Journal of Educational Statistics
Winter, 1980, Volume 5, Number 4 pp. 309-335

A COMPARISON OF THE POWER OF WILCOXON'S RANK-SUM STATISTIC TO THAT OF STUDENT'S t STATISTIC UNDER VARIOUS NONNORMAL DISTRIBUTIONS

R. Clifford Blair and James J. Higgins

University of South Florida

Key words: Power, t test, Wilcoxon's rank sum test

ABSTRACT

Computer generated Monte Carlo techniques were used to compare the power of Wilcoxon's rank-sum test to the power of the two independent means t test for situations in which samples were drawn from (1) uniform, (2) Laplace, (3) half-normal, (4) exponential, (5) mixed-normal, and (6) mixed-uniform distributions. Sample sizes studied were (n_1, n_2) = (3,9), (6,6), (9,27), (18,18), (27,81), and (54,54).

It was concluded that (1) generally speaking, the Wilcoxon statistic held very large power advantages over the t statistic, (2) asymptotic relative efficiencies were reasonably good indicators of the relative power of the two statistics, (3) results obtained from smaller samples were often markedly different from the results obtained from larger samples, and (4) because of the narrow ranges of population shapes and sample sizes investigated in some widely cited previous studies of this type, the conclusions reached in those studies must now be deemed questionable.

BACKGROUND

Although nonparametric statistical tests enjoyed some popularity among educational and psychological researchers during the 1950's (Glass, Peckham, & Sanders, 1972), attitudes concerning the usefulness of such procedures have changed markedly since that time. This change in attitude is reflected by Glass et al. (1972) who characterize the 1950's movement toward nonparametrics as "unnecessary" and "unproductive." These authors go on to imply that researchers who use such procedures are not doing so on the basis of an informed decision, but rather, are simply caught up in a "herd" psychology.

Unlike Glass et al. (1972), Guilford and Fruchter (1978) seem to feel that nonparametric tests may be of some very limited use in analyzing research data, but go on to admonish the reader that "Where there is any choice...we should prefer a parametric test, except where a quick, rough test will do." (p. 212)

Why do these authors and so many others discourage the use of nonparametric tests? First, it is argued that, although nonnormal data may be encountered with some frequency in educational and psychological research, the commonly used t and F tests are quite insensitive to this violation of their underlying assumptions, thereby making the use of nonparametric tests unnecessary (Boneau, 1960; Glass et al., 1972). Second, it is often argued that nonparametric tests are less powerful than parametric tests, thereby making them the less desirable alternative (Gay, 1976; Guilford & Fruchter, 1978; Kerlinger, 1973; Popham & Sirotnick, 1973). Although the first part of the rationale outlined above is questionable to some degree (Bradley, 1978), it is the second part, that is, the assertion that parametric tests are more powerful than nonparametric tests, that gives rise to the focus of this study.

THE PROBLEM

In education and psychology, the most commonly used two sample test for shift is, of course, the Student t test. A major reason for its popularity lies in the fact that it is said to be (a) robust to deviations of populations from normality, and (b) more powerful than nonparametric counterparts that might be used in its stead (Boneau, 1960; 1962). Thus, researchers who face the task of analyzing data that have been drawn from populations whose shapes are nonnormal or unknown, are assured that the *t* test is still the most appropriate procedure.

Generally unrecognized, or at least not made apparent to the reader, is the fact that the *t* test's claim to power superiority rests on certain optimal power properties that are obtained under normal theory. Thus, when the shape of the sampled population(s) is unspecified, there are no mathematical or statistical imperatives to ensure the power superiority of this statistic. Unfortunately, not much is known of the relative power performance of the *t* test and its nonparametric counterparts when samples of various sizes are drawn from a wide variety of population shapes. Such information would, however, be very useful in choosing an appropriate test when population shapes are nonnormal or unknown. The present study is designed, therefore, to assess the relative power of the *t* test and its most popular nonparametric counterpart (Bradley, 1972), Wilcoxon's rank sum test, under a wide variety of sample size and population shape combinations.

RELEVANT LITERATURE

Sampling experiments, mathematical calculations, and asymptotic theory have all been used to demonstrate the fact that the *t* test and Wilcoxon's test have nearly equivalent power when samples are drawn from normally distributed populations (Dixon, 1954; Hodges & Lehmann, 1956; Lehmann, 1975; Neave & Granger, 1968). The slight power advantage that is obtained in this circumstance is, of course, in favor of the *t* test.

An interesting and potentially important asymptotic result was obtained by Hodges and Lehmann (1956) who demonstrated that while the asymptotic relative efficiency (or Pitman efficiency) of the Wilcoxon test relative to the *t* test can be as high as infinity, it can never be lower than .864. Commenting on this result, Hodges and Lehmann state that:

> To the extent that the above concept of efficiency adequately represents what happens for the sample sizes and alternatives arising in practice, this result shows that the use of the Wilcoxon test instead of the Student's *t* test can never entail a serious loss of efficiency for testing against shift. (On the other hand, it is obvious...that the Wilcoxon test may be infinitely more efficient than the *t* test.) (p. 356)

Unfortunately, asymptotic relative efficiencies are calculated under a rather unrealistic set of assumptions prompting Bradley (1968, p. 58) to state: "No experimenter takes infinitely large samples, and virtually no one is interested in power to reject hypotheses that differ only infinitesimally from the null hypotheses."

Boneau (1962) used computer generated Monte Carlo techniques to study the relative power of the *t* and Wilcoxon statistics when sampling is from normal, rectangular, and exponential distribution. He concluded that, "In general the *t* test is more powerful than the Mann-Whitney *U* (Wilcoxon) test, but never by much." In addition, he implied that the asymptotic results obtained by Hodges and Lehmann may not carry over to the situation in which sample sizes are finite.

Blair, Higgins, and Smitley (1980) used computer simulation to study the relative power of the two tests at hand under the exponential distribution. They concluded that the small sample sizes employed by Boneau (1962) had led him to a faulty conclusion and that the Wilcoxon test attains large power advantages over the *t* test under the exponential distribution.

