6], the cumulative distribution function of U may be approximated by evaluating the standard normal cumulative distribution function as follows:

$$P(U \leq u) \doteq F(x) = \int_{-\infty}^{x} (1/2\pi)^{1/2} \exp(-v^2/2)\,dv \qquad (2.4)$$

where x now denotes the standardized value of u given by

$$x = \frac{u - E(U)}{\sqrt{\mathrm{Var}(U)}} = \frac{u - mn/2}{\sqrt{mn(N+1)/12}} \,. \qquad (2.5)$$

In general, the cumulative distribution function of any one of the statistics may be approximated in the same manner. Letting T denote any one of the statistics, then $P(T \leq t) \doteq F(t)$ where F(t) is as given in formula (2.4) and

$$x = \frac{t - E(T)}{\sqrt{\mathrm{Var}(T)}} = \frac{t - E(T)}{\sqrt{mn(N+1)/12}} \,.$$

Employing the usual correction for continuity, the standardized u-value becomes

$$x_c = \frac{u - E(U) + 1/2}{\sqrt{\mathrm{Var}(U)}} = \frac{u - mn/2 + 1/2}{\sqrt{mn(N+1)/12}} \,. \qquad (2.6)$$

Thus, F(x) will be referred to as the uncorrected normal approximation when x is as given by formula (2.5) and referred to as the corrected normal approximation when x is as given by formula (2.6).

The Edgeworth Approximation

As discussed earlier, Fix and Hodges [4] sought to improve the normal approximation by using the Edgeworth series to terms of order $1/m^2$. Due to the accuracy obtained by the Edgeworth approximation to terms of order 1/m (see Verdooren [10]), both versions of the approximation are considered in this research.

Fix and Hodges [4] give the Edgeworth approximation to terms of order 1/m to be

$$P(U \leq u) \doteq F_1(x) = F(x) + e^{(3)}_{n,m} f^{(3)}(x) \tag{2.7}$$

where F(x) is the normal approximation as given by formula (2.4), and x is either the uncorrected or the corrected standardized u-value as given by formula (2.5) or (2.6). The term $e^{(3)}_{n,m}$ is the Edgeworth coefficient where $e^{(3)}_{n,m} = (\mu_4/\mu_2^2 - 3)/4!$ and $\mu_k = E(U - mn/2)^k$. Mann and Whitney [6] give the values of μ_k for k = 2 and 4 as $\mu_2 = mn(N+1)/12$ and

$$\mu_4 = mn(N+1)[5(m^2n + mn^2) - 2(m^2 + n^2) + 3mn - 2N]/240.$$

Thus, Fix and Hodges [4] give the Edgeworth coefficient to be

$$e^{(3)}_{n,m} = -[m^2 + n^2 + mn + N]/[20mn(N+1)]. \tag{2.8}$$

The term $f^{(3)}(x)$ is the third derivative of the standard normal density function; thus,

$$f^{(3)}(x) = -(2\pi)^{-\frac{1}{2}}(x^3 - 3x)\exp(-x^2/2). \tag{2.9}$$

The Edgeworth approximation to terms of order $1/m^2$ is given by Fix and Hodges [4] to be

$$P(U \leq u) \doteq F_2(x) = F_1(x) + e^{(5)}_{n,m} f^{(5)}(x) + e^{(7)}_{n,m} f^{(7)}(x) \qquad (2.10)$$

where

$$e^{(5)}_{n,m} = (\mu_6/\mu_2^3 - 15\mu_4/\mu_2^2 + 30)/6! \qquad (2.11)$$

and

$$e^{(7)}_{n,m} = 35(\mu_4/\mu_2^2 - 3)^2/8!. \qquad (2.12)$$

Mann and Whitney [6] give μ_6 to be as follows:

$$\mu_6 = mn(N+1)[35m^2n^2(m^2+n^2) + 70m^3n^3 + P(m,n)]/4032$$

where

$$\begin{aligned} P(m,n) = & -42mn(m^3+n^3) - 14m^2n^2N + 16(m^4+n^4) \\ & - 52mn(m^2+n^2) - 43m^2n^2 + 32(m^3+n^3) \\ & + 14mnN + 8(m^2+n^2) + 16mn - 8N. \end{aligned}$$

After substitution and simplification of the above, Fix and Hodges [4] give

$$e^{(5)}_{n,m} = \frac{[2(m^4+n^4) + 4mn(m^2+n^2) + 6m^2n^2 + 4(m^3+n^3) + 7mnN + (m^2+n^2) + 2mn - N]}{210m^2n^2(N+1)^2}, \qquad (2.13)$$

and

$$e^{(7)}_{n,m} = \frac{(m^2 + n^2 + mn + N)^2}{800\ m^2n^2(N+1)^2}. \qquad (2.14)$$

It follows that

$$f^{(5)}(x) = -(2\pi)^{-1/2}(x^5 - 10x^3 + 15x)\exp(-x^2/2) \text{ and} \tag{2.15}$$

$$f^{(7)}(x) = -(2\pi)^{-1/2}(x^7 - 21x^5 + 105x^3 - 105x)\exp(-x^2/2). \tag{2.16}$$

CHAPTER III

PROCEDURES AND ACCURACY

Calculation of the Exact Cumulative Probabilities

The exact cumulative probabilities of the Mann-Whitney U statistic under the null hypothesis were calculated for the lower half of the distribution, $m \leq n \leq 30$. It is not necessary to calculate the upper half of the distribution as the U statistic is symmetric about its mean for all m and n. Although a non-recurrence procedure exists for constructing the exact cumulative probabilities (see van der Vaart [9]), its application is limited due to the tedium involved in the computations. The method used here involved the recurrence relation given by Owen [8]. He found that

$$F_{n,m}(u) = F_{n-1,m}(u-m) + F_{n,m-1}(u) \qquad (3.1)$$

where $F_{n,m}(u)$ denotes the cumulative frequency of the event $U \leq u$ where the sample from the first population is of size m and the sample from the second population is of size n. The following conditions are given for this recurrence relation:

$$F_{n,m}(u) = 0 \text{ for all } u < 0,$$

$$F_{n,o}(u) = F_{o,m}(u) = 1 \text{ for } u = 0,$$

$$F_{n,o}(u) = F_{o,m}(u) = 0 \text{ for } u \geq 1, \text{ and}$$

$$F_{n,m}(u) = F_{m,n}(u) \text{ for all } u.$$

The recurrence relation as stated above can be derived from the probability recurrence relation given by Mann and Whitney [6].

For each value of u, $0 \leq u \leq mn/2$, the normal approximation was calculated using formula (2.4) employing both the uncorrected and corrected standardized u-value as given by formulas (2.5) and (2.6). In the same manner, using both the uncorrected and corrected standardized u-value, the Edgeworth approximations to terms of order $1/m$ and to terms of order $1/m^2$ were calculated as given by formulas (2.7) and (2.10), respectively. The exact cumulative probabilities and the six approximations as described above were calculated for $m \leq n$ = 15, 17, 20, 22, 25, 27, and 30.

To measure the effectiveness of each of the approximations, the percent relative error (PRE) was calculated for each approximate probability obtained where

$$\text{PRE} = \frac{\text{(Approximated Probability)} - \text{(Exact Probability)}}{\text{(Exact Probability)}} \cdot 100. \tag{3.2}$$

A positive value obtained for the PRE represents an over-estimation by the approximation, and a negative value represents an under-estimation by the approximation. The behavior of the percent relative error function for each of the approximations is summarized in the following chapter.

Computational Accuracy

All computations were performed on the IBM 360, Model 65 Computer at the Oklahoma State University Computer Center. All computations were performed in double precision, and the DERF/DERFC algorithm [3] was used to evaluate the standard normal distribution function F(x).

