DOCUMENT RESUME

ED 319 773 TM 014 980

AUTHOR	Luh, Wei-Ming; Olejnik, Stephen
TITLE	Two-Stage Sampling Procedures for Comparing Means When Population Distributions Are Non-Normal.
PUB DATE	Apr 90
NOTE	24p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Comparative Analysis; Mathematical Models; Sample Size; *Sampling; *Statistical Distributions
IDENTIFIERS	*Nonnormal Distributions; Population Parameters; Power (Statistics); Type I Errors

ABSTRACT

Two-stage sampling procedures for comparing two population means when variances are heterogeneous have been developed by D. G. Chapman (1950) and B. K. Ghosh (1975). Both procedures assume sampling from populations that are normally distributed. The present study reports on the effect that sampling from non-normal distributions has on Type I error rates, statistical power, and sample size requirements. Four factors were manipulated in the simulation study: (1) distribution shape; (2) degree of variance heterogeneity; (3) initial sample size; and (4) difference in population means. For each condition, 1,000 replications were performed, and the frequency of rejecting the null hypothesis was recorded. The results indicate that Ghosh's procedure is less sensitive to non-normal distributions, but can be liberal when sampling from distributions that are skewed and have a small initial sample size. Average sample sizes needed remained constant across distribution shapes, but greater variability was found with heavy-tailed distributions. Moderate to large sample sizes at the first stage of sampling can reduce the overall total sample size needed and can minimize the inflated Type I error rate in situations where the sampled distributions are extremely non-normal. (Author/TJH)

Two-Stage Sampling Procedures for Comparing Means When Population Distributions Are Non-Normal

Wei-Ming Luh

Stephen Olejnik

University of Georgia

Paper presented at the American Educational Research Association meeting

April 19, 1990.

ABSTRACT

Two-stage sampling procedures for comparing two population means when variances are heterogeneous have been developed by Chapman (1950) and Ghosh (1975). Both procedures assume sampling from populations that are normally distributed. The present study reports on the effect sampling from non-normal distributions has on the type I error rates, statistical power, and sample size requirements. Four factors were manipulated in the simulation study: (1) distribution shape, (2) degree of variance heterogeneity, (3) initial sample size, and (4) difference in population means. The results indicate that Ghosh's procedure is less sensitive to non-normal distributions but can be liberal when sampling from distributions that are skewed and the initial sample size is small. Average sample sizes needed remain constant across distribution shapes but greater variability was found with heavy-tailed distributions. Moderate to large sample sizes at the first stage of sampling can reduce the overall total sample size needed and can minimize the inflated type I error rate in situations where the sampled distributions are extremely non-normal.

Applied researchers generally rely on Student's t-test or the ANOVA F-test when testing the null hypothesis of equal population means. These procedures assume that the observations are normally and independently distributed with constant variance σ^2. If the assumptions are met, the t and F tests are unbiased and uniformly the most powerful (UMP). However, violating homogeneity of variance assumption can seriously affect the type I error rate and power these procedures (Box, 1954; Glass, Peckham, & Sanders, 1972; Brown & Forsythe, 1974). Violations of the equal variance assumption is particularly serious when sample sizes are unequal but several studies have shown that test of mean equality may not be robust to variance heterogeneity even when sample sizes are equal (Ramsey, 1980; Rogan & Keselman, 1977; Wilcox, Charlin, and Thompson, 1986). The degree of variance heterogeneity, the number of populations being compared, and the sample size are factors affecting the robustness of the tests.

In addition to affecting the type I error rate, variance heterogeneity affects the statistical power of the analysis (Wilcox, et. al., 1986). Power and sample size estimation procedures assume variance equality. When the population variances differ, the actual statistical power can be less than that desired and required sample size may be underestimated.

