DOCUMENT RESUME

ED 449 221		TM 032 366
AUTHOR	Locker, Lesley Ann	
TITLE	Some "Modern" Statistics: A Primer and Demonstration.	
PUB DATE	2001-02-01	
NOTE	14p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (New Orleans, LA, February 1-3, 2001).	
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)	
EDRS PRICE	MF01/PC01 Plus Postage.	
DESCRIPTORS	*Educational Research; Research Methodology; Researchers; Social Science Research; *Statistical Analysis	
IDENTIFIERS	*Trimmed Means	

ABSTRACT

Researchers in behavioral science have traditionally used "classical" statistics (e.g., mean and standard deviation) in analyzing data and reporting the results of their studies. However, it has been argued that classical statistical methods do not always represent the population well when analyzing sampling data, resulting in reduced statistical significance for many studies. Problems tend to arise when outliers (unusual scores) are drawn from a sample of the population, and distributions are skewed or heavy-tailed. The most common "modern" methods of statistical analysis are "Winsorized" (named after the statistician Charles Winsor) and "Trimmed" means. Both of these modern methods censor the outlying scores of the sample to allow for the mean to characterize the population more accurately. Most researchers, however, are still unaware or have limited knowledge of modern statistics and their benefits. Perhaps new awareness can be attained through a more concrete definition of the differences between "classical" and "modern" statistics. Sole reliance on "classical" methods will continue to reduce the number of statistically significant findings by researchers. (Contains 1 table and 11 references.) (Author/SLD)

Running Head: Some "Modern" Statistics: A Primer and Demonstration

Some "Modern" Statistics: A Primer and Demonstration

Lesley Ann Locker

Texas A&M University, 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA, February 1, 2001.

Abstract

Researchers of behavioral science have traditionally used "classical" statistics (e.g., mean and standard deviation) in analyzing data and reporting the results of their studies. However, it has been argued that "classical" statistical methods do not always represent the population well when analyzing sampling data, resulting in reduced statistical significance for many studies. Problems tend to arise when outliers (unusual scores) are drawn from a sample of the population, and distributions are skewed or heavy-tailed. The most common "modern" methods of statistical analysis are "Winsorized" (named after the statistician Charles Winsor) and "trimmed" means. Both of these "modern" methods censor the outlying scores of the sample to allow for the mean to more accurately characterize the population. Most researchers, however, are still unaware or have limited knowledge of modern statistics and their benefits. Perhaps new awareness can be attained through a more concrete definition of the differences between "classical" and "modern" statistics. Sole reliance on "classical" methods will continue to reduce the number of statistically significant findings by researchers.

Some "Modern" Statistics:

A Primer and Demonstration

Researchers of behavioral science have traditionally used "classical" statistics (e.g., mean and standard deviation) in analyzing data and reporting the results of their studies. These "classical" approaches to statistics are the same ones being taught to up-and-coming researchers and scientists today, with little regard to the more modern schools of thought. "Modern" statistical methods that have been promulgated over the past 30 years may prove to be more effective in analyzing data drawn from nonnormal samples.

"Classical" statistics are dependent on the mean, $\underline{M}$. Standard deviation ($\underline{SD}$), the coefficient of skewness ($\underline{S}$), and the coefficient of kurtosis ($\underline{K}$) all rely on the mean. They are demonstrated as follows:

$\underline{SD}_x = ((\Sigma\ (\underline{X}_i - \underline{M}_x)^2) / (\underline{n} - 1))^{.5} = ((\Sigma\ \underline{x}_i^2) / (\underline{n} - 1))^{.5}$;

Coefficient of $\underline{S}$kewness$_x$ $(\underline{S}_x) = (\Sigma\ [\underline{X}_i - \underline{M}_x)/ \underline{SD}_x]^3) / \underline{n}$; and

Coefficient of $\underline{K}$urtosis$_x$ $(\underline{K}_x) = ((\Sigma\ [\ (\underline{X}_i - \underline{M}_x)/ \underline{SD}_x]^4) / \underline{n}) - 3$.

