DOCUMENT RESUME

ED 438 315		TM 030 626
AUTHOR	VandenBosch, Dawn	
TITLE	A Primer Demonstration of Some Modern Statistics.	
PUB DATE	2000-01-29	
NOTE	17p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Dallas, TX, January 27-29, 2000).	
PUB TYPE	Information Analyses (070) -- Reports - Descriptive (141) -- Speeches/Meeting Papers (150)	
EDRS PRICE	MF01/PC01 Plus Postage.	
DESCRIPTORS	*Scores; *Statistical Analysis	
IDENTIFIERS	Mean (Statistics); *Trimmed Means	

ABSTRACT

"Modern" statistics may generate more replicable characterizations of data because, at least in some respects the influence of more extreme and less representative scores are minimized. Explaining both trimmed and Winsorized (C. Winsor) statistics, this paper uses examples to show the effects concretely. Although the paper focuses on illustrations involving Winsorized or trimmed means, these "modern" procedures can also be used to yield improved estimates of other statistics, such as standard deviations or correlations. (Contains 1 table and 11 references.) (Author/SLD)

A Primer Demonstration of

Some Modern Statistics

Dawn VandenBosch

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, January 29, 2000.

Abstract

"Modern" statistics may generate more replicable characterizations of data, because at least in some respects the influences of more extreme and less representative scores are minimized. The present paper explains both trimmed and Winsorized statistics, and uses examples to concretely show the effects. Although the present paper focuses on illustrations involving Winsorized or trimmed means, these "modern" procedures can also be employed to yield improved estimates of other statistics, such as standard deviations or correlations.

All "classical" statistics are centered about the arithmetic mean, $\underline{M}$. For example, the standard deviation ($\underline{SD}$), the coefficient of skewness ($\underline{S}$), and the coefficient of kurtosis ($\underline{K}$) are all moments about the mean, respectively:

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

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

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

Similarly, the Pearson product-moment correlation invokes deviations from the means of the two variables being correlated:

$$\underline{r}_{XY} = \frac{(\Sigma\ (\underline{X}_i - \underline{M}_X)(\underline{Y}_i - \underline{M}_Y))\ /\ \underline{n}-1}{(\underline{SD}_X * \underline{SD}_Y)}$$

The problem with "classical" statistics all invoking the mean is that these estimates are notoriously influenced by atypical scores (outliers), partly because the mean itself is differentially influenced by outliers. Lind and Zumbo (1993) noted that a single outlier can greatly bias the statistics with the product being partially invalid results that are at least somewhat misleading. Furthermore, the standard analysis of variance, Pearson product-moment correlations, and least squares regression can have low power with even a small deviation from normality (Wilcox, 1998; Wilcox, 1994). With regard to the Pearson-product moment correlation Wilcox (1998) argued

> A basic problem with correlation is that it is not resistant. That is, a single unusual value, or a small change in many values, can affect a Pearson product-moment correlation to the point that one fails to detect associations that are revealed when more modern methods are used.

Additionally, confidence intervals and measures of effect size can be extremely misleading with a slight deviation in normality (Wilcox, 1998).

Another problem arises with regard to the mean as the classical measure of location when conducting a random effects ANOVA test. The problem with regard to the ANOVA is the assumption of equal variances (Wilcox, 1994). The assumption of equal variances, when violated, distorts the Type I error rate which also indicates a possible problem with regard to power. The other problem concerns the assumption of normality.

There are many problems when assuming that a distribution is normal. According to Wilcox (1997),

> To believe in the normal distribution implies that only two numbers are required to tell us everything we need to know about the probabilities associated with a random variable: the population mean (μ) and the population variance (σ^2).

Micceri (1989) noted,

> Today's literature suggests a trend toward distrust of normality; however, this attitude frequently bypasses psychometricians and educators.

Micceri (1989) stressed that although the arithmetic mean, standard deviation, and Pearson product moment have not been proven to be robust in nonnormal distributions, textbooks and the literature are inundated with the normality assumption (Micceri, 1989).

The traditional Student's t test and Welch's test (1947) also have problems with regard to robustness when working with nonnormal distributions. Generally, the independent samples t test is robust to Type I error when sample sizes are equal but nonrobust otherwise (Sawilowsky & Blair, 352). Wilcox (1998) has also found a problem with regard to Student's *t* test. The population variance is not robust meaning, small changes in the distribution tails can alter the variance of the population. Small departures can inflate the population variance and furthermore the standard error of the mean can become inflated and hence lower power (Keselman, Kowalchuk & Lix, 1998; Wilcox, 1998).

