Original Article

Comparison of sample and re-sampling techniques in the estimation of third trimester obstetric parameters

Arumugam P[1*], Keshav Singh[2], Rajendran A K[3], Kasthuri A[4]

[1]Professor of Biostatistics, [2,4]Associate Professor, [3]Professor and HOD, Department of Community Medicine, Sri Muthukumaran Medical College and Research Institute, Chennai-600 069, Tamil Nadu, INDIA.
Email: pitchaiarumugam47@gmali.com, drkessingh@gmail.com, akrajendran@rediffmail.com, kasthumohan@gmail.com

Abstract

Background: Obstetricians interested in estimating the third trimester parameters, can find the outliers by estimating the crucial boundaries of these parameters. The estimation may be made by two approaches namely random sampling and re-sampling techniques (bootstrap sampling). **AIM:** Comparison between random sample and Bootstrap sample parameters **Objectives:** To calculate the statistics of the two procedures by means of averages. To estimate the population parameters by both methods. To interpret the difference statistically. **Methodology:** A random sample of 240 III trimester ante-natal mothers was selected as first 20 singleton pregnant mothers through the months of January to December 2014and among them 154 and 86 mothers were primi para and multi para respectively obtained from the records of a teaching hospital in Tamil Nadu. The Ultra Sona Graphed statistics like Head Circumference (HC), Abdominal Circumference (AC), and Placental Thickness (PT) were collected. The statistics and parameters of primi mothers (154) were computed by both techniques and the difference between the variance was inferred by variance ratio test with the help of IBM SPSS Version 20. **Results:** The following obstetric statistics viz. mean median SD and SE calculated by sample and Bootstrap methods are furnished. **Discussion:** The difference between the variances were not statistically significant (P>0.05). Hence the difference between Std. errors were also not statistically significant (P>0.05). And thus the parameters will also be not statistically significant (P>0.05). **Conclusion:** As the sample size of 154 was large, the estimated parameters of both techniques may not be significant. The Bootstrap sample parameter may be a more appropriate estimation in respect of small samples.

Keywords: Random sample, Bootstrap sample, SE, estimation, Parameters.

[*]**Address for Correspondence:**
Dr P Arumugam, Professor of Biostatistics, Department of Community Medicine, Sri Muthukumaran Medical College and Research Institute, Chikkarayapuram, Mangadu, Chennai-600 069, Tamil Nadu, INDIA.
Email: pitchaiarumugam47@gmali.com

Received Date: 07/06/2016 Revised Date: 17/07/2016 Accepted Date: 02/08/2016

Access this article online

Quick Response Code:

Website: www.statperson.com

DOI: 06 August 2016

INTRODUCTION

An inferential statistics is dealt with two parts namely estimation and test of significance. The estimation is an essential tool in the context of infinite population to find out the averages and such averages will fall within a range of two limits. The two limits are called confidence limits and the interval within the limits is called Confidence Interval (CI). The confidence intervals are being calculated as follows. The standard error is multiplied by the respective confidence coefficient and the product is subtracted from the average. The subtracted value of the average is called lower limit and the product added with the average is called upper limit. The standard error (SE) is the variance (SD^2) divided by the sample size (n). The SE is the determining factor of the confidence interval (CI). In this study the SE is being computed by two approaches namely random or probability sampling and Bootstrap sampling. The random or probability sample is 10^{th} century procedures of sampling technique and Bootstrap sampling is 20^{th}

How to site this article: Arumugam P, Keshav Singh, Rajendran A K, Kasthuri A. Comparison of sample and re-sampling techniques in the estimation of third trimester obstetric parameters. *International Journal of Recent Trends in Science and Technology* August 2016; 20(1): 41-45 http://www.statperson.com (accessed 08 August 2016).

century re sampling procedures. The re-sampling technique is taking as many samples as required within the same sample by using replacement method.

Random sample

The random sample is defined as the best procedure for selecting such a sample is to use probability. In probability sampling the subjects are selected randomly in such a way that the researcher knows the probability of selecting each member of the population. Random selection implies that each member of the population as a whole or of subgroups of the population has an equal chance of being selected[1].

