DOCUMENT RESUME

ED 062 388 TM 001 326

AUTHOR Ramseyer, Gary C.; Tcheng, Tse-Kia

TITLE The Robustness of the Studentized Range Statistic to Violations of the Normality and Homogeneity of Variance Assumptions.

PUB DATE Apr 72

NOTE 14p.; Paper presented at the annual meeting of the American Educational Research Association (Chicago, Ill., April 1972)

EDRS PRICE MF-$0.65 HC-$3.29

DESCRIPTORS Analysis of Variance; Confidence Testing; *Hypothesis Testing; *Statistical Analysis; *Statistics; *Tests of Significance

IDENTIFIERS *q Statistic; Type I Error

ABSTRACT

The present study was directed at determining the extent to which the Type I Error rate is affected by violations in the basic assumptions of the q statistic. Monte Carlo methods were employed, and a variety of departures from the assumptions were examined. (Author)

THE ROBUSTNESS OF THE STUDENTIZED RANGE STATISTIC TO VIOLATIONS OF THE NORMALITY AND HOMOGENEITY OF VARIANCE ASSUMPTIONS[1]

Gary C. Ramseyer and Tse-Kia Tcheng

Illinois State University

Multiple comparison procedures in recent years have earned a prominent role in the analysis and interpretation of experimental research in the behavioral sciences. Most of these procedures are designed to either test individual contrasts between means after the null hypothesis of no treatment differences in ANOVA has been rejected or to test a selected set of mean contrasts which are of apriori interest to an investigator in an experiment. Three popular techniques which have primarily been employed for the first purpose are the Tukey WSD method (1953), the Newman-Keuls test (Keuls, 1952; Newman, 1939) and the Duncan multiple range test (1955). All of these tests have as their parent statistic the studentized range statistic q (Pearson & Hartley, 1943; Student, 1927) defined by

$$q\ (k,\ df_2) = \frac{\overline{X}_L - \overline{X}_S}{\sqrt{\frac{ms_w}{n}}}$$

where $\overline{X}_L$ = the largest of a set of k group means

$\overline{X}_S$ = the smallest of a set of k group means

df_2 = the degrees of freedom for ms_w

n = the sample size for each group

This statistic is distributed exactly as q with parameters k and df_2 if the following assumptions are satisfied: (1) The overall null hypothesis $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ is true (2) Samples are independently selected at random (3) Populations are normally distributed and (4) Populations are equally variable.

[1]Paper presented at the annual meeting of the American Educational Research Association, Chicago, April 1972.

It is generally conceded that the q statistic is less powerful overall than the corresponding F statistic (Winer, 1962), but this finding assumes normal distributions with equal variances. Surprisingly, when the above assumptions are violated the robustness of q with respect to both power and Type I error is relatively unknown (Games, 1971). Petrinovich and Hardyck (1969) do offer limited evidence that q is robust under either non-normality or unequal variances but their work was restricted to exponential populations and did not consider various simultaneous violations of the two assumptions. In view of the wealth of studies available that confirm the robustness of the t and F statistics (See for example Boneau, 1960; Box, 1954; Donaldson, 1968; Norton, 1952) and the extensive usage of the q statistic in conjunction with multiple comparisons in ANOVA, it would appear that empirical investigations of the latter statistic are long overdue. The present study was therefore directed at determining the extent to which Type I error rate is affected by violations in the basic assumptions of the q statistic. Monte Carlo methods were employed and a variety of departures from the assumptions were examined.

Method

First, a sampling distribution of q with the assumptions inviolate {i.e., populations sampled were normally distributed with mean 0 and variance 1 denoted by N(0,1)} was simulated on an IBM 360/50 computer by generating 2000 values of the statistic. This was done four times using initially 3 groups (k=3) with 5 scores in each group (n=5) and then the following three pairings: k=3, n=15; k=5, n=5; k=5, n=15. These four combinations furnished a fairly representative set of df_2-values ranging from a rather small value of 12 to a moderately large value of 70. In each set of 2000 values the percentage of q's exceeding the theoretical tabled 95th and 99th percentiles for the appropriate k and df_2 were determined. Since all the assumptions were satisfied, the long run expected values of these observed percentages were 5% and 1% respectively. Hence, these observed quantities served as a valuable indicator of the pure chance discrepancy one could expect between the nominal and obtained Type I error rates. Boneau (1960) in his study on t reported that discrepancies as large as 1% above or below the nominal 5% error rate are not uncommon when 2000 values of the statistic were calculated with all assumptions satisfied.