The Pearson product-moment correlation coefficient is dependent on the mean also, as it relies on the deviations from the mean when correlating two variables.

$$\underline{r}_{xy} = \frac{(\Sigma\ (\underline{X}_1 - \underline{M}_x)\ (\underline{Y}_1 - \underline{M}_y)) / \underline{n} - 1}{(\underline{SD}_x * \underline{SD}_y)}$$

But, as is learned even in the first doctoral statistics class, the mean is heavily pulled toward any outlier scores. This influence disproportionately distorts the mean and all statistics invoking deviations from the mean. One way to resolve this problem is to utilize statistics that are less susceptible to outlier influences and departures from normality, or that do not invoke deviations from the mean. This paper is an overview of some of these options.

Problems with Classical Statistics

It has been pointed out that "classical" statistical methods do not always represent the population well when analyzing sampling data. Problems may arise when outliers (unusual scores) are drawn from a sample of the population, and distributions are skewed or heavy-tailed. According to Wilcox (1998), "a more accurate description of standard hypothesis-testing methods is that they are robust when there are no differences" (p. 300). In other words, only when variance is low can "classical" statistics provide an accurate portrayal of the population being examined.

With regard to power and accurate probability coverage, Wilcox (1998) stated that, "standard ANOVA and regression methods are affected by three characteristics of data that are commonly seen in applied work: skewness, heteroscedasticity (unequal variances among groups), and outliers" (p. 301). Utilizing traditional statistical approaches is not a problem providing that the sampling distribution is normal. As Wilcox (1998) noted, as the population variance goes up, power will go down. Outliers, or unusual scores however, can greatly impact the mean and subsequently all other statistics that rely on the mean, thus decreasing power and increasing the likelihood for Type I errors.

In terms of statistical significance testing, Thompson (1999) asserted, "statistical significance tests evaluate the probability of a given set of statistics occurring, assuming that the sample came from a population exactly described by the null hypothesis, given the sample size" (p. 20). As Thompson also pointed out, that because most researchers are not able to secure truly random samples of the population, some statisticians have argued that statistical significance tests should not be used. However, he further suggested that, "statistical tests may be reasonable if there are grounds to believe that the score sample of convenience is expected to be reasonably representative of a population" (p. 20, 1999).

Wilcox (1998) pointed out that problems occur when using the traditional Student's t test on heavy-tailed and skewed distributions when comparing groups. The population variance and standard error of the mean can inflate as a result of small departures from normality, thus decreasing power (Kesselman, Kowalchuk & Lix, 1998; Wilcox, 1998). This may result in the loss of potential correlations appearing uncorrelated due to the nonnormal distribution. Indeed, throughout the General Linear Model (Thompson, 2000), because all analyses are correlational and departures from normality or outliers impact GLM results, effect sizes are attenuated whenever classical statistics are used and methodological assumptions are not met perfectly.

"Modern" statistics minimize or avoid these problems through additional non-classical manipulation of the data. Wilcox (1998) asserted, "An important point is that modern methods do not assume or require that distributions are mixed normals. Rather, mixed normals illustrate the very general concern that very small departures from normality can inflate the population standard deviation" (p. 302). Modern methods allow nonnormal sample distributions to appear more similar to the normal population.

Why Not Discard Outliers?

It may seem that the most effective way to deal with unusual scores that have a distorting effect on our statistics and decrease power is to simply discard the outliers.

According to Wilcox (1998), a common approach to this problem is to identify outliers, toss them out and apply standard statistical significance test methods to the remaining data. Lind and Zumbo (1993) described this method as 'outlier identification'. Wilcox (1998) stated, "this approach fails because it results in using the wrong standard error" (p. 305) and is therefore not recommended.

The first problem with discarding scores is loss of randomness. When discarding is applied, the data set can no longer be considered random and results become biased. Thus, one compromises any conclusions that may have been drawn regarding causality. If the researcher decides to discard data beyond a specific point, such as 3 standard deviations above or below the mean, this implies that the mean and standard deviation have already been determined and thus are manipulated and now biased by the researcher.

Another disadvantage is impracticality, as many data sets are so large that many cases must be discarded (Lind & Zumbo, 1993). When establishing data cutoff points, Lind and Zumbo (1993) further considered this process to be a waste of time, because setting the cutoff points to low may result in the disposal of valuable data, while setting them too high may result in the retention of scores that should have been thrown out. Thus time is usually a factor to be considered in most research projects. If researchers had more time to devote to these projects, this time would be better invested in the collection of more data.