Toothaker (1972) used computer simulation to draw samples from normal, uniform, and skewed populations in order to compare the power of the two statistics under discussion. Sample sizes used in this study were ≤ 5 and the results obtained were much the same as those reported by Boneau (1962); that is, there was little difference between the powers of the two tests.

Neave and Granger (1968) drew samples of size $n_1 = n_2 = 20$ and $n_1 = 20$, $n_2 = 40$ from a population that is formed by the super position of two normal distributions. After comparing the power of the statistics of interest, they concluded that the Wilcoxon statistic is "much superior" to the *t* statistic under the particular nonnormal distribution that they studied. In this study, the difference in proportions of null hypotheses rejected by the two tests was as high as .12 with the Wilcoxon having the larger proportion.

The literature reviewed above is confusing in that it presents what appears to be conflicting pictures of the relative power of the two tests. The asymptotic results of Hodges and Lehmann (1956) suggest that, while the Wilcoxon test may be much more powerful than the t test, the t test can never show more than a modest advantage over the Wilcoxon test. But, as Bradley (1968) has warned, asymptotic results must be suspect because of the unrealistic assumptions underlying their calculation. Added to this warning is the fact that Boneau (1962) has denied, on the basis of his sampling experiments, the utility of the Hodges and Lehmann finding. Results obtained from Toothaker (1972) seem to support the Boneau (1962) position. On the other hand, Blair et al. (1980) have questioned the usefulness of the finding of Boneau (1962) and Toothaker (1972) by pointing out that (a) sample sizes employed by these two authors were smaller than those commonly found in educational and psychological research, and (b) results obtained from small samples may be very different from those obtained with more moderate-sized samples.

THE PRESENT STUDY

The general purpose of the present study was to determine whether the t test or Wilcoxon's test is typically the more powerful procedure when samples are drawn from a wide variety of population shapes. In order to accomplish this in a manner that will be most useful to educational and psychological researchers, two distinct voids in the present literature had to be filled.

First, as was mentioned earlier, previous studies have often considered sample sizes that in educational and psychological research contexts would be characterized as very small (≤ 5) or very large (infinite). Therefore, this study considered a more moderate range of sample sizes. Second, Bradley (1977) has criticized previous studies for considering too marrow a range of distribution shapes. Therefore, this study dealt with a larger variety of population shapes than is found in previous studies of this type.

The present study used, as its primary means of investigation, computer generated Monte Carlo methods. Through this technique, samples of various sizes were drawn from populations with known characteristics. The two statistics of interest were calculated on the drawn samples, tests of significance were carried out, and the reject/fail-to-reject decision was recorded. Because the populations were constructed to simulate the situation in which the null hypothesis was not true, the proportion of samples that resulted in a rejection of the null hypothesis was a statement of the power of the test. (In this study all power functions are one-tailed.) Details of the populations studied and the simulation techniques employed are given below.

The first population investigated was the uniform (or rectangular) distribution whose functional form is as follows:

$$f(x) = 1, \ 0 \leq x \leq 1.$$

This population was included in the study because it represents one extreme in the family of symmetric power distributions, and as such, is a good example of a light-tailed symmetric distribution. The asymptotic relative efficiency of the Wilcoxon to the *t* test is 1.0 under this distribution.

The second population studied was the Laplace (or double exponential) distribution whose functional form is as follows:

$$f(x) = \tfrac{1}{2} \exp\{-|x-\mu|\}, -\infty < x < \infty.$$

It was included in this study because it represents one extreme (the opposite extreme of the uniform distribution) in the family of symmetric power distributions and, as such, is a good example of a heavy-tailed symmetric distribution. The asymptotic relative efficiency of the Wilcoxon to the *t* test is 1.5 under this distribution.

The third population studied was the truncated (or half) normal distribution whose functional form is as follows:

$$f(x) = (2/\sigma^2\pi)^{\frac{1}{2}} \exp\{-(x-\mu)^2/2\sigma^2\}, \quad x \geq \mu.$$

This may be thought of as the upper half of a normal distribution, and as such, is a good example of a nonsymmetric distribution whose tail descends at the same rate as would be found in a normal curve function. This function is also useful in modeling data that is gathered in connection with certain compensatory education programs. The asymptotic relative efficiency of the Wilcoxon to the *t* test is approximately 1.2 under this distribution.

The fourth population studied was the exponential distribution whose functional form is as follows:

$$f(x) = e^{-(x-\mu)} \qquad x \geq \mu$$

This may be thought of as the upper half of the Laplace distribution and, as such, is a good example of a heavy-tailed nonsymmetric distribution. The asymptotic relative efficiency of the Wilcoxon to the t test is 3.0 under this distribution.

The fifth population studied was the mixed normal whose functional form is as follows:

$$f(x) = \frac{.95\ e^{-x^2/2}}{\sqrt{2\ \pi}} + \frac{.05\ e^{-(x-33)^2/200}}{10\sqrt{2\ \pi}} \qquad -\infty < x < \infty.$$

This distribution was included because it represents a radical departure from normality that nonethless appears to model data collected in certain social science research contexts (Allport, 1934; Bradley, 1977). At the same time, it is a good example of a highly skewed population. The asymptotic relative efficiency of the Wilcoxon to the t test is approximately 45.0 under this distribution.

The last population studied was the mixed uniform whose functional form is as follows:

$$f(x) = \begin{cases} .4, & 0 \le x \le 1, \\ \frac{.6}{39}, & 1 \le x \le 40. \end{cases}$$

This distribution was included for essentially the same reasons outlined in connection with the fifth population. The asymptotic relative efficiency of the Wilcoxon to the t test is approximately 58.0 under this distribution.

The six populations described above are extremely diverse in terms of both skew and kurtosis, thereby providing a broad base for the present study.

Sample sizes investigated in connection with each of the six populations were: $(n_1, n_2) = (3,9), (6,6), (9,27), (18,18), (27,81)$, and $(54,54)$. The sequence of events in the simulation were as follows: (1) Two independent samples of sizes n_1 and n_2 were selected from the population being studied. (2) A constant was added to the scores of the designated "treatment" group (i.e., the group having sample size n_1), thus simulating the condition under which $\mu_1 > \mu_2$. (3) The t and Wilcoxon statistics were computed for the two sammples. (4) The calculated statistics were compared with the appropriate critical values and the reject/fail-to-reject decision was recorded.

