A Review of Experimentwise Type I Error: Implications for Univariate Post Hoc and for Multivariate Testing.

Although many researchers are familiar with "testwise" alpha, "experimentwise" Type I error rates are also important and must be considered in many research situations. Experimentwise error rates can inflate rapidly when researchers use multiple univariate tests. Both analysis of variance (ANOVA) post hoc and multi-variate methods incorporate a correction for experimentwise error. Researchers ought to understand experimentwise error if they are to understand what post hoc tests really do and an important rationale for multi-variate methods. (Contains 4 tables and 13 references.) (Author/SLD)
A Review of Experimentwise Type I Error:
Implications for Univariate Post Hoc and for Multivariate Testing

Dan Altman
Texas A&M University 77843-4225

ABSTRACT

Experimentwise error rates can rapidly inflate when researchers use multiple univariate tests. Both (a) ANOVA post hoc and (b) multivariate methods incorporate a correction for experimentwise error. Researchers ought to understand experimentwise error if they are to understand (a) what post hoc test really do and (b) an important rationale for multivariate methods.
Experimentwise Error

A REVIEW OF EXPERIMENTWISE TYPE I ERROR:
IMPLICATIONS FOR UNIVARIATE POST HOC AND FOR MUTIVARIATE TESTING

Researchers are wary of making a Type I error. In order to guard against doing that, researchers set alpha to be small. However, some researchers, focus only on "testwise" alpha, and are unaware of the "experimentwise" alpha and the importance of not inflating "experimentwise" Type I error rates. This 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, experimentwise error concerns are one reason why multivariate tests are almost always vital in educational research (Fish, 1988; Thompson, 1999), so researchers ought to understand experimentwise error if they are to understand an important rationale for multivariate methods.

Experimentwise Error

Researchers are cognizant of the possibility of rejecting a null hypothesis ($H_0$) even when the $H_0$ is true. This is called a "testwise" Type I error. Researchers set an alpha ($\alpha$) level a priori at a small near-zero value to protect against testwise Type I errors. If the alpha level is set at .01 of statistical significance, one percent of the time the null will be falsely rejected. In this case, the null is rejected even though the null may be true in the population.
Most researchers are familiar with "testwise" alpha ($\alpha_{Tw}$). However, while "testwise" alpha refers to the probability of making a Type I error for a given hypothesis test, "experimentwise" (or "familywise" -- see Maxwell, 1992, p. 138) error rate refers to the probability of having made a Type I error anywhere within a set of hypothesis tests (Thompson, 1994). "Experimentwise" error rate inflates when a number of hypotheses are tested (e.g., two or more dependent variables) at the same alpha level within a given study (Love, 1988).

"Experimentwise" error rate equals "testwise" error rate when only one hypothesis is tested for a given group of people in a study. However, when more than one hypothesis is being tested in a given study with only one sample, the two error rates may not be equal (Thompson, 1994). This occurs as Type I errors from each individual tested hypothesis build off each other, causing a highly inflated experimentwise error rate. Huberty and Morris (1989, p. 306) referred to this as "probability pyramiding."

Given the number of hypotheses being tested, the inflation of experimentwise error rates can be quite serious, as emphasized by Morrow and Frankiewicz (1979).

Experimentwise and testwise error rates are equal given the presence of multiple hypothesis tests (e.g., two or more dependent variables) in a single sample study only if the hypotheses (or the dependent variables) are perfectly correlated (or independent). This is so by reason that, for example, when one has perfectly correlated hypotheses, one actually is still only testing a single hypothesis. Therefore, it can be said that
two factors effect 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 tested (Thompson, 1994).

**Bonferroni Formula for \( \alpha_{EW} \)**

"Experimentwise" error rate inflation is at its maximum when multiple dependent variables (e.g., multiple hypothesis tests) in a single sample study are perfectly uncorrelated (Fish, 1988). When this occurs, the experimentwise error (\( \alpha_{EW} \)) rate can be calculated. This is done using what is called the Bonferroni inequality (Love, 1988):

\[
\alpha_{EW} = 1 - (1 - \alpha_{Tw})^k,
\]

where \( k \) is the number of perfectly uncorrelated hypotheses or variables being tested at a given testwise alpha level (\( \alpha_{Tw} \)).

