Estimation of P-Value: Bootstrap a Viable Alternative to Permutation Approach

[1*]**Ikpotokin O. and** [2]**Oyegue F.O.**

[1]**Department of Mathematics and Statistics,**
Ambrose Alli University, Nigeria
[2]**Department of Mathematics,**
University of Benin, Nigeria

Abstract

In an attempt to ensure that the probability of a Type I error is exactly α, the p-value obtained through the permutation approach turns out to be the most reliable especially when complete enumeration is feasible for small sample sizes. The problem with the permutation approach has been high computationally demanding in terms of space and time complexities especially for fairly large sample sizes when complete enumeration is not feasible. The bootstrap approach is adopted as a viable alternative to the permutation approach especially when the numbers of observations are fairly large for a two - sample testing problem. Illustrative implementations are achieved and leads to the computation of accurate p-values.

Keywords: bootstrap, permutation, Type 1 error, p-values, observations, replications

1.0   Introduction

One of the fundamental goals in statistics is making inferences about the population mean, μ. Data are collected by sampling from the population in order to draw inference about the population based on the information contained in the sample. Theoretical approaches are known to be inaccurate when the underlying assumptions are violated, the usual case in practice. This leads to the adoption of the bootstrap and permutation approaches which is a statistical tool that based its inference upon repeated sampling within the same sample without considering the assumption of normality [1- 6]. Fisher, the founder of classical statistical testing developed the permutation test, as noted in [7-10]. The greatest virtue of permutation testing is its accuracy [11]. These tests are computationally intensive [12, 8]. Permutation test is a test procedure in which data are randomly re-assigned without replacement so that the statistic of interest is calculated [13].

Odiase and Ogbonmwan [9] made it possible to obtain unconditional exact permutation distribution by ensuring that a complete enumeration of all the distinct permutations of any two sample experiment is achieved, especially when sample sizes are small. However, it is computationally demanding to generate all possible distinct rearrangements in term of space and time complexities even with the latest and vary fast processor speed of available personal computers when the volumes of figures involved are so large. For instance, if m = 20 and n = 20, there are 137,846,528,820 permutations. As a result of these difficulties, numerous alternative methods have been proposed, and one of such is the bootstrap method. This work shall adopt the bootstrap approach because it is computationally less demanding in term of space and time complexities to generate large bootstrap samples.

The Bootstrap was found to be a powerful tool for statistical inference that substitutes's raw computing power for theoretical analysis [14, 15]. Hesterberg [16] and Teknomo [17] illustrated the one sample bootstrap re-sampling. Berger [13] used the bootstrap methods as sampling with replacement to establish confidence intervals. Demiralp et al [18] used the bootstrap regression for assessing the confidence in the result of a search for causal dependencies. Chatterjee and Qiu [19] estimate the bootstrap control limits for a statistical process control when both the in-control and out-of-control distributions of the process are unknown. Sen et al [20] provide a set of sufficient conditions for the consistency of any bootstrap method by bootstrapping from a smoothed version. In this study, consideration is given to the bootstrap distribution of two sets of observation. A practical issue, however, is whether a

*Corresponding author: **Ikpotokin O.,** E-mail: ikpotokinosayomore@yahoo.co.uk, Tel. +234 8039289270

permutation method provides better control over the Type I error probability, versus a bootstrap method, when complete enumeration is feasible for small to moderate sample sizes. In section 2, bootstrap algorithms developed for two sample experiment to generate bootstrap samples shall be discussed.

2.0   The bootstrap algorithms for two sample experiment

Suppose a two sample experiment is drawn from the same population distribution. If the first set denotes x_1, x_2, x_3, x_4, x_5 and the second set denotes y_1, y_2, y_3, and observed that $m = 5$ and $n = 3$, a pool of the observations into a single sample yield:

$$\theta = (x_1, x_2, x_3, x_4, x_5 \ y_1, y_2, y_3) \qquad (1)$$

The number of bootstrap samples with replacement from the original sample (1) is:

$$\theta^* = (x_1^*, y_2^*, y_3^*, x_5^*, y_2^*, x_3^*, x_2^*, x_2^*) \qquad (2)$$

Note that the actual process involved in obtaining (2) is that, randomly assign them to an ordered $(x_1^*, \dots, x_m^*, y_1^*, \dots, y_n^*)$. So that an x_i^* can also be assigned the value of y_j in the original sample, and vice versa.