In addition, the .05 one-sided critical values and probability levels reported by Wilcoxon, Katti, and Wilcox [14] were verified for the values of m and n given above. However, in comparison to the corrected normal and Edgeworth (to terms of order 1/m) approximations given by Verdooren [10], several inconsistencies arose in the Edgeworth approximations. The differences were never by more than .0001, and it is felt that the inconsistencies arose as a result of computational techniques. For example, Verdooren presented his results in 1963, but the DERF/DERFC algorithm [3] used in this research to evaluate the standard normal distribution function was not available until 1969.

CHAPTER IV

SUMMARY OF FINDINGS

It is the purpose of this chapter to summarize the general behavior of the PRE incurred as a function of the exact cumulative probability α being approximated for each of the approximations considered. This summary will be made with the help of graphs which will also serve to illustrate the fact that the magnitude of PRE incurred tends to be a decreasing function of either m or n for fixed α. Also, attention is given to the question of whether or not to include the usual correction for continuity. The reader is asked to keep in mind that each PRE "curve" represents a series of discrete points.

To enhance the readability of the section on Edgeworth approximations, the Edgeworth approximation to terms of order $1/m$ will be denoted by EE1, and the Edgeworth approximations to terms of order $1/m^2$ will be denoted by EE2.

As tables are readily available which give the lower-tail critical values of W or U for $\max(m, n) \leq 20$, the approximations are discussed for $\min(m, n) \geq 20$ and $\max(m, n) \leq 30$. It is felt that under these restrictions on m and n, a sufficient amount of data was generated on which to base this paper.

The Normal Approximation

The general behavior of the PRE incurred as a function of the exact probability may be seen in Figure 1 for the uncorrected normal approximation. Note the change in the PRE scale at 10 percent in Figure 1(a) and the change in scale from Figure 1(a) to Figure 1(b). The PRE is extremely large for small exact probability levels, and decreases very rapidly to zero corresponding to an exact probability level between .015 and .020. The PRE then becomes negative, bottoms out for α between .09 and .10, and then decreases in absolute value to a value between 0 and -1.1 percent at $\alpha = .50$. For example, when $m = n = 20$, the PRE for the uncorrected normal approximation is almost 100 percent for $\alpha = .0005$, becomes negative at $\alpha = .0158$, bottoms out to -3.5 percent at $\alpha = .0913$, and then decreases in absolute value to -1.1 percent at $\alpha = .50$.

As is illustrated by Figure 1, the magnitude of PRE corresponding to the uncorrected normal approximation is a decreasing function of either m or n for fixed α. An exception to this statement is noted in the neighborhood of $\alpha = .015$ where the PRE changes sign from positive to negative. Although only three sets of sample size are presented in Figure 1, intermediate sample sizes were graphed by the author, and the general behavior of the PRE function is consistent with the behavior exemplified by Figure 1.

Figure 2 compares the behavior of the PRE incurred using the uncorrected normal approximation to that using the corrected normal approximation for fixed m and n. As is suggested by Figure 2, it is advantageous to use the usual correction for continuity for α greater than about .025 and omit it otherwise. Within the scope of this study, a reasonable rule of thumb would be to omit the correction for continuity only for values of the statistic which do not exceed the .025 level critical value. It is worth noting that the exact probability level at which the corrected normal approximation becomes advantageous over the uncorrected approximation increases for increasing sample sizes. For example, when $m = n = 20$ the corrected normal approximation becomes more advantageous at $\alpha = .0245$, at $\alpha = .0253$ when $m = n = 25$ and at $\alpha = .0266$ when $m = n = 30$.

Another point of interest is the probability level at which the uncorrected and corrected normal approximations change sign. The uncorrected normal approximation changes sign at about the probability level $\alpha = .017$, and the corrected normal approximation changes sign at about the probability level $\alpha = .042$. For example, the uncorrected normal approximation changes sign at the probability level $\alpha = .0157$ when $m = n = 20$, at about $\alpha = .0173$ when $m = n = 25$, and at about $\alpha = .0183$ when $m = n = 30$. The corrected normal approximation changes sign at about the probability level $\alpha = .0418$ when $m = n = 20$, at about $\alpha = .0416$ when $m = n = 25$, and at about $\alpha = .0416$ when $m = n = 30$.

It was also observed that the PRE for the normal approximation over the entire distribution decreases in magnitude when the sample size m increases while n and α remain fixed. This can be seen graphically in Figure 3 for the corrected normal approximation for $m = 20, 25, 30$ and $n = 30$. (Of course, this result also extends to the uncorrected normal approximation.) For all possible exact probability levels, the effectiveness of either normal approximation increases as one of the sample sizes increases while the other sample size remains fixed.

The Edgeworth Approximation

One would expect many of the properties of the normal approximation discussed in the previous section to extend to the Edgeworth approximation, the reason being that the first term of the Edgeworth approximation is the normal approximation. However, somewhat different results were obtained. These results will be discussed in this section.

The general behavior of the PRE function for the Edgeworth approximation is quite different from the PRE function for the normal approximation. In general, its behavior depends upon whether or not the usual correction for continuity is employed. The PRE for the normal approximation is quite large for extremely small exact probability levels. In contrast, either corrected Edgeworth approximation gives PRE's which are negative with moderately large magnitudes for extremely small probability levels. The algebraic values of the PRE function increase rapidly to a maximum near $\alpha = .002$ having become positive near $\alpha = .0006$. The PRE function for EE1 then slowly decreases and changes sign in a neighborhood of $\alpha = .05$. It attains its minimum algebraic value in a neighborhood of $\alpha = .075$, then decreases in magnitude to almost zero for $\alpha = .50$. After attaining its maximum positive PRE near $\alpha = .002$, the PRE function for EE2 slowly decreased to almost zero (but, remaining positive) as the exact probability level increases from about $\alpha = .002$ to $\alpha = .50$.

Not only is there a significant difference in the general behavior of the PRE for the normal and the Edgeworth approximations, there is also a significant difference in the magnitude of the PRE for each. Verdooren [10] in 1963 compared the corrected normal approximation as given by formula (2.4) to the corrected EE1 as given by formula (2.7). Each approximation employed the usual correction for continuity as given by formula (2.6). The comparisons were made for six selected sets of values of m and n and at the significance levels $\alpha = .001, .005, .010,$

.025, .050, and .10. From his results, it can be clearly seen that the normal approximation "is subject to large percentage errors at the higher significance levels, while use of the correction term in $f^{(3)}(x)$ improves the accuracy considerably." One needs only to compare the PRE "curves" of the normal approximations as seen on Figures 1 and 2 to the PRE "curves" of EE1, Figure 4. For example, when m = n = 20 and $\alpha = .005$, the PRE incurred by the uncorrected normal approximation is about 37 percent, 43 percent for the corrected normal approximation, but only about .32 percent for the corrected EE1.

As discussed in Chapter I, it seems intuitively appealing to compare the PRE incurred by the corrected EE1 to the PRE incurred by the uncorrected EE2. Fix and Hodges [3] state that the <u>corrected</u> EE2 is accurate to about four decimal places. Thus, we would expect the uncorrected EE2 to be accurate to at most four decimal places. As can be seen by the results given by Verdooren [9], the corrected EE1 is also accurate to about four decimal places. Comparing the two approximations brings a rather interesting result to light. The magnitude of PRE incurred by the uncorrected EE2 is at least six times greater than that incurred by the corrected EE1 (see Figure 5). Thus, it is more advantageous to employ the usual correction for continuity in EE1 than to compute the additional terms required for the uncorrected EE2.

It was surprising to note that in comparing the three uncorrected approximations, the smallest magnitudes of PRE are associated with EE1 for $.005 \leq \alpha \leq .01$ and with the normal approximation for $.01 \leq \alpha \leq .04$. For $\alpha > .04$, the PRE for EE1 and EE2 are virtually identical, and their magnitudes of PRE are at least two-thirds of the corresponding magnitudes for the normal approximation.