After 5,000 repetitions of the above sequence, the value of the added constant (i.e., $\mu_1 - \mu_2$) was increased and the process repeated. This was continued until a wide range of the respective power functions were obtained. (Data were generated by means of the GGUSN and GGUS3 subroutines of the International Mathematical and Statistical Laboratories (1977) computer package.)

It should be noted that critical _t_ values were obtained by simulating the null distribution of the _t_ statistic for all sample sizes under each of the six population shapes. The critical _t_ value chosen for a particular power comparison was the _t_ value whose associated probability was equal to the probability associated with the corresponding critical value of the Wilcoxon test. For example, if a particular critical Wilcoxon value had an associated probability of .048, then the _t_ value chosen was the one that, for the particular distribution being studied, also had an associated probability of .048. This procedure was necessary because, under the Neyman and Pearson (1933) concept of power, comparisons of this type must be made at the same level of significance.

RESEARCH QUESTIONS

Questions specifically addressed by this study are listed below.

1. In the case of moderate sample sizes (operationally defined as $n_1 + n_2 = 36$ and $n_1 + n_2 = 108$), does Wilcoxon's test tend to be more powerful than the _t_ test under some distributions?

2. In the case of moderate sample sizes, does the _t_ test tend to be more power than Wilcoxon's test under some distributions?

3. Given circumstances in which Wilcoxon's test is more powerful than the _t_ test and vice versa, do the magnitudes of the power advantages differ for the two tests?

4. When samples are of moderate sizes, do asymptotic relative efficiencies provide an adequate indication as to which of the two tests being studied is the more powerful under a particular distribution?

5. Are the results obtained from small samples (operationally defined as $n_1 + n_2 = 12$) generalized to the moderate sample size situation?

RESULTS AND CONCLUSIONS

The amount of data generated by this study make publication of all results impractical. Therefore, data are represented in two summary forms. First, the one-tailed power functions of the two statistics as calculated under each of the six populations, are presented graphically in Figures 1 through 6. These graphs depict situations in which $n_1 = n_2$ and $\alpha \approx .025$.

TABLE I

Maximum Power Advantages Attained by the *t* and Wilcoxon Statistics at Various Sample Size/Significance Level Combinations for Samples Drawn From Uniform Distributions

		Level of Significance			
n_1,n_2	Statistic	.005	.010	.025	.050
3,9	w	.00	.00	<.01	<.01
	t	.13	.13	.09	.06
6,6	w	.02	.01	.01	.01
	t	<.01	.01	.05	.06
9,27	w	<.01	.01	<.01	.01
	t	.07	.08	.07	.05
18,18	w	.00	.00	<.01	<.01
	t	.09	.07	.05	.04
27,81	w	.00	.01	.00	.00
	t	.07	.05	.05	.04
54,54	w	.00	.01	<.01	.00
	t	.07	.04	.04	.03

TABLE II

Maximum Power Advantages Attained by the *t* and Wilcoxon Statistics at Various Sample Size/ Significance Level Combinations for Samples From Laplace (Double Exponential) Distributions

n_1,n_2	Statistic	Level of Significance .005	.010	.025	.050
3,9	w	.00	.01	.02	.04
	t	.13	.07	.04	.02
6,6	w	<.01	<.01	.01	.01
	t	.07	.05	.04	.01
9,27	w	.09	.08	.08	.07
	t	.00	<.01	.00	.00
18,18	w	.10	.10	.09	.10
	t	.00	.00	.00	.00
27,81	w	.17	.17	.12	.12
	t	.00	.00	.00	.00
54,54	w	.17	.14	.15	.15
	t	.00	.00	.00	.00

TABLE III

Maximum Power Advantages Attained by the *t* and Wilcoxon Statistics at Various Sample Size/ Significance Level Combinations for Samples Drawn From Truncated Normal Distributions

n_1,n_2	Statistic	Level of Significance .005	.010	.025	.050
3,9	w	.08	.02	.05	.07
	t	.00	<.01	<.01	.00
6,6	w	.00	.01	<.01	.01
	t	.10	.04	.03	.02
9,27	w	.07	.08	.08	.14
	t	<.01	<.01	.00	.00
18,18	w	.05	.06	.05	.04
	t	<.01	.00	.00	.00
27,81	w	.14	.11	.11	.11
	t	.00	<.01	.00	.00
54,54	w	.12	.11	.09	.09
	t	.00	.00	.00	.00

TABLE IV

Maximum Power Advantages Attained by the t and Wilcoxon Statistics at Various Sample Size/Significance Level Combinations for Samples Drawn From Exponential Distributions

n_1,n_2	Statistic	Level of Significance .005	.010	.025	.050
3,9	w	.02	.02	.06	.12
	t	.00	.03	.01	.00
6,6	w	.02	<.01	.02	.05
	t	.09	.11	.05	.01
9,27	w	.27	.28	.29	.30
	t	<.01	.00	.00	.00
18,18	w	.17	.19	.22	.21
	t	.00	.00	.00	.00
27,18	w	.44	.42	.37	.33
	t	.00	.00	.00	.00
54,54	w	.36	.35	.32	.29
	t	.00	.00	.00	.00

TABLE V

Maximum Power Advantages Attained by the t and Wilcoxon Statistics at Various Sample Size/Significance Level Combinations for Samples Drawn From Mixed Normal Distributions

n_1,n_2	Statistic	.005	.010	.025	.050
3,9	w	.08	.05	.03	.30
	t	.08	.16	.17	.00
6,6	w	.19	.18	.20	.30
	t	.14	.15.	.12	.02
9,27	w	.75	.74	.73	.71
	t	.00	.00	.00	.00
18,18	w	.68	.63	.61	.58
	t	.00	.00	.00	.00
27,81	w	.94	.92	.89	.85
	t	.00	.00	.00	.00
54,54	w	.89	.88	.84	.79
	t	.00	.00	.00	.00