To obtain $\hat{\theta}^* = \bar{x}^* - \bar{y}^*$, separate (2) into two parts according to the size of x and y from the original sample, so that (3) becomes

$$x^* = (x_1^*, y_2^*, y_3^*, x_5^*, x_2^*) \ and \ y^* = (x_3^*, y_2^*, x_2^*) \qquad (3)$$

Calculate the group means and the mean difference as follows:

$$\bar{x}^* = \frac{1}{m}(x_1^* + y_2^* + y_3^* + x_5^* + x_2^*) = \frac{1}{5}\sum_{i=1}^{5} x_i^* \qquad (4)$$

$$\bar{y}^* = \frac{1}{n}(x_3^* + y_2^* + x_2^*) = \frac{1}{3}\sum_{j=1}^{3} y_j^* \qquad (5)$$

$$\hat{\theta}^* = \frac{1}{5}\sum_{i=1}^{5} x_i^* - \frac{1}{3}\sum_{j=1}^{3} y_j^* = \bar{x}^* - \bar{y}^* \qquad (6)$$

Repeat process (2) to (6) B times to obtain bootstrap configurations. To obtain the bootstrap p-value(P_{boot}) for mean difference of left-tailed test, count the number of mean difference ($\hat{\theta}^*$) from the bootstrap data that is less than or equal to the mean difference ($\hat{\theta}$) of the original data and then divide by the number of bootstraps sample performed, and vice versa for a right-tailed test.

$$\text{i.e.}(P_{boot}) = \frac{1}{B} \#\{\hat{\theta}^* \le \hat{\theta}\} \ \ or \ \ \frac{1}{B} \#\{\hat{\theta}^* \ge \hat{\theta}\} \qquad (7)$$

On the other hand, five steps considered by Good [8] were followed by Odiase and Ogbonmwan [9] to generate a complete enumeration of all the distinct permutations of the experiment required. When this is achieved, the statistic of interest is computed for each of the permutation. The test statistic in this case is the mean difference and is denoted by:

$$\hat{\theta}_{pem} = \bar{x} - \bar{y} \qquad (8)$$

and can be expressed as:

$$\hat{\theta}_{pem} = \frac{1}{m}\sum_{i=1}^{m} x_i - \frac{1}{n}\sum_{j=1}^{n} y_j \qquad (9)$$

The distribution that put probability $1/\left(\frac{(N)!}{n! \, m!}\right)$ on each one of these statistic is called the permutation distribution of $\hat{\theta}_{pem}$. To obtain the permutation p-value(P_{perm}), calculate the fraction of times where the observed mean difference exceeds the mean difference in the permutations.

$$(P_{perm}) = \frac{1}{\frac{N!}{n! \, m!}} \#\{\hat{\theta}_{pem} \le \hat{\theta}\} \qquad (10)$$

In what follows in section 3 is the presentation of data and results from practical applications of both the bootstrap and permutation approaches.

3.0   Presentation of Data

Appendix 1 represent the daily pocket money (in naira) collected randomly by 10 students from Mathematics and 12 students from Statistics in the Department of Mathematics, Ambrose Alli University, Ekpoma Edo State, Nigeria. If the population variances of the student's daily pocket money are the same for the two options, test the hypothesis that the difference between the average daily pocket money received by the entire students in the Department is N50, for $\alpha = 0.05$ level of significance. Appendix II denotes the data from Freund [21], page 592, exercise 16.30. An examination designed to measure basic knowledge of American history was given to random samples of freshmen at two major universities, and their grades. Test at the 0.05 level of significance the null hypothesis that there is no difference in the average knowledge of American history between freshmen entering the two universities.

