CONFIDENCE INTERVAL ESTIMATION USING BOOTSTRAPPING METHODS AND MAXIMUM LIKELIHOOD ESTIMATE

Siti Fairus Mokhtar, Zahayu Md Yusof and Hasimah Sapiri

Siti Fairus Mokhtar[a]

Department of Mathematical Sciences, Universiti Teknologi Mara, Malaysia[a]

4 - 5 AUGUST 2021

CONTENTS

Introduction

According to Petty (2012), confidence interval (CI) is an interval estimate of a parameter of a population (e.g., a mean) calculated from a sample drawn from the population.

- Confidence level

Important:

-in statistical inferences about a population parameter and

-evaluate the accuracy of its estimator

- more information than point estimates (Das, 2019)

95% confidence interval using the sample's mean and standard deviati
assuming a normal distribution

Objectives

- to present the steps to calculate confidence interval using SPSS
- to calculate and maximum likelihood estimate, percentile bootstrap, and bias-corrected and accelerated methods.

Methodology

Bootstrap methods

- **Efron (1979), which can be used to increase the sample size by applying nonparametric resampling.**
- **The samples are used to test the statistical characteristics of the unknown distribution, such as mean, variance, standard deviation, and confidence interval (Zhang et al., 2019).**
- **Bootstrap is not commonly used because this method is complex to calculate (Doğan, 2017).**
- **The advantages: valid for small samples, and it is a convenient tool.**

Methodology

Bootstrap methods

According to Thai et al., (2013), let B be the number of bootstrap samples to be drawn from the original dataset, a general bootstrap algorithm is:

1.Generate a bootstrap sample by resampling from the data and/or from the estimated model (Sample n elements with replacement from original sample data)
2.Obtain the estimates for all parameters of the model for the bootstrap sample eg. mean, median etc.
3. Repeat steps 1-2 B times to obtain the bootstrap distribution of parameter estimates and then compute mean, standard deviation, and 95% confidence interval of this distribution

Replication is very important
Diciccio and Efron (1996) - 2000
Efron & Tibshirani, 1994 - than n^n

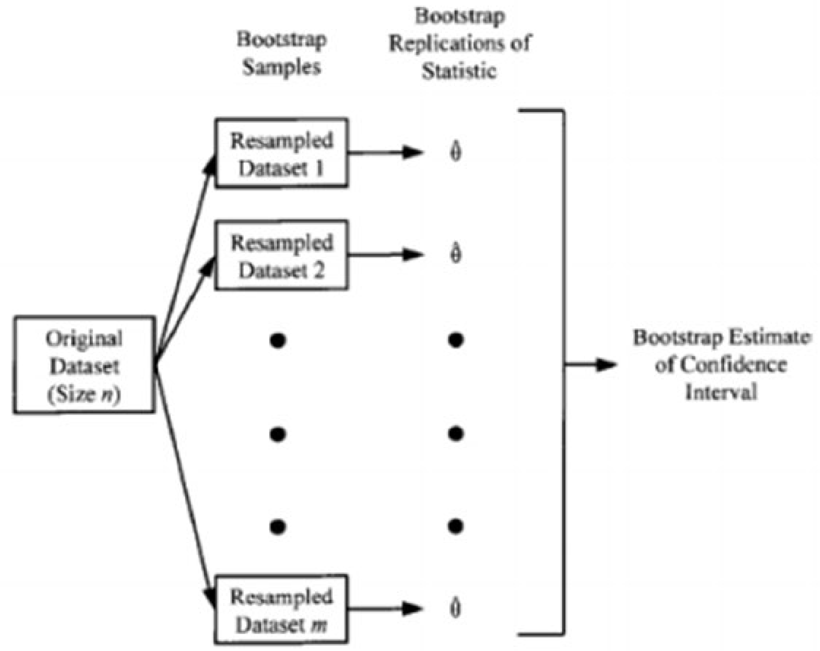


Figure 1: Description of the steps in bootstrapping.