One solution to the problem of variance heterogeneity has been to modify the test statistics and the corresponding degrees of freedom to consider the degree of variance inequality. Procedures developed by Welch (1937), and Brown and Forsythe (1974) are well-known examples of this approach. They are however only approximate tests since the risk of a type I error will not be exactly equal to the nominal significance level. In addition, Wilcox, Charlin, and Thompson (1986) have reported that these approximate tests still can provide liberal hypothesis tests if the ratio of standard deviations is as large as 4 to 1. Wilcox (1987a) has argued that such differences in variances occur regularly in social science research. A further problem with these procedures is that they do not give researchers precise control over power nor can accurate minimal sample sizes be determined (Wilcox, 1984).

An alternative solution to the heterogeneous variance problem requires sampling observations from the populations of interest in two steps or stages. These procedures require the researcher to randomly select a sample from each population to estimate the variances and then depending on the initial results additional data may be required from a second sample. Population means are estimated using data from both stages but confidence intervals and hypothesis tests use variance estimates from only the first sampling stage. This two-stage sampling approach is a compromise between the fixed sample size techniques like the t-test or ANOVA F-ratio and pure sequential tests such as the sequential probability ratio test (Mood, Graybill & Boes, 1974). Although several two-stage sampling procedures have been developed (Hewett & Spurrier, 1983) not all of the techniques allow the researcher to control statistical power. The present study focuses on two procedures, one developed by Chapman (1950) and the other developed by Ghosh (1975) to compare two population means. Both procedures provide control over the statistical power and both are based on Stein's (1945) two-stage sampling procedure for testing $H_0: \mu = \mu_0$.

Chapman (1950) extended Stein's approach to the estimation of the difference in two population means. His confidence interval is computed as the following:

$$(\tilde{X}_1 - \tilde{X}_2) \pm A.$$

Where: A is the acceptable margin of error (or half width of the confidence interval);

$\tilde{X}_g = (1-b_g k_g)\bar{X}_{1g} + (b_g k_g)\bar{X}_{2g}$, $g=1,2$;

$\bar{X}_{1g}$, $\bar{X}_{2g}$ are the mean observations for group g from the initial sample and second sample respectively, $g=1,2$;

$b_g = \{1+[N_{1g}(N_{Tg}d-s^2_{1g})]^{1/2}/(k_g s^2_{1g})^{1/2}\}/N_{Tg}$, $g=1,2$;

$d = (A/C)^2$ where C is the $1-\alpha/2$ percentage point of the difference between two independent sample t random variables;

N_{1g} is the initial sample size randomly selected from treatment group g, $g=1,2$;

N_{Tg} is the total sample size needed from group g and is determined from: $\max\{N_{1g}+1, (s^2_{1g}/d)^*+1\}$, $g=1,2$;
* indicates the integer value of the ratio,

s^2_{1g} is the unbiased estimate of the sample variance from the initial sample of N_{1g} observations;

k_g is the additional number of observations needed in the second sample ($k_g = N_{Tg}-N_{1g}$), $g=1,2$.

For example, suppose a researcher wanted to compare two treatment conditions. A .95 confidence interval was of interest with a margin of error no greater than 2 points. If an initial sample size of 5 individuals were selected at random from each population with $\bar{X}_1=15$, $s_1=2$ and $\bar{X}_2=12$, $s_2=4$. Then:

C taken from tables provided by Wilcox (1987a) with v_1=4 and v_2=4 equals 3.107;

$$d = (2/3.107)^2$$
$$= .4144$$

$$N_{T1} = \max\{5+1, [4/.4144]^*+1\}$$
$$= 10$$

$$N_{T2} = \max\{5+1, [16/.4144]^*+1\}$$
$$= 39$$

$$k_1 = 5, \; k_2 = 34$$

$$b_1 = \{1+[5((10)(.4144)-4)]^{1/2}/[(5)(4)]^{1/2}\}/10$$
$$= .119$$

$$b_2 = \{1+[5((39)(.4144)-16)]^{1/2}/[(34)(16)]^{1/2}\}/39$$
$$= .0266.$$

Suppose the mean of the additional 5 subjects from treatment group 1 equalled 14 and the mean of the additional 34 subjects from treatment group 2 equalled 13. The confidence interval would then be computed as:

$$\tilde{X}_1 = [1-(.119)(5)]15 + (.119)(5)14$$
$$= 14.405$$

$$\tilde{X}_2 = [1-(.0226)(34)]12 + (.0266)(34)13$$
$$= 12.904$$

$$(14.405 - 12.904) \pm 2$$

$$1.501 \pm 2$$

$$(-.499, 3.501).$$

Thus there would be insufficient evidence to reject the null hypothesis of no difference between treatment population means.

Bishop and Dudewicz (1978) have extended this approach to the single factor multiple group design and Tamhane (1977) has modified Chapman's technique for multiple comparisons analyses in the multiple group design.

Ghosh (1975) approached the problem a little differently. He suggests sampling an equal number of observations from each population of interest and to calculate the confidence interval based on the following formula:

$$(\bar{X}_1 - \bar{X}_2) \pm t_{v,1-\alpha/2}\sqrt{(s_z^2 / N_T)}$$

Where $\bar{X}_g$ are the mean observations across the first and second sampling stages from each treatment population, g=1,2

$t_{v,1-\alpha/2}$ is the 1-α/2 percentile point of Student's t-distribution with v degrees of freedom

v equals the first stage sample size minus 1, N_1-1.

s_z^2 is the variance of the difference scores of the randomly matched samples from the two populations of interest

N_T is the total sample size selected from each population across the first and second stages. It is determined from:

$N_T = \max \{N_1, [(s_z^2/d)^* + 1]\}$,

N_1 is the initial sample size at the first stage of sampling,

d is chosen by such that:

$Pr\{[T<t_{v,1- /2}-(\delta/\sqrt{d})] + [T>t_{v,1-\alpha/2} + (\delta/\sqrt{d})]\} = 1-\beta$;

is the minimal difference between population means thought to be important.

When δ >0 d can be determined by solving for T in $Pr(T>t_{v,1-\alpha/2}-(\delta/\sqrt{d})$ and when δ <0 d can be approximated using $Pr(T>t_{v,1-\alpha/2}+(\delta/\sqrt{d})$ (Wilcox, 1987b).

For example, suppose a researcher is interested in comparing the effectiveness of two treatments and has decided that a difference of 2 points between population means was of practical significance. If the researcher administers each treatment to 6 individuals with the following results:

T_1	T_2	$z=T_1-T_2$
10	12	-2
16	12	4
12	7	5
14	11	3
18	12	6
11	15	-4

The variance of the difference scores equals 4.05. If the significance level is set equal to .05 and power was to equal .8 then for $Pr(T>t-\delta/\sqrt{d})=.8$, $T=-.920$ and d would be determined as:

$$-.920 = 2.571- 2/\sqrt{d}$$

$$d =.328.$$

The total number of subjects needed in the second stage from each population would equal:

$$\max(6,[4.05/.328]^{*}+1)=13.$$

Since 6 observations had been collected, 7 additional observations from each treatment condition is needed. Suppose the mean score from all 13 observations from treatment 1 equalled 13.5 and the mean of the 13 observations from treatment 2 equalled 11.5. Then the confidence interval would be determined as:

$$(13.5 - 11.5) \pm 2.571 \sqrt{4.05/13}$$

$$2 \pm 1.435$$

$$(.564, 3.435)$$

Since 0 is not included in the interval estimate the researcher would conclude that there is sufficient evidence to reject the null hypothesis.