For example, if four perfectly uncorrelated hypotheses (or dependent variables) are tested using data from a single sample, each at the \( \alpha_{Tw} = .01 \) level of statistical significance, the experimentwise Type I error rate will be:

\[
\begin{align*}
\alpha_{EW} &= 1 - (1 - \alpha_{Tw})^k \\
&= 1 - (1 - .01)^4 \\
&= 1 - (.99)^4 \\
&= 1 - (.99(.99)(.99)(.99)) \\
&= 1 - .960596 \\
\alpha_{EW} &= .039404.
\end{align*}
\]

Thus, for a study testing four perfectly uncorrelated dependent variables, each at the \( \alpha_{Tw} = .01 \) level of statistical significance, the probability is .039404 (or 3.9404%) that one or more null hypotheses will be incorrectly rejected within the study. However, knowing this will not inform the researcher as
to which one or more of the statistically significant hypotheses
is a Type I error. Table 1 provides an illustration of these
calculations for several $\alpha_{\text{FW}}$ levels. This table also illustrates
how quickly $\alpha_{\text{FW}}$ can become inflated.

Witte (1985) 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 a coin is tossed
only once, then the probability of a head on the one toss ($\alpha_{\text{FW}}$),
and of at least one head within a set ($\alpha_{\text{SW}}$) consisting of one
toss, will both equal 50%.

If the coin is tossed three times, the "testwise"
probability of a head on each toss is still 50%, i.e., $\alpha_{\text{TW}} = .50$
(not .05). The Bonferroni inequality is a literal fit to this
example situation (i.e., that is, a literal analogy), because the
coin's behavior on each flip is literally uncorrelated with the
coin's behavior on previous flips. In other words, the coin does
not alter its behavior on any given flip as a result of its
behavior on any previous flip.

Thus, the "experimentwise" probability ($\alpha_{\text{SW}}$) that there will
be at least one head in the whole set of three flips will be
exactly:

$$\alpha_{\text{SW}} = 1 - (1 - \alpha_{\text{FW}})^k$$
$$= 1 - (1 - .50)^3$$
$$= 1 - (.50)^3$$
$$= 1 - (.125000)$$
$$= .875000.$$

Table 2 illustrates these concepts more concretely. In the
table are listed eight equally likely outcomes for sets of three
coin flips. Of the eight sets of three flips, seven involve one or more Type I error, defined in this example as a heads. According to the Bonferroni inequality, 7/8 equals .875000, as expected.

As stated earlier, the above example is a literal fit for the Bonferroni inequality because the behavior of the coin on a given flip is uncorrelated with the behavior of the coin on any other flip. The exact $\alpha_{\text{EW}}$ can be determined using the Bonferroni inequality formula if the hypotheses or variables are perfectly uncorrelated. This formula is not necessary when the hypotheses are perfectly correlated because the $\alpha_{\text{EW}}$ and the $\alpha_{\text{TW}}$ equal each other.

However, in most studies hypotheses are neither perfectly uncorrelated nor perfectly correlated, and rather are partially correlated. For such studies, the actual experimentwise error rate will range somewhere between the computed experimentwise error rate (see above) and the testwise error rate, but may never really be known (Fish, 1988; Love, 1988; Morrorw & Frankiewicz, 1979).

Also, the $\alpha_{\text{EW}}$ inflation can be quite severe given the number of hypotheses tested and the level of correlation. Therefore, the power to reject can be low (Olejnik, Li, Supattathum, & Huberty, 1997). In other words, with multiple univariate follow-up tests at the original $\alpha_{\text{TW}}$ level (e.g., .05), the $\alpha_{\text{EW}}$ is inflated to statistical significance even if no statistical significance is found anywhere in the study. In order to
Experimentwise Error

compensate for this, researchers apply a "correction." This is called the "Bonferroni correction."

