DOCUMENT RESUME

ED 450 157 TM 032 347

AUTHOR	Davis, Brandon
TITLE	Experimentwise Type I Error: What Are They and How Do They Apply to Both Univariate Post Hoc and Multivariate Testing?
PUB DATE	2001-02-00
NOTE	11p.; 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	Error of Measurement; *Multivariate Analysis
IDENTIFIERS	*Experimentwise Error; Post Hoc Tests; *Type I Errors; Univariate Analysis

ABSTRACT

This paper reviews the concept of experimentwise Type I error. While "testwise" alpha refers to the probability of making a Type I error for a single hypothesis test, "experimentwise" error refers to the probability of having made a Type I error anywhere within the study. Experimentwise error concerns are the basis for two common statistical practices (i.e., analysis of variance (ANOVA) post hoc tests and multivariate tests), and researchers will not understand these two applications if the basic concept of experimentwise error is not understood. First, ANOVA post hoc tests implicitly incorporate a correction for experimentwise error, using adjustments similar to the Bonferroni correction. Second, experimentwise error concerns are one reason why multivariate tests are almost always vital in educational research. (Author/SLD)

Running head: EXPERIMENTWISE

Experimentwise Type I Error: What are They and How do They Apply to Both Univariate Post Hoc and Multivariate Testing?

Brandon Davis

Texas A&M University 77843-4225

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

Abstract

The present paper reviews the concept of experimentwise Type I error. While "testwise" alpha refers to the probability of making a Type I error for a single hypothesis test, "experimentwise" error rate refers to the probability of having made a Type I error anywhere within the study. Experimentwise error concerns are the basis for two common statistical practices (i.e., ANOVA post hoc tests and multivariate tests), and researchers will not understand these two applications if the basic concept of experimentwise error is not understood. First, ANOVA post hoc tests implicitly incorporate a correction for experimentwise error, using adjustments similar to the "Bonferroni correction. Second, experimentwise error concerns are one reason why multivariate tests are almost always vital in educational research.

It is clear that, "Whenever multiple statistical tests are carried out in inferential data analysis, there is a potential problem of 'probability pyramiding'" (Huberty & Morris, 1989, p. 306). Using a .05 significance level, although acceptable for one statistical test, if used in successive statistical tests, would likely result in a damagingly high Type I error rate across the entire study. And as Morrow and Frankiewicz (1979) emphasized, it is also clear that in some cases the inflation of experimentwise error rates can be quite serious.

The present paper reviews experimentwise Type I error. The concept is fundamentally important in two respects. First, ANOVA post hoc tests implicitly incorporate a correction for experimentwise error; if this correction is not understood, the researcher does not understand post hoc tests themselves. Second, because experimentwise error concerns are one reason why multivariate tests are almost always vital in educational research (Fish, 1988; Thompson, 1999), researchers ought to understand experimentwise error if they are to understand an important rationale for multivariate methods.

Most researchers are familiar with testwise alpha, which is set by the researcher as an acceptable probability of making a Type I error. But while testwise alpha refers to the probability of making a Type I error for a single hypothesis test, experimentwise error rate refers to the probability of having made a Type I error anywhere within the study. When only one hypothesis is tested for a given group of people in a study, the experimentwise error rate will exactly equal the testwise error rate. But when more than one hypothesis is tested in a given study with only one sample, the two error rates may

not be equal. Regardless of the number of hypotheses, the experimentwise error rate is always greater than, or equal to, the testwise error rate.

Given the presence of multiple hypothesis tests (e.g., two or more dependent variables) in a single study with a single sample, the testwise and the experimentwise error rates will still be equal only if the hypotheses (or the dependent variable) are perfectly correlated. Logically, the correlation of the dependent variables will impact the experimentwise error rate, because, for example, when one has perfectly correlated hypotheses, in actuality, one is still only testing a single hypothesis. In the case where dependent variables are neither perfectly correlated nor uncorrelated, the experimentwise error rate will be somewhere between the testwise error rate and the computed experimentwise error rate. Thus, the true experimentwise error rate is difficult to estimate in situations where the hypotheses are not perfectly correlated or uncorrelated. In summary, two factors impact the inflation of experimentwise Type I error: (a) the number of hypotheses tested using a single sample of data, and (b) the degree of correlation among the dependent variables or the hypotheses being tested.