Evolution of Random Sample

The historical evidences like Uthiramerur (Kanchipuram Dist, Tamil Nadu, India) Inscriptions are giving some evidences for random selection. The Chola King Parantakan-I ruled Uthiramerur in 907-955 AD. During the period the members of ward committee of Uthiramerur village was elected by selection process Kudavolai system [2]. The inscription explains the election system of ward members as follows. The system is called Kudavolai system (Kudam means pot and Volai means palm leaves). Each village was divided into 30wards. From each ward one member was elected by random selection. The member to be elected should have the qualifications such as were (1) Age- in between 35-70 years, (2) Immovable property -minimum 1.54 acres of tax payable land and house, (3) Education- literate in Mantras and able to teach others. The period of tenure was 1 year. And there were 21 disqualifications. All the eligible candidates names were written in palm leave of each ticket and make it as packet of that ward. Thirty such packets were prepared and kept with the custody of village electoral officer nominated by the King (Royal Order). All the 30 packets would bring to the great meeting, which was held at the Great Assembly [Fig-1], including the young and old members, shall be convened. In the midst of the temple priests' one of them, who happens to be the eldest, shall stand up and lift that pot looking upwards so as to be seen by all people. One ward, i.e., the packet representing it, shall be taken out by any young boy standing close, who does not know what is inside, and shall be transferred to another empty pot and shaken. From this pot [Fig-2] one ticket shall be drawn by the young boy [Fig-3] and made over to the arbitrator 3. The arbitrator would receive the ticket in the palm with opening of five fingers. He read out the name written in the ticket and all the priests assembled in the hall follow to read name. The name thus read out shall be put down (and accepted). Similarly one man shall be chosen for each of the thirty wards. The process of selection was purely simple random selection by lottery method. In this election process wards were clusters and the population was the target population. The eligible contestant's packet was the sampling frame. Now also we are using the same technique as simple random sampling in unbiased true representative of the target population.

Development of Random Sample

The Kuda Olai system is the origin of random sampling. After that, the random selection was used in Germany during 1553 for testing Mines. The system was called ("**Stich probe**") a spoonful of mines and it is similar to testing of rice in boiling pot 4. In 1602-1662 the Father of Demography John Graunt made the systematic analysis of vital events and estimated the population of England by selecting sample 5. Next to John Graunt, Pierre

Simon Laplace (1749 – 1827) made an attempt to estimate the French population by selecting 30 clusters distributed over the area of France. He used 30 clusters and sampling was the non random sampling. But he used the central limit theorem and proved that his estimator was normal distribution [6]. Nicolaas Struyck estimated the world population to be 500 million and a death rate of 35 per 1000 population with an hourly 2000 deaths [7]. The first separated Netherlands statistician Lobatto applied probability in estimation of true value[8]. Anders Kiaer (1838-1919), the first Norway Director of Statistics was first to use sampling in collection of proposed retirement and sickness insurance scheme throughout Norway[9]. This method was called Representative Method. This method was criticized by Bowley and emphasized the importance of random sampling in 1906. Bowley proved that the large samples selected from the population, the estimates have a normal distribution approximately[10]. Up to 1934 the two methods of purposive and random samples were existing. In 1934, the Polish scientist Jerzy Newman introduced the concept of confidence interval in estimation of population parameters. He had established the random or probability sampling was superior to representative sampling[11].

R. A. Fisher introduced randomization and probabilities in agriculture experiments[12]. Mahalanobis (1950) described that the costs of sample survey with an acceptable level of precision are only about 10% of a complete enumeration[13]. The same opinion was thrown by R.A. Fisher and stated that sampling was a more scientific method than complete enumeration. Around 1950 more number of renowned statisticians was positively opined sampling is better than census method[14].

Bootstrap sampling

In view of minimizing the cost involved in random sampling, Bradley Efron in 1979 published a paper by taking more number of samples was taken in single small sample[15].

What is bootstrapping? The name bootstrapping came from the phrase "To lift himself up by his bootstraps" [Fig-4][16 & 17]. The method was called as simple random sampling with replacement without more cost. Traditionally, we are adopting the simple random sample by using either with or without replacement procedure. In respect of without replacement, it is not possible to adopt uniform probability. For adopting the uniform probability, the replacement method is being adopted. In this procedure the reoccurrence of items are disallowed. But in bootstrapping procedure the reoccurrence of items are allowed to get the required sample size which is equivalent to the original random sample size. Such a sample is possible for to take as many samples are required [18]. The advantage of bootstrapping method was its simplicity and calculations of Standard Error (SE) and confidence intervals. At the same time it will not give general finite- sample guarantees as a disadvantage. It is being used for computations some analytical statistic like averages and variance. The basic methodology was adopted by Monte Carlo algorithm for a case re sampling procedures.