Chapman's procedure has the advantage of allowing sample sizes to be unequal and Ghosh's procedure has the potential advantage of not requiring the researcher to sample additional subjects. Both procedures however assume that populations sampled have normal distributions. Wilcox (1985) has examined the effect non-normality has on Chapman's procedure when it is applied to the single factor multiple groups design as suggested by Bishop and Dudewicz (1978). The results of his study indicated that small departures from normality has little effect on type I errors or statistical power. With large departures from normality the type I error rate can be inflated. Increasing the initial sample size can reduce the effect of non-normality but cannot totally eliminate a liberal test in extreme non-normal distributions.

The purpose of the present study is to examine the effect sampling from non-normal populations has on Ghosh's procedure in terms of type I error rates and statistical power. A second purpose is to study the effect non-normality has on total sample sizes needed in both Ghosh and Chapman's procedure. Finally, the initial sample size used in the first stage is studied as a moderator variable to determine the influence the initial sample size has on type I errors, statistical power, and total sample size.

Method

Computer simulation technique was employed. The data were generated and analyzed by the SAS Proc Matrix routine (1985). Four conditions were manipulated: (1) initial sample sizes, (2) population variances, (3) distribution form, and (4) difference in population means. A summary of the first three factors is provided in Table 1. All combinations of initial sample size, variances, and distribution forms were included. Population means either did not differ, in order to study type I error rates or they differed by one point. For the variance conditions studied, the one point mean difference corresponds to a difference of one pooled standard deviation. The level of statistical significance for all analyses was set equal to .05. Finally two levels of statistical power were considered, power of .8 and .9.

Normal random variables were generated with a mean of zero and variance of one by using the normal random number generating function (RANNOR) in SAS. The grand mean was set equal to 10. Non-normal distributions were created by using a polynomial transformation suggested by Fleishman (1978):

$$Y = a + bx + cx^2 + dx^3 ,$$

where $x \sim N(0,1)$, and a, b, c, and d modify the skewness and kurtosis when mean and variance are unchanged.

For each condition, 1000 replications were performed and the frequency of rejecting the null hypothesis was recorded. In addition, the average, minimum, maximum, and the variance of the required sample sizes were recorded. Empirical estimates of Type I error rates above two standard errors from the nominal significance level are considered as evidence that the procedures are not robust to the normality assumption.

Results

Type I Error. Stem and leaf plots are reported in Table 2 for the empirical type I error rates obtained for Chapman and Ghosh's procedures. The 96 values reported are collapsed across all distribution conditions considered and the desired power levels of .8 and .9. When initial sample size was small both procedures had fairly sizable number of conditions where the type I error rate was unacceptably high (greater than .063). The increased type I error rate for Ghosh's procedure was found primarily in skewed distributions. Twenty-two of the 25 empirical rejection rates were obtained from distributions that had skew of at least .75. Chapman's procedure on the other hand, had an unacceptable rate of rejection when distributions were leptokurtic. Forty-nine of the 50 unacceptable rejection rates were obtained from distributions that had kurtosis of at least 2.75.

When the initial sample size was increased to 16 the number of conditions where unacceptable rejection rates observed was reduced in half (11 conditions) for Ghosh's procedure. Again it was extreme skewness which affected the type I error rate. Chapman's procedure had 36 conditions where the type I error rate was unacceptable. Leptokurtic distributions were the primary source of the inflated type I errors.

Although not reported in Table 2, when the initial sample size was increased to 31, only 4 conditions had unacceptable type I error rates for Ghosh's procedure and they were scattered across all distribution shapes and variance ratios. Chapman's procedure resulted in 14 conditions where the type I error rate was unacceptably high and they found with extreme leptokurtic distributions.

Statistical Power. Table 3 resents the empirical power estimates for both techniques when the theoretical power was .8 and .9. Neither non-normality nor variance inequality seriously affected the statistical power of the tests. It might be noted that because of the way sample size was determined for Ghosh's test, statistical power was expected to be slightly greater than the nominal value. Chapman's procedure allows for the exact control of statistical power, so we anticipated the power estimates to be .8 or .9.