TABLE VI

Maximum Power Advantages Attained by the t and Wilcoxon Statistics at Various Sample Size/ Significance Level Combinations for Samples Drawn From Mixed Uniform Distributions

n_1,n_2	Statistic	Level of Significance .005	.010	.025	.050
3,9	w	.05	.00	.02	.03
	t	.02	.04	.06	.05
6,6	w	.01	.01	.04	.08
	t	.16	.16	.13	.09
9,27	w	.06	.08	.09	.14
	t	.05	.04	.04	.02
18,18	w	.16	.14	.17	.20
	t	.06	.06	.04	.03
27,81	w	.29	.27	.34	.37
	t	.00	.00	.00	.00
54,54	w	.31	.34	.38	.44
	t	.01	.01	.01	.00

TABLE VII

Maximum Power Advantages Attained by the t and Wilcoxon Statistics When Small and Moderate Sized Samples Are Drawn From Certain Nonnormal Distributions

Distribution	Statistic	Small Samples	Moderate Samples
Uniform	w	.00	.01
	t	.13	.09
Laplace	w	.04	.17
	t	.13	.00
Truncated Normal	w	.08	.14
	t	.10	.00
Exponential	w	.12	.44
	t	.11	.00
Mixed Normal	w	.30	.94
	t	.17	.00
Mixed Uniform	w	.08	.44
	t	.16	.06

TABLE VIII

Frequency of Occurrence of the Maximum Power Advantages of the *t* and Wilcoxon Tests

	W		*t*	
	f	percent	f	percent
x = .00	27	19	76	53
.00 < x $\le$ .05	26	18	38	26
.05 < x $\le$ .10	23	16	18	13
.10 < x $\le$ .20	28	19	12	8
.20 < x $\le$ .30	11	.8	0	0
.30 < x $\le$ .40	10	7	0	0
.40 < x $\le$ .50	3	2	0	0
.50 < x $\le$ .60	0	0	0	0
.60 < x $\le$ .70	4	3	0	0
.70 < x $\le$ .80	5	3	0	0
.80 < x $\le$ .90	5	3	0	0
.90 < x $\le$ .100	2	1	0	0

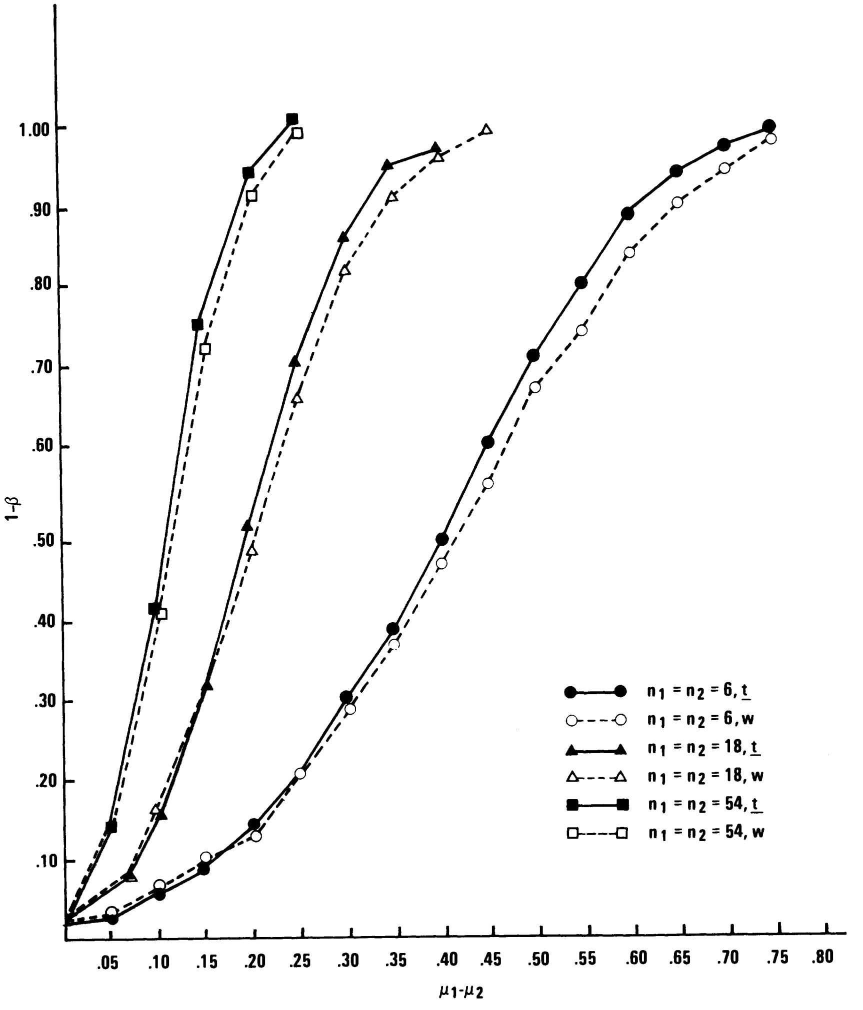

Figure 1

One-Tailed Power Function of the Two Independent Means t Test and Wilcoxon's Rank Sum Test for Samples Drawn from a Uniform Distribution. $\alpha \simeq .025$.