Bonferroni Correction

The Bonferroni correction compensates for the inflation by dividing the original $\alpha_{TW}$ by the number of $k$ hypotheses in the study yielding a new $\alpha_{TW}^*$ (Maxwell, 1992; Thompson, 1994):

$$\alpha_{TW}^* = \frac{\alpha_{TW}}{k}.$$  

Each individual post hoc test then utilizes the $\alpha_{TW}^*$ in order to maintain the $\alpha_{SW}$ at an appropriate level. Table 3 illustrates how the Bonferroni correction is utilized in order to maintain the $\alpha_{SW}$ at an appropriate level. However, this table also illustrates how the use of the Bonferroni correction has the potential for severe loss in power (Olejnik, Li, Supattathum, & Huberty, 1997).

Post Hoc Analysis

After using an ANOVA omnibus test to analyze overall differences in a multi-group study with more than two groups, many researchers use "post hoc" (also called "a posteriori," "unplanned," or "unfocused") tests to determine which group means differ for each set of pairs or combinations of groups. All comparisons/contrasts only test whether exactly two means are equal. There are two kinds of comparisons: simple and complex. Although all contrasts test the equality of exactly two means, simple and complex contrasts differ as regards the permissible ways in which the two means are created. Put simply, "simple" contrasts compare the dependent variable means of two groups using the existing levels of a way, without any combinations of
any levels. "Complex" contrasts, on the other hand, include all possible "simple" contrasts, but also include means computed by aggregating data across levels of the way.

For example, let's presume that a researcher did a one-way three-level ANOVA in which there were 10 people in each of the three groups of car owners: (a) Ford, (b) Nissan and (c) Rolls Royce. The dependent variable might be satisfaction with one's car. For this design three "simple" contrasts of mean levels of satisfaction are possible:

\[ M_{FORD} (n = 10) = M_{NISSAN} (n = 10); \]
\[ M_{FORD} (n = 10) = M_{ROLLS} (n = 10); \]
\[ M_{NISSAN} (n = 10) = M_{ROLLS} (n = 10). \]

The "complex" contrasts include these simple contrasts, plus the following three "uniquely complex" contrasts:

\[ M_{FORD} (n = 10) = M_{NISSAN} \text{ or } M_{ROLLS} (n = 20); \]
\[ M_{NISSAN} (n = 10) = M_{FORD} \text{ or } M_{ROLLS} (n = 20); \]
\[ M_{ROLLS} (n = 10) = M_{FORD} \text{ or } M_{NISSAN} (n = 20). \]

Table 4 illustrates these combinations for both three- and four-level one-way ANOVA problems. As Table 4 makes clear, as the number of levels gets larger, the number of simple contrasts gets larger, but the number of complex contrasts gets exponentially larger.

For each comparison, simple or complex, there are specific post hoc tests used. For simple comparisons the Tukey method, also called the HSD (honestly significant difference) test, is often used. For complex comparisons the Scheffé method is often used (Hinkle, Wiersman, & Jurs, 1998). Each of these method utilizes an analogue to the Bonferroni correction in order to maintain the \( \alpha_{EW} \) at the a priori \( \alpha \) level.
Tukey

The Tukey method is likely the most recommended and used procedure for controlling Type I error when making simple comparisons. The original Tukey method is based on Studentized range statistics, which takes into account the number of means being compared, adjusting for the total number of tests to make all simple comparisons. Later revisions of the Tukey method have demonstrated its robustness to violations of normality and homogeneity assumptions (Barnette, 1998). The Tukey method is also relatively insensitive to skewness. The Tukey method is not affected too much by many varied conditions. The exception to that is with the variability of the population means. Keselman (1976) found that the Tukey method is more powerful for the maximum variability of the population means. This is logical given that under this condition the magnitude of simple comparisons is largest. However, with larger sample sizes, the Tukey tends to lose relative power.

Scheffé

The Scheffé method is designed to analyze all possible comparisons (Sato, 1996). Therefore, the Scheffé method is used for complex or multiple comparisons. The Scheffé's infinite intersectional nature is its greatest strength and its greatest weakness. It is strong because it can analyze all possible comparisons. Klockars and Hancock (1998), however, assert that researchers are not always interested in many of the comparisons Scheffé makes. Because it is designed to test so many multiple comparisons, the Scheffé method is extremely conservative. The
Scheffé methods suffers loss of power for some researchers because it is so conservative (Sato, 1996).

