Bonferroni correction

From Wikipedia, the free encyclopedia

In statistics, the **Bonferroni correction** is a method used to counteract the problem of multiple comparisons. It is named after Italian mathematician Carlo Emilio Bonferroni for the use of Bonferroni inequalities,[1] but modern usage is often credited to Olive Jean Dunn, who described the procedure in a pair of articles written in 1959 and 1961.[2][3]

Contents

Informal introduction

Statistical inference logic is based on rejecting the null hypotheses if the likelihood of the observed data under the null hypotheses is low. The problem of multiplicity arises from the fact that as we increase the number of hypotheses being tested, we also increase the likelihood of a rare event, and therefore, the likelihood of incorrectly rejecting a null hypothesis (i.e., make a Type I error).

The Bonferroni correction is based on the idea that if an experimenter is testing m hypotheses, then one way of maintaining the familywise error rate (FWER) is to test each individual hypothesis at a statistical significance level of $1/m$ times the desired maximum overall level.

So, if the desired significance level for the whole family of tests is α, then the Bonferroni correction would test each individual hypothesis at a significance level of α/m. For example, if a trial is testing $m = 8$ hypotheses with a desired $\alpha = 0.05$, then the Bonferroni correction would test each individual hypothesis at $\alpha = 0.05/8 = 0.00625$.

Statistically significant simply means that a given result is unlikely to occur if the null hypothesis is true (i.e., no difference among groups, no effect of treatment, no relation among variables).

Definition

Let $H_1, ..., H_m$ be a family of hypotheses and $p_1, ..., p_m$ their corresponding p-values. The familywise error rate is the probability of rejecting at least one true H_i; that is, to make at least one type I error. The Bonferroni Correction states that rejecting the null hypothesis for all $p_i \leq \frac{\alpha}{m}$ controls the FWER. The proof follows from Boole's inequality:

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

This control does not require any assumptions about dependence among the p-values.[4]

Extensions

Generalization

Rather than testing each hypothesis at the α/m level, the hypotheses may be tested at any combination of levels that add up to α, provided that the level of each specific test is determined before looking at the data. For example, for two hypothesis tests, an overall α of .05 could be maintained by conducting one test at .04 and the other at .01.

Confidence intervals

Bonferroni correction can be used to adjust confidence intervals. If we are forming m confidence intervals, and wish to have overall confidence level of $1 - \alpha$, we can adjust each individual confidence interval to the level of $1 - \frac{\alpha}{m}$.

Alternatives

There are other alternatives 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. However, unlike the Bonferroni procedure, these methods do not control the per-family Type I error rate (the expected number of Type I errors per family).[5]

Criticisms

The Bonferroni correction can be somewhat conservative if there are a large number of tests and/or the test statistics are positively correlated. The correction also comes at the cost of increasing the probability of producing false negatives, and consequently reducing statistical power.

Another criticism concerns the concept of a family of hypotheses. There is not a definitive consensus on how to define a family in all cases. As there is no standard definition, test results may change dramatically, only by modifying the way we consider the hypotheses families.

All of these criticisms, however, apply to adjustments for multiple comparisons in general, and are not specific to the Bonferroni correction.

See also

- Bonferroni inequalities
- Familywise error rate
- Holm–Bonferroni method
- Multiple testing

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 (1959). "Estimation of the Medians for Dependent Variables". *Annals of Mathematical Statistics* **30** (1): 192–197. doi:10.1214/aoms/1177706374. JSTOR 2237135.
3. Dunn, Olive Jean (1961). "Multiple Comparisons Among Means" (PDF). *Journal of the American Statistical Association* **56** (293): 52–64. doi:10.1080/01621459.1961.10482090.
4. Goeman, Jelle J.; Solari, Aldo (2014). "Multiple hypothesis testing in genomics". *Statistics in Medicine* **33** (11). doi:10.1002/sim.6082.
5. 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.

Further reading

- Abdi, H. (2007). "Bonferroni and Šidák corrections for multiple comparisons". In Salkind, N. J. *Encyclopedia of Measurement and Statistics* (PDF). Thousand Oaks, CA: Sage.
- Manitoba Centre for Health Policy (2008). "Concept: Multiple Comparisons".
- Dunn, O. J. (1961). "Multiple Comparisons Among Means". *Journal of the American Statistical Association* **56** (293): 52–64. doi:10.1080/01621459.1961.10482090.
- 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.
- Dunnett, C. W. (1964). "New tables for multiple comparisons with a control". *Biometrics* **20** (3): 482–491. doi:10.2307/2528490. JSTOR 2528490.
- Perneger, Thomas V. (1998). "What's wrong with Bonferroni adjustments". *British Medical Journal* **316** (7139): 1236–1238. doi:10.1136/bmj.316.7139.1236. See also the Rapid Response to this suggesting much of it is mistaken.
- Shaffer, J. P. (1995). "Multiple Hypothesis Testing". *Annual Review of Psych* **46**: 561–584. doi:10.1146/annurev.ps.46.020195.003021.

- 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.
- 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.
- Hochberg, Yosef (1988). "A Sharper Bonferroni Procedure for Multiple Tests of Significance" (PDF). *Biometrika* **75** (4): 800–802. doi:10.1093/biomet/75.4.800.

External links

- School of Psychology, University of New England, New South Wales, Australia, 2000, http://www.une.edu.au/WebStat/unit_materials/c5_inferential_statistics/bonferroni.html
- Weisstein, Eric W., "Bonferroni correction" (http://mathworld.wolfram.com/BonferroniCorrection.html), *MathWorld*.
- Bonferroni, Sidak online calculator (http://www.quantitativeskills.com/sisa/calculations/bonfer.htm)
- Explanation of p-value correction methods under the context of differential gene expression analysis (http://www.silicongenetics.com/Support/GeneSpring/GSnotes/analysis_guides/mtc.pdf)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Bonferroni_correction&oldid=696518591"

Categories: Multiple comparisons | Hypothesis testing