When the dependent variables or hypotheses tested using a single sample of data are perfectly uncorrelated, the experimentwise error rate (α_{EW}) can be calculated. This is done using what is called the Bonferroni inequality (Love, 1988):

$$\alpha_{EW} = 1 - (1 - \alpha_{TW})^{K},$$

where $\underline{K}$ is the number of perfectly uncorrelated hypotheses being tested at a given testwise alpha level (α_{TW}). Love (1988) presented the mathematical proof that this formula is correct.

For example, if three perfectly uncorrelated hypotheses (or dependent variables) are tested using data from a single sample, each at the α_{TW}=.05 level of statistical significance, the experimentwise Type I error rate will be:

$$\begin{aligned}
\alpha_{EW} &= 1 - (1 - \alpha_{TW})^{K} \\
&= 1 - (1 - .05)^{3} \\
&= 1 - (.95)^{3} \\
&= 1 - (.95(.95)(.95)) \\
&= 1 - (.9025(.95)) \\
&= 1 - .857375 \\
\alpha_{EW} &= .142625
\end{aligned}$$

Thus, for a study testing three perfectly uncorrelated dependent variables, each tested at the α_{TW}=.05 level of statistical significance, the probability is .142625 (or 14.2625%) that one or more null hypotheses will be incorrectly rejected within the study. Most unfortunately, knowing this will not inform the researcher (a) as to which one or more of the statistically significant hypotheses is, in fact, a Type I error, or (b) as to exactly how many Type I errors are being made.

These concepts may be too abstract to be readily grasped. Luckily, Witte (1985, p. 236) explains the two error rates using an intuitively appealing example involving a coin toss. If the toss of heads is equated with a Type I error, and if the coin is tossed only once, then the probability of a head on the one toss (α_{TW}), and of at least one head within the set (α_{EW}) of one toss, will both equal 50%. Because there is only one toss of the coin, or hypothesis, the experimentwise and testwise error rates for this example are equal.

If the coin is tossed three times, rather than only once, the testwise probability of a head on each toss is still exactly 50%, i.e., α_{TW}=.50 (not .05). The Bonferroni inequality is a literal fit to this example situation (i.e., is a literal analogy rather than a figurative analogy) because the coin's behavior on each flip is perfectly uncorrelated with the coin's behavior on previous flips, assuming the coin is fair. That is, a coin is not aware of its behavior on previous flips and does not alter its behavior on any single flip given some awareness of its previous behavior.

Thus, the experimentwise probability (α_{EW}) that there will be <u>at least</u> one head in the whole set of three flips will be exactly:

$$
\begin{aligned}
\alpha_{EW} &= 1 - (1 - \alpha_{TW})^K \\
&= 1 - (1 - .50)^3 \\
&= 1 - (.50)^3 \\
&= 1 - (.50(.50)(.50)) \\
&= 1 - (.2500(.50)) \\
&= 1 - .125000 \\
\alpha_{EW} &= .875000
\end{aligned}
$$

Table 1 illustrates these concepts in a more concrete fashion. There are eight equally likely outcomes for sets of three coin flips. These are listed in the table. Seven of the eight equally likely sets of three flips involve one or more Type I error, defined in this example as a heads. And 7/8 equals .875000, or 87.5%, as expected, according to the Bonferroni inequality.

Researchers control testwise error rates by picking small values, usually 0.05, for the testwise alpha. Experimentwise error rates can be limited by employing multivariate

Table 1

All Possible Families of Outcomes for a Fair Coin Flipped Three Times

	Flip # 1	2	3	
1.	T	T	T	
2.	H	T	T	p of 1 or more H's (TW error analog)
3.	T	H	T	in set of 3 flips = 7/8 = 87.5%
4.	T	T	H	or
5.	H	H	T	where TW error analog = .50,
6.	H	T	H	EW p $= 1 - (1 - .5)^3$
7.	T	H	H	$= 1 - (.5)^3$
8.	H	H	H	$= 1 - .125 = .875$
p of H on each flip	50%	50%	50%	

statistics to test omnibus hypotheses as against lots of discrete univariate hypotheses. As shown in the illustration, using several univariate tests could lead to an extremely high experimentwise error rate if left uncontrolled.