Next, the variance assumption was violated. This was accomplished for k=3 and n=5 by generating 2000 q's based on normally distributed populations with means of 0 and variances of 1,1, and 2. This represented a rather moderate departure from the variance assumption and certainly one that would be encountered quite frequently in the behavioral sciences. The procedure was then repeated with variances of 1,1, and 4 (a rather extreme violation). Additional sampling distributions of q were generated blending similar variance violations with the other three combinations of k and n. For example, when k=5 and n=5 the variances used were 1,1,1,2,2 and 1,1,1,4,4. In all situations, the nominal and observed error rates were compared.

The normality assumption was then violated. For this phase of the study three distributions in standardized form were employed as populations: the positively skewed exponential, the negatively skewed exponential, and the rectangular {i.e., $E^+(0,1)$, $E^-(0,1)$ and $R(0,1)$}. In order to generate random numbers distributed according to the above characteristics, the computer first sampled from the rectangular distribution of the random variable r in the interval from 0 to 1. These results were then converted to the desired variates by the following transformations:

$$x = -\ln r - 1 \qquad \text{for } E^+(0,1)$$

$$y = \ln r + 1 \qquad \text{for } E^-(0,1)$$

$$z = \frac{r - .5}{\sqrt{\frac{1}{12}}} \qquad \text{for } R(0,1)$$

where ln r = the natural logarithm of r

x, y, and z = the desired standardized variates

Sets of 2000 q's were generated and error rates were compared for situations in which the populations were all $E^+(0,1)$ or all $R(0,1)$ under each of the four k and n combinations. Other sampling distributions were produced using distributions that were not all identical as the underlying populations. That is, $N(0,1)$, $E^+(0,1)$ and $R(0,1)$ were introduced together as an underlying population pattern and $E^+(0,1)$, $E^+(0,1)$ and $E^-(0,1)$ were introduced as another pattern. While the occurrence of this latter configuration in practice would indeed be rare, it was nevertheless included for the intrinsic purpose of exploring the effects of oppositely skewed distributions.

In the final phase of the study the variance and normality assumptions were violated simultaneously in a multitude of ways and the error rates were compared. Particular importance was attached to this segment since simultaneous violations are the rule rather than the exception in the real world. The number of different possible violations under these conditions, however, could easily have become unmanageable. Thus, only situations were considered that incorporated the extreme variances of 1,1, and 4 (or 1,1,1,4, and 4) into the population patterns of the preceding phase of the study.

Results

When the assumptions were satisfied, the observed Type I error rates for the nominal 5% level were 5.1%, 5.9%, 5.2% and 4.8% respectively for the four conditions $k=3$, $n=5$; $k=3$, $n=15$; $k=5$, $n=5$; $k=5$, $n=15$. The 1% error rates were 1.3%, 1.1%, 1.6% and 1.0% respectively. The error rate of 5.9% for $k=5$ and $n=5$ would seem to confirm Boneau's statement (1960) that observed rates may deviate as much as 1% from the nominal 5% value when the assumptions are fulfilled. All in all, however, these results not only justify the random sampling procedure used but reaffirm one's faith in the mathematically determined tabled values of q.

Table 1 presents the observed 5% and 1% error rates under violations of the variance assumption. Introduction of the moderate violation of 1,1, and 2 (or 1,1,1,2, and 2 when k=5) had no distinguishable effect on the observed error rates. The extreme variance violation of 1,1, and 4 (or 1,1,1,4, and 4 when k=5) produced 5% rates ranging from a low of 5.9% to a high of 6.9% for the four k and n conditions. The 1% rates ranged from 1.6% to 2.0%. It thus appears that a violation this severe may typically only produce increments as high as 2% and 1% above the nominal 5% and 1% levels respectively.

When the populations were equally variable but all positively skewed exponentials or all rectangular the observed error rates for the most part dropped slightly below the nominal rates. Table 2 indicates that the 5% rates ranged from 3.8% to 4.5% for the exponential populations under the four sampling conditions and from 4.2% to 5.5% for rectangular populations. Similarly, the 1% rates varied from .5% to .9% for exponential populations and from .8% to 1.3% for rectangular populations. Hence, these particular identical non-normal populations seem to have a negligible effect on the Type I error rate.