Multivariate Methods

Multivariate methods are designed for multiple outcome variables. As Huberty and Morris (1989) noted, multivariate methods ask, "Are there any overall effects present?" This questioning, or this philosophy, best honors the reality from which data are collected. That is, if data are collected from samples upon which there are many influences, or variables, then it is logical to use a statistical method that is designed to take those variables into account simultaneously (Thompson, 1994).

Because multivariate methods are designed for multiple outcome variables, multivariate methods require only one omnibus test to determine if any differences exist. This is in contrast to univariate methods, which require many tests, thus increasing the likelihood of making erroneous decisions. For this reason alone, multivariate methods should be used when multiple outcome variables are of concern.

Summary

Although many researchers are familiar with "testwise" alpha, "experimentwise" Type I error rates are also important, and must be considered in many research situations. Testing multiple hypotheses with a single sample of data can radically inflate the "experimentwise" Type I error rate.

The present paper has explained how this inflation can be avoided in various research situations. First, it was explained that ANOVA post hoc tests implicitly incorporate a hidden analog
of the "Bonferroni correction" to avoid Type I error rate inflation. Second, it was noted that multivariate statistics are frequently employed by researchers to control "experimentwise" errors that would otherwise occur by conducting several ANOVA's or regression analyses with a single sample of data.
References


Table 1

Experimentwise Error Inflation Rates

<table>
<thead>
<tr>
<th>$\alpha_{TW}$ Tests</th>
<th>$\alpha_{EW}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - ( 1 - .01 ) ** 1</td>
<td>= 0.01</td>
</tr>
<tr>
<td>1 - ( 0.99 ) ** 1</td>
<td>= 0.01</td>
</tr>
<tr>
<td>1 - 0.99</td>
<td>= 0.01</td>
</tr>
<tr>
<td>1 - ( 1 - .001) ** 10</td>
<td>= 0.009955</td>
</tr>
<tr>
<td>1 - ( 1 - .001) ** 20</td>
<td>= 0.019811</td>
</tr>
<tr>
<td>1 - ( 1 - .001) ** 30</td>
<td>= 0.029569</td>
</tr>
<tr>
<td>1 - ( 1 - .001) ** 40</td>
<td>= 0.039230</td>
</tr>
<tr>
<td>1 - ( 1 - .001) ** 50</td>
<td>= 0.048794</td>
</tr>
<tr>
<td>1 - ( 1 - .01 ) ** 10</td>
<td>= 0.0561792</td>
</tr>
<tr>
<td>1 - ( 1 - .01 ) ** 20</td>
<td>= 0.1820931</td>
</tr>
<tr>
<td>1 - ( 1 - .01 ) ** 30</td>
<td>= 0.2602996</td>
</tr>
<tr>
<td>1 - ( 1 - .01 ) ** 40</td>
<td>= 0.3310282</td>
</tr>
<tr>
<td>1 - ( 1 - .01 ) ** 50</td>
<td>= 0.3949939</td>
</tr>
<tr>
<td>1 - ( 1 - .05 ) ** 10</td>
<td>= 0.401263</td>
</tr>
<tr>
<td>1 - ( 1 - .05 ) ** 20</td>
<td>= 0.641514</td>
</tr>
<tr>
<td>1 - ( 1 - .05 ) ** 30</td>
<td>= 0.785361</td>
</tr>
<tr>
<td>1 - ( 1 - .05 ) ** 40</td>
<td>= 0.871488</td>
</tr>
<tr>
<td>1 - ( 1 - .05 ) ** 50</td>
<td>= 0.923055</td>
</tr>
</tbody>
</table>
Table 2