Distributions can be skewed, have heavy tails, and as stated before, random samples often have outliers. When a distribution is skewed it is more difficult is to get an accurate confidence interval for the mean causing problems when hypothesis testing. Keselman et al. (1998) and Wilcox (1994) noted that the population mean as a measure of location is questionable when the distribution is skewed. On a more applied level, when a distribution is skewed questions arise in terms of what a typical participant is in the study (Wilcox, 1997). And as stated before, power can be low when comparing groups due to outliers and heavy tailed distributions. There are methods for hypothesis testing that when a distribution is normal has good power and when a distribution is not normal still has good power compared to classical methods that are based on means (Wilcox, 1997).

The mean is a measure of location, which reflects what the typical participant is like. The mean fails typically in doing this because it is heavily influenced by the tails of the distribution. In other words, very few participants in the tail of a distribution can influence the mean profoundly. Thus, the mean is not representing the typical participants in the study. Hogg (1974) also stressed that the mean is an extremely bad estimator of the middle of a symmetrical distribution.

In response to these common problems, researchers have developed alternative "modern statistics" that can be employed to mitigate these problems. At the time his article was written, Hogg (1974) stressed the use of "modern" statistics amid the flurry of research in the area of robustness since the 1960's. The purpose of the present paper is to describe some of these "modern statistics".

Outlier Detection and Deletion

Before discussing "modern" methods some may ask why outliers cannot be simply discarded and the rest of the data be using standard hypothesis testing methods (Wilcox, 1998)? One of the most common approaches to solving the problem on outliers is to basically search the data and when outliers are identified throw them out. This is typically called outlier identification (Lind & Zumbo, 1993). There are several disadvantages to this method and it's use is not recommended.

The most obvious problem is the question of randomness of the data after cleaning the data. Removing unwanted data creates bias and compromises the pure randomness of the data set. Subsequently, any conclusions that are made about causality are compromised.

Another disadvantage is that many data sets are too large and this method would be impractical (Lind & Zumbo, 1993). Also, rejecting data that falls outside three standard deviations from the mean implies the mean and standard deviation are biased. In other words, the mean and standard deviation originate in the same data set that includes the outliers thus making the accuracy of this method somewhat circuitous. Finally, setting the cutoff too low will increase the chances of throwing out valuable data on the other hand, while setting the cutoff too high may result in keeping data points that are meant to be thrown out which is a waste of time because there will still be outliers that were meant to be rejected (Lind & Zumbo, 1993).

"Modern" Statistics

"Winsorizing"

One of the "modern" methods "Winsorizes" (recommended by Charles P. Winsor) the score distribution by substituting less extreme values for more extreme values (Thompson, 1999). The procedure begins by ordering the samples data points by magnitude (Sachs, 1982). Then, the outlier is replaced by the value next to it. For the values 21, 34, 29, 20, 25, 99 they would be ordered as 20, 21, 25, 34, 99. Then, the extreme value would be replaced by the adjacent value because it is regarded as unreliable (i.e., 20, 21, 25, 34, 34). Both sides of the scores in order can also be Winsorized (Thompson, 1999). In a set of 20 skewed scores the 4th score (e.g., 140) may be substituted for scores 1 through 3, and in the other tail the 17th score (e.g., 312) may be substituted for scores 18 through 20. Symbolically, the Winsorized mean is represented by:

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

The mean of this Winsorized distribution in Table 1 (e.g., $\underline{M}_{X'}$ = 210.10) thus becomes less extreme than the original value (e.g., $\underline{M}_X$ = 225.25). The effect of Winsorizing is to give less weight to the values in the tails while at the same time allowing more attention to be paid to the data in the middle (Wilcox, 1997). By transforming the tails, the Winsorized mean is closer to the central portion of the distribution.

