Technical Note

An open-source software program for performing Bonferroni and related corrections for multiple comparisons

Kyle Lesack, Christopher Naugler

Faculty of Medicine, Bachelor of Health Sciences Program, Room G503, O’Brien Centre for the BSc, 3330 Hospital Drive N.W. Calgary, Alberta T2N 4N1, 1, Departments of Pathology and Laboratory Medicine, University of Calgary and Calgary Laboratory Services, C414, Diagnostic and Scientific Centre, 9, 3535 Research Road NW, Calgary AB Canada T2L 2K8

E-mail: *Christopher Naugler - Christopher.naugler@cls.ab.ca
*Corresponding author

Received: 07 September 2011 Accepted: 18 November 2011 Published: 26 December 2011

This article may be cited as:
Lesack K, Naugler C. An open-source software program for performing Bonferroni and related corrections for multiple comparisons. J Pathol Inform 2011;2:52.
Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2011/2/1/52/91130

Copyright: © 2011 Lesack K. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Increased type I error resulting from multiple statistical comparisons remains a common problem in the scientific literature. This may result in the reporting and promulgation of spurious findings. One approach to this problem is to correct groups of \( P \)-values for “family-wide significance” using a Bonferroni correction or the less conservative Bonferroni-Holm correction or to correct for the “false discovery rate” with a Benjamini-Hochberg correction. Although several solutions are available for performing this correction through commercially available software there are no widely available easy to use open source programs to perform these calculations. In this paper we present an open source program written in Python 3.2 that performs calculations for standard Bonferroni, Bonferroni-Holm and Benjamini-Hochberg corrections.

Key words: Bonferroni correction, software program, type I error

BACKGROUND

When multiple hypotheses are tested in a single experiment, the risk of type I error is increased and with it the risk of promulgating spurious “significant” findings.\(^1\)\(^2\)\(^3\) The likelihood of obtaining a false positive result increases proportional to the number of tests performed. For example, the probability of obtaining at least one false positive result when performing 10 tests is given by

\[
1 - P(A) = 1 - 0.95^{10} = 0.4013
\]

where \( P(A) \) is the confidence level of the test.

Although the problems associated with multiple testing are well known, numerous studies still fail to correct their reported \( P \)-values. For instance, Bennett et al. found that only between 60% and 74% of the neuroimaging articles published in several major journals corrected for multiple comparisons.\(^4\) Similarly, a study performed by Austin et al. also demonstrated that the failure to account for multiple testing resulted in statistically significant, yet implausible results.\(^5\) In both cases the results were no longer significant after correcting for multiple testing.

The lack of attention paid to this problem in the pathology literature stands in stark contrast to its recognition in other fields such as ecology where there has been intense interest for over two decades since the seminal publication by Rice.\(^6\) That being said, even within the field of ecology this topic still engenders debate.\(^7\) A systematic exploration of this problem in the pathology literature has not been undertaken; however we have previously reported on a convenience sample of 800 publications from the pathology literature in 2003, of which 37 presented multiple comparisons. Twenty
One of these 37 did not attempt to control for increased type I error due to multiple comparisons.\textsuperscript{[9]}

One means of reducing the type I error from multiple testing is the Bonferroni correction, which controls the family-wise error rate (FWER). The FWER is the probability of type I error among the entire set of hypotheses.

The Bonferroni correction is calculated as follows:

\[ P_{\text{corrected}} = \frac{P_{\text{original}}}{n} \]  

where \( n \) is the number of hypotheses tested. There is a lack of consensus as to what actually represents a “family” of statistical tests; however it has been suggested that if it is appropriate to place multiple \( P \)-values in the same table, it may be appropriate to correct all values in that table for multiple comparisons.\textsuperscript{[6]}

Because the Bonferroni correction is conservative with regard to statistical power, other methods of correcting for multiple testing have been developed. Another method that controls for the FWER is the Bonferroni-Holm correction.\textsuperscript{[9]}

The Bonferroni-Holm correction is calculated as follows:

\[ P_{\text{corrected}} = \frac{P_{\text{original}}}{\min(n-k+1,1)} \]  

where \( n \) is the number of hypotheses tested, and \( k \) is the ordered rank of the uncorrected \( P \)-values (from smallest \( P \)-value to largest \( P \)-value).