Sample Size. The average total sample size needed remained constant across distribution shapes and both procedures required approximately the same total number of observations. Table 4 presents the average, minimum, and maximum sample size needed across the eight distribution forms studied when power equalled .9 and .8. The values listed for Ghosh's procedure are the number of observations needed from each population. When the initial sample size equalled 6 and desired power was .9, the sample size need for the entire study was approximately 68.

The values listed for Chapman's procedure are the number of observations needed from the population with the greatest variability. To determine the total sample size needed for Chapman's procedure it would be necessary to add the number of observations needed from the less variable population. So when the population variances were equal to 1, for example, Chapman's procedure would require approximately 34 observations from each population for a total sample size of 68 when the initial sample size was 6 and desired power equalled .9. If the population variances differed by a ratio of 11 to 1, approximately 63 observations would be needed from the more variable population but the initial 6 observations would be sufficient data for the less variable population. The total sample size across both groups would then equal approximately 69 observations. For different variance ratios the total sample size needed remained constant at approximately 68 observations.

As the initial sample size increased to 16, the total sample size needed decreased by almost 25 percent when power equalled .9 and by almost 20 percent when power equalled .8. But if the initial sample size equalled 31 no additional observations were needed for Ghosh's procedure for the given effect size and power requirements.

In addition to the average sample size needed across distribution forms, the sample size variability within a distribution is of some interest. Values reported in table 5 are the standard deviations of the distribution of sample sizes across the 1000 replications for each distribution form. Distributions 4 through 8 were heavy-tailed distributions. Distribution 2 was the light-tailed distribution. These results combined with those reported in table 4 indicate that although the average sample size needed remained constant across distribution form, the sample size needed are less stable with heavy tailed distributions than normal or light-tailed distributions.

It also appears that with Chapman's procedure, the stability of the sample sizes needed is a function of the variance ratio. However, as the population variance increases the sample size needed also increases. When this factor is considered by examining the coefficient of variation, variance ratio does not appear to appreciably affect the stability of the needed sample size.

Conclusions

The two-stage sampling procedures developed by Chapman (1950) and Ghosh (1975) provide valid tests for comparing population means with specified statistical power when variances differ but only when population distributions are normal or do not depart greatly from normality. Inflated type I error rates can result when the initial sample size is small and the population distributions are moderately non-normal. Distribution shape however affects the two procedures differently. Chapman's procedure appears to be affected by heavy-tailed distributions. Even when the initial sample size was large, the type I error rates were inflated. Our results are consistent with those reported by Wilcox (1985) which showed that the two-stage procedure developed by Bishop and Dudewicz's (1978), based on Chapman's approach, can provide a liberal hypothesis test when populations are leptokurtic. The present results which show that Wilcox's findings generalize down to two group studies have important implications for the two-stage multiple comparison procedure developed by Tamhane (1978). Our results indicate that Tamhane's procedure many not be appropriate with heavy-tailed distributions since it is also based on Chapman's approach.

Non-normality also affects Ghosh's procedure, but not to the same degree as Chapman's. In addition the type of non-normality which affect's Ghosh's test differs from Chapman's. Asymmetric distributions led to inflated type I error rates for Ghosh's procedure but the problem could be minimized by increasing the initial sample size. This result suggests that a two-stage multiple comparison test based on Ghosh's technique may be a useful development.

Statistical power for the two procedures was not reduced as a result of sampling from non-normal distributions. And perhaps surprisingly, the total sample size needed was not affected by the distribution form. Distribution form however did affect the stability of the required sample sizes. With heavy-tailed distributions the variability of sample sizes was greater than with normal or light-tailed distributions.

