ISSN 1684-8403
Journal of Statistics
Volume 24, 2017. pp. 159-165

Estimation of mean CGPA using Bootstrapping: A Study of GC University Lahore Students

Azeem Ali[1]*, Farah Hassan* [2]*, Iqra Afzal* [3]*, Farrukh Shehzad*[4] *and Muhammad Akbar Ali Shah*[5]

Abstract

In surveys, respondents whether avoid to answer or give incorrect information about sensitive questions. Both of the cases cause for biasness in estimates. The main focus of this study is to tackle the problem of biasness in estimate through Bootstrapping technique. Mean CGPA of the students of GC University Lahore is estimated and compared with that of the subpopulations.

Keywords

Bootstrapping technique, CGPA, Biasness

1. Introduction

Collecting information through surveys is a common practice in observational research. The information through these surveys fails to produce unbiased results when a questionnaire contains some sensitive questions. A question is said to be sensitive if the respondent hesitates to answer correctly. Normally questions about income, age, smoking, etc. are considered as sensitive. Sensitivity of questions varies for different group of people. For example, information about age is a sensitive question for female same as about income for males. Similarly collecting information about Grades or CGPA is a sensitive for students.

[1] Department of Statistics, GC University Lahore, Pakistan.
Email: syedazeemali@gmail.com
[2] Department of Statistics, GC University Lahore, Pakistan.
Email: farahhassan30@yahoo.com
[3] Department of Statistics, GC University Lahore, Pakistan.
[4] COMSAT Institute of Information Technology, Lahore, Pakistan.
[5] Department of Statistics, The Islamia University of Bahawalpur, Pakistan.

Efron (1979) introduced Bootstrap method who was inspired by Jackknife technique. A Bootstrap method is a resampling technique specially used to compute variance and bias estimation.

The idea behind Bootstrap method is to resample (with replacement) from the sample data for the purpose of approximating the sampling distribution of a statistic. The sample statistics are then computed on each of the Bootstrap samples (usually a few thousand). A histogram of the set of these computed values is referred to as the Bootstrap distribution of the statistic. Any statistic can be estimated for this sampling distribution and relates to the data under study.

One standard choice for an approximating sampling distribution is the empirical distribution function of the experimental data. Bootstrap method works when a set of observations are assumed to be independent and identically distributed. Following this technique a number of resamples from a dataset are drawn with replacement and of equal size. Bootstrapping deals with misinterpretations caused by the sample not adequately representing population under study.

In this study, mean CGPA of students of Government College University Lahore is estimated. Bootstrapping technique is applied to minimize biasness caused by incorrect information about sensitive questions.

This paper is organized in four sections. Section 1 consists of introduction and background. Section 2 contains materials of the study and method of research. Section 3 summarizes results obtained from collected data and discussion. Section 4 includes conclusion drawn from results.

2. Material and methods

The population under study consists of the currently registered 9901 undergraduate and postgraduate students at GC University Lahore. A sample of 100 students is drawn to estimate mean CGPA of overall students. A self-explanatory questionnaire is designed to collect the data about the subject under study.

Several probability distributions are applied on selected sample in order to check best fitted distribution. Five of the following distributions named as Skew Normal distribution (Azzalini, 1985), Nakagami distribution, Rice distribution (Rice, 1945), Tsallis Q Gaussian Distribution (Leeuwen and Maassen, 1995) and Von Mises distribution (Best and Fisher, 1979), provided relatively good and closer fit. Probability density functions and relevant properties of these distributions are furnished in the Table 1.

3. Result and discussion

The collected data from 100 students then cross matched with the original CGPA taken from the examination branch. Only 41 students provide their correct CGPA, while 48 students report incorrect CGPA and the information of remaining 12 students were not matched with that of the examination branch.

From the data of 41 students (who have given correct information), 1000 random numbers generated through Bootstrapping using Mathematica version 9.0. Five of the distributions, which provide good fit, tested for Goodness of fit for this data. Empirical/estimated graphs of c.d.f. and p.d.f. for these distributions are furnished in Table 2 whereas Goodness of fit results is summarized in Table 3.

