International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Novel Hybrid Resampling Approach for Analyzing Immune Thrombocytopenia

Dr. A. Jagadish Kumar[1], Prof (Retd.) A.V. Dattatreya Rao[2]

[1]Biostatistician, Dept. of Biostatistics & Statistical Programming, Sanofi Health Care India Private Limited, Hyderabad, Telangana, India
[2]Retired Professor of Statistics, Dept. of Statistics, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India

Abstract
A Platelet Variability Index (PVI), reflecting both platelet count fluctuations and thrombocytopenia severity, distinguishes immune thrombocytopenia (ITP) from other thrombocytopenic causes (Na Li et al, 2021). Higher PVI scores correlated with definite ITP, even at initial assessment. Prior studies have used various metrics (e.g., average Standard Deviation, Coefficient of Variation (CV), Median Absolute Deviation, Mean Absolute Error) to represent individual patient platelet levels. This work focuses on the average CV. Given ITP's rarity and the fact that extreme variations in platelet counts are a key characteristic of the disease, we hypothesize that the distribution of average CVs may be modeled by an extreme value distribution. Specifically, we employ the Gumbel distribution, commonly used to model the maximum or minimum of a set of independent and identically distributed random variables, which is appropriate for capturing the extreme fluctuations in platelet counts observed in ITP. We then apply Bootstrapping, Jackknife resampling, and their convex combination to analyze the CV data.

Keywords: Rare Blood Disorder (RBD), Immune Thrombocytopenia (ITP), Gumbel distribution Bootstrap, Jackknife, Convex combination

1.INTRODUCTION

In this study, we analyzed platelet count data from patients with immune thrombocytopenia (ITP), a rare blood disorder. For each patient in both the treatment and control groups, we calculated the average Coefficient of Variation (CV) from platelet counts measured at multiple time points. Given the substantial variability typically observed in ITP platelet counts, we modeled the average CVs for both groups using the Gumbel distribution, an appropriate choice for extreme value distributions. To account for the smaller samples made available in the life situations we used Bootstrapping and Jackknife resampling techniques to compare the group means.

2.Methodology

Bootstrapping, a resampling technique, was employed to assess the stability and variability of our findings. By repeatedly resampling (with replacement) from the original dataset, we generated a large number of bootstrap samples. For each sample, the statistic of interest mean difference was calculated for creating a distribution of estimates. This allowed us to derive robust standard errors and confidence intervals, providing a non-parametric measure of uncertainty that does not rely on distributional assumptions. The bootstrap approach is particularly useful in situations where traditional analytical methods are challenging due to small sample sizes, complex model specifications, or deviations from distributional assumptions, offering a more reliable assessment of the true population parameter. In this study, bootstrapping (with 1000 resamples) was crucial for evaluating the confidence intervals of the interaction effect, assessing difference between Test and Reference treatments, thereby strengthening the validity of our conclusions.

The jackknife, a resampling technique similar in spirit to the bootstrap, was utilized to evaluate the stability and bias of our estimates. Unlike bootstrapping, which resamples with replacement, the jackknife creates new samples by systematically leaving out one observation at a time. For each of these "leave-one-out" samples, the statistic of interest was recalculated. This process yielded a set of pseudo-values, which were then used to estimate the bias and standard error of the original statistic. The jackknife is particularly useful for reducing bias in estimates and providing robust standard errors, especially when dealing with small sample sizes or complex estimators. In this study, the jackknife (with n = 140 for Test and n=70 observations) was employed to assess the difference between Test and Reference treatments platelet counts, providing a more accurate and reliable assessment of mean difference. This approach allowed us to strengthen the robustness of our findings and mitigate the potential influence of outliers or specific data points on our conclusions.

To leverage the strengths of both bootstrapping and jackknife resampling, we explored a convex combination of their respective estimators. This approach recognizes that bootstrapping provides robust standard errors and confidence intervals, particularly in complex scenarios, while the jackknife is effective in bias reduction. By combining the bootstrap and jackknife estimates, we aimed to achieve a balance between these desirable properties. Specifically, the combined estimator was calculated as a weighted average of the bootstrap and jackknife estimates, with the weights (λ and 1- λ, respectively) chosen to optimize a specific criterion, such as minimizing the mean squared error. This convex combination allowed us to potentially improve upon the performance of either method used in isolation. In this study, this hybrid approach was particularly relevant because the average difference between platelet counts of Test and Reference treatment groups. By carefully selecting the weighting parameter λ, we sought to minimize the variance of our estimator while controlling for potential bias and maximizing statistical power of the estimate, ultimately leading to more precise and reliable estimates of average difference between Test and Reference treatment groups and enhancing the overall robustness of our conclusions.