METHODOLOGY

A random sample of 240 III trimester ante-natal mothers was selected as first 20 singleton pregnant mothers through the months of January to December 2014 from the obstetric records of the hospital. Among them 154 and 86 were primi and multi paras respectively. The data were obtained from the records of a teaching hospital in Tamil Nadu. The Ultra Sona Graphed statistics like Head Circumference (HC), Abdominal Circumference (AC), and Placental Thickness (PT) were collected. The statistics and parameters of primi mothers (154) were computed by both techniques and the difference between the variance was inferred by variance ratio test since the standard errors are the estimator of the population parameter like mean. The above statistical procedures were performed with the help of statistical package IBM SPSS statistics -20. The P- values less than or equal to 0.05 ($P \leq 0.05$) were considered as statistically significant.

RESULTS

The statistics viz. mean median SD and SE were calculated for random sample and Bootstrap sampling methods and the results are furnished. The parameters of the mean were estimated @ 95% CI in both methods for the obstetric variables like Head Circumference, Abdominal Circumference and Placental Thickness.

Table 1: The obstetrics statistics and parameter computed by random sample

Variable	Mean	Median	SD	SE	Confidence Interval (95%)	
H C	308.2468	312.0	21.0301	1.6947	304.8988	311.5947
A C	302.9935	309.0	30.2649	2.4388	298.1754	307.8116
P T	34.6299	35.0	3.9390	0.3174	34.0028	35.2569

Table 2: The obstetrics statistics and parameter computed by Bootstrap sample

Variable	Mean	SE	Bias	SE	Confidence Interval(95%)	
H C	308.2468	1.6946	-0.0173	1.6461	305.0067	311.2597
A C	302.9935	2.4388	-0.0512	2.3555	298.1642	307.7313
P T	34.6299	0.3174	-0.0086	0.3096	33.9870	35.2401

Table 3: Difference between the Std. Errors of sample and Bootstrap sample

Variable	Random sample variance	Bootstrap variance	F	Significance
Head Circumference	442.266	417.307	1.059	P>0.05
Abdominal Circumference	915.967	854.496	1.071	P>0.05
Placental Thickness	15.516	14.762	1.051	P>0.05

The table-1 describes the HC, AC and Placental thickness in terms of averages, SD, SE and 95% confidential intervals of the means by random sample method. The means of the variables were estimated. In the estimation, the SE of the mean was taken in to consideration. The estimated parameters of the populations' HC, AC and Placental thickness were 304.8988 to 311.5947, 298.1754 to 307.8116 and 34.0028 to 35.2569 @ 95% CI respectively. Similarly, the above statistics were calculated by bootstrap sampling method and the results were tabulated in table-2. The estimated parameters of the populations' HC, AC and Placental thickness were 305.0067 to 311.2597, 298.1642 to 307.7313, and 33.9870 to 35.2401@ 95% CI respectively. The difference between parameters was the consequence of the changes in the calculation of SE by two procedures. The difference between the standard errors of both procedures was tested for significance. The calculated variances of both procedures of HC, AC and Placental thickness were tabulated in table-3. The differences between the random and bootstrap methods of HC, AC and Placental thickness were not statistically Significant ($P>0.05$).

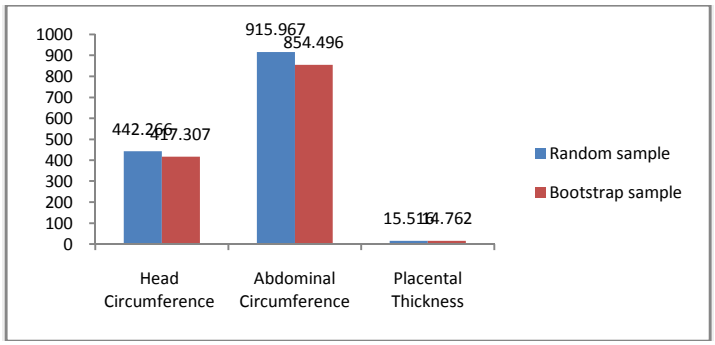

Figure 5: Difference between the variances of both procedures of HC, PC & PT

Figure 1: The Election office at Uthiramerur of Kanchipuram Dist. (907-955 AD) Facade of Vaikundaperumal Temple - General Village Assembly (Maha sabha) of Uttaramerur- Chaturvedimangalam