The estimated and observed values of probability and cumulative distributions are close at left tail but not at center and right tail. From Table 3, p-values of three statistics for all five probability distributions at 10% level of significance yields insignificant results except Pearson χ^2-test for Tsallis Q Gaussian distribution. The results show that all five distributions provide good fit for collected data. A probability distribution will be preferred for fitting a data if p-values of Anderson-Darling, Cramer-Von Mises and Pearson χ^2 test are closer to 1. In this study, Skew Normal and Nakagami distributions jointly provide best fit.

In order to cross validate estimated result, three randomly selected departments (Statistics, Mathematics and Economics) are taken as sub populations. The estimated means are compared with that of these sub populations in Table 4-5.

4. Conclusion

In this study, Bootstrapping technique is applied to minimize biasness due to sensitive information provided by the students of GC University Lahore about their CGPA. A random sample of 100 students is selected for this purpose. Collected information is cross checked from examination branch and only correct information is used for computation.

Five of the probability distributions were selected and their Goodness of Fit is checked for artificially generated populations using Bootstrapping technique. Skew Normal and Nakagami distributions provide best fit for the given data.

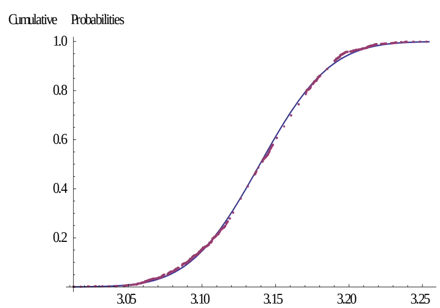
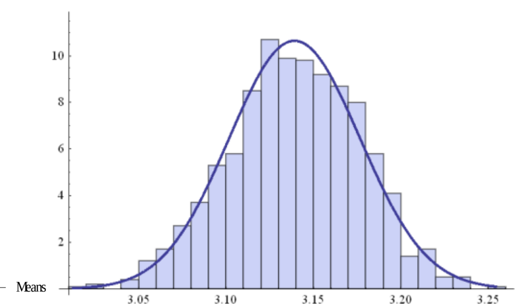
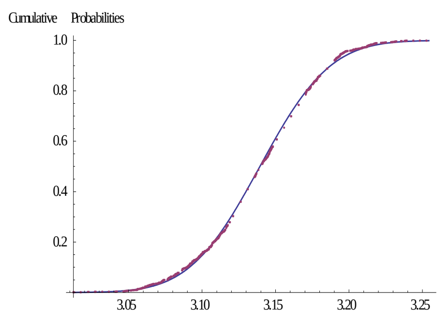
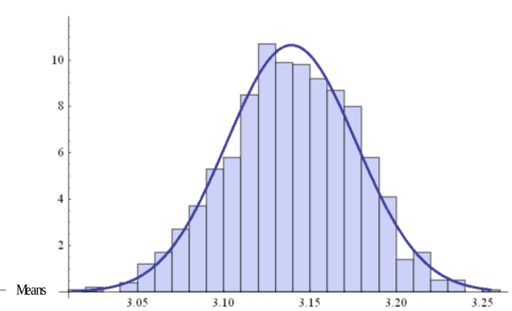
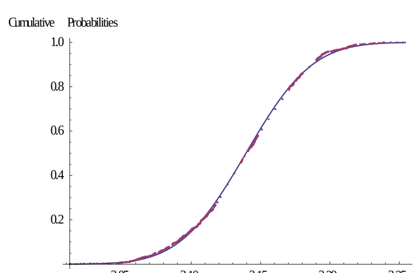
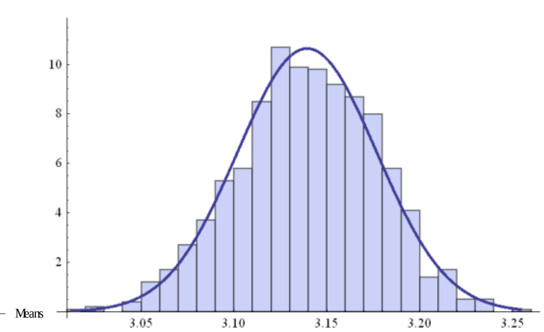
Finally, mean and bias estimated by selective distributions compared with that of three sub populations. We concluded that smaller sample gives less biased estimate from large population when sensitive information is collected. Bootstrapping technique can be applied in similar study of other areas.