All Possible Families of Outcomes

for a Fair Coin Flipped Three Times

<table>
<thead>
<tr>
<th>Flip #</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>p of 1 or more H's (TW error analog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>p of 1 or more H's (TW error analog)</td>
</tr>
<tr>
<td>2.</td>
<td>H</td>
<td>T</td>
<td>T</td>
<td>p of 1 or more H's (TW error analog)</td>
</tr>
<tr>
<td>3.</td>
<td>T</td>
<td>H</td>
<td>T</td>
<td>p of 1 or more H's (TW error analog)</td>
</tr>
<tr>
<td>4.</td>
<td>T</td>
<td>T</td>
<td>H</td>
<td>p of 1 or more H's (TW error analog)</td>
</tr>
<tr>
<td>5.</td>
<td>H</td>
<td>H</td>
<td>T</td>
<td>p of 1 or more H's (TW error analog)</td>
</tr>
<tr>
<td>6.</td>
<td>H</td>
<td>T</td>
<td>H</td>
<td>p of 1 or more H's (TW error analog)</td>
</tr>
<tr>
<td>7.</td>
<td>T</td>
<td>H</td>
<td>H</td>
<td>p of 1 or more H's (TW error analog)</td>
</tr>
<tr>
<td>8.</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>p of 1 or more H's (TW error analog)</td>
</tr>
</tbody>
</table>

p of H on each Flip | 50% 50% 50%
Table 3

Experimentwise Error Rate Without and With
The Application of the Bonferroni Correction

<table>
<thead>
<tr>
<th>Number of Hypotheses</th>
<th>$1 - (1 - \alpha_{TW})^k$</th>
<th>$\alpha_{EW}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No Bonferroni Correction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>$1 - (1 - .05)^{10}$</td>
<td>.40126</td>
</tr>
<tr>
<td>50</td>
<td>$1 - (1 - .05)^{50}$</td>
<td>.92306</td>
</tr>
<tr>
<td>100</td>
<td>$1 - (1 - .05)^{100}$</td>
<td>.99408</td>
</tr>
<tr>
<td><strong>Bonferroni Correction</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>$.05/10 = .00500</td>
<td>.04889</td>
</tr>
<tr>
<td>50</td>
<td>$.05/50 = .00100</td>
<td>.04879</td>
</tr>
<tr>
<td>100</td>
<td>$.05/100 = .00050</td>
<td>.04878</td>
</tr>
</tbody>
</table>

Note. All original $\alpha_{TW}$ for equations in Table 3 are at the .05 level.
Table 4
List of Simple and Complex Contrasts for One-way $k=3$ and $k=4$ ANOVA

Design Contrasts

$k=3$ levels

Simple $[3 (3 - 1)] / 2 = 6 / 2 = 3$
1 vs 2
1 vs 3
2 vs 3

Complex $3 + 3 = 6$
Simple
1 vs 2
1 vs 3
2 vs 3

Uniquely complex
1 vs 2,3
2 vs 1,3
3 vs 1,2

$k=4$ levels

Simple $[4 (4 -1)] / 2 = 12 / 2 = 6$
1 vs 2
1 vs 3
1 vs 4
2 vs 3
2 vs 4
3 vs 4

Complex $6 + 15 = 21$
Simple
1 vs 2
1 vs 3
1 vs 4
2 vs 3
2 vs 4
3 vs 4

Uniquely complex
1, 2 vs 3
1, 2 vs 4
1, 3 vs 4
2, 1 vs 3
2, 1 vs 4
2, 3 vs 4
3, 1 vs 3
3, 1 vs 4
3, 2 vs 4
Experimentwise Error

4, 1 vs 2
4, 1 vs 3
4, 2 vs 3
1, 2 vs 3, 4
1, 3 vs 2, 4
1, 4 vs 2, 3
I. DOCUMENT IDENTIFICATION:

Title: A REVIEW OF EXPERIMENTWISE TYPE I ERROR: IMPLICATIONS FOR UNIVARIATE POST HOC AND FOR MULTIVARIATE TESTS

Author(s): DAN ALTMAN

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

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

__________________________

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 1

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

__________________________

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 2A

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only.

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

__________________________

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 2B

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.

Signature: ____________________________

Printed Name/Position/Title: DAN ALTMAN

Organization/Address: TAMU Dept Educ Psyc

College Station, TX 77843-4225

Telephone: 409/7845-1335

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

PREVIOUS VERSIONS OF THIS FORM ARE OBSOLETE.