Table 2 also reports the Type I error rates when non-identical distributions were sampled. For the patterns involving $N(0,1)$, $R(0,1)$ and $E^{+}(0,1)$ the 5% rates ranged from 4.2% to 4.6% and the 1% rates were from .7% to .8%. These values again are systematically below the nominal values but represent very mild departures from expectation. Introduction of oppositely skewed distributions {i.e., patterns involving $E^{+}(0,1)$, $E^{+}(0,1)$ and $E^{-}(0,1)$} produced rather surprising results. In all four sampling conditions, the observed rates were below the nominal rates but the smallest 5% rate was 3.4% and the smallest 1% rate was .6%. Intuitively, one would expect this type of normality violation to have a far greater effect on Type I error rate.

The observed Type I error rates for a variety of simultaneous violations of the normality and variance assumptions are given in Table 3. Since the variance violation of 1,1, and 2 produced rates almost identical to those obtained when the assumptions were satisfied, only the extreme variance violation of 1,1, and 4 (or 1,1,1,4,4 when k=5) was considered in this phase. When the population patterns exemplified by $E^+(0,1)$, $E^+(0,1)$ and $E^+(0,4)$ were used, the 5% rates ranged from 6.9% to 8.2% and the 1% rates from 2.0% to 2.3%. The patterns characterized by R(0,1), R(0,1) and R(0,4) yielded rates from 6.1% to 8.2% for the 5% level and from 1.5% to 2.9% for the 1% level. Thus distributions that are all exponential or all rectangular under the extreme variance violation appear to generate Type I error rates that reach at most only the 8% and 3% neighborhoods for the nominal 5% and 1% levels respectively.

Fourteen situations were examined that involved the extreme variance violation with non-identical populations. When the normal, exponential, and rectangular distributions were used within the same pattern (six situations in Table 3), the maximum observed rates were 7.7% and 2.7% respectively for the 5% and 1% levels. Except for two notable exceptions, the eight situations involving oppositely skewed exponentials within the same pattern produced parallel results. The two exceptions {i.e., $E^+(0,1)$, $E^+(0,4)$ and $E^-(0,4)$ for n=5 and n=15} resulted in observed rates that were surprisingly close to their nominal values. This occurrence so amazed the authors that both situations were rerun on the computer. The second run produced 5% rates of 5.4% and 5.7% respectively for the situations and 1% rates of 1.4% and 1.2% respectively. Hence, the original results appear to be no fluk or quirk of chance. It should be pointed out that these two situations actually arose by accident. The variances of 1,4, and 4 for the respective populations was intended to be 1,1, and 4 which, of course, was routinely used throughout the study. The former set of variances essentially reflects the same degree of departure from the assumption but when combined with the given population sequence produces two oppositely skewed distributions with the same variance. The latter set of variances, on the other hand, results in two oppositely skewed distributions with different variances.

Finally, an attempt was made to assess the role of the sample size (n) and the number of groups (k) in the distortion of Type I error rate under the various violations. Two trends were noticeable when all 46 situations of the three tables were examined. In the cases in which three populations were sampled (k=3), a situation with a sample size of 5 tended to produce a larger deviation from the 5% nominal rate than a corresponding situation with a sample size of 15. Also when the cases involving a sample size of 15 were considered, a situation involving 5 populations tended to produce a larger deviation from the 5% nominal rate than a corresponding situation involving 3 populations. No trends were discernable at the 1% level.

Discussion

Multiple comparison techniques based on the studentized range statistic currently enjoy intuitive appeal among research practitioners in the behavioral sciences. The present study has unveiled yet another attractive property. It appears that q, like t and F, withstands remarkably well violations of the homogeneity of variance and normality assumptions when Type I error rate is the criterion. The extreme variances of 1,1, and 4 for normal populations (Table 1) produced error rates up to only 6.9% and 2.0% for the nominal 5% and 1% levels respectively. Violations of only the normality assumption using exponential and rectangular distributions (Table 2) resulted in rates systematically but negligibly below the nominal levels. In the 16 situations considered in this phase, the smallest observed error rates were 3.4% and .5% for the nominal 5% and 1% levels respectively. Twenty-two simultaneous violations of both assumptions (Table 3) led to maximum rates of 8.2% and 2.9% respectively.