We compared the average sample size needed in the present simulation study with the required sample size needed for the independent samples t-test when the populations sampled are normal and have equal variances. For the effect size considered in the present study (1 standard deviation unit), the independent sample t-test would require 22 and 17 observations per group for power of .9 and .8 respectively (Cohen, 1977). The Ghosh procedure required an average of 33.5 and 25.7 observations per group when initial sample size was 6. If the initial sample size was 16 the average sample size needed was 25.4 and 20.6 for .9 and .8 power respectively. The two-stage procedure typically requires more observations than the independent samples t-test but as the initial sample size increases the difference in sample size requirements diminishes. We have examined this relationship for different effect sizes and found it consistent. Additional analyses comparing the two-stage procedures with the independent samples t-test may be a useful exercise.

Two-stage sampling procedures have been available for some time but have rarely been used by researchers in the social sciences. Only recently (Wilcox, 1987b) have they been presented in textbooks used by students in introductory or intermediate statistical methods classes. Although there may be some practical restrictions in applying the procedures, the results of the present study as well as others indicate that the techniques can be useful. As with all statistical procedures the indiscriminate use of the techniques can result in invalid conclusions. Variance heterogeneity may no longer be a problem but distribution form should be considered carefully.

References

Bishop, T. A. & Dudewicz, E. J. (1978). Exact analysis of variance with unequal variances: Test procedures and tables. Technometrics, 20, 419-430.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. effect of inequality of variance in the one-way classification. Analysis of Mathematical Statistics, 25, 290-302.

Brown, M. B., & Forsythe, A. B. (1974). The small sample behavior of some statistics which test the equality of several means. Technometrics, 16, 129-132.

Chapman, D. G. (1950). Some two-sample tests. Annals of Mathematical Statistics, 21, 601-606.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Fleishman, A. I. (1978). A method for simulating non-normal distribution. Psychometrika, 43, 521-532.

Ghosh, B. K. (1975). A two-stage procedure for the Behrens-Fisher Problem. Journal of the American Statistical Association, 70, 457-462.

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

Hewett, J., & Spurrier, J. (1983). A survey of two-stage tests of Hypotheses: Theory and application. Communications in Statistics--Theory and Methods, 12, 2307-2435.

Mood, A. M., Graybill, F. A. & Boes, D. C. (1974). Introduction to the Theory of Statistics (3rd ed.). New York: McGraw-Hill.

Ramsey, P. H. (1980). Exact type I error rates for robustness of Student's t-test with unequal variances. Journal of Educational Statistics, 5, 337-349.

Rogan J. C. & Keselman, H. J. (1977). Is the ANOVA F-test robust to variance heterogeneity when sample sizes are equal?: An investigation via a coefficient of variation. American Educational Research Journal, 14, 493-498.

SAS Institute (1985). SAS user's Guide: Statistics. Cary, NC: Author.

Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. Annals of Mathematical Statistics, 16, 243-258.

Tamhane, A. C. (1977). Multiple comparisons in model I one-way ANOVA with unequal variances. Communications in Statistics- Theory and Methods, A6, 15-32.

Welch, B. L. (1937). The significance of the difference between two means when the population variances are unequal. Biometrika, 29, 350-362.

Wilcox, R. R. (1984). A review of exact hypothesis testing procedures (and selection techniques) that control power regardless of the variances. British Journal of Mathematical and Statistical Psychology, 37, 34-48.

Wilcox, R. R. (1985). The effects of skewness and kurtosis on three heteroscedastic ANOVA procedures that control power. Journal of Organizational Behavior and Statistics, 2, 29-36.

Wilcox, R. R. (1987a). New designs in analysis of variance. Annual Review of Psychology, 38, 29-60.

Wilcox, R. R. (1987b). New statistical procedures for the social sciences: Modern solutions to basic problems. New Jersey: Lawrence Erlbaum.

Wilcox, R., Charlin, V., & Thompson, K. (1986). New Monte Carlo results on the robustness of the ANOVA F, W and F* statistics. Communications in Statistics--Simulation and Computation, 15, 933-944.

Table 1.

Summary of conditions investigated.