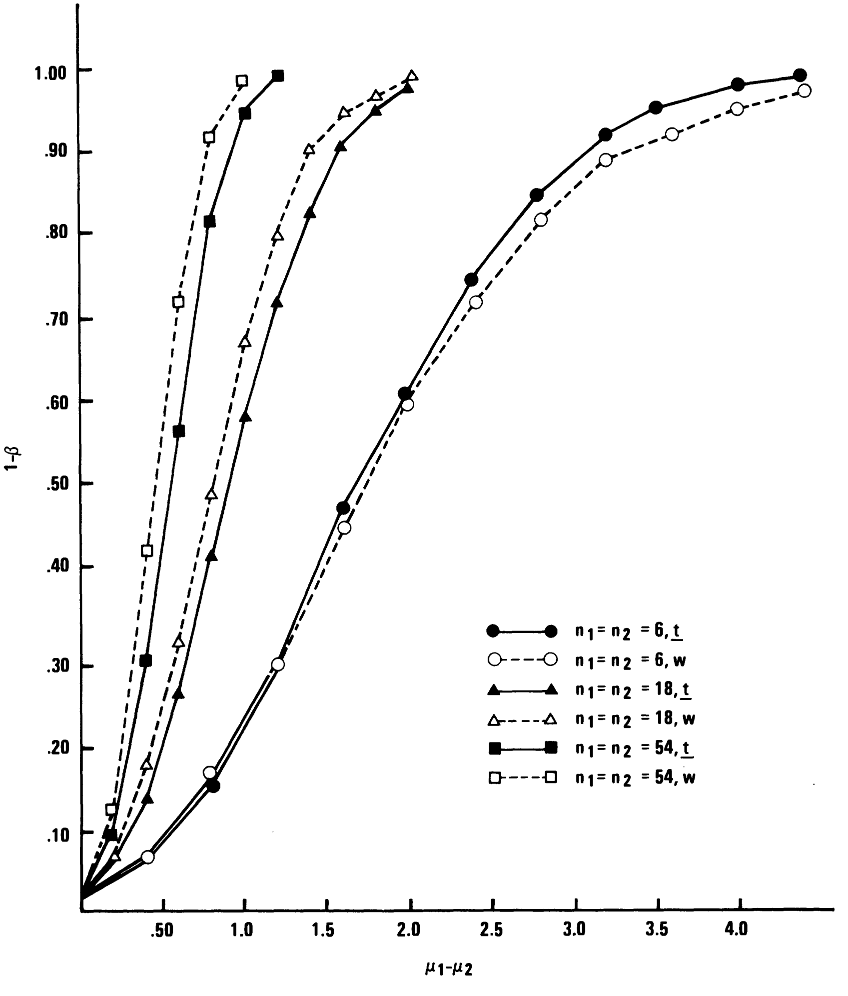

Figure 2

One-Tailed Power Functions of the Two Independent Means $\underline{t}$ Test and Wilcoxon's Rank Sum Test for Samples Drawn From a Laplace Distribution. $\alpha \simeq .025$.

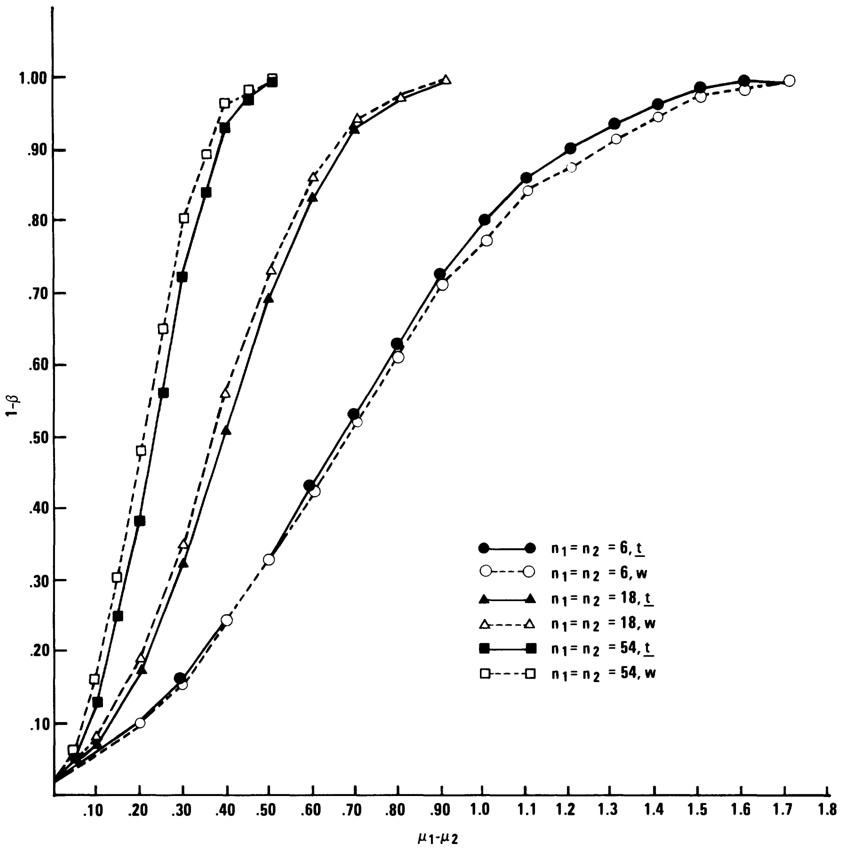

Figure 3

One-Tailed Power Functions of the Two Independent Means $\underline{t}$ Test and Wilcoxon's Rank Sum Test for Samples Drawn From a Truncated Normal Distribution. $\alpha \simeq .025$.

