

Calculating and reporting Holm-Bonferroni corrected p-values

Jessica Heeman

When performing frequentist statistical tests, p-values should be corrected for multiple comparisons. A simple, and not over-conservative (like Bonferroni correction) is the Holm-Bonferroni correction. For an explanation and examples, see:

https://en.wikipedia.org/wiki/Holm%E2%80%93Bonferroni_method

To report the results of your tests with Holm-Bonferroni corrections, you might want to report the adjusted p-values ($\tilde{p}$). The wiki page, however, seems to contain an error on this procedure in the Extensions section. That is, the function ‘max()’ needs at least two arguments otherwise there is no need for this function. So the correct rule should be:

$$\tilde{p}_{(i)} = \max_{j \leq i} \{ \tilde{p}_{(i-1)}, (m - j + 1)p_{(i)} \}, \text{ where } \{x\}_1 \equiv \min(x, 1)$$

The procedure starts with sorting the p-values resulting from all your tests from lowest to highest. Starting with the lowest, the $\tilde{p}$ can now be calculated using the second argument within the ‘max()’ function:

$$\tilde{p}_{(i)} = (m - i + 1)p_{(i)}$$

After computing the $\tilde{p}$ there are still two stop-rules that need to be applied consecutively in order to get the Holm-Bonferroni corrected p-values.

- (1) if $\tilde{p}_{(i)} > 1$ then $\tilde{p}_{(i)} = 1$ and
- (2) if $\tilde{p}_{(i)} < \tilde{p}_{(i-1)}$ then $\tilde{p}_{(i)} = \tilde{p}_{(i-1)}$

Summary for computing adjusted p-values:

- 1 perform all tests
- 2 sort the resulting p-values from lowest to highest
- 3 apply: $\tilde{p}_{(i)} = (m - i + 1)p_{(i)}$
 m = total number of tests
 i = rank of the current test
- 4 Stop at the first not-significant p-value and apply following rules
 - a) if $\tilde{p}_{(i)} > 1$ then $\tilde{p}_{(i)} = 1$
 - b) if $\tilde{p}_{(i)} < \tilde{p}_{(i-1)}$ then $\tilde{p}_{(i)} = \tilde{p}_{(i-1)}$

Example of reporting these adjusted values:

“Post-hoc one-tailed t-tests against zero showed that in the presence of a Close distractor, saccade endpoints in the Audiovisual condition ($M = 1.67^\circ$, $SD = .98$) deviated significantly *towards* the distractor (Audiovisual: $t(8) =$

5.113, $\tilde{p} = .006$, $d = 2.41$). In the presence of a Remote distractor the Audiovisual distractor ($M = -.55^\circ$, $SD = .44$) evoked a significant deviation away from the distractor ($t(8) = -3.692$, $\tilde{p} = .03$, $d = 1.768$). All p-values were adjusted ($\tilde{p}$) according to the Holm-Bonferroni method for multiple comparisons (Holm, 1979).”

Also see:

http://nebc.nerc.ac.uk/courses/GeneSpring/GS_Mar2006/Multiple%20testing%20corrections.pdf

<http://www.talkstats.com/showthread.php/40049-How-to-present-data-after-Holm-Bonferroni-Correction>