To evaluate the performance of various resampling techniques in the context of skewed data, we simulated two independent groups, a Test group (n=140) and a Reference group (n=70), from a Gumbel distribution with location parameter 0 and scale parameter 1. This distribution was chosen due to its common occurrence in modeling extreme value events and its inherent skewness, which poses challenges for traditional statistical methods. We then applied and compared the performance of bootstrapping, jackknife, and a convex combination of their respective estimators to estimate the difference in means between the two groups. Bootstrapping (with 1000 resamples) provided robust standard errors and confidence intervals, accounting for the non-normality of the Gumbel distribution. The jackknife, by systematically leaving out one observation at a time, was used to assess and potentially reduce bias in the mean difference estimate. Finally, a convex combination of the bootstrap and jackknife estimators was explored, aiming to leverage the strengths of both methods: the robust standard errors from bootstrapping and the bias reduction capabilities of the jackknife. This combined approach sought to optimize the trade-off between variance and bias, potentially yielding more accurate and reliable estimates of the treatment effect in this skewed and potentially imbalanced sample setting.

3.Results

Table 1: Sampling Characteristics of Bootstrap and Jackknife methods

	Bootstrap		Jackknife		True Parameter
	Test	**Reference**	**Test**	**Reference**	
Mean	0.5763 (0.0009)	0.3782 (0.1989)	0.6776 (0.1004)	0.4917 (0.0855)	0.5772
SD	0.0991 (1.1834)	0.1390 (1.1435)	0.0093 (1.2732)	0.0156 (1.2669)	1.2825
Median	0.5782 (0.2117)	0.3764 (0.0099)	0.6783 (0.3118)	0.4942 (0.1277)	0.3665

Note: The numbers within the brackets are Absolute Bias values. SD: Standard Deviation

1. The absolute bias of mean for the reference group is found to be minimum for Jackknife estimator.
2. The absolute bias of mean, SD and Median for the test group are found to be minimum for Bootstrap estimator.
3. The absolute bias of SD and Median for the reference group are found to be minimum for Bootstrap estimator.

Activity 1: Testing significance of mean effects under test and reference

i) Bootstrap

ii) Jackknife

Activity 2: Having observed the absolute bias of mean under reference group for Bootstrap is slightly on the higher side when compared to that of Jackknife, the test of significance of difference of means is conducted based upon the convex combination of Bootstrap and Jackknife.

Activity 3: Power of the test under the above two activities are computed and tabulated as follows.

Table 2: Power of the tests for Activities 1 & 2

Bootstrap	**Jackknife**	**Convex Combination: λ*J+(1- λ)*B**		
		$\lambda = 0.25$	$\lambda = 0.5$	$\lambda = 0.75$
100.00%	99.99%	100.00%	100.00%	100.00%

B: Bootstrap Statistic; J: Jackknife Statistic; λ: Coefficient of Convex Combination

4.Conclusions

Data pertaining to Immune Thrombocytopenia in general we have a small sample, because it is rare blood disorder. This kind of rare phenomenon can be modeled using any extreme value distribution such as Gumbel. Carrying data analysis based on a single sample pertaining to Test and Reference treatment groups may not give satisfactory decision for making efficiency of test drug compare to reference drug in clinical trials. Therefore, resampling procedures such as Jackknife and Bootstrap can be employed for better understanding the underlying sampling distribution hence the statistical inference has to be carried out. In case of testing the significance of difference of means of test vs reference, either Bootstrap or Jackknife can be used as a trial and error. If neither of these procedures is satisfactory in terms of power of the test procedure, then a convex combination of these two estimators can be tried to improve upon power of the tests. From the statistical theory it is well known that both Bootstrap and Jackknife estimators are consistent and will follow asymptotically normal. For this reason we are suggesting these logical steps to follow by practitioners.

Our experimentation and the observations found are presented as follows:

Since we assumed Immune Thrombocytopenia is a Rare Blood Disorder, for the purpose of illustration we generated a sample of 140 and 70 respectively for test and reference groups using standard Gumbel distribution with location parameter is 0 and scale parameter is 1 without loss of generality. Taking Level of Significance as 5% we conducted test for significance of difference of means between test vs reference treatment groups and found to be equally efficient under the resampling methods of Bootstrap and Jackknife with respect to power of the test as evidenced by results shown in table 2.

References

1. Na Li, et. al., 2021 "Platelet variability Index: A measure of platelet count fluctuations in patients with Immune Thrombocytopenia" , blood Advances, Vol.5, Number 20.
2. Zufferey A, Kapur R, Semple JW. "Pathogenesis and therapeutic mechanisms in immune thrombocytopenia (ITP)". J Clin Med. 2017; 6(2):16.
3. Arnold DM, Nazy I, Clare R, et al. "Misdiagnosis of primary immune thrombocytopenia and frequency of bleeding: lessons from the McMaster ITP Registry." Blood Adv. 2017;1(25):2414-2420.
4. Terrell DR, Beebe LA, Neas BR, Vesely SK, Segal JB, George JN. "Prevalence of primary immune thrombocytopenia in Oklahoma". Am J Hematol. 2012; 87(9):848-852
5. Jun Shao, Dongsheng Tu,"The Jackknife and Bootstrap" ,Text book published by Springer Series in Statistics, Oct 2012.
6. Michael R.Chernick, "Bootstrap Methods", A Guide for practitioners and Researchers, Wiley series, 2008
7. Cooper N, Bussel J. "The pathogenesis of immune thrombocytopaenic purpura". Br J Haematol. 2006; 133(4):364-374.
8. A.C.Davison, D.V.Hinkley, "Bootstrap Methods and their Applications", Cambridge Series in Statistical and Probabilistic Mathematics, 1997