The advantage of using the continuity correction is quite evident and will be discussed in more detail later in this section. In the case of the corrected EE1 when m = n = 20, the PRE is about zero at α = .0004, attains its maximum (for $\alpha > .0004$) of .91 percent at α = .0015, and attains its minimum algebraic value of -.035 percent at α = .09. Thus, for $\alpha > .0004$, the PRE for the corrected EE1 is always less than 1 percent. In the case of the corrected EE2 when m = n = 20, the PRE is about zero at α = .0006 and reaches its absolute maximum of .09 percent at α = .009. Thus, for $\alpha > .0006$, the PRE for the corrected EE2 is always less than .1 percent (see Figure 4). Undoubtedly, the magnitude of the PRE for the Edgeworth approximations is quite small in comparison to the magnitude of the PRE for the normal approximation.

As would be expected (and is demonstrated in Figure 6), the PRE function for the Edgeworth approximation decreases for increasing sample sizes. Although only the corrected EE2 is considered in Figure 6, the result is general whether EE1 or EE2 is used and whether or not the correction for continuity is employed in the approximation. Likewise, only three particular sets of sample sizes are considered in Figure 6, but the result extends to all intermediate values of m and n.

As mentioned previously in this section, the advantage of using the usual correction for continuity in the Edgeworth approximation is quite evident. This holds true for both EE1 and EE2. The magnitude of PRE incurred by the uncorrected Edgeworth approximation will always exceed the magnitude of PRE incurred by the corrected Edgeworth approximation. Graphically, this can be seen in Figure 7 which compares the uncorrected EE2 to the corrected EE2 for m = n = 20 and m = n = 30. In fact, for m = n = 20 and $\alpha \geq .00013$, the PRE incurred by the corrected EE2 will always be less than one-tenth the PRE incurred by the uncorrected EE2.

Due to the significant advantage of using the continuity correction in the Edgeworth approximation, the comparison of EE1 to EE2 will be limited to only the corrected versions of each. Figure 4 compares EE1 to EE2 (both corrected) for $m = n = 20$ and $m = n = 30$. It is evident that for $\alpha \geq .0005$ the PRE incurred by EE1 will always exceed the PRE incurred by EE2 everywhere except in a small neighborhood of the exact probability level at which EE1 changes sign. The exact probability level at which EE1 changes sign is $\alpha = .0428$ for $m = n = 20$, $\alpha = .0421$ for $m = n = 25$, and $\alpha = .0415$ for $m = n = 30$.

General Guidelines

This section is intended to be a guide for the researcher in selecting the approximation which meets his requirements and/or limitations. However, it is the responsibility of the researcher to determine the accuracy desired, and thus the amount of effort he is willing to put forth in computing a particular probability level.

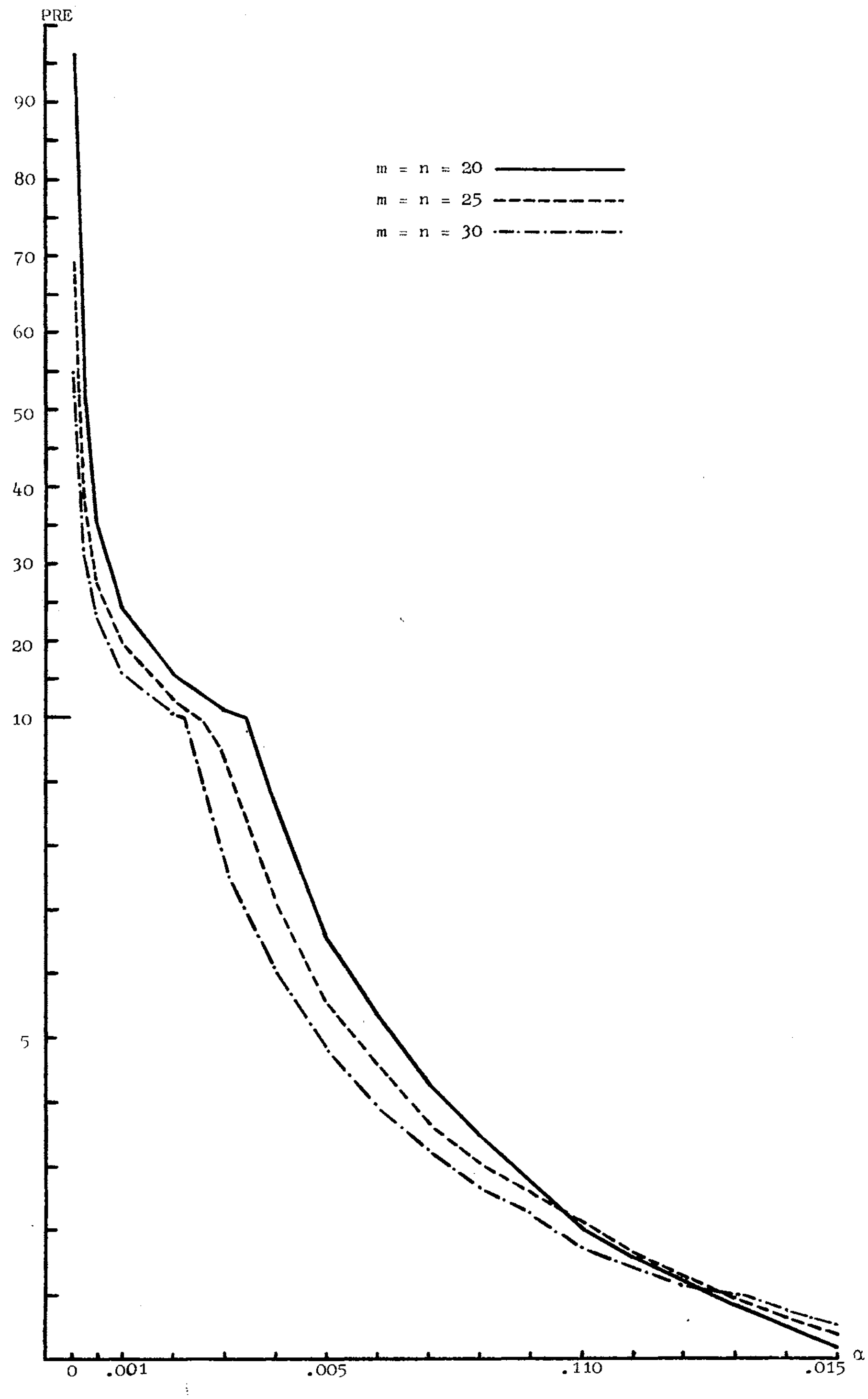
After reading most of the literature on the Wilcoxon-Mann-Whitney test procedure, it becomes apparent that there are no compelling reasons for preferring the Wilcoxon text statistic over the Mann-Whitney test statistic, or vice versa. For example, the test procedure has been discussed in terms of the Wilcoxon test statistic by Verdooren [10], White [11], Wilcoxon [12 and 13], and Wilcoxon, Katti, and Wilcox [16], whereas it has been discussed in terms of the Mann-Whitney test statistic by Buckle, Kraft, and van Eeden [1], Conover [2], Fix and Hodges [4], Mann and Whitney [6], Milton [7], and van der Vaart [9]. It was shown in Chapter II the relations between the Wilcoxon and the Mann-Whitney test statistics are simple as they are identically distributed except for a shift in the scale values. For these reasons, selection of a particular test statistic is based largely on the personal preference of the researcher.

As was shown in this chapter, the PRE incurred by the normal approximation is quite large for small exact probability levels. Thus, the normal approximation should be used only when the Edgeworth approximation is too difficult to compute. It would seem that in this era of advanced electronics (calculators and computers) this would seldom be the case. However, should the researcher decide to use the normal approximation, several points should be kept in mind. For exact probability levels greater than .025, always use the corrected normal approximation. Otherwise, use the uncorrected normal approximation. For $\alpha > .025$, the PRE incurred by the corrected normal approximation will never exceed 1.5 percent in absolute value ($\min(m, n) \geq 20$). If the researcher desires only to solve for a critical value of U for which to reject his null hypothesis, the inverse procedure may be applied to the normal approximation. This procedure is particularly appealing at two exact probability levels. At $\alpha = .017$, the PRE incurred by the uncorrected normal approximation is close to zero. The same is true for the PRE incurred by the corrected normal approximation at $\alpha = .042$.