A Look at Some "Modern" Statistics

The most common "modern" methods of statistical analysis are "Winsorized" (named after the statistician Charles Winsor) and "trimmed" means. Both of these "modern" methods censor the outlying scores of the sample to allow for the mean to more accurately characterize the population. Without such censorship, results that were otherwise statistically significant may be deemed nonsignificant. Wilcox (1998) even suggested that discoveries have potentially been lost due to researchers ignoring modern statistical methods.

Winsorized Means

The "winsorize" method substitutes extreme values with less extreme values in a score distribution. To utilize this method, one begins by ordering the data points, or scores, by magnitude (Sachs, 1982). Any outliers, on either end of the tails, may be replaced by less extreme values nearest that outlying score. For example, in a sampling distribution of 5 scores--1,2,3,4,10--the researcher may choose to "winsorize" this distribution by changing the outlying score of 10 by replacing that score with a score of 4 as it deviates less from the mean and was the next nearest score to the outlying score. A mean of 2.8 may be more representative of the population that a mean of 4 because the score 10 departs so far from the other scores of the sample. The "winsorized" mean is represented symbolically as:

$$\overline{X}_w = 1/n \, \Sigma \, W_i$$

As evidenced by the "Winsorized" distribution in Table 1, the mean becomes less extreme than the original value. Winsoring allows for less weight to be given to the outlying scores in the tails, while yielding greater attention to the scores in the middle (Wilcox, 1997). By utilizing this method, the new Winsorized mean better represents the majority of the scores in the distribution.

Trimmed Means and M Estimators

In using this "modern" approach, the researcher "trims" the more extreme scores resulting in a "trimmed" mean (or trimmed SD, trimmed r, etc.). To compute a trimmed mean, one simply removes a percentage of the highest and lowest scores and averages the remaining values. The percentage of scores to be trimmed, however, is determined in advance. "Ten percent trimming" indicates that 10% of the highest and 10% of the lowest scores have been removed from the sample data and the remaining scores are averaged to find the mean.

To compute the sample "trimmed" mean, take the data from the random sample X_1, $X_2, \ldots, X_n$, letting $X_1 < X_2 < \ldots < X_n$ be written in ascending order (Wilcox, 1997). Then choose the desired amount of trimming, for instance $\gamma = 20\%$ and proceed by eliminating 20% of the highest and lowest scores (g) from the data set. Following this process, average the remaining data points:

$$\overline{X}_t = \frac{X(g+1) + \ldots + X(n-g)}{n - 2g}$$

The researcher chooses the percentage of scores (γ) to be trimmed, and the remainder of scores will be used to calculate the trimmed mean. If γ is too small, however, the statistics will still be influenced by the outliers and if γ is to large, the standard error may be inflated compared to the standard error of the sample mean. As recommended by Wilcox (1997), the "trim" (γ) should be between 0 to .25, with .20 being optimal. According to Wilcox (1998), "the more one trims, the more outliers one can have among *n* randomly sampled observations without getting relatively high standard errors" (p. 304). When n=50 and 10% trimming is used, as many as 5 outliers (10% of the sample size) may exist without inflating the standard error, where 6 outliers may cause problems.

In heavy tailed distributions, power increases as γ increases (Wilcox, 1994), because the trimmed population mean (μ_t) can be more similar to the bulk of the data in a skewed distribution. In a normal distribution, however, power decreases.

M estimators, however, first determine which scores are outliers, then adjustments to the data are made through trimming (Wilcox, 1998). M estimators allow for the possibility of no trimming or even asymmetric trimming (the trimming of only one tail). Wilcox (1998) did caution, however, that trimming only one tail may lead to technical difficulties that should be handled with special techniques.

Summary

As has been demonstrated, "modern" statistics may produce more accurate characterizations of data, because the influence of the scores least representative of the data are eliminated from the data set (Thompson, 1999). Outlying scores are least likely to be drawn in the first place and thus unlikely to be replicated in the future. Extreme scores may be drawn again in the future, but it is unlikely they will be the same as the outlying scores drawn in the original sample.