Methodology

Two types of confidence interval estimation which are

- **Likelihood ratio method is based on maximum likelihood estimation (MLE)**
  - Used in SD and built in SD software
  - The likelihood ratio method – computationally efficient but required strong assumptions about data
  - Assumptions are frequently violated in SD
  - Advantage of likelihood ratio – convenience & easy using software
  - Assumptions that are often violated

  The confidence interval construction is based on asymptotic normality of the MLE (Kreutz et al., 2013).

- **Traditional CI**
- **Bootstrapping (percentile bootstrap method and BCa)**

Methodology

Confidence interval estimation which are

- Traditional CI (Dogan, 2017)

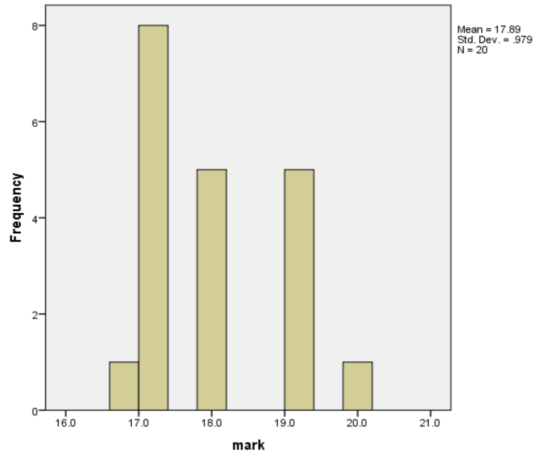
$$\bar{x} \pm t_{\alpha/2} * \left(\frac{s}{\sqrt{n}}\right) \qquad \bar{x} \pm 1.96 * \left(\frac{s}{\sqrt{n}}\right)$$

- Bootstrapping (percentile bootstrap method and BCa)

- The different methods available for estimating bootstrap confidence intervals for estimated parameters (Mesabbah et al., 2015).

Real Data Example

20, 19, 17, 18, 17, 17, 17, 17, 17, 18, 18, 19, 17, 19, 18, 17, 19, 18, 19 and 16.8



Findings

Table 1. percentile bootstrap (SPSS Output)

Descriptive Statistics

		Statistic	Bootstrap[a]			
			Bias	Std. Error	95% Confidence Interval	
					Lower	Upper
mark	N	20	0	0	20	20
	Mean	17.890	.001	.214	17.490	18.300
	Std. Deviation	.9787	-.0345	.1195	.7016	1.1760
Valid N (listwise)	N	20	0	0	20	20

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Table 2. BCa bootstrap (SPSS Output)

Descriptive Statistics

		Statistic	Bootstrap[a]			
			Bias	Std. Error	BCa 95% Confidence Interval	
					Lower	Upper
mark	N	20	0	0	.	.
	Mean	17.890	-.012	.212	17.540	18.240
	Std. Deviation	.9787	-.0308	.1133	.8013	1.1014
Valid N (listwise)	N	20	0	0	.	.

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Findings

Methods	Point estimation		Confidence interval	Interval length
	Estimate	Dev.		
Traditional CI	17.89	0.9787	[17.5116, 18.2684]	0.7568
percentile bootstrap	17.89	0.9787	[17.490, 18.300]	0.8100
BCa	17.89	0.9787	[17.540, 18.240]	0.7000

Based on the bootstrap method, the original samples were randomly sampled n=1000 times with replacement, and 1000 bootstrap samples were obtained.

The point estimation of the original sample is estimated to be $\hat{\mu} = 17.89$, and the 95% confidence interval for percentile bootstrap is between 17.490 and 18.300.

BCa bootstrap, the point estimation of the original sample is estimated to be $\hat{\mu} = 17.89$, and the 95% confidence interval between 17.540 and 18.240.

Conclusion

-3 methods

-advantages of bootstrapping are assumptions on bootstrap are less restrictive, and more easily checked, than the assumptions on MLE.

-The bootstrap also can be applied to situations where MLE may be difficult or impossible to find.

THANK YOU

2021 ICMS
INTERNATIONAL CONFERENCE ON COMPUTING, MATHEMATICS AND STATISTICS