Initial Sample N	Variance ratio	σ_1^2/σ_2^2	Form Skew	Kurtosis	Distribution	
6	1:1	1/1	0.0	0.0	Normal	1
16	1:3	.5/1.5	0.0	-1.0	Platykurtic	2
31	1:7	.25/1.75	0.75	0.00	Skewed	3
	1:11	.167/1.837	0.0	2.75	Leptokurtic	4
	1:15	.125/1.875	0.0	3.75	Leptokurtic	5
	1:19	.1/1.9	0.75	3.75	Skewed/Leptokurtic	6
			1.25	3.75	Skewed/Leptokurtic	7
			1.75	3.75	Skewed/Leptokurtic	8

Table 2.

Stem-and-leaf plot of empirical type I error rates across different distribution shapes and variance ratios for Ghosh and Chapman's two-stage tests when α=.05, 1- =.8 and .9

Chapman		Ghosh
	unit=0.001, 51 represents 0.051	
N_1=6		
	12*	013
9	11.	5
31	11*	13
99887	10.	5
2	10*	
	9.	5
221	9*	
75	8.	
4300	8*	
99877666	7.	567
443211111	7*	1113
9988886655555	6.	66777888
443322100	6*	000012233344
9988777666555	5.	5555567777778999
4431111	5*	0000001122233344444
99996555	4.	5555666667777788899
4333220	4*	0122
9986	3.	9
	3*	
N_1=16		
	9.	
3	9*	
966	8.	6
200	8*	023
9865	7.	9
433000000	7*	0
998877776655555	6.	6899
422111000	6*	01111222234
99999999888887766666	5.	5556666777888899
4444322221111100	5*	0111222233334444
999887	4.	555555666677788999
42200	4*	112222233344444
999	3.	7889
4	3*	0134
	2.	9
	2*	

Note. Values included in the leaf are rounded to the thousandths place.

Table 3.

Empirical power for Ghosh and Chapman's two-stage tests when initial sample sizes are six per group and nominal power equals .9 and .8

Test	Distribution	Variance Ratio 1:1	1:3	1:7	1:11	1:15	1:19
		Power = .90					
Ghosh							
	Normal (0,0)	.91	.90	.91	.92	.92	.91
	Platykurtic (0,-1)	.93	.93	.92	.91	.92	.90
	Skewed (0.75, 0)	.90	.92	.93	.94	.95	.94
	Leptokurtic (0,2.75)	.91	.92	.91	.91	.92	.92
	Leptokurtic (0,3.75)	.93	.93	.91	.92	.91	.93
	Skew/Leptokurtic (.75, 3.75)	.92	.92	.92	.93	.92	.94
	Skew/Leptokurtic (1.25, 3.75)	.91	.93	.93	.93	.9[illegible]	.93
	Skew/Leptokurtic (1.75, 3.75)	.93	.95	.94	.95	.95	.96
Chapman							
	Normal (0,0)	.89	.90	.92	.90	.90	.90
	Platykurtic (0,-1)	.90	.91	.92	.90	.91	.90
	Skewed (0.75, 0)	.89	.91	.91	.91	.92	.92
	Leptokurtic (0,2.75)	.88	.90	.87	.90	.89	.90
	Leptokurtic (0,3.75)	.89	.90	.89	.91	.89	.89
	Skew/Leptokurtic (0.75, 3.75)	.89	.87	.90	.89	.89	.89
	Skew/Leptokurtic (1.25, 3.75)	.86	.88	.89	.91	.90	.89
	Skew/Leptokurtic (1.75, 3.75)	.85	.88	.91	.93	.92	.93
		Power = .80					
Ghosh							
	Normal (0,0)	.83	.82	.83	.80	.85	.82
	Platykurtic (0,-1)	.82	.84	.84	.80	.82	.83
	Skewed (0.75, 0)	.83	.86	.86	.85	.86	.84
	Leptokurtic (0,2.75)	.84	.83	.83	.84	.84	.84
	Leptokurtic (0,3.75)	.83	.84	.85	.84	.85	.82
	Skew/Leptokurtic (0.75, 3.75)	.83	.87	.86	.84	.84	.88
	Skew/Leptokurtic (1.25, 3.75)	.82	.85	.86	.85	.89	.88
	Skew/Leptokurtic (1.75, 3.75)	.85	.86	.88	.88	.87	.88
Chapman							
	Normal (0,0)	.82	.83	.81	.81	.81	.81
	Platykurtic (0,-1)	.80	.83	.81	.82	.84	.83
	Skewed (0.75, 0)	.79	.80	.85	.83	.85	.83
	Leptokurtic (0,2.75)	.81	.81	.80	.79	.80	.82
	Leptokurtic (0,3.75)	.79	.80	.78	.81	.81	.73
	Skew/Leptokurtic (0.75, 3.75)	.81	.81	.84	.82	.82	.82
	Skew/Leptokurtic (1.25, 3.75)	.78	.80	.82	.83	.84	.84
	Skew/Leptokurtic (1.75, 3.75)	.78	.79	.83	.85	.87	.83