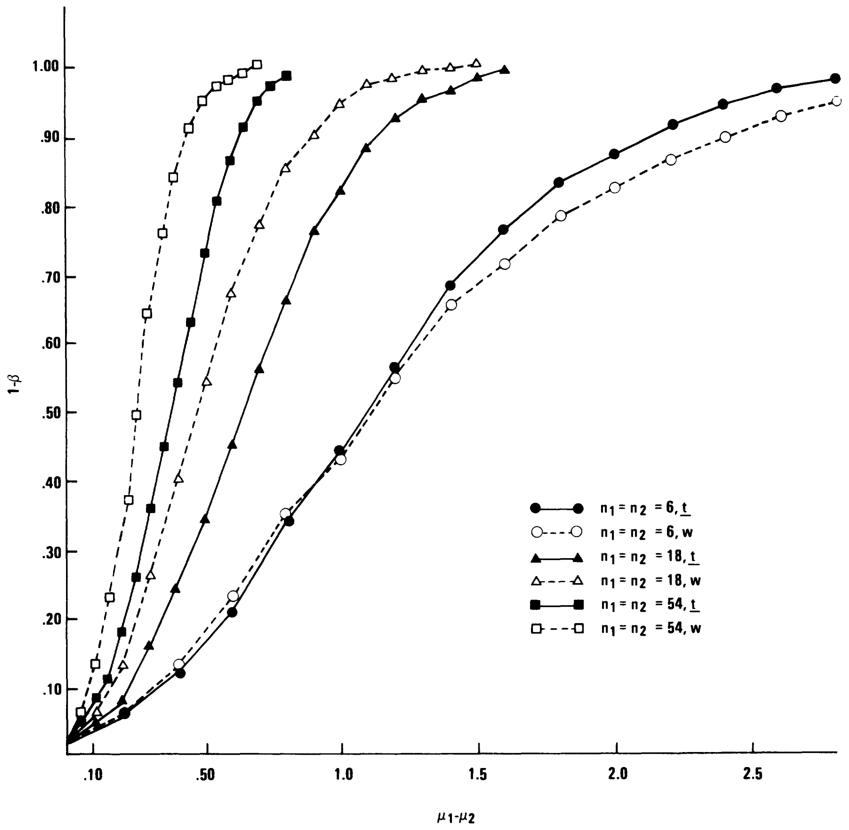

Figure 4

One-Tailed Power Functions of the Two Independent Mean $\underline{t}$ Test and Wilcoxon's Rank Sum Test for Samples Drawn From an Exponential Distribution. $\alpha \simeq .025$

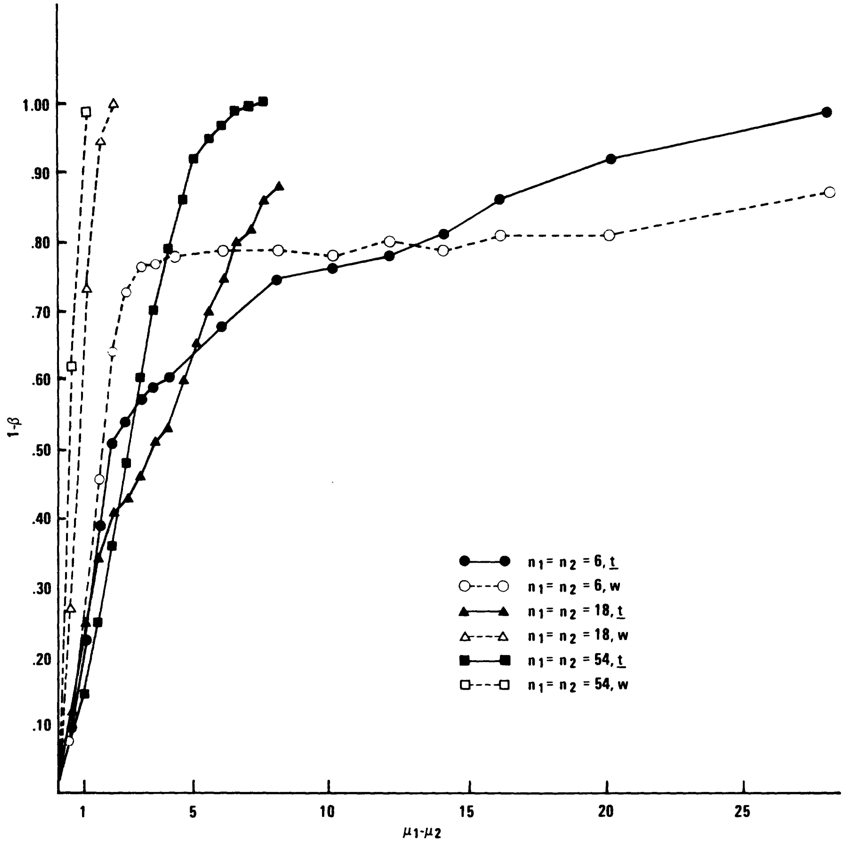

Figure 5

One-Tailed Power Functions of the Two Independent Means $\underline{t}$ Test and Wilcoxon's Rank Sum Test for Samples Drawn from a Mixed Normal Distribution. $\alpha \simeq .025$.

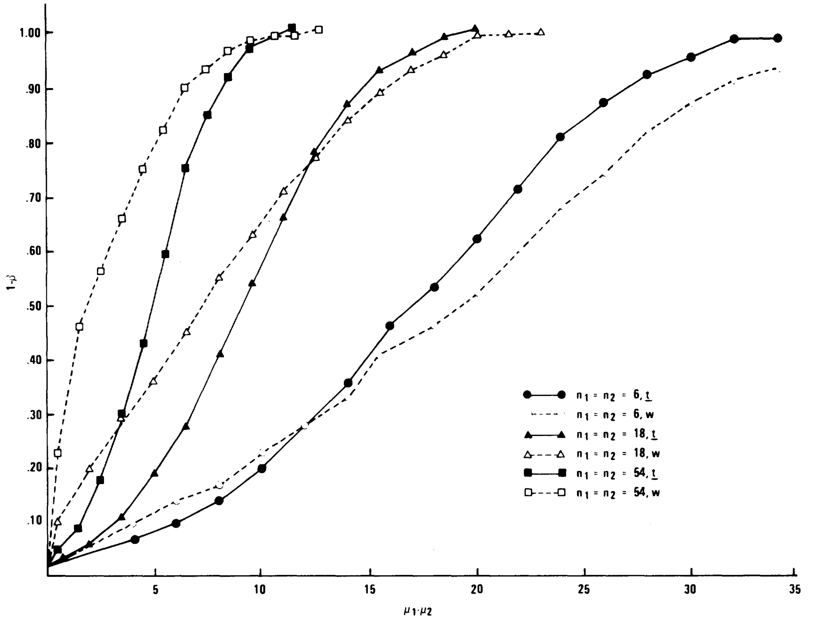

Figure 6

One-Tailed Power Functions of the Two Independent Means t Test and Wilcoxon's Rank Sum Test for Samples Drawn From a Mixed Uniform Distribution. $\alpha \simeq .025$.

Tables I through VI show the largest power advantages attained by each statistic for each sample size and significance level combination. Power advantage refers to the quantity obtained when the proportion of hypotheses rejected by the less powerful statistic is subtracted from the proportion of rejections by the more powerful statistic with both proportions being calculated at a particular value of $\mu_1 - \mu_2$. Thus, the largest power advantage of a given statistic was obtained by considering all values of $\mu_1 - \mu_2$ for which the given statistic held the advantage.