3.1   Results from the Bootstrap Approach

The bootstrap algorithms developed in section 2 was translated to Visual Basic code to obtain the bootstrap mean difference generated as presented in Table 1. The bootstrap sample generated was replicated 3000 times (B = 3000). The observed mean difference from the original data is calculated as 22.75, its probability value is 0.084 and Figure 1 displayed the bootstrap distribution.

Table 1: Bootstrap Mean Difference

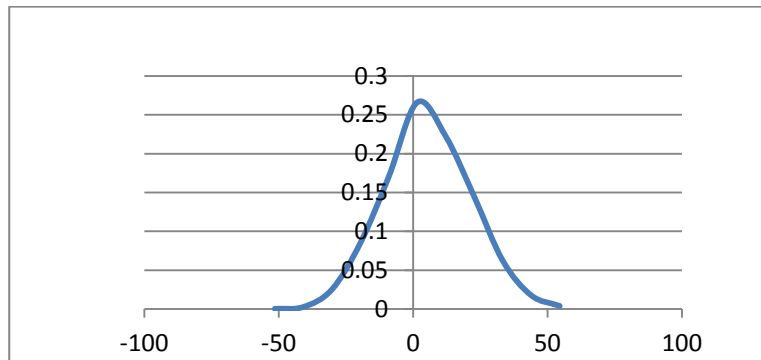| Mean Difference	Probability	Frequency
-51.4592	0.000333	1
-40.8593	0.002667	8
-30.2594	0.025	75
-19.6595	0.084667	254
-9.05966	0.170333	511
1.540226	0.266	798
12.14011	0.222333	667
22.74	0.144667	434
33.33989	**0.062333**	**187**
43.93977	**0.017667**	**53**
54.53966	**0.004**	**12**
	Sum Freq	**3000**

Figure 1: Bootstrap Distribution of Mean Difference

3.2   Results from the Permutation Approach

This approach is based on the assumption that all observations are exchangeable [8 – 9]. For the data on student's daily pocket money, the number of distinct permutation is $\frac{(22)!}{10!\,12!} = 646,646$. The observed mean difference from the original data is calculated as 22.75, its probability value is 0.097851 and Figure 2 displayed the permutation distribution.

Table 2: Permutation Mean Difference

| Mean Difference	Probability	Frequency
-57	1.55E-06	1
-56.0833	1.55E-06	1
-55.1667	3.09E-06	2
.	.	.
.	.	.
.	.	.
22.75	**9.17E-03**	**5930**
23.66667	**8.61E-03**	**5570**
24.58333	**8.01E-03**	**5182**
.	.	.
.	.	.
.	.	.
57.58333	**7.73E-06**	**5**
58.5	**4.64E-06**	**3**
59.41667	**3.09E-06**	**2**
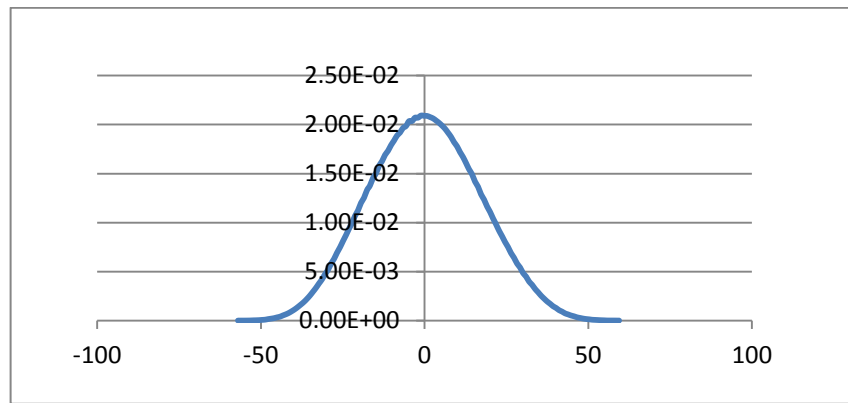
Figure 2: Permutation Distribution of Mean Difference.

