WIKIPEDIA

Bonferroni correction

In statistics, the **Bonferroni correction** is one of several methods used to counteract the problem of multiple comparisons.

Contents

Background

The Bonferroni correction is named after Italian mathematician Carlo Emilio Bonferroni for its use of Bonferroni inequalities.[1] Its development is often credited to Olive Jean Dunn, who described the procedure's application to confidence intervals.[2][3]

Statistical hypothesis testing is based on rejecting the null hypothesis if the likelihood of the observed data under the null hypotheses is low. If multiple hypotheses are tested, the chance of observing a rare event increases, and therefore, the likelihood of incorrectly rejecting a null hypothesis (i.e., making a Type I error) increases.[4]

The Bonferroni correction compensates for that increase by testing each individual hypothesis at a significance level of α/m, where α is the desired overall alpha level and m is the number of hypotheses.[5] For example, if a trial is testing $m = 20$ hypotheses with a desired $\alpha = 0.05$, then the Bonferroni correction would test each individual hypothesis at $\alpha = 0.05/20 = 0.0025$.

Definition

Let $H_1, \ldots, H_m$ be a family of hypotheses and $p_1, \ldots, p_m$ their corresponding p-values. Let m be the total number of null hypotheses and m_0 the number of true null hypotheses. The familywise error rate (FWER) is the probability of rejecting at least one true H_i, that is, of making at least one type I error. The Bonferroni correction rejects the null hypothesis for each $p_i \leq \frac{\alpha}{m}$, thereby controlling the FWER at $\leq \alpha$. Proof of this control follows from Boole's inequality, as follows:

$$\text{FWER} = P\left\{\bigcup_{i=1}^{m_0}\left(p_i \leq \frac{\alpha}{m}\right)\right\} \leq \sum_{i=1}^{m_0}\left\{P\left(p_i \leq \frac{\alpha}{m}\right)\right\} = m_0\frac{\alpha}{m} \leq m\frac{\alpha}{m} = \alpha.$$

This control does not require any assumptions about dependence among the p-values or about how many of the null hypotheses are true.[6]

Extensions

Generalization

Rather than testing each hypothesis at the α/m level, the hypotheses may be tested at any other combination of levels that add up to α, provided that the level of each test is determined before looking at the data.[7] For example, for two hypothesis tests, an overall α of 0.05 could be maintained by conducting one test at 0.04 and the other at 0.01.

Confidence intervals

The Bonferroni correction can be used to adjust confidence intervals. If one establishes m confidence intervals, and wishes to have an overall confidence level of $1-\alpha$, each individual confidence interval can be adjusted to the level of $1-\frac{\alpha}{m}$.[2][3]

Alternatives

There are alternative ways to control the familywise error rate. For example, the Holm–Bonferroni method and the Šidák correction are universally more powerful procedures than the Bonferroni correction, meaning that they are always at least as powerful. Unlike the Bonferroni procedure, these methods do not control the expected number of Type I errors per family (the per-family Type I error rate).[8]

Criticism

With respect to FWER control, the Bonferroni correction can be conservative if there are a large number of tests and/or the test statistics are positively correlated. [9]

The correction comes at the cost of increasing the probability of producing false negatives, i.e., reducing statistical power.[10][9] There is not a definitive consensus on how to define a family in all cases, and adjusted test results may vary depending on the number of tests included in the family of hypotheses. Such criticisms apply to FWER control in general, and are not specific to the Bonferroni correction.

References

1. Bonferroni, C. E., Teoria statistica delle classi e calcolo delle probabilità, Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 1936
2. Dunn, Olive Jean (1958). "Estimation of the Means for Dependent Variables". *Annals of Mathematical Statistics*. **29** (4): 1095–1111. doi:10.1214/aoms/1177706374 (https://doi.org/10.1214%2Faoms%2F1177706374). JSTOR 2237135 (https://www.jstor.org/stable/2237135).