It is without doubt that the Edgeworth approximations are more efficient than the normal approximation. The most important result to bear in mind is that the usual correction for continuity will always increase the efficiency of either Edgeworth expansion. Also, it will always be more advantageous to use EE2 over EE1. The best possible approximation obtainable is the corrected EE2. For exact probability levels greater than .005, the PRE incurred by the corrected EE2 will never exceed 0.09 percent for $20 \leq m \leq n$.

In view of the accuracy observed for the corrected Edgeworth approximation to terms of order $1/m^2$, it would seem highly unlikely that the expense involved in extending the tabulations of cumulative probabilities and critical values for W_x and U_x could be justified. This is particularly true since the approximations discussed in this thesis are quite easy to program for calculation by a computer.

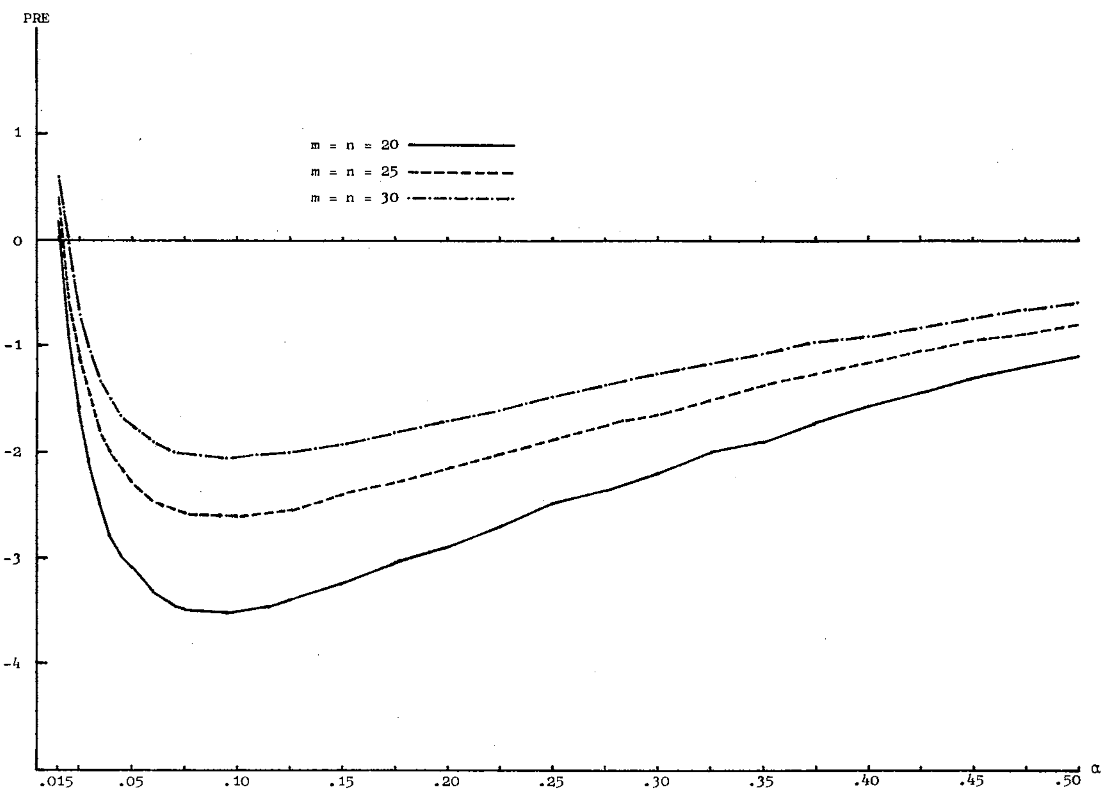
(a) $0.0005 \leq \alpha \leq 0.015$



NOTE: The change in the PRE scale at 10 percent

Figure 1. Percent Relative Error (PRE) of the Uncorrected Normal Approximation With Respect to Probability Level (α) for $m = n = 20, 25, 30$

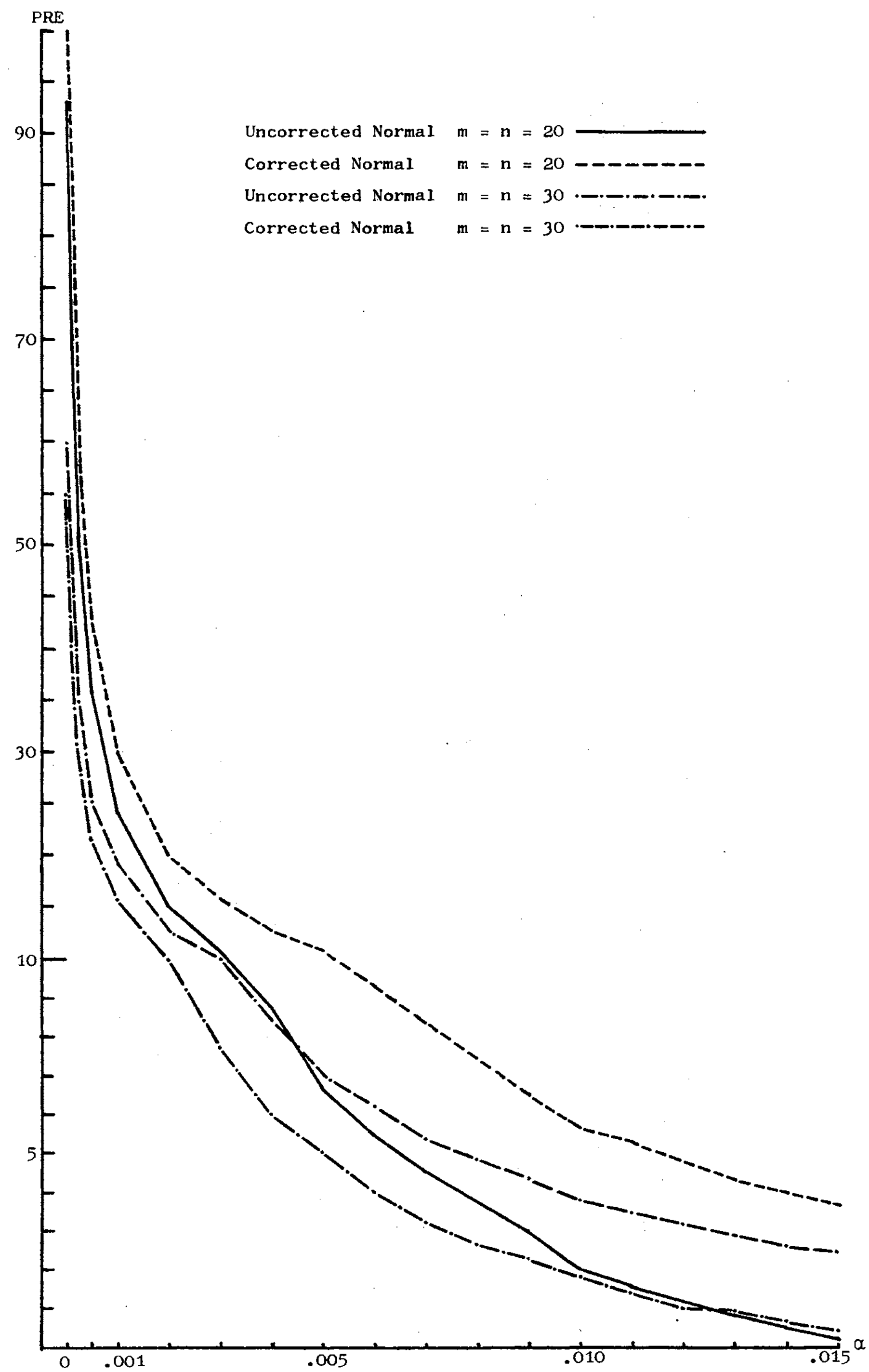
(b) $0.015 \leq \alpha \leq 0.50$



NOTE: There is a change in the α scale from Figure 1(a) to Figure 1(b).