Tables VII and VIII are further distillations of the data contained in Tables I-VI. Table VII gives maximum power advantages attained by the two statistics for small- and moderate-sized samples by distribution. Table VIII is explained later.

Attention is now turned to the previously stated research questions. Because of their similarity, research questions 1 and 2 will be addressed jointly.

Figures 1-6, Tables I-VI, and in a more succinct fashion, Table VII indicate that clear patterns of power superiority emerge when the moderate sample size case is considered across the six distributions. Specifically, the _t_ test holds power superiority under the uniform distribution, while the Wilcoxon dominates under the other five distributions. A minor exception occurred when samples of sizes $n_1 + n_2 = 36$ were drawn from the mixed uniform distribution. In this situation the _t_ test shows definite, though modest, power advantages over some ranges of the power functions. This intermittent advantage virtually disappears for samples of sizes $n_1 + n_2 = 108$ and the Wilcoxon test becomes the more powerful test over the full range of power functions except where both functions approach 1.0

From the results outlined about it can be concluded that the answers to research questions 1 and 2 are both in the affirmative.

Figures 1 through 6 as well as Tables I through VII give the impression that, in those circumstances where the _t_ test is the more powerful statistic, the magnitude of its power superiority is typically quite modest. On the other hand, in those circumstances where the Wilcoxon is the more powerful statistic, the magnitude of its power superiority is oftentimes quite large. For example, while Figure 1 indicates that that the _t_ test is typically the more powerful statistic under the uniform distribution, Tables I and VII show that the magnitude of that advantage never exceeded .13 and was usually about half that amount. On the other hand, Figure 5 as well as Tables V and VII indicate not only that the Wilcoxon test is the more powerful statistic under the mixed normal distribution, but also that the magnitude of the advantage can reach as high as .94 with maximum advantage in the .60 to .80 range being common. Further insights pertaining to research question 3 can be gained from Table VIII.

Because four levels of significance were investigated for each combination of sample sizes, and because six combinations of sample sizes were investigated for each of the six populations, there were 4 x 6 x 6 = 144 pairs of entries in Tables I through VI. Table VIII classifies the 144 entries for each of the two statistics by their magnitudes. For example, Table VIII indicates that of the 144 entries for the Wilcoxon test in Table I through VI, 7 percent of these entries are in the range $.30 < x \leq .40$ where x is the value of the maximum power advantage.

Table VIII indicates that for 19 percent of the power function comparisons, the Wilcoxon test never held an advantage. The comparable figures for the _t_ test are a much larger 53 percent. It is also intresting to note that only 8 percent of the maximum power advantages of the _t_ test exceed .10, while the comparable figure for the Wilcoxon statistic was 46 percent. While the _t_ test never showed maximum power advantages greater than .20, some 27 percent of the Wilcoxon's entries exceeded this figure.

Summarizing the results related to research question 3, it appears that while the _t_ test is sometimes more powerful than the nonparametric procedure, the magnitude of that advantage is never very large and is usually quite modest. On the other hand, the Wilcoxon test often shows power advantages that are _very_ large. As a result, research question 3 must also be answered in the affirmative. Attention is now turned to research question 4.

As was mentioned previously, the asymptotic relative efficiency of the Wilcoxon test relative to the _t_ test is unity under the uniform distribution. This would indicate that the two tests have equivalent power under this distribution. In contrast to this expectation, however, the t test held a definite, though modest, power advantage when moderate-sized samples were taken from this distribution. It can be concluded, therefore, that the asymptotic relative efficiency is slightly misleading in this set of circumstances.

The asymptotic relative efficiency of approximately 1.2 obtained under the truncated normal distribution would suggest a slight power advantage for the Wilcoxon statistic. As was noted earlier, the Wilcoxon test did show a power advantage under this distribution leading to the conclusion that the asymptotic relative efficiency was a good indicator in this case.

As was noted earlier, an asymptotic relative efficiency of 1.5 is obtained under the Laplace distribution, suggesting a power advantage for the Wilcoxon statistic. In addition, it might be expected that the magnitude of the advantage obtained under this distribution would be slightly larger than that obtained under the truncated normal distribution. Comparison of Tables II and III support these expectations.

The asymptotic relative efficiency of 3 obtained under the exponential distribution suggests that the power advantage obtained under this distribution would be larger than that associated with the Laplace distribution. This supposition is fully supported by the data in Tables II and IV.

The asymptotic relative efficiency of 45 obtained under the mixed normal distribution suggests that the advantage here would be substantially larger than that obtained under the exponential distribution. Again, this expectation is fully supported by the data in Tables IV and V.

The two statistics of interest have an asymptotic relative efficiency of approximately 58 under the mixed uniform distribution. This efficiency would suggest that the power advantage of the nonparametric test would be greater under this distribution than under the mixed normal distribution. Comparisons of the data in Tables V and VI indicated that while the Wilcoxon is generally the more powerful statistic under this distribution, the magnitude of its advantage tends to be far less than that attained under the mixed normal distribution. It can be concluded, therefore, that while asymptotic relative efficiency is a good indicator as to which of the two statistics is the more powerful under the mixed uniform distribution, it is somewhat misleading as an indicator of the magnitude of that advantage.

Based on the six distributions investigated, it can be concluded that, in general, asymptotic relative efficiencies often provide an adequate indication as to which of the two tests investigated is the more powerful under a particular distribution when samples are of moderate sizes. It should be noted, however, that these efficiencies are not unerring in this regard, as was demonstrated with the uniform distribution. Research question 4 can thus be answered with a qualified yes. Attention is now turned to research question 5.

Small sample sizes were operationally defined, for the purposes of this study, as $n_1 + n_2 = 12$. Table I indicates that, with some exceptions, the $\underline{t}$ test was the more powerful statistic when samples were small and were drawn from uniform distributions. This is essentially the same result as was obtained with moderate-sized samples.

Table II indicates that, with some exceptions, the $\underline{t}$ test was the more powerful test when samples were small and were drawn from Laplace distributions. This result is contrary to that obtained with moderate-sized samples where the Wilcoxon test dominated.