3. Dunn, Olive Jean (1961). "Multiple Comparisons Among Means" (http://sci2s.ugr.es/keel/pdf/algorithm/articulo/1961-Bonferroni_Dunn-JASA.pdf) (PDF). *Journal of the American Statistical Association*. **56** (293): 52–64. CiteSeerX 10.1.1.309.1277 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.309.1277). doi:10.1080/01621459.1961.10482090 (https://doi.org/10.1080%2F01621459.1961.10482090).
4. Mittelhammer, Ron C.; Judge, George G.; Miller, Douglas J. (2000). *Econometric Foundations* (https://books.google.com/books?id=fycmsfkK6RQC&pg=PA73). Cambridge University Press. pp. 73–74. ISBN 978-0-521-62394-0.
5. Miller, Rupert G. (1966). *Simultaneous Statistical Inference* (https://books.google.com/books?id=4C7VBwAAQBAJ&printsec=frontcover#v=onepage&q=Bonferroni&f=false). Springer. ISBN 9781461381228.
6. Goeman, Jelle J.; Solari, Aldo (2014). "Multiple Hypothesis Testing in Genomics". *Statistics in Medicine*. **33** (11): 1946–1978. doi:10.1002/sim.6082 (https://doi.org/10.1002%2Fsim.6082). PMID 24399688 (https://pubmed.ncbi.nlm.nih.gov/24399688).
7. Neuwald, AF; Green, P (1994). "Detecting patterns in protein sequences". *J. Mol. Biol*. **239** (5): 698–712. doi:10.1006/jmbi.1994.1407 (https://doi.org/10.1006%2Fjmbi.1994.1407). PMID 8014990 (https://pubmed.ncbi.nlm.nih.gov/8014990).
8. Frane, Andrew (2015). "Are per-family Type I error rates relevant in social and behavioral science?". *Journal of Modern Applied Statistical Methods*. **14** (1): 12–23. doi:10.22237/jmasm/1430453040 (https://doi.org/10.22237%2Fjmasm%2F1430453040).
9. Moran, Matthew (2003). "Arguments for rejecting the sequential Bonferroni in ecological studies". *Oikos*. **100** (2): 403–405. doi:10.1034/j.1600-0706.2003.12010.x (https://doi.org/10.1034%2Fj.1600-0706.2003.12010.x).
10. Nakagawa, Shinichi (2004). "A farewell to Bonferroni: the problems of low statistical power and publication bias" (https://academic.oup.com/beheco/article/15/6/1044/206216). *Behavioral Ecology*. **15** (6): 1044–1045. doi:10.1093/beheco/arh107 (https://doi.org/10.1093%2Fbeheco%2Farh107).

Further reading

- Dunnett, C. W. (1955). "A multiple comparisons procedure for comparing several treatments with a control". *Journal of the American Statistical Association*. **50** (272): 1096–1121. doi:10.1080/01621459.1955.10501294 (https://doi.org/10.1080%2F01621459.1955.10501294).
- Dunnett, C. W. (1964). "New tables for multiple comparisons with a control" (https://semanticscholar.org/paper/888b68b0713879ced708ad45dc7cfdbe11108b3b). *Biometrics*. **20** (3): 482–491. doi:10.2307/2528490 (https://doi.org/10.2307%2F2528490). JSTOR 2528490 (https://www.jstor.org/stable/2528490).
- Shaffer, J. P. (1995). "Multiple Hypothesis Testing". *Annual Review of Psychology*. **46**: 561–584. doi:10.1146/annurev.ps.46.020195.003021 (https://doi.org/10.1146%2Fannurev.ps.46.020195.003021). hdl:10338.dmlcz/142950 (https://hdl.handle.net/10338.dmlcz%2F142950).
- Strassburger, K.; Bretz, Frank (2008). "Compatible simultaneous lower confidence bounds for the Holm procedure and other Bonferroni-based closed tests". *Statistics in Medicine*. **27** (24): 4914–4927. doi:10.1002/sim.3338 (https://doi.org/10.1002%2Fsim.3338). PMID 18618415 (https://pubmed.ncbi.nlm.nih.gov/18618415).
- Šidák, Z. (1967). "Rectangular confidence regions for the means of multivariate normal distributions". *Journal of the American Statistical Association*. **62** (318): 626–633. doi:10.1080/01621459.1967.10482935 (https://doi.org/10.1080%2F01621459.1967.10482935).

- Hochberg, Yosef (1988). "A Sharper Bonferroni Procedure for Multiple Tests of Significance" (http://www-stat.wharton.upenn.edu/~steele/Courses/956/Resource/MultipleComparision/Hochberg88.pdf) (PDF). *Biometrika*. **75** (4): 800–802. doi:10.1093/biomet/75.4.800 (https://doi.org/10.1093%2Fbiomet%2F75.4.800).

External links

- Bonferroni, Sidak online calculator (http://www.quantitativeskills.com/sisa/calculations/bonfer.htm)
- Multiple Testing Corrections in GeneSpring and Gene Expression data (http://nebc.nerc.ac.uk/courses/GeneSpring/GS_Mar2006/Multiple%20testing%20corrections.pdf)