The two exceptional situations of Table 3 (i.e., $E^+(0,1)$, $E^+(0,4)$ and $E^-(0,4)$ for n=5 and n=15) are worthy of additional comment. As indicated in the table and further supported by replication, the observed error rates associated with these situations were very close to the nominal 5% and 1% levels. The cause of this strange occurrence is open to speculation. One possible explanation lies in the opposing forces that are operating in these violations. That is, oppositely skewed exponentials depress the error rate and unequal variances elevate the rate. When this phenomenon is considered along with a coincidental blend of the particular variance magnitudes and the placement of the two equal variances in the oppositely skewed distributions, it is conceivable that some sort of rare balance was achieved. Some support for this conjecture was gained when another violation was constructed which incorporated an even more extreme variance set of 1,9, and 9 into the same distributions. Here for n=5, the observed rates jumped to 8.5% and 3.0% for the 5% and 1% nominal levels respectively. In this case, it appears that the severity of the variance violation has overwhelmed the combined effect of the other forces. For comparative purposes, the variances 1,1, and 9 were employed with the same distributions using n=5 (an equivalent variance violation with the oppositely skewed distributions having different variances). The resulting error rates were much larger -- 11.7% and 4.9% respectively. The principle that emerges from these findings is that when two overall variance violations are equivalent, the presence of two oppositely skewed distributions with <u>equal</u> variances within one pattern represents a less serious violation than the presence of two oppositely skewed distributions with <u>unequal</u> variances within the other pattern. Moreover, this effect seems to be more pronounced for k=3 than for k=5 (see the last four situations in Table 3).

Although this study has investigated quite extensively the robustness of q when Type I error is the criterion, much more research is needed on this popular but little understood statistic. For example, additional work is necessary on the robustness of q when power is the criterion. Also the effect of unequal group sample sizes on Type I error and power needs to be examined. This problem is of prime importance because the assumption of equal n's has always been a serious limitation in the application of the studentized range statistic. Another factor which merits some thought is the effect of kurtosis on the robustness of q. This study did not consider various bell-shaped non-normal distributions with varying degrees of kurtosis.

References

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

Box, G. E. P. 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. Annals of Mathematical Statistics, 1954, 25, 290-302.

Donaldson, T. S. Robustness of the F-test to errors of both kinds and the correlation between the numerator and denominator of the F-ratio. Journal of the American Statistical Association, 1968, 63, 660-676.

Duncan, D. B. Multiple range and multiple F tests. Biometrics, 1955, 11, 1-42.

Games, P. A. Multiple comparisons of means. American Educational Research Journal, 1971, 8, 531-565.

Keuls, M. The use of the "studentized range" in connection with an analysis of variance. Euphytica, 1952, 1, 112-122.

Newman, D. The distribution of range in samples from a normal population, expressed in terms of independent estimate of a standard deviation. Biometrika, 1939, 31, 20-30.

Norton, D. W. An empirical investigation of some effects of non-normality and heterogeneity on the F-distribution. Unpublished doctoral dissertation, State University of Iowa, 1952. Cited by E. E. Lindquist, Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin, 1953.

Pearson, E. S. & Hartley, H. O. Tables of the probability integral of the studentized range. Biometrika, 1943, 33, 89-99.

Petrinovich, L. F. & Hardyck, C. D. Error rates for multiple comparison methods: Some evidence concerning the frequency of erroneous conclusions. Psychological Bulletin, 1969, 71, 43-54.

Student. Errors of routine analysis. Biometrika, 1927, 19, 151-164.

Tukey, J. W. The problem of multiple comparisons. Unpublished manuscript, Princeton University, 1953.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

TABLE 1

Observed Type I Error Rates Under Violations of the Variance Assumption

Sample Conditions	Population Pattern	5% Rate	1% Rate
k=3, n=5	N(0,1); N(0,1); N(0,2)	5.4%	1.2%
k=3, n=15	N(0,1); N(0,1); N(0,2)	5.0%	1.1%
k=5, n=5	three N(0,1); two N(0,2)	6.1%	1.1%
k=5, n=15	three N(0,1); two N(0,2)	5.6%	1.2%
k=3, n=5	N(0,1); N(0,1); N(0,4)	6.5%	2.0%
k=3, n=15	N(0,1); N(0,1); N(0,4)	5.9%	1.6%
k=5, n=5	three N(0,1); two N(0,4)	6.9%	1.8%
k=5, n=15	three N(0,1); two N(0,4)	6.9%	2.0%

