How we Deal with Small Data?

GHICA Manuela[1], BĂNCESCU Irina[2], UDEANU Denisa Ioana[3]

[1] Department of Biostatistics, Faculty of Pharmacy, University of Medicine and Pharmacy "Carol Davila", 020956 Bucharest, (ROMANIA)
[2] National Institute for Economic Research "Costin C. Kiriţescu", Romanian Academy, (ROMANIA)
[3] Department of Clinical Laboratory and Food Safety, Faculty of Pharmacy, University of Medicine and Pharmacy "Carol Davila," 020956 Bucharest, (ROMANIA)
Emails: manuela.ghica@umfcd.ro, irina_adrianna@yahoo.com, denisaudeanu@gmail.com

Abstract

Most commonly used statistical methods in analyzing and interpreting clinical trial data are based on a specific distribution of the data, such as normal, binomial and exponential models. The alternative is to choose nonparametric methods based on rank, but with the disadvantage that for small size samples, they may give erroneous results. In this article, we refer an accurate statistical method based on computational simulations in order to characterize and compare as efficiently as possible clinical trial studies having few values. The bootstrap technique is a statistical method that does not depend on any data distribution hypothesis and can be considered adequate to estimate some essential parameters in data analysis by different types of bootstrap confidence intervals. The effect-size statistics and its confidence intervals are also presented, being useful when we compare group means.

Keywords: Simulation, Bootstrap method, Confidence intervals, Effect size, Nonnormal data

Introduction

Computational statistics along with statistical software packages can describe and shape any biomedical phenomena using bootstrap replication methods [1]. With the help of these bootstrap techniques, it is possible to estimate using confidence intervals any statistical parameter belonging to a statistical population which does not respect the hypotheses of classical inferential statistics.

Challenges occur when data distribution is uncertain and fails to be approximated with a normal one and, last but not least, when we deal with insufficient number of data, thus not being able to obtain a known statistical distribution of the data at a reasonable risk. The hypotheses of inferential statistics require a selection volume of at least 30 values, which disagrees with the ethical code and the economic conditions of animal experiment studies [2]. Thus, the research on the number of animals necessary to obtain credible results is pervasive [3] but, also, the statistical approach must be modern, based on computational simulations [4].

The bootstrapping technique, also called resampling, is based on random replications having sample extractions returns (that is, the same individual may be encountered several times in the bootstrap set and also in any other bootstrap sample), from the dataset under analysis, replications having the same size as the original dataset. This method assumes that the original dataset is randomly extracted from an infinite population and is representative of it.

This bootstrap resampling technique is also used in bioequivalence studies of formulations to obtain a confidence interval for the difference of two treatments, one tested and one reference [5].

This article is divided into four sections. In the first section, we analyze the need to use computational techniques when the hypotheses of inferential parametric statistics are not accomplished. The second section describes several models of bootstrap confidence intervals and the assumptions they imply. In the next section, we amplify the new statistics approach by exposing the

effect size measures both under normality conditions and when this assumption is not met. The last section is reserved for conclusions.

Confidence intervals

In this section, we describe by comparison several bootstrap confidence intervals that can be built computationally using software R.

The first bootstrap interval is the 95% normal confidence interval which is used in classical inferential statistics. Except for the standard error that is estimated based on a replicated bootstrap distribution, all the items of the study remain unchanged, the known quantile being 1.96. This type of normal confidence interval can be used if the statistical bootstrap distribution is approximately normal and symmetric, and the sample estimator is unbiased.

The second bootstrap presented interval is the 95% percentile confidence interval. This is a confidence interval that uses the lower limits of 2.5% and upper 97.5%, and not only. These percentiles correspond to the confidence interval of 95%, otherwise, the percentages vary. This type of interval can be used if the replicated statistical bootstrap distribution is approximately symmetric, and the parameter sample estimator is unbiased.

The basic confidence interval is based on a method that corrects the bootstrap distribution asymmetry and recalculates the upper and lower margins to identify the desired 95% confidence interval. This interval is based only on the assumption of an unbiased sample estimator and no longer implies any assumption for the replicated bootstrap distribution.

The studentized confidence interval is constructed by generating for each replicated sample the estimated parameter (mean for example) and associated standard error. Thus, the replicated statistic of the t-student form is constructed, and for this, we apply the bootstrap percentile confidence interval with the 0.025 and 0.975 quantiles. This confidence interval is based only on the assumption that the standard error of the estimated parameter can be computationally generated.

The last bootstrap confidence interval, Bias-Corrected and Accelerated Interval, BCA, can be built without any prior assumption, being the most commonly used bootstrap interval. The construction method for this interval corrects the bootstrap replicated distribution asymmetry and the uncertain variance of each replicated sample and uses the 0.025 and 0.975 quantiles.

All methods of constructing bootstrap confidence intervals allow the use of non-normal and asymmetric distributions. The sample size is a factor that can contribute to higher accuracy of results and much narrower confidence intervals in order to estimate as accurately and efficiently as possible the unknown parameters in the research.

Effect size

The results of testing statistical hypotheses are inferior in information because of the dichotomic response by which a hypothesis is accepted or rejected. Confidence intervals offer more information, but the answers required in a particular research to accept or not certain treatments can be offered along with the magnitude of differences or similarities between them [6].

In statistical studies, there are many measures used to quantify the magnitude of differences. We present two of the most widely used and robust statistical effect sizes. One of the most used statistical measures that quantify the difference between two datasets is Cohen effect size, denoted by d, recommended only in the case of two normally distributed datasets having homogeneity of variances [7]. If normality and homogeneity assumptions are not verified, other effect sizes may be applied, but the one proposed by Hogarty and Kromrey, denoted by d_r, and developed by Algina is the one that offers the same metrics as Cohen effect size being robust [10].