Paradoxically, although the use of several univariate tests in a single study can lead to too many null hypotheses being spuriously rejected, as reflected in inflation of the experimentwise error rate, it is also possible that the failure to employ multivariate methods can lead to a failure to identify statistically significant results which actually exist. In other words, lowering the testwise error rate in univariate testing to control the experimentwise error rate (called the Bonferroni correction) increases the chance of not rejecting hypotheses that are in fact false, or of Type II errors. The Bonferroni correction involves using a new testwise alpha level, $\alpha_{TW}*$, estimated, for example, by dividing α_{TW} by the number of K hypotheses in the study. This approach attempts to control the experimentwise Type I error rate by reducing the testwise error rate level. For example, if

three hypotheses were being tested and the desired experimentwise error rate was .05, the new testwise alpha would be:

$$\begin{aligned}
\alpha_{EW} &= 1 - (1 - \alpha_{TW}*)^K \\
.05 &= 1 - (1 - \alpha_{TW}*)^3 \\
-.95 &= -(1 - \alpha_{TW}*)^3 \\
.95 &= (1 - \alpha_{TW}*)^3 \\
.95^{1/3} &= 1 - \alpha_{TW}* \\
.983 &= 1 - \alpha_{TW}* \\
-.017 &= -\alpha_{TW}* \\
.017 &= \alpha_{TW}
\end{aligned}$$

The new testwise alpha is reduced to .017 in order to keep the experimentwise error rate at .05. It is in these cases that some statistically significant results will not be identified if univariate methods are used to analyze data. Fish (1988), Maxwell (1992), and Thompson (1999) provide data sets illustrating this equally disturbing possibility. This means that the so-called Bonferroni correction is not a satisfactory solution to this problem.

In addition, the use of the Bonferroni correction does not address the second (and more important) reason why multivariate methods are so often vital. As Thompson (1999) argued, “Multivariate methods are often vital in behavioral research simply because multivariate methods best honor the reality to which the researcher is purportedly trying to generalize” (p. 21). Testing the effects of each variable in isolation provides little information because the variable is not isolated in reality. Depending on the researcher’s values, more valuable information may be obtained by determining the

contribution made by each variable in relation to other variables that are also present in research as well as in reality. Even with the Bonferroni correction, univariate methods usually still remain unsatisfactory.

References

Fish, L.J. (1988). Why multivariate methods are usually vital. Measurement and Evaluation in Counseling and Development, 21, 130-137.

Huberty, C.J, & Morris, J.D. (1989). Multivariate analysis versus multiple univariate analysis. Psychological Bulletin, 105, 302-308.

Love, G. (1988, November). Understanding experimentwise error probability. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville. (ERIC Document Reproduction Service No. ED 304 451)

Morris, J.R., & Frankiewicz, R.G. (1979). Strategies for the analysis of repeated and multiple measures designs. Research Quarterly, 50, 297-304.

Maxwell, S. (1992). Recent developments in MANOVA applications. In B. Thompson (Ed.), Advances in social science methodology (Vol. 2, pp. 137-168). Greenwich, CT: JAI Press.

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)

Witte, R.S. (1985). Statistics (2nd ed.). New York: Holt, Rinehart and Winston.

TM032347

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: EXPERIMENTWISE TYPE I ERRORS: WHAT ARE THEY AND HOW DO THEY APPLY TO BOTH UNIVARIATE POST HOC AND MULTIVARIATE TESTING?

Author(s): BRANDON DAVIS

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

BRANDON DAVIS

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: Brandon Davis	Position: RES ASSOCIATE
Printed Name: BRANDON DAVIS	Organization: TEXAS A&M UNIVERSITY
Address: TAMU DEPT EDUC PSYC COLLEGE STATION, TX 77843-4225	Telephone Number: 979/845-1335
	Date: 1/23/01

OVER