

# Bonferroni and Holm approximations for Šidák and Holland–Copenhaver $q$ -values

Roger B. Newson  
National Heart and Lung Institute, Imperial College London  
London, United Kingdom  
r.newson@imperial.ac.uk

## Abstract.

We describe the use of the Bonferroni and Holm formulas as approximations for Šidák and Holland–Copenhaver formulas when precision issues are encountered, especially with  $q$ -values corresponding to very small  $p$ -values.

**Keywords:** st0001, qqvalue, smileplot, multproc, multiple-test procedure, familywise error rate, Bonferroni, Šidák, Holm, Holland, Copenhaver.

## 1 Introduction

Frequentist  $q$ -values for a range of multiple-test procedures are implemented in Stata using the package `qqvalue`, downloadable from SSC (Newson (2010)). The Šidák  $q$ -value for a  $p$ -value  $p$  is given by  $q_{\text{sid}} = 1 - (1 - p)^m$ , where  $m$  is the number of multiple comparisons (Šidák (1967)). It is a less conservative alternative to the Bonferroni  $q$ -value, given by  $q_{\text{bon}} = \min(1, mp)$ . However, the Šidák formula may be incorrectly evaluated by a computer to zero when the input  $p$ -value is too small to give a result lower than 1 when subtracted from 1, which is the case for  $p$ -values of  $10^{-17}$  or less, even in double precision. Zero  $q$ -values are logically possible as a consequence of zero  $p$ -values, but, in this case, they may be over-liberal. This liberalism may possibly be a problem in the future, given the current technology-driven trend of exponentially-increasing multiple comparisons and the human-nature-driven problem of ingenious data-dredging. We present a remedy for this problem, and discuss its use in computing  $q$ -values and discovery sets.

## 2 Methods for $q$ -values

The remedy used by the SSC packages `qqvalue` and `parcest`, is to substitute the Bonferroni formula for the Šidák formula for such small  $p$ -values. This works because the Bonferroni and Šidák  $q$ -values converge in ratio as  $p$  tends to zero. To prove this, note that, for  $0 \leq p < 1$ ,

$$dq_{\text{bon}}/dp = m \quad \text{and} \quad dq_{\text{sid}}/dp = m(1 - p)^{m-1} \quad (1)$$

and that the Šidák/Bonferroni ratio of these derivatives is  $(1 - p)^{m-1}$ , which is 1 if  $p = 0$ . By L'Hôpital's rule, it follows that the ratio  $q_{\text{sid}}/q_{\text{bon}}$  also tends to 1 as  $p$  tends

to zero.

A similar argument shows that the same problem exists with the  $q$ -values output by the Holland–Copenhaver procedure (Holland and Copenhaver (1987)). If the  $m$  input  $p$ -values, sorted in ascending order, are denoted  $p_i$  for  $i$  from 1 to  $m$ , then the Holland–Copenhaver procedure is defined by the formula

$$s_i = 1 - (1 - p_i)^{m-i+1} \quad (2)$$

where  $s_i$  is the  $i$ th  $s$ -value. (In the terminology of Newson (2010),  $s$ -values are truncated at 1 to give  $r$ -values, which are in turn input into a step-down procedure to give the eventual  $q$ -values.) The remedy used by `qqvalue` here is to substitute the  $s$ -value formula for the procedure of Holm (1979), which is

$$s_i = (m - i + 1)p_i \quad (3)$$

whenever  $1 - p_i$  is evaluated as 1. This also works because the two  $s$ -value formulas converge in ratio as  $p_i$  tends to zero. Note that the Holm procedure is derived from the Bonferroni procedure using the same step-down method as is used to derive the Holland–Copenhaver procedure from the Šidák procedure.