Cohen Effect size for mean difference

In cases where we want to compare two treatment groups using Cohen effect size index, the assumption of normality for the empirical distributions involved should be respected [7]. This is calculated on the same principle as the t-student statistic, exception being the sample mean involved and the standard deviation that can be estimated by the bootstrap method. Cohen effect size index is given by

$$d = \frac{\bar{x}_T - \bar{x}_C}{s_p}$$

where $\bar{x}_T$ and $\bar{x}_C$ are the estimated mean of two groups, and the pooled variance is given by

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

n_1, n_2 are the sizes of two groups and s_1^2, s_2^2 the estimated variances, respectively.

The Cohen value *d* can describe differences in percentages between T (treatment) and C (control) treatments. If the index is 0, then 50% of control group C values are below the mean of the T group.

It is considered that we have a small effect between treatments if d <0.2, a moderate effect if d <0.5, and otherwise the effect is assumed to be large.

The Cohen *d* value is superior to the p-value of a statistical hypothesis, but it is also important to calculate an afferent confidence interval for greater accuracy, especially when the sample size is small. If 95% of confidence intervals contain the value of zero, then we have a non-significant difference between the two treatments [8].

The confidence interval proposed by Hedge and Olkin in 2014 assume normality and homogeneity of data. This confidence interval is $(d - 1.96\sigma(d), d + 1.96\sigma(d))$, where $\sigma(d)$ represents the Cohen standard deviation given by

$$\sigma(d) = \sqrt{\frac{n_1 + n_2}{n_1 * n_2} + \frac{d^2}{2(n_1 + n_2)}}$$

and n_1, n_2 are the sizes of two groups [9].

Effect size for non-normal data

For situations where assumptions of normality and homogeneity are not meet, an estimator of Cohen's effect size is used [11]. Applying the same construction idea as in the case of Cohen's index, the means of the two data sets are replaced with 20% truncated means $\bar{y}_2$ si $\bar{y}_1$ and the standard deviation is replaced with a 20% Winsorized variance s_W^2. That is

$$d_r = 0.642\left(\frac{\bar{y}_2 - \bar{y}_1}{s_W}\right)$$

where

$$s_W = \sqrt{\frac{(n_1 - 1)s_{W_1}^2 + (n_2 - 1)s_{W_2}^2}{n_1 + n_2 - 2}}$$

n_1, n_2 are the sizes of two groups and $s_{W_1}^2$, $s_{W_1}^2$ are 20% Winsorized variances. These variances are calculated as a common variance, but for modified datasets as follow. The data are ordered, 10% of it is deducted from the smallest values and 10% is deducted from the highest values and then replaced with the closest neighbor value. The scaling constant of 0.642 is required to reduce the influence of eliminating 20% of the initial dataset values.

In this situation, a bootstrap confidence interval is preferable, given that parametric hypothesis of normality is unlikely to be true. For each bootstrap sample, a bootstrap value of the effect size is determined, and so we get an ordered dataset by which we can choose one of the bootstrap ranges proposed in the first section.

The effect size is a value that gives information about the magnitude of the differences between treatments, being a dimensionless measure, property that reduces discrepancies between the different measurement units that may occur between different treatments. Also, the magnitude of the effect also offers the opportunity to embrace more independent research into a meta-analysis that provides overall more general and complex conclusions [6].

Conclusions

Parametric model testing such as t-test, ANOVA, or regression, is based on a few simple assumptions: independence of variables, normality of sample distributions or uniformity of residues.

When these requirements fail, we can either use rank tests or, more advanced, bootstrap computational procedures. Estimating confidence intervals and effect size based on bootstrap principles are new statistics that can be integrated into a meta-analysis study which allows us to drawn complex conclusions on the impact of various drugs on the population.

REFERENCES

1. Efron, B., Tibshirani, RJ. (1993). An Introduction to the Bootstrap. Chapman and Hall: London.
2. Fitts, D. A. (2011). Ethics and Animal Numbers: Informal Analyses, Uncertain Sample Sizes, Inefficient Replications, and Type I Errors. Journal of the American Association for Laboratory Animal Science: JAALAS, 50(4), pp. 445-453.
3. Charan, J., & Kantharia, N. D. (2013). How to calculate sample size in animal studies? Journal of Pharmacology & Pharmacotherapeutics, 4(4), pp. 303-306.
4. Bolton, S., Bon, C. (2004). Pharmaceutical Statistic. Practical and Clinical Applications, Marcel Dekker, Inc., New York Basel.
5. Patterson, S., Jones B. (2006). Bioequivalence and Statistics in Clinical Pharmacology, Published in Chapman & Hall/CRC Taylor & Francis Group.
6. Li, J.CH. (2016). Effect size measures in a two-independent-samples case with nonnormal and nonhomogeneous data, Behavior Research Methods 48(4), pp. 1560-1574.
7. Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.
8. Lee, D. K. (2016). Alternatives to P-value: confidence interval and effect size. *Korean Journal of Anesthesiology*, 69(6), pp. 555-562.
9. Hedge LV, Olkin I. (2014). Statistical methods for meta-analysis. Orlando: Academic Press Inc.
10. Hogarty, K. Y., & Kromrey, J. D. (2001). We've been reporting some effect sizes: Can you guess what they mean? Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
11. Algina, J., Keselman, H. J., & Penfield, R. D. P. (2005). An alternative to Cohen's standardized mean difference effect size: A robust parameter and confidence interval in the two independent groups case. Psychological Methods, 10, pp. 317-328.