TABLE 2

Observed Type I Error Rates Under Violations of the Normality Assumption

Sample Conditions	Population Pattern	5% Rate	1% Rate
k=3, n=5	All populations $E^+(0,1)$	3.8%	.7%
k=3, n=15	All populations $E^+(0,1)$	4.5%	.5%
k=5, n=5	All populations $E^+(0,1)$	4.2%	.9%
k=5, n=15	All populations $E^+(0,1)$	4.5%	.8%
k=3, n=5	All populations R(0,1)	4.8%	1.2%
k=3, n=15	All populations R(0,1)	4.8%	.8%
k=5, n=5	All populations R(0,1)	5.5%	1.3%
k=5, n=15	All populations R(0,1)	4.2%	1.1%
k=3, n=5	N(0,1); R(0,1); $E^+(0,1)$	4.3%	.8%
k=3, n=15	N(0,1); R(0,1); $E^+(0,1)$	4.4%	.8%
k=5, n=5	N(0,1); two R(0,1); two $E^+(0,1)$	4.2%	.8%
k=5, n=15	N(0,1); two R(0,1); two $E^+(0,1)$	4.6%	.7%
k=3, n=5	$E^+(0,1)$; $E^+(0,1)$; $E^-(0,1)$	3.4%	.7%
k=3, n=15	$E^+(0,1)$; $E^+(0,1)$; $E^-(0,1)$	4.6%	.8%
k=5, n=5	three $E^+(0,1)$; two $E^-(0,1)$	4.4%	.9%
k=5, n=15	three $E^+(0,1)$; two $E^-(0,1)$	3.9%	.6%

TABLE 3

Observed Type I Error Rates Under Simultaneous Violations

of the Variance and Normality Assumptions

Sample Conditions	Population Pattern	5% Rate	1% Rate
k=3, n=5	$E^+(0,1)$; $E^+(0,1)$; $E^+(0,4)$	8.1%	2.0%
k=3, n=15	$E^+(0,1)$; $E^+(0,1)$; $E^+(0,4)$	6.9%	2.3%
k=5, n=5	three $E^+(0,1)$; two $E^+(0,4)$	8.2%	2.0%
k=5, n=15	three $E^+(0,1)$; two $E^+(0,4)$	7.4%	2.2%
k=3, n=5	R(0,1); R(0,1); R(0,4)	7.8%	2.6%
k=3, n=15	R(0,1); R(0,1); R(0,4)	6.1%	1.5%
k=5, n=5	three R(0,1); two R(0,4)	8.2%	2.9%
k=5, n=15	three R(0,1); two R(0,4)	7.4%	2.4%
k=3, n=5	N(0,1); R(0,1); $E^+(0,4)$	7.1%	1.8%
k=3, n=15	N(0,1); R(0,1); $E^+(0,4)$	6.8%	2.1%
k=5, n=5	N(0,1); two R(0,1); two $E^+(0,4)$	7.1%	2.1%
k=5, n=15	N(0,1); two R(0,1); two $E^+(0,4)$	7.7%	2.0%
k=5, n=5	N(0,1); R(0,1); R(0,4); $E^+(0,1)$; $E^+(0,4)$	7.2%	2.3%
k=5, n=15	N(0,1); R(0,1); R(0,4); $E^+(0,1)$; $E^+(0,4)$	7.6%	2.7%
k=3, n=5	$E^+(0,1)$; $E^+(0,4)$; $E^-(0,4)$	5.1%	1.5%
k=3, n=15	$E^+(0,1)$; $E^+(0,4)$; $E^-(0,4)$	5.5%	1.2%
k=3, n=5	$E^+(0,1)$; $E^+(0,1)$; $E^-(0,4)$	7.0%	2.2%
k=3, n=15	$E^+(0,1)$; $E^+(0,1)$; $E^-(0,4)$	7.2%	2.8%
k=5, n=5	two $E^+(0,1)$; $E^+(0,4)$; $E^-(0,1)$; $E^-(0,4)$	6.3%	1.9%
k=5, n=15	two $E^+(0,1)$; $E^+(0,4)$; $E^-(0,1)$; $E^-(0,4)$	7.3%	2.6%
k=5, n=5	three E(0,1); two $E^-(0,4)$	7.3%	2.2%
k=5, n=15	three E(0,1); two $E^-(0,4)$	7.7%	2.0%