"Trimming"

Another "modern" alternative "trims" the more extreme scores and then computes a "trimmed" mean. The sample trimmed mean is computed by taking a random sample $X_1, \ldots, X_n$ and let $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ be the data written in ascending order (Wilcox, 1997). The desired amount of trimming is chosen by the researcher (γ can equal 10% or any other logical value). For example, 20% trimming means that 20% of the largest percent is removed from the data points and 20% of the smallest percent is also removed from the data points (Wilcox, 1997; Wilcox, 1998). After the removal of the largest and smallest data points (*g*) the data points that remain are then averaged:

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

The value for trimming the mean (γ) is chosen by the researcher. This poses another problem. For example, if γ is too small, the trimmed mean can still be overly influenced by outliers. However, if γ is too large, the standard error can be too large compared to the standard error of the sample mean. Wilcox (1997) recommends a γ from 0 to .25 with .20 being the optimal percentage.

Wilcox (1994) notes that as γ increases, power also increases for heavy tailed distributions. As stated before as the trimmed population mean, μ_t, can be closer to most of the data in a skewed distribution (Wilcox, 1994). However, power decreases when distributions are normal. In this example, .15 of the distribution is trimmed from each tail. The resulting mean in (e.g., $\underline{M}_{X-}$ = 203.29) is thereby closer to the median of the original distribution, which has remained 183.50.

Estimation

Another "modern" method uses M-estimators to determine whether a data point is an outlier (Wilcox, 1998). If outliers are detected empirically then adjustments are made by trimming (Wilcox, 1998). M-estimators allow for one-tailed trimming even though two-tailed trimming is best when the ultimate goal is the accuracy of confidence intervals (Wilcox, 1998).

Summary

In theory, "modern" statistics may generate more replicable characterizations of data, because at least in some respects the influence of more extreme scores, which are less likely to be drawn in future samples from the tails of a non-uniform (non-rectangular or non-flat) population distribution, has been minimized (Thompson, 1999). In other words, one is less likely to draw these data points from the population in future samples because they are so extreme. Future samples may draw outliers but it is probably unlikely or they will be different scores from previous samples drawn.

Wilcox (1998) stressed that psychology journals are inundated with nonsignificant findings that would have been statistically significant if only more "modern" methods were used. "Classical" statistics like the mean can be severely affected by a single outlier, thus affecting the types of analyses we use such as Students *t*, Pearson product-moment correlation and ANOVA. Since the 1960's more robust techniques have been found to work for nonnormality (Hogg, 1974; Wilcox, 1998). Winsorizing and trimming the mean do not even require a computer and undergraduates can even learn to do these "modern" methods by hand. The problem is that most researchers are unaware or, if they are aware, still do not make use of these "modern" methods. Concretely describing the differences between "classical" and "modern" statistics can help to bring about awareness of new methods. Reliance on old methods can seriously flaw the type of research that is currently being produced as it has in the past thirty years. It is necessary to implement these new methods. A way to begin could be to help applied scientists understand the flaws in the old concepts and advances in new 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.

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

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin, 105, 156-166.

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

Sawilowsky, S. S. & Blair, R. C. (1992). A more realistic look at the robustness and Type II error properties of the *t* test to departures from normality. Psychological Bulletin, 111, 352-360.

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)

Welch, B. L. (1947). The generalization of 'Students' problem when several different population variances are involved. Psychometrika, 34, 28-35.

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

Examples of Two "Modern" Statistics: Winsorized and Trimmed Mean

Id	X	X'	X-
1	133	140	---
2	135	140	---
3	137	140	---
4	140	140	140
5	145	145	145
6	160	160	160
7	163	163	163
8	169	169	169
9	174	174	174
10	179	179	179
11	188	188	188
12	199	199	199
13	220	220	220
14	246	246	246
15	262	262	262
16	289	289	289
17	312	312	312
18	355	312	---
19	399	312	---
20	500	312	---
M	225.25	210.10	203.29
Md	183.50	183.50	183.50
SD	100.58	66.90	54.41
S	1.43	.56	.85
K	1.63	-1.34	-.42

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM030626

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: A PRIMER DEMONSTRATION OF SOME MODERN STATISTICS

Author(s): Dawn VandenBosch

Corporate Source:

Publication Date: 1/29/00

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 media, and sold through the ERIC Document Reproduction Service (EDRS). 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 and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
[X]	[ ]	[ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic 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.

Sign here, → please

Signature: X Dawn Vandenbosch	Printed Name/Position/Title: Dawn Vandenbosch	
Organization/Address: TAMU Dept Educ Psyc College Station, TX 77843-4225	Telephone: 409/845-1335	FAX:
	E-Mail Address:	Date: 1/25/00

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com