Wilcox (1998) argued that many important findings might have been lost due to researcher's limited knowledge of the benefits of using "modern" statistical methods. Outliers, however, do have an important impact on the mean and related statistics, and decrease power for statistical significance testing (Wilcox, 1998). A single oulier can adversely affect "classical" statistics such as the mean, having a subsequent influence on the Students t, standard deviation, coefficient of skewness, coefficient of kurtosis, Pearson product-moment correlation, and ANOVA. Hogg (1974) and Wilcox (1998) have demonstrated that the more robust techniques, promulgated since the 1960s, have been proven to work well with nonnormal distributions.

Computer software has been developed for use of modern methods, but may also be calculated by hand easily. Most researchers, however, are still unaware or have limited knowledge of modern statistics and their benefits. Perhaps new awareness can be attained through a more concrete definition of the differences between "classical" and "modern" statistics. Sole reliance on "classical" methods will continue to reduce the number of statistically significant findings by researchers. Understanding of the limitations of "classical" methods should encourage researchers to consider more "modern" methods.

References

Hogg, R. V. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and theory. Journal of the American Statistical Association, 69, 909-922.

Keselman, H. J., Kowalchuk, R. K. & Lix, L. M. (1998). Robust nonorthogonal analyses revisited: An update based on trimmed means. Psychometrika, 63, 145-163.

Keselman, H. J., Kowalchuk, R. K. & Lix, L. M. (1998). Multiple comparison procedures for trimmed means. Psychological Methods, 3, 123-141.

Lind, J. C. & Zumbo, B. D. (1993). The continuity principle in psychological research: An introduction to robust statistics. Canadian Psychology, 34, 407-414.

Lunneborg, C. E. (1999). Data analysis by resampling: Concepts and applications. Pacific Grove, CA: Duxbury.

Sachs, L. (1982). Applied statistics: A handbook of techniques. New York: Springer-Verlag.

Thompson, B. (1999, April). Common methodology mistakes in educational research, revisited, along with a primer on both effect sizes and the bootstrap. Invited address presented at the annual meeting of the American Educational Research Association Montreal. (ERIC Document Reproduction Service No. ED 429 110)

Thompson, B. (2000). Canonical correlation analysis. In L. Grimm & P. Yarnold (Eds.), Reading and understanding more multivariate statistics (pp.285-316). Washington, DC: American Psychological Association.

Wilcox, R. R. (1994). A one-way random effects model for trimmed means. Psychometrika, 59, 289-306.

Wilcox, R. R. (1997). Introduction to robust estimation and hypothesis testing. San Diego: Academic Press.

Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? American Psychologist, 53, 300-314.

Table 1

Two Illustrative "Modern" Statistics

Id	**X**	**X'**	**X-**
1	430	433	--
2	431	433	--
3	432	433	--
4	433	433	433
5	435	435	435
6	438	438	438
7	442	442	442
8	446	446	446
9	451	451	451
10	457	467	457
11	465	465	465
12	474	474	474
13	484	484	484
14	496	496	496
15	512	512	512
16	530	430	530
17	560	560	560
18	595	560	--
19	649	560	--
20	840	560	--
M	500.00	480.10	473.07
Md	461.00	461.00	461.00
SD	100.27	49.34	38.98
S	2.40	0.72	1.04
K	6.54	-1.08	0.30

Table reproduced with permission by Bruce Thompson from Thompson, B. (1999, April). Common methodology mistakes in educational research, revisited, along with a primer on both effect sizes and the bootstrap. Invited address presented at the annual meeting of the American Educational Research Association Montreal. (ERIC Document Reproduction Service No. ED 429 110)

TM032366

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE
(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: SOME "MODERN" STATISTICS: A PRIMER AND DEMONSTRATION	
Author(s): LESLEY ANN LOCKER	
Corporate Source:	Publication Date: 2/1/01

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] ← Sample sticker to be affixed to document

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
LESLEY ANN LOCKER
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document → [ ]

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

or here
Permitting reproduction in other than paper copy.

Level 2

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature:	Position: RES ASSOCIATE
Printed Name: LESLEY ANN LOCKER	Organization: TEXAS A&M UNIVERSITY
Address: TAMU DEPT EDUC PSYC COLLEGE STATION, TX 77843-4225	Telephone Number: 979/845-1335
	Date: 1/30/01

OVER