Table III shows a rather mixed pattern of power advantages for the situation in which samples are small and drawn from truncated normal distributions. When samples were of sizes 3 and 9, it was the Wilcoxon test that dominated, but it was the $\underline{t}$ test that showed superior power when samples were of sizes 6 and 6. This contrasts with the moderate sample size situation where the advantage was with the Wilcoxon test for both balanced and unbalanced data.

Table IV also shows a mixed pattern of power advantages for the situation in which samples are small and drawn from exponential distributions. In this situation, each of the tests dominated in a given set of circumstances. This contrasts with the moderate sample size situation where the Wilcoxon clearly dominated.

Table V again shows a mixed pattern of advantages for the case of small sample sizes. As was noted previously in regard to other distributions, the mixed pattern gives way to clear domination by the nonparametric test when samples are of moderate sizes.

Table VI shows that, with some exceptions, it was the $\underline{t}$ test that showed power dominance when samples were small and drawn from mixed uniform distributions. This contrasts with the moderate sample size situation where, with some exceptions that occurred when $n_1 + n_2 = 36$, it was the Wilcoxon statistic that attained power dominance.

From the results outlined above, it can be concluded that the results obtained from small sample studies that compare the power of the two statistics in question do not, typically, generalize to situations involving samples of moderate sizes. In fact, conclusions reached on the basis of small sample studies are oftentimes in direct opposition to those reached on the basis of moderate sample size studies.

COMMENTS

Perhaps the most important consequence of this study is the fact that it raises serious questions about the validity of some of the more "authoritative" literature dealing with the relative usefulness of parametric and nonparametric procedures.

For example, Boneau (1960), in one of the most widely cited articles on the subject, has maintained that the *t* test rather than a nonparametric test should be employed when the population shape is not normal. Boneau (1960) based his position on the assertion that the *t* test is robust, in terms of Type I errors, to population nonnormality. But as this study has demonstrated, a researcher may choose to use a nonparametric counterpart of the *t* test, not only because of the advantage of obtaining a stable Type I error rate but also because of large advantages gained in terms of relative Type II error rates. This same logic can be used to refute the arguments of Glass et al. (1972) who strongly condemned the use of nonparametric tests but, like so many others, failed to identify and investigate the issue of relative power.

In addition, this study further strengthens the position taken by Blair et al. (1980) that Boneau (1962) erred seriously in basing his assessment of the relative power of the two tests in question on experiments that employed, for the most part, small sample sizes ($n_1 = n_2 = 5$). Boneau (1962) concluded on the basis of his small sample studies that the *t* test tends to be more powerful than Wilcoxon's test in the nonnormal case and implied, by his conclusion, that asymptotic relative efficiencies may not be useful in the case of finite sample sizes. It should be noted that

Boneau's (1962) conclusions are quite contrary to those reached here but are in accord with the conclusions that would have been reached if this study had considered only small samples.

Finally, the conclusion that much of the conventional wisdom related to this topic is flawed leads to the further conclusion that much more research is needed.

REFERENCES

Allport, F. H. The J-curve hypothesis of conforming behavior. Journal of Social Psychology, 1934, 5, 141-183.

Blair, R. C., Higgins, J. J., & Smitley, W. D. S. On the relative power of the U and *t*-tests. British Journal of Mathematical and Statistical Psychology, 1980, 33(1), 114-120.

Boneau, C. A. The effects of violations of assumptions underlying the *t*-test. Psychological Bulletin, 1960, 57, 49-64.

Boneau, C. A. A comparison of the power of the *U* and t-tests. Psychological Review, 1962, 69, 246-256.

Bradley, J. V. Distribution-free statistical tests. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Bradley, J. V. Nonparametric statistics. In R. E. Kirk (Ed.), Statistical issues: A reader for the behavioral sciences. Monterey, Calif.: Brooks/Cole, 1972.

Bradley, J. V. A common situation conducive to bizarre distributional shapes. The American Statistician, 1977, 31, 147-150.

Bradley, J. V. Robustness? British Journal of Mathematical and Statistical Psychology, 1978, 31, 144-152.

Dixon, W. J. Power under normality of several nonparametric tests. Annals of Mathematical Statistics, 1954, 25, 610-614.

Gay, L. R. Educational research: Competencies for analysis and application. Columbus, Ohio: Charles E. Merrill, 1976.

Glass, G. V , Peckham, P. D., & Sanders, J. R. Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. Review of Educational Research, 1972, 42, 237-288.

Guilford, J. P., & Fruchter, B. Fundamental statistics in psychology and education. New York: McGraw-Hill, 1978.

Hodges, J. L., & Lehmann, E. L. The efficiency of some nonparametric competitors of the t-test. Annals of Mathematical Statistics, 1956, 27, 324-335.

International Mathematical and Statistical Libraries. Houston, Tex.: International Mathematical and Statistical Libraries, Inc., 1977.

Kerlinger, F. N. Foundations of behavioral research (2nd ed.). New York: Holt, Rinehart, & Winston, 1973.

Lehmann, E. L. Nonparametrics. San Francisco: Holden-Day, 1975.

Neave, H. R., & Granger, C. W. J. A Monte Carlo study comparing various two-sample tests for differences in mean. Technometrics, 1968, 10, 509-522.

Neyman, J., & Pearson, E. S. On the problem of the most efficient tests of statistical hypotheses. Transactions of the Royal Society of London, Series A, 1933, 231, 289-337.

Popham, W. J., & Sirotnik, K. A. Educational statistics use and interpretation (2nd ed.). New York: Harper & Row, 1973.

Toothaker, L. E. An empirical investigation of the permutation t-test as compared to Student's t-test and the Mann-Whitney U-test. Madison, Wis.: Wisconsin Research and Development Center for Cognitive Learning, 1972.

AUTHORS

Blair, R. Clifford. Address: College of Education, University of South Florida, Tampa, FL 33620. Title: Assistant Professor. Degrees: B.A., M.A., Ph.D., University of South Florida. Specialization: Educational Statistics.

Higgins, James J. Address: Statistics Department, Kansas State University, Manhattan, KS 66506. Title: Associate Professor. Degrees: B.S. University of Illinois, M.S. Illinois State University, Ph.D. University of Missouri. Specialization: Mathematics Statistics.