3.3   Results showing the advantages of Bootstrap over the permutation Approach

The second data set demonstrate the advantages of bootstrap method over the permutation method in terms of time and space complexities. The observed mean difference from the original data is calculated as 12.25, its probability value is 0.00533 and Figure 3 displayed the bootstrap distribution.

Table 3: Bootstrap Mean Difference on second data

Mean Difference	Probability	Frequency
-19.351	0	0
-16.1919	0.000667	2
-13.0328	0.004333	13
-9.8737	0.020333	61
-6.7146	0.069	207
-3.5555	0.147	441
-0.3964	0.253	759
2.7627	0.239333	718
5.9218	0.153333	460
9.0809	0.083	249
12.24	0.024667	74
15.3591	0.004667	14
18.5182	0.000667	2
	Sum Freq	**3000**
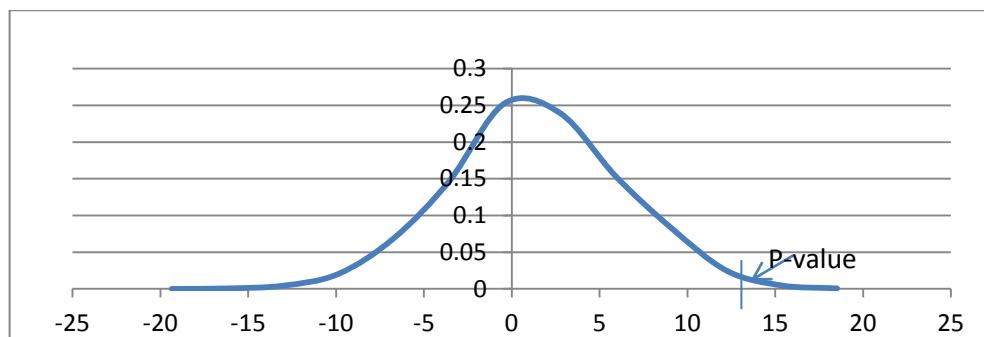
Figure 3: Bootstrap Distribution of Mean Difference

3.4 Results from the Permutation Approach

However, permutation approach for the set of data became impossible because the complete enumeration involved is very large, (i.e. m = 20 and n = 20, then $\frac{(N)!}{n!\,m!} = \frac{(40)!}{20!\,20!} = 137{,}846{,}528{,}820$. Nevertheless, discussions of results and their various conclusions are presented in sections 4 and 5 respectively.

4.0 Discussion of Results

A performance study of the two approaches was done using two empirical data sets. Bootstrap samples are generated 3000 times from the data in appendix I. Analysis carried out for bootstrap mean difference is presented in Tales 1 and the plot which exhibited the distribution of a normal curve is shown in Fig. 1. Similarly, Table 2 present the analysis carried out for permutation mean difference and the plot which displayed the distribution of a normal curve is shown in Fig. 2. Estimated p-values from both the bootstrap and permutation mean differences are 0.084 and 0.097851 respectively. These compared with $\alpha = 0.05$ level of significant (0.084 and 0.098 > 0.025) revealed that the null hypothesis (H_0) should not be rejected. However, results of the data in appendix II presented the advantages of bootstrap method over the permutation method in terms of time and space complexities. Analysis carried out for bootstrap mean difference is presented in Tables 3 and the plot which exhibited the distribution of a normal curve is shown in Fig. 3. Estimated p-value from the permutation became difficult because of the enumeration involved is very large, i.e. 137,846,528,820. Thus, estimated p-value from the bootstrap approach when compared with $\alpha = 0.05$ level of significant (0.004333 < 0.025); revealed that the null hypothesis (H_0) must be rejected.

5.0 Conclusion

In general, a method of computing the bootstrap p-value is given and the resultant p-value ensured that the probability of making a Type 1 error is approximately α. With a small sample size, the estimated p-value indicate that the bootstrap method is generally approximate to the exact p-value obtain from the permutation method. As the sample size gets larger, situations are found where the bootstrap method performs better than the permutation method.