Figure 1. (Continued)

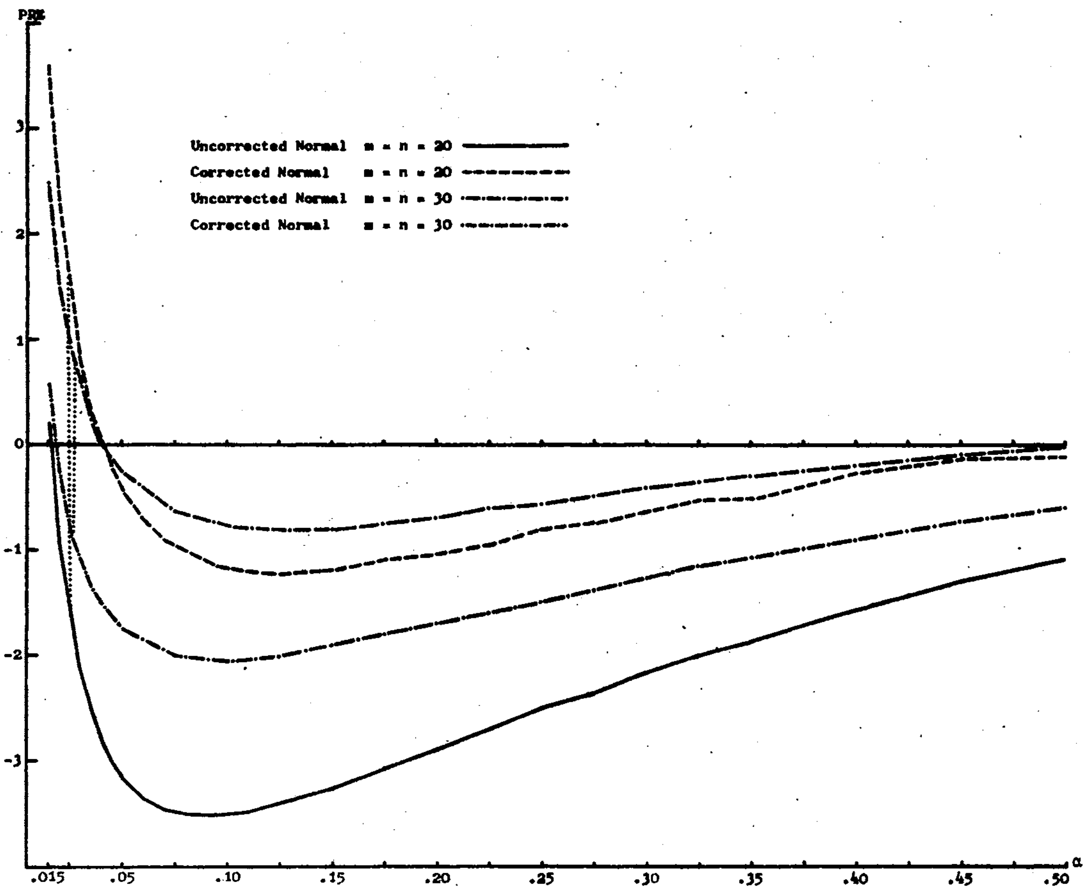
(a) $0.0005 \leq \alpha \leq 0.015$



NOTE: There is a change in the PRE scale at 10 percent.

Figure 2. Percent Relative Error (PRE) of the Uncorrected and Corrected Normal Approximation With Respect to Probability Level (α) for m = n = 20, 30

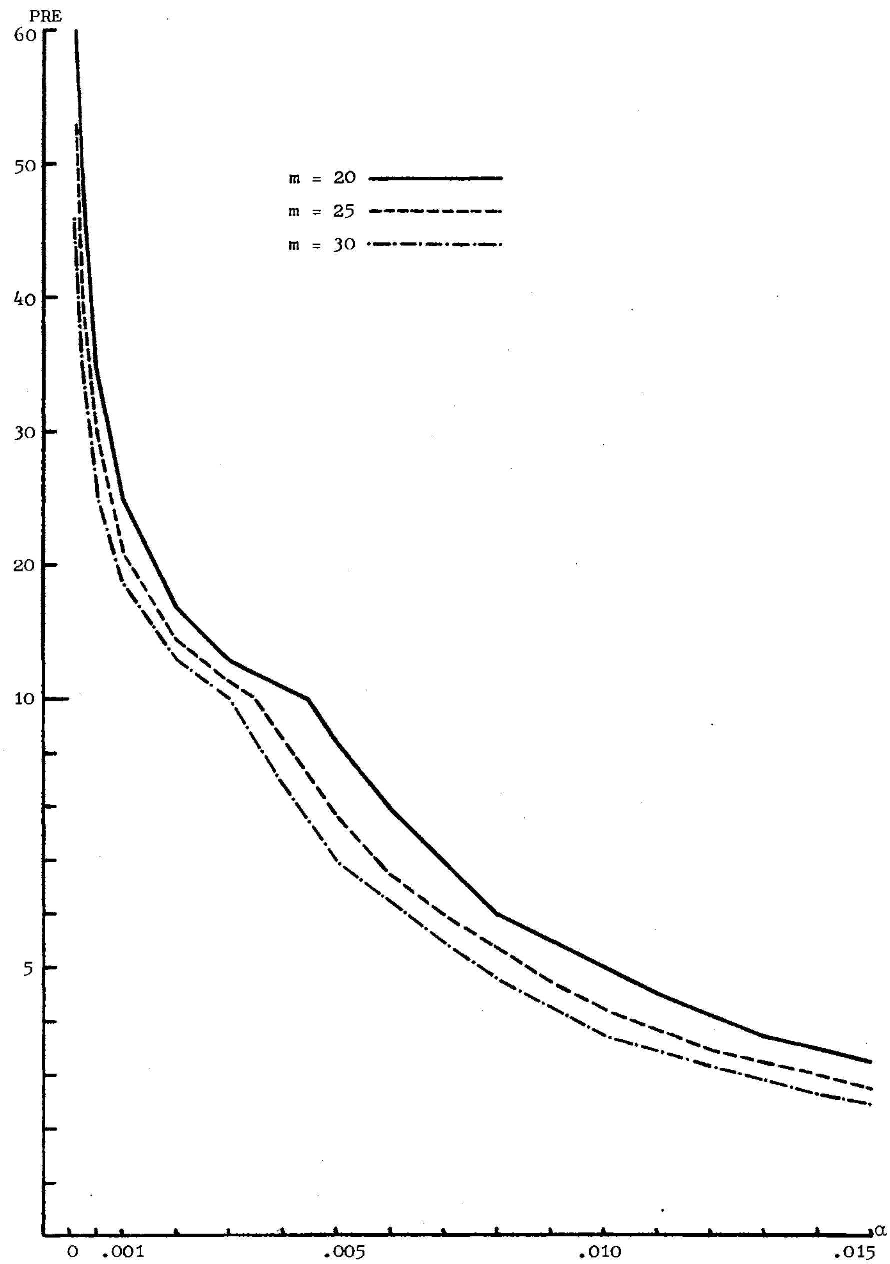
(b) $0.015 \leq \alpha \leq 0.50$



NOTE: There is a change in the α scale from Figure 2(a) to Figure 2(b)

Figure 2. (Continued)