Table 1: Probability density functions and relevant properties of these distributions

Distribution	**p.d.f.**	**Parameters**	**Mean and S.D.**
Skew Normal	$\frac{1}{\beta}\sqrt{\frac{2}{\pi}}e^{-\frac{(x-\mu)^2}{2\beta^2}}\int_{-\infty}^{\alpha(\frac{x-\mu}{\beta})} e^{-\frac{t^2}{2}}dt$	μ --- Location β --- Scale α --- Shape	$\mu+\alpha\delta\sqrt{\frac{2}{\pi}}$ $\sqrt{\sigma^2\left(1-\frac{2\delta^2}{\pi}\right)}$ where $\delta=\frac{\beta}{\sqrt{1+\beta^2}}$
Nakagami	$\frac{2\alpha^\alpha}{\Gamma(\alpha)\beta^\alpha}x^{2\alpha-1}\exp\left(-\frac{\alpha}{\beta}x^2\right)$	β --- Scale α --- Shape	$\frac{\Gamma(\beta+\frac{1}{2})}{\Gamma(\beta)}\left(\frac{\alpha}{\beta}\right)^{\frac{1}{2}}$ $\sqrt{\alpha\left(1-\frac{1}{\beta}\left(\frac{\Gamma\left(\beta+\frac{1}{2}\right)}{\Gamma(\beta)}\right)^2\right)}$
Rice	$\frac{x}{\beta^2}\exp\left(\frac{-\left(x^2+\mu^2\right)}{2\sigma^2}l_o\left(\frac{x\mu}{\sigma^2}\right)\right)$	μ --- Location β --- Scale	$\sigma\sqrt{\frac{\pi}{2}}l_{\frac{1}{2}}\left(\frac{-\alpha^2}{2\alpha^2}\right)$ $\sqrt{2\alpha+\mu^2-\frac{\pi\alpha}{2}l^2_{\frac{1}{2}}\left(\frac{-\mu^2}{2\alpha}\right)}$

Tsallis Q Gaussian	$\frac{\sqrt{\beta}}{C_\alpha} e_\alpha\left(-\beta x^2\right)$	μ --- Location α --- Shape	0 for α < 2 otherwise ∞ $\sqrt{\frac{1}{\beta(5-3\alpha)}}$
Von Mises	$\frac{e^{\kappa\cos(x-\mu)}}{2\pi I_o(\kappa)}$	μ --- Location Ƙ > 0	μ $\sqrt{1-I_1(\kappa)/I_o(\kappa)}$

Table 2: Empirical, estimated graph of c.d.f. and p.d.f.