Appendix I:
Data on daily pocket money (in naira) collected randomly by 10 students in Mathematics and 12 students in Statistics
Mathematics Students (X):50, 100, 75, 80, 90, 120, 80, 160, 120, 40
Statistics Students (Y): 45, 55, 30, 50, 110, 70, 50, 65, 40, 55, 75, 180
Source: Department of Mathematics, Ambrose Alli University, Ekpoma, Edo State, Nigeria

Appendix II:
Data on average grade knowledge of American history between freshmen entering the two universities
University A: 77, 72, 58, 92, 87, 93, 97, 91, 70, 98, 76, 90, 62, 69, 90, 78, 96, 84, 73, 80
University B: 89, 74, 45, 56, 71, 74, 94, 88, 66, 62, 88, 63, 88, 37, 63, 75, 78, 34, 75, 68
Source: Freund (1992), page 592, exercise 16.30

References
[1] Efron B. (1979): Bootstrap methods; another look at the jackknife. The Annals of Statistics, 7, 1-26
[2] Efron B. (1982): The jackknife, the bootstrap and other resampling plans. Society of Industrial and Applied Mathematics, CBMS-NSF Monographs, 38, 1 – 3.
[3] Efron B. and Tibshirani R.J.(1993): An introduction to the bootstrap. New York; Chapman and Hall.
[4] Lunneborg C.E. (2000): Data analysis by resampling. Concept and Applications, Pacific Grove, CA: Duxbury.
[5] Fox J. (2002): Bootstrapping Regression Models. Appendix to an R and S-plus Companion to Applied Regression @ http://fox.mht
[6] Xu K.M. (2005): Using the Bootstrap Method for a Statistical Significance Test of Differences between Summary Histograms. Monthly weather Review (online)
[7] Agresti A. (1992): A Survey of Exact Inference for Contingency Tables. Statistical Science 7(1), 131-177.
[8] Good P. (2000): Permutation Tests: A practical guide to resampling methods for testing hypotheses, (2nd ed.) New York: Springer-Verlag.
[9] Odiase J.I. and Ogbonmwan S.M.(2005): An algorithm for generating unconditional exact permutation distribution for a two-sample experiment. Journal of Modern Applied Statistical Methods. 4(1): 319-332.
[10] Odiase J.I. and Ogbonmwan S.M. (2007): Exact Permutation algorithm for paired observations: The challenge of R.A. Fisher. Journal of Mathematics and Statistics 3(3):116-121.
[11] Fahoome G. (2002): Twenty nonparametric statistics and their large sample applications. Journal of Modern Applied Statistical Methods 1(2), 248-268.
[12] Edginton E.S. (1995): Randomization tests. New York. M. Dekker.

[13]   Berger (2006): Introduction to Resampling Techniques @ http://wfs.cgu.edu/bergerd/papers/resampling.xls

[14]   Hinkley D.V. (1988), Bootstrap methods. J.R.Stat. Soc. B50, 321-337.

[15]   Hesterberg T.C., Monaghan S., Moore D.S., Clipson A., and Epstein R. (2003): Bootstrap methods and Permutation tests. W.H. Freeman, New York.

[16]   Hesterberg (1997): Fast bootstrapping by combining importance sampling and concomitants. Computing Science and Statistics, 29(2), 72-78

[17]   Teknomo K. (2006): Bootstrap Sampling Tutorial @ http://people.revol.edu.com/kard/tutorial/bootstrap

[18]   Demiralp S, Hoover K.D. and Perez S.J. (2008): A Bootstrap Method for Identifying and Evaluating a Structural Vector Auto-regression, Oxford Bulletin of Economics and Statistics 70, 509-533.

[19]   Chatterjee S. and Qiu P. (2009): Distribution – free cumulative sum control charts using bootstrap based control limits. Annals of Applied Statistics, 3(1): 349 – 369.

[20]   Sen B., Benerjee M. and Woodroofe M. (2010): Inconsistency of Bootstrap: The Grenander Estimator, submitted to the Annals of Statistics, @ http://www.stat.isa.umich.edu/~michaelw

[21]   Freund T.E. (1992): Mathematical Statistics 5th edition. Prentice Hall, Englewood Cliffs, New Jersey.