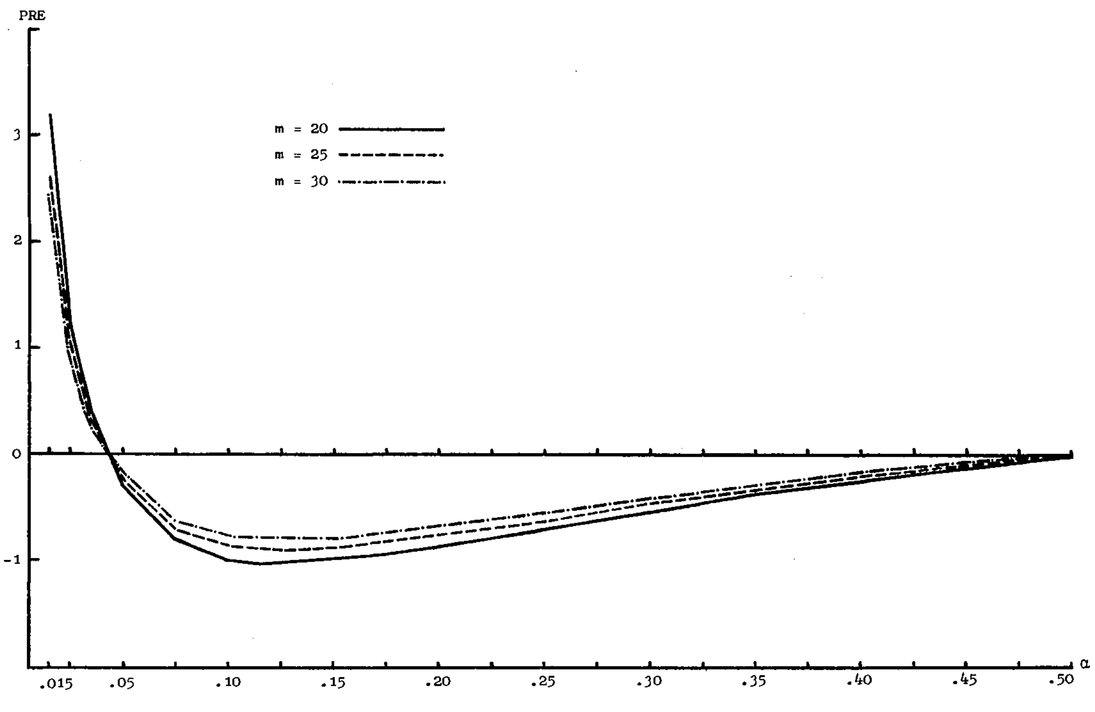
(a) $0.0005 \leq \alpha \leq 0.015$



NOTE: There is a change in the PRE scale at 10 percent.

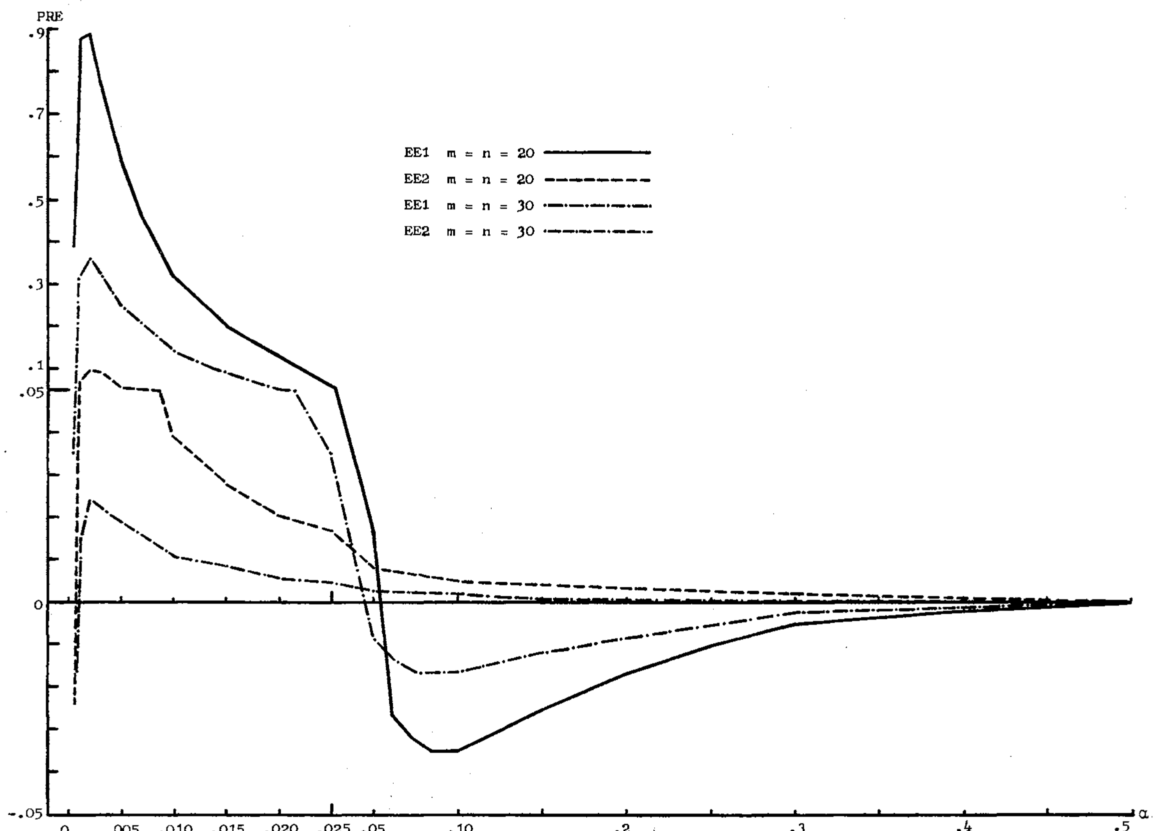
Figure 3. Percent Relative Error (PRE) of the Corrected Normal Approximation With Respect to Probability Level (α) for Fixed $n = 30$ and $m = 20, 25, 30$

(b) $0.015 \leq \alpha \leq 0.50$



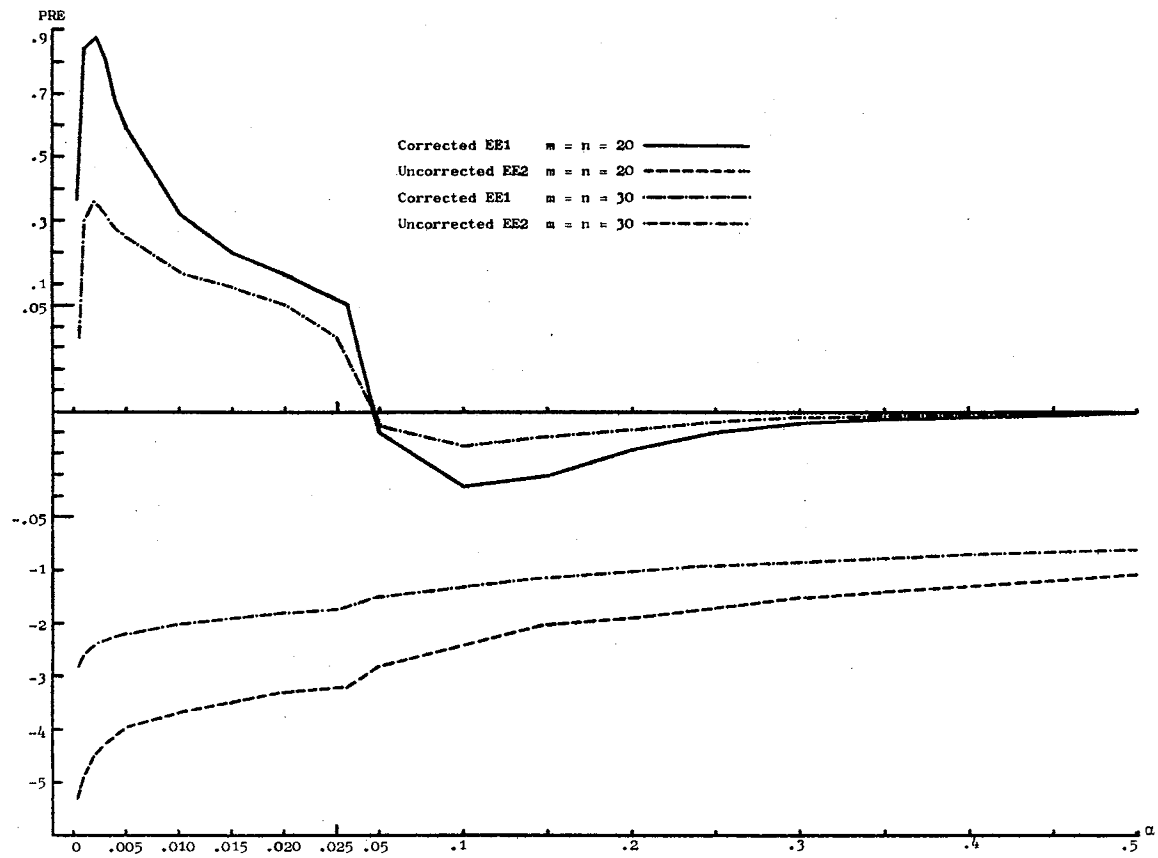
NOTE: There is a change in the α scale from Figure 3(a) to Figure 3(b)