Rather than controlling for the probability of one or more type I errors in the entire experiment, some of the more recent approaches to the multiple testing problem have focused on controlling the false discovery rate (FDR) in the experiment. By controlling the proportion of type I errors, this has the advantage of further increasing the statistical power of the algorithm, and is especially suitable when conducting numerous hypothesis tests.\textsuperscript{[10,11]}

The Benjamini-Hochberg method\textsuperscript{[32]} is a commonly used way to control the FDR of an experiment. It is calculated as follows:

\[ P_{\text{corrected}} = \frac{P_{\text{original}} \cdot n}{k} \]  

where \( n \) is the number of hypotheses tested, and \( k \) is the rank of the uncorrected \( P \) value.

Several commercial statistical software packages are capable of performing one or more of these corrections as well as at least one open-source program (GNU R); however the cost of the commercial packages, and the learning curves involved, may discourage researchers from using these programs. Online tools are also available (e.g., http://www.quantitativeskills.com/sisa/calculations/bonfer.htm) but are limited in scope and available options and rely on continued access to the publisher’s website.

**“Bonferroni Calculator” software**

Using the open-source programming language Python v 3.2, we developed a program capable of performing Bonferroni, Bonferroni-Holm, and Benjamini-Hochberg corrections for any number of \( P \)-values. The user is prompted for a set of \( P \)-values and the desired significance (alpha) level. From the main menu the user may choose to display the results of the desired correction to the screen, or to export the corrected \( P \) values to the hard disk (text and csv file types). The source code is available free as a supplementary file to this article (which may serve as a literature reference for the program). A copy of the source code may also be obtained by email from the corresponding author. The program requires the free programming language Python 3.2 which is capable of running on Microsoft Windows, MAC OS, and Linux/Unix operating systems. It may be downloaded from http://www.python.org/getit/releases/3.2/.

The program is available for free by emailing the senior author at christopher.naugler@els.ab.ca. Detailed instructions and a FAQ are available at https://sites.google.com/site/christophernaugler/. To use the Bonferroni Calculator software, place the files “Bonferroni Calculator.py” and “Lesack and Naugler.txt” in a folder on your hard drive. In windows, the program will run from the command line by double clicking on the “Bonferroni Calculator.py” icon; however the preferred method is to right click on the icon and select “Edit with IDLE” from the dropdown list. Press F5 to run the software, and then maximize the size of the window. Follow the instructions on the screen. If the option is selected to save the results to files, these will be found in the same folder as the “Bonferroni Calculator.py” icon. The program is also available from the authors as a stand-alone executable file.

**REFERENCES**

1. Koch G, Gansky M. Statistical considerations for multiplicity in confirmatory protocols. Drug Inf J 1996;30:523-33.
2. Bender R, Lange S. Adjusting for multiple testing-when and how? J Clin Epidemiol 2001;54:343-9.
3. Karr A, Young SS. Deming, data and observational studies. Significance 2011;8:116-120.
4. Bennett CM, Baird AA, Miller MB, Woford GL. Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: an argument for multiple comparisons correction. J Serendipitous Unexpected Results 2010;1:1-5.
5. Austin PC, Mamdani MM, Juurlink DN, Hux JE. Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health. J Clin Epidemiol 2006;59:964-9.
6. Rice WR. Analyzing tables of statistical tests. Evolution 1989;43:223-5.
7. Nakagawa S. A farewell to Bonferroni: the problems of low statistical power and publication bias. Behav Ecol 2004;15:1044-5.
8. Zheng Z, Naugler C. Type I error in pathology papers, prevalence and effect on publication citations. Poster Presentation, Canadian Association of Pathologists Annual Scientific Meeting, Montreal, PQ, Jul 11-15 2010.
9. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat 1979;6:65-70.
10. Garcia LV. Escaping the Bonferroni iron claw in ecological studies. Oikos 2004;105:657-63.
11. Wit E, McClure J. Statistics for microarrays: Design, Analysis, and Inference. 1st ed. Hoboken, New Jersey: John Wiley and Sons; 2004. p.195.
12. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B 1995;57:289-300.