Figure 2: The Election Box

Figure 3: The boy selecting a Palm Leaf

Figure 4: The symbol of bootstrap

DISCUSSIONS

The head circumference variance was 442.266 in random sample and the bootstrap sample was 417.307 and it was observed as lesser than the random sample. The lesser quantity of variance was reflected in the population parameter estimation of head circumference. The estimated head circumferences were also not statistically significant (P>0.05). The variances of both methods of abdominal circumference were 915.967 and 854.496. The random sample calculated variance was greater than the bootstrap method and the same was not statistically significant (P>0.05). The variances of both methods of placental thickness were 15.516 and 14.762. The random sample calculated variance was greater than the bootstrap method and the same was not statistically significant (P>0.05). The random sample was adopted in Tamil Nadu during 10th century by the Chola king Parantaka Cholan I in the election process of ward members by creating 30 clusters of so called Chadurvedi Mangalam. Now it is being called as Uthiramerur, the birth place of democracy[19]. Cluster sampling involves the random selection of naturally occurring groups or areas and then the selection of individual elements from the chosen groups or areas[20]. The literature simply defined the Cluster (Area) Random sampling to follow these steps 1. Divide population in to clusters (wards) 2. Randomly sampled clusters (all wards) and 3. Measure all units within the sampled clusters (qualified candidates) [21]. During 18th century, Pierre Simon Laplace attempted to estimate the France population by collecting 30 clusters. But he had not adopted random sampling. He estimated the France population standing on the steps of De Moivere (1667-1754)[22]. Laplace argued that the 30 and above took normal distribution based on the central limit theorem. Historically thirty was considered as cut point for small and large sample. The William Gosset's Students "t" test probability table, the normality of 95% ($t_{\alpha/2}$ – confidence coefficient) will take only after 120 and up to ∞ as 1.96 ($Z_{\alpha/2}$ – confidence coefficient)[23] and it is

presumed that the sample may not be random sampling. Because of that the 't' probability curve is widened in the X axis and flat in the top of the curve.

CONCLUSION

The random sample which is either simple or restricted we are taking the cut point is 30 in size for considering large sample since it will take normality and continuity. According to William Gosset, the sample size 120 and above as large sample irrespective of the selection randomly or not and it takes continuity and normality. The sample which is going to be bootstrap sampling it was not known whether it was a random sample or non random sample. A small in size may be manipulated to more number of samples to form a sampling distribution. The manipulation process may be cumbersome and it needs full computer assistant as well as statistical soft ware and its essential knowledge. In this study, it was proved that there was no significant difference between the two methods of parameter estimation. By considering the simplicity and a sample size 120 or more, irrespective of either random or non random may be considered for estimation of population parameter @ 95% CI with confidential coefficient of Z.

ACKNOWLEDGEMENT

The authors acknowledged the Archeology Department of India for using their Inscriptions in this study and Miss. P. Harini for rendering her support in job works.

REFERENCES

1. Educational Research, Fundamentals for the Consumer; II Ed. JAMES H. MCMILLAN. Virginia Commonwealth University, Chapter-4 Page- 87.
2. Muthusamy, R. Uthiramerur Inscriptions on Chola Kudavolai Election System. Internet- www. File///C:/ Documents.
3. Face book logo: Internet- www. File///C:/ Documents.
4. Jelke Bethlehem, The rise of survey sampling, Discussion Paper Statistics Netherlands. The Hague/Heerlen, 2009. Pages:6,
5. Ibid (4)- page-6.
6. Ibid (4)- page-6
7. Ibid (4)- page-7
8. Ibid (4)- page-8
9. Ibid (4)- page-10
10. Ibid (4)- page-12
11. Ibid (4)- page-14
12. Ibid (4)- page-16
13. Ibid (4)- page-16
14. Ibid (4)- page-19
15. Kesar Singh and Minge Xie, Bootstrap: A Statistical Method, Rutgers University.
16. Mooney & Duval, Bootstrapping, A nonparametric Approach to Statistical Inference. Series on Quantitative Applications in the Social Sciences, Sage University Papers Seminar, General Statistics, 25 October 2012.
17. Wikipedia. Why the Name Bootstrapping?
18. Bootstrap Methods AND Permutation Tests. W. H. Freeman and Company New York- Chapter -18.2: Page-7.
19. Saravanan Iyer, Uthiramerur Sri Sundara Varadhar: Hind Pedia.
20. Ibid (1) Page-90.
21. Ibid (3)
22. Ibid (4) - page-8.
23. Frerichs, R.R. Rapid Surveys (unpublished), © 2008, Chapter-3. Simple Random Sampling, Page: 12.

Source of Support: None Declared
Conflict of Interest: None Declared