Distribution	Graphs of c.d.f. and p.d.f.	
Skew Normal		
Nakagami		
Rice		

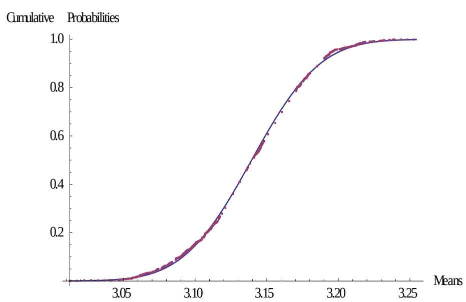
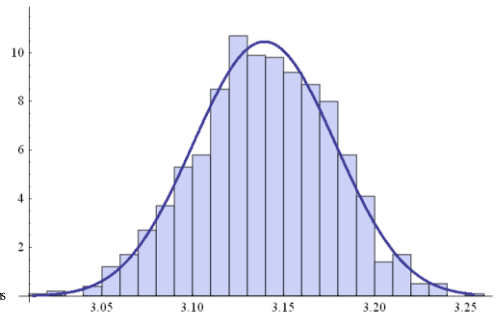
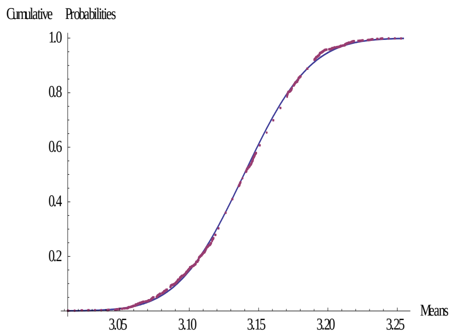
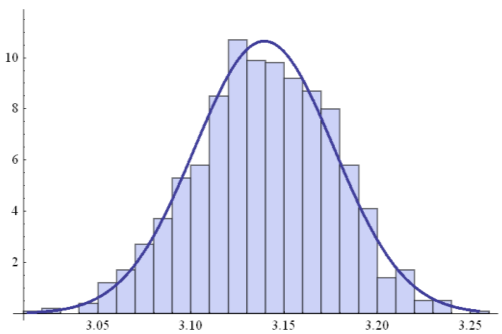
Tsallis Q Gaussian	 
Von Mises	 

Table 3: Goodness of Fit tests

Distribution	Properties	Tests	"Statistics"	P- value
Skew Normal	Mean—3.023 S.D—0.033	Anderson-Darling	0.666	0.587
		Cramer-Von Mises	0.091	0.629
		Pearson χ^2	31.93	0.419
Nakagami	Mean—3.139 S.D—0.037	Anderson-Darling	0.666	0.587
		Cramer-Von Mises	0.091	0.629
		Pearson χ^2	31.93	0.419
Rice	Mean—3.139 S.D—0.036	Anderson-Darling	50.00	0.25
		Cramer-Von Mises	70.6	0.21
		Pearson χ^2	85.4	0.17
Tsallis Q Gaussian	Mean—3.139 S.D—0.036	Anderson-Darling	76.4	0.20
		Cramer-Von Mises	87.2	0.19
		Pearson χ^2	89.7	0.10
Von Mises	Mean—3.139 S.D—0.037	Anderson-Darling	47.23	0.34
		Cramer-Von Mises	85.47	0.23
		Pearson χ^2	74.94	0.13

Table 4: Comparison of Estimated and Sub Population CGPA

Distribution	**Estimated CGPA**	**Overall CGPA**	**Maths Depart.**	**Stats Depart.**	**Economics Depart.**
Skew Normal	3.02268	3.14	3.13	3.15	3.14
Nakagami	3.139				
Rice					
Tsalli Q Gassian					
Von Mises					

Table 5: Bias in Mean CGPA

Distribution	**Estimated CGPA**	**Overall Bias**	**Maths Depart.**	**Stats Depart.**	**Economics Depart.**
Skew Normal	3.02268	0.1173	0.10732	0.12732	0.1173
Ranks		4	3	5	4
Nakagami	3.139	0.001	0.009	0.011	0.001
Rice					
Tsalli Q Gassian					
Von Mises					
Ranks		1	2	3	1

References

1. Azzalini, A. (2005). The Skew-Normal distribution related multivariate families. *Scandinavian Journal of Statistics*, **32(2)**, 159-188.
2. Best, D. and Fisher, N. (1979). Efficient simulation of the von Mises distribution. *Applied Statistics*, **28**, 152–157.
3. Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, **7(1)**, 1–26.
4. Rice, S. O. (1945), Mathematical analysis of random noise. *Bell System Technical Journal*, **24**, 46–156.
5. Leeuwen, V. H., Maassen, H. (1995). A Q-deformation of the Gauss distribution. *Journal of Mathematical Physics*, **36(9)**, 43-47.