Figure 3. (Continued)



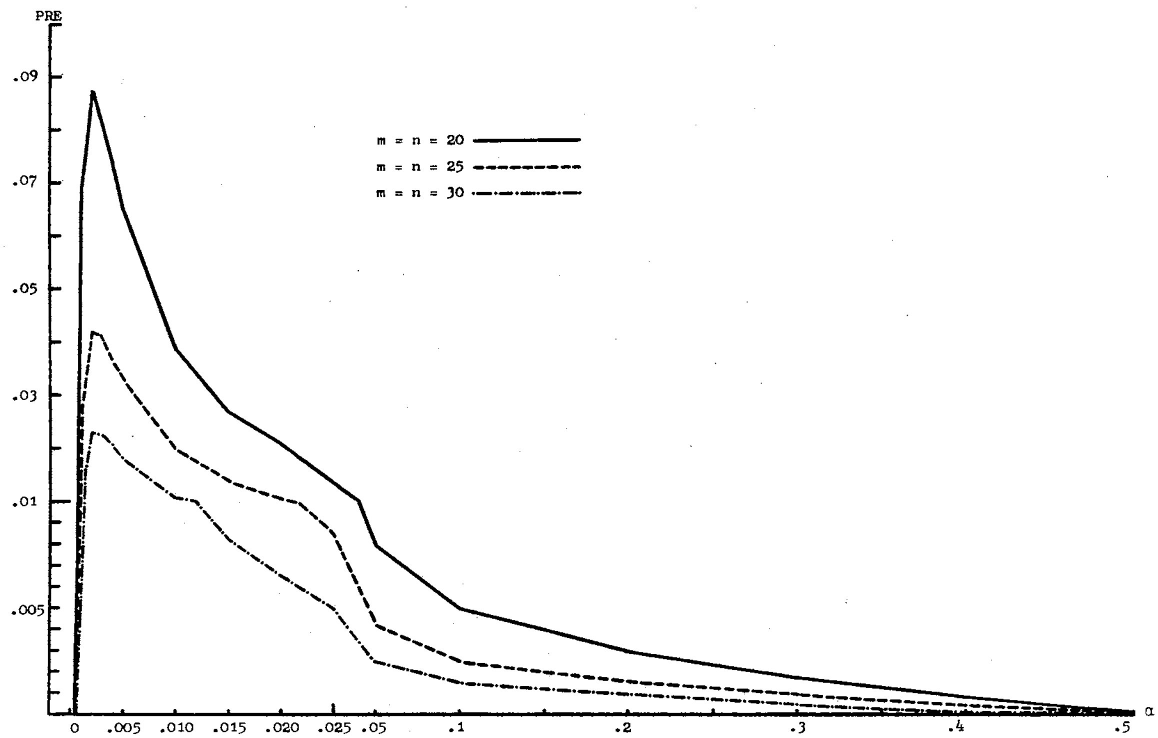
NOTE: There is a change in the PRE scale at .05 percent and the change in the α scale at .025.

Figure 4. Percent Relative Error (PRE) of the Corrected EE1 and EE2 With Respect to Probability Level (α) for m = n = 20, 30



NOTE: There is a change in the PRE scale at -.05 and .05 percent and the change in the α scale at .025.

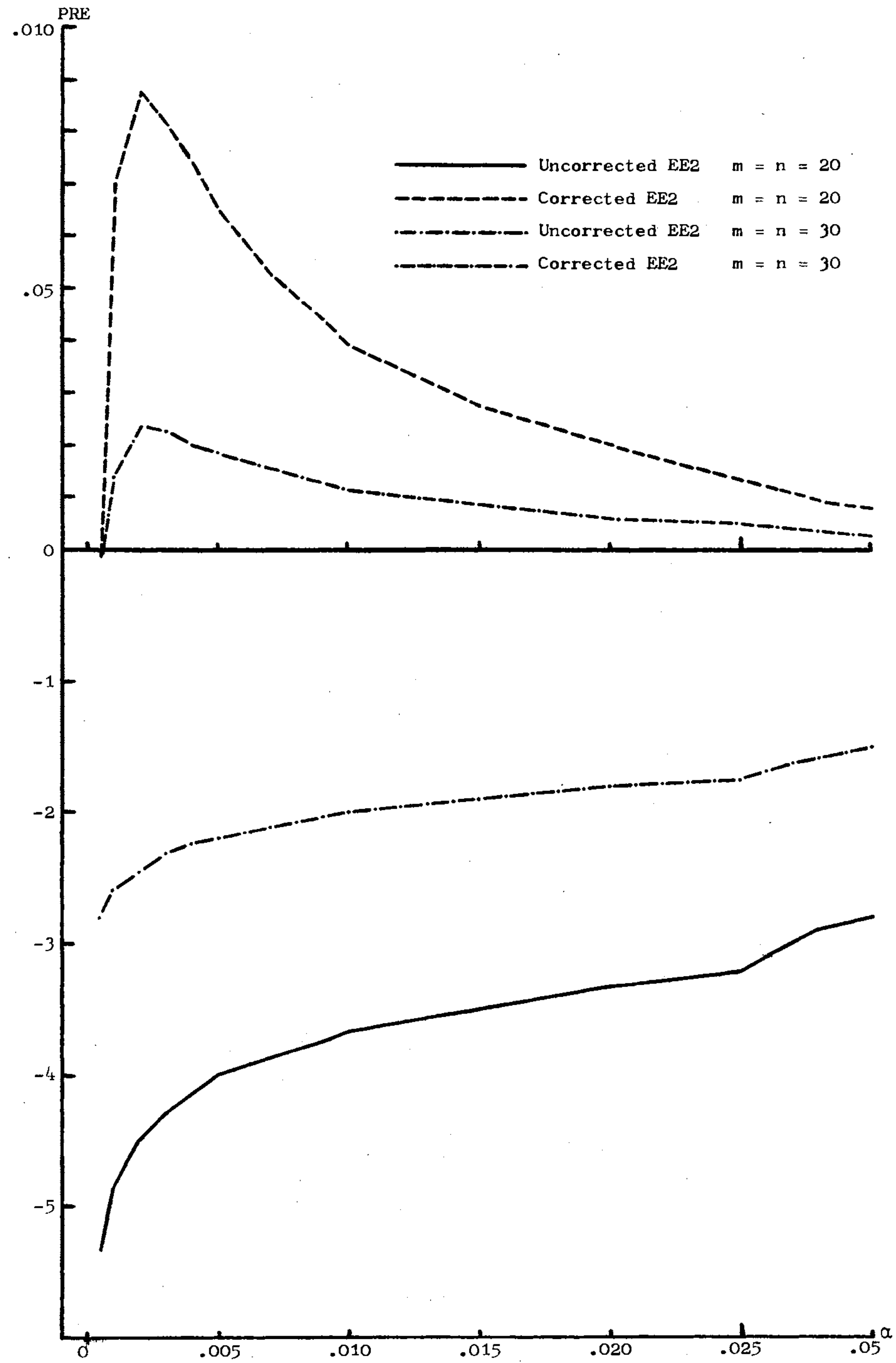
Figure 5. Percent Relative Error (PRE) of the Corrected EE1 and Uncorrected EE2 With Respect to Probability Level (α) for m = n = 20, 30



NOTE: There is a change in the PRE scale at .01 percent and the change in the α scale at .025.