Note: Power estimates are rounded to the hundredths place.

Table 4.

Average, minimum, and maximum mean of sample size requirements across different variance ratios for Ghosh's test and on the group with large variance for Chapman's test across all distribution forms.

			Power=.9			Power=.8		
			Ave	Min	Max	Ave	Min	Max
$N_1=6$	Ghosh		33.5	33.0	33.9	25.4	25.1	25.9
	Chapman	1:1	34.1	32.2	35.2	26.6	25.9	28.2
		1:3	50.7	48.0	51.9	39.6	37.5	42.4
		1:7	59.3	57.9	61.3	45.7	43.5	48.5
		1:11	62.9	59.9	65.1	48.3	46.0	51.4
$N_1=16$	Ghosh		25.7	25.3	26.1	20.6	19.8	21.2
	Chapman	1:1	26.6	25.7	27.0	23.1	21.9	23.9
		1:3	39.6	35.5	38.3	32.1	31.8	32.7
		1:7	45.7	42.3	44.1	36.7	35.1	37.5
		1:11	48.7	44.7	46.7	38.5	38.2	39.5

Table 5.

Standard deviation of sample size requirements across different variance ratios for Ghosh's test and on the group with large variance for Chapman's two-stage test.

				Distribution[a]							
Power	N_1	Test		1	2	3	4	5	6	7	8
.90	6	Ghosh		21[b]	17	21	28	30	30	29	30
		Chapman	1:1	22	17	20	26	33	29	36	30
			1:3	32	24	32	47	46	48	51	48
			1:7	38	28	38	55	56	60	57	61
			1:11	38	29	40	61	68	57	64	54
	16	Ghosh		8	7	7	10	12	12	12	12
		Chapman	1:1	8	6	8	12	14	14	14	13
			1:3	13	10	13	18	21	20	21	22
			1:7	15	11	16	24	25	23	25	24
			1:11	16	12	16	23	24	27	29	27
.80	6	Ghosh		15	13	15	20	23	22	23	22
		Chapman	1:1	16	12	12	22	28	27	25	26
			1:3	25	18	23	32	38	39	40	37
			1:7	29	22	27	40	42	38	48	43
			1:11	30	23	29	44	50	48	43	51
	16	Ghosh		5	4	5	7	8	9	8	8
		Chapman	1:1	6	5	6	9	11	11	11	11
			1:3	11	8	11	16	19	19	19	18
			1:7	13	10	13	18	21	21	21	20
			1:11	14	10	14	21	23	22	21	23

Note. [a]1. Normal (0,0), 2. Platykurtic (0,-1), 3. Skewed (0.75, 0), 4. Leptokurtic (0,2.75), 5. Leptokurtic (0,3.75), 6. S/L (0.75, 3.75), 7. S/L (1.25, 3.75), 8. S/L (1.75, 3.75).

[b]Numbers are rounded.