### 3 Methods for discovery sets

The SSC package `smileplot` (Newson and the ALSPAC Study Team (2003)) also implements a range of multiple-test procedures, using two modules `multproc` and `smileplot`. However, instead of outputting  $q$ -values, `smileplot` outputs a corrected critical  $p$ -value threshold, and a corresponding discovery set, defined as the subset of input  $p$ -values at or below the corrected critical  $p$ -value. The Šidák corrected critical  $p$ -value corresponding to an uncorrected critical  $p$ -value  $p_{\text{unc}}$  is given by  $c_{\text{sid}} = 1 - (1 - p_{\text{unc}})^{1/m}$ , and may be over-conservative, if wrongly evaluated to zero. In this case, the quantity that might be wrongly computed as 1 is  $(1 - p_{\text{unc}})^{1/m}$ . When this happens, `smileplot` substitutes the Bonferroni corrected critical  $p$ -value  $c_{\text{bon}} = p_{\text{unc}}/m$ . However, this is a slightly less elegant remedy in this case, because the quantity  $(1 - p_{\text{unc}})^{1/m}$  is usually evaluated to 1 because  $m$  is large, and not because  $p_{\text{unc}}$  is small.

To study the behavior of the Bonferroni approximation for large  $m$ , we define  $\lambda = 1/m$ , and note that

$$dc_{\text{bon}}/d\lambda = p_{\text{unc}} \quad \text{and} \quad dc_{\text{sid}}/d\lambda = -\ln(1 - p_{\text{unc}})(1 - p_{\text{unc}})^\lambda \quad (4)$$

implying (again by L'Hôpital's rule) that, in the limit, as  $\lambda$  tends to 0, the Šidák/Bonferroni ratio of the two derivatives (and therefore of the two corrected thresholds) tends to  $-\ln(1 - p_{\text{unc}})/p_{\text{unc}}$ . This quantity is not as low as 1, but is 1.150728, 1.053605, 1.025866 and 1.005034 if  $p_{\text{unc}}$  is 0.25, 0.10, 0.05 and 0.01, respectively. Therefore, the Bonferroni approximation in this case is still slightly conservative for a very large number of multiple comparisons over a range of commonly-used uncorrected critical  $p$ -values, but is less conservative than the value of 0 that would otherwise be computed.

This argument is easily generalized to the Holland–Copenhaver procedure. In this case, `smileplot` initially calculates a vector of  $m$  candidate critical  $p$ -value thresholds, using the formula

$$c_i = 1 - (1 - p_{\text{unc}})^{1/(m-i+1)} \quad (5)$$

for  $i$  from 1 to  $m$ , and selects the corrected critical  $p$ -value, corresponding to a given uncorrected critical  $p$ -value, from these candidates, using a step-down procedure. If the quantity  $(1 - p_{\text{unc}})^{1/(m-i+1)}$  is evaluated as 1, then `smileplot` substitutes the corresponding Holm critical  $p$ -value threshold

$$c_i = p_{\text{unc}}/(m - i + 1) \quad (6)$$

which again is conservative as  $m - i + 1$  becomes large (corresponding to the smallest  $p$ -values from a large number of multiple comparisons), but less conservative than the value of 0 that would otherwise be computed.

It is argued in Newson (2010) that  $q$ -values are an improvement on discovery sets, because, given the  $q$ -values, different members of the audience can apply different input critical  $p$ -values, and derive their own discovery sets. The technical precision issue presented here may be one more minor reason for preferring  $q$ -values to discovery sets.

## 4 Acknowledgements

I would like to thank Tiago V. Pereira of the University of São Paulo, Brazil for drawing my attention to this precision issue with the Šidák and Holland–Copenhaver procedures.

## 5 References

- Holland, B. S., and M. D. Copenhaver. 1987. An improved sequentially rejective Bonferroni test procedure. *Biometrics* 43: 417–423.
- Holm, S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6: 65–70.
- Newson, R., and the ALSPAC Study Team. 2003. Multiple-test procedures and smile plots. *Stata Journal* 3(2): 109–132.
- Newson, R. B. 2010. Frequentist  $q$ -values for multiple-test procedures. *Stata Journal* 10(4): 568–584.
- Šidák, Z. 1967. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association* 62: 626–633.

### About the author

Roger B. Newson is a Lecturer in Medical Statistics at Imperial College London, UK, working principally in asthma research. He wrote the packages `parmeta`, `qqvalue` and `smileplot`.