Figure 6. Percent Relative Error (PRE) of the Corrected EE2 With Respect to Probability Level (α) for m = n = 20, 25, 30

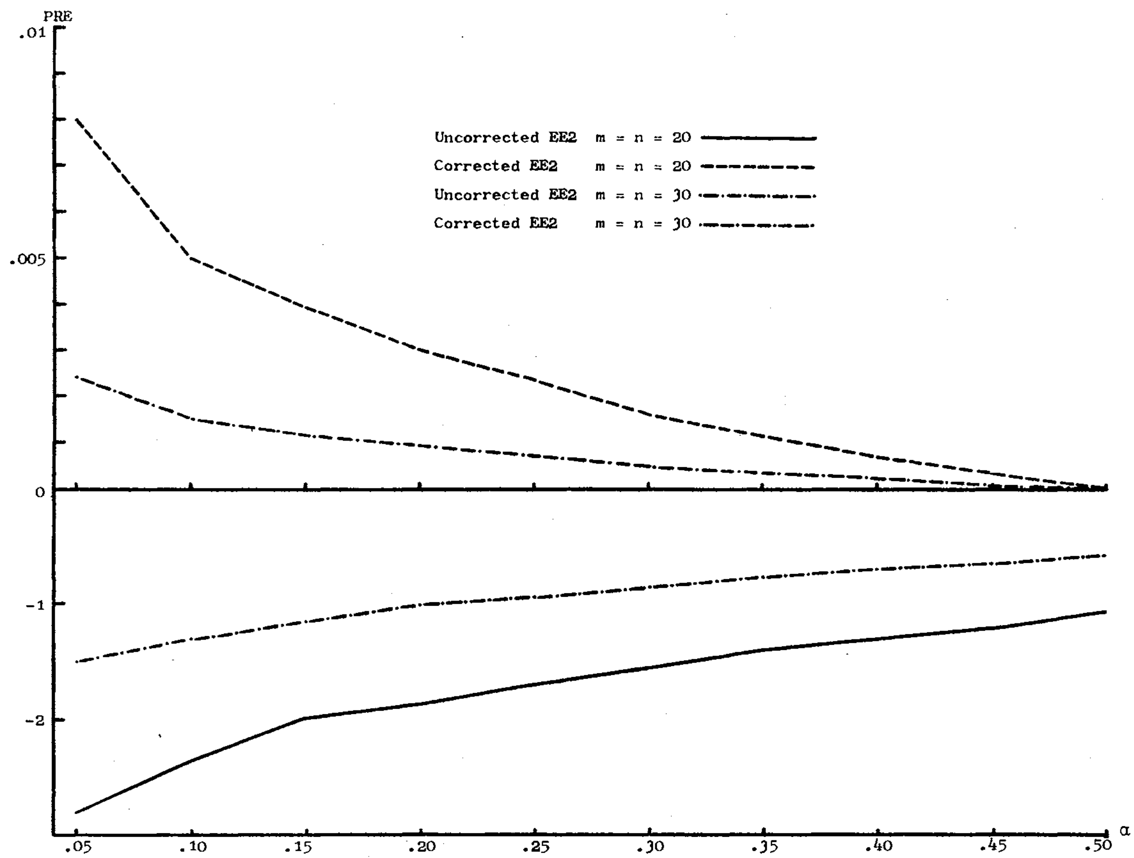
(a) $0.0005 \leq \alpha \leq 0.05$



NOTE: There is a change in the PRE scale at 0 percent.

Figure 7. Percent Relative Error (PRE) of the Uncorrected and Corrected EE2 With Respect to Probability Level (α) for m = n = 20, 30

(b) $0.05 \leq \alpha \leq 0.50$



NOTE: There is a change in the PRE scale at 0 percent and the change in the α scale from Figure 7(a) to Figure 7(b).

Figure 7. (Continued)

BIBLIOGRAPHY

[1] Buckle, N., C. Kraft, and C. van Eeden. "An Approximation to the Wilcoxon-Mann-Whitney Distribution." Journal of the American Statistical Association, 64 (1969), pp. 591-599.

[2] Conover, W. J. Practical Nonparametric Statistics. New York: John Wiley & Sons, Inc., 1971.

[3] IBM Systems Reference Library. "Mathematical and Service Subprograms." (Form GC28-6818-0). White Plains, New York: IBM Technical Publications Department, July, 1969.

[4] Fix, E., and J. L. Hodges, Jr. "Significance Probabilities of the Wilcoxon Test." Annals of Mathematical Statistics, 26 (1955), pp. 301-312.

[5] Jacobson, J. E. "The Wilcoxon Two-Sample Statistic: Tables and Bibliography." Journal of the American Statistical Association, 58 (1963), pp. 1086-1103.

[6] Mann, H. B., and D. R. Whitney. "On a Test of Whether One of Two Random Variables is Stochastically Larger Than the Other." Annals of Mathematical Statistics, 18 (1947), pp. 50-60.

[7] Milton, R. C. "An Extended Table of Critical Values for the Mann-Whitney (Wilcoxon) Two-Sample Statistic." Journal of the American Statistical Association, 59 (1964), pp. 925-934.

[8] Owen, D. B. Handbook of Statistical Tables. Reading, Mass.: Addison-Wesley Publishing Co., Inc., 1972.

[9] van der Vaart, H. R. "A Closed Expression for Certain Probabilities in Wilcoxon's Two Sample Test." Experimentia, 12 (1956), pp. 14-15.

[10] Verdooren, L. R. "Extended Tables of Critical Values for Wilcoxon's Test Statistic." Biometrika, 50 (1963), pp. 177-186.

[11] White, C. "The Use of Ranks in a Test of Significance for Comparing Two Treatments." Biometrics, 8 (1952), pp. 33-41.

[12] Wilcoxon, F. "Individual Comparison by Ranking Methods." Biometrics, 1 (1945), pp. 80-83.

[13] Wilcoxon, F. "Probability Tables for Individual Comparisons by Ranking Methods." Biometrics, 3 (1947), pp. 119-122.

[14] Wilcoxon, F., S. K. Katti, and R. A. Wilcox. Critical Values and Probability Levels for the Wilcoxon Rank Sum Test and the Wilcoxon Signed Rank Test. Pearl River, N. Y.: American Cyanamid Co. and the Florida State University, 1963. See, also, Selected Tables in Mathematical Statistics, Vol. 1, ed. H. L. Harter and D. B. Owen (Chicago, 1970).

VITA

Ralph Edwin Porter Jr.

Candidate for the Degree of

Master of Science

Thesis: NORMAL AND EDGEWORTH APPROXIMATIONS TO THE DISTRIBUTION OF THE WILCOXON-MANN-WHITNEY STATISTIC

Major Field: Statistics

Biographical:

Personal Data: Born in Kittaning, Pennsylvania, September 13, 1949, the son of Mr. and Mrs. Ralph Edwin Porter Sr.

Education: Graduated from Northwestern High School, Darlington, Pennsylvania, in May, 1967; received the Bachelor of Science degree from Oklahoma State University, Stillwater, Oklahoma, with a major in Mathematics in December, 1972; completed requirements for the Master of Science degree at Oklahoma State University in July, 1974.

Professional Experience: Worked as a research laboratory technician at the St. Joseph Lead Company, Monaca, Pennsylvania, 1968-69; special electronics technician for the United States Air Force from May, 1969 to January, 1971.