Methodology of Bootstrapping Technique with Application to Inflation Rate

منهجية تقنية التمهيد مع تطبيق معدل التضخم

Abstract

Sampling methods is one of the most important topics in Statistics which developed in recent years. Since, all traditional methods use equations that estimate the sampling distribution for a specific sample statistic when the data follow a particular distribution. Unfortunately, formulas for all combinations of sample statistics and data distributions do not exist! For example, there is no known sampling distribution for medians, which makes bootstrapping the perfect analyses for it. This paper aims to introduce the concept and Methodology of bootstrap methods in statistics and focuses on applications in regression analysis. These applications contrast two forms of bootstrap re-sampling in regression; these techniques require fewer assumptions and offer greater accuracy and insight than do standard methods in many problems. Inflation rate is one of the most important topics raised in the present time, specially after the crises faced in the world (i.e. COVID-19 pandemic, the war of Russian-Ukraine and the crises of food and Energy around world), so inflation rate was selected to apply the bootstrapping techniques using regression method and compare the results with the original sample data using R-Package.

Key words: Re-Sampling techniques – Bootstrapping – Regression – Inflation Rate
المستخلص:

تعتبر طرق أخذ العينات من أهم الموضوعات في الإحصاء والتي تم تطويرها في السنوات الأخيرة. منذ ذلك الحين ، تستخدّم جميع الطرق التقليدية المعادلات التي تقترن توزيع العينات لعينة إحصائية محددة عندما تتبع البيانات توزيعاً معيناً. لسوء الحظ، لا توجد صيغ لجميع مجموعات إحصائيات العينات وتوزيعات البيانات! على سبيل المثال، لا يوجد توزيع عينات معروف للمتوسطات، مما يجعل التمهيد هو التحليلات المثالية له. تهدف هذه الورقة إلى تقديم مفهوم ومنهجية طرق التمهيد في الإحصاء والتركيز على التطبيقات في تحليل الانحدار. تقارن هذه التطبيقات شكلين من أشكال إعادة أخذ عينات التمهيد في الانحدار؛ تتطلب هذه الأساليب افتراضات أقل وتوفر دقة وبصرية أكبر من الأساليب القياسية في العديد من المشكلات. يعد معدل التضخم أحد أهم الموضوعات التي أثيرت في الوقت الحاضر، خاصة بعد الأزمات التي يواجهها العالم (مثل وباء COVID-19، الحرب الروسية الأوكرانية وأزمات الغذاء والطاقة حول العالم)، لذلك تم اختيار معدل التضخم لتطبيق تقنيات التمهيد باستخدام طريقة الانحدار ومقارنة النتائج مع بيانات العينة الأصلية باستخدام R-Package.

الكلمات المفتاحية: تقنيات إعادة أخذ العينات - الإقلاع - الانحدار - معدل التضخم

1. Introduction

The basic idea of bootstrapping is that inference about a population from sample data can be modeled by re-sampling the sample data and performing inference about a sample from re-sampled data. The bootstrap procedure was first suggested by Julian Simon in 1969. Efron (1979) coined the term "bootstrap", and developed this method in the statistical literature since in 1979.

Varian (2005) defined the Bootstrapping technique as a sampling with replacement from observed data to estimate the variability in a statistic of interest.

As an example, assume we are interested in the average (or mean) height of people worldwide. We cannot measure all the people in the global population, so instead we sample only a tiny part of it, and measure that. Assume the sample is of size $n$; that is; we measure the heights of $n$ individuals. From that single sample, only one estimate of the mean can be obtained. In order to reason about the population, we need some sense of the variability of the mean that we have computed. The simplest bootstrap method involves taking the original data set of $n$ heights, and, using a computer, sampling from it to form a new sample (called
a 'resample' or bootstrap sample) that is also of size \( n \). The bootstrap sample is taken from the original by using sampling with replacement, assuming \( n \) is sufficiently large, for all practical purposes there is virtually zero probability that it will be identical to the original "real" sample. This process is repeated a large number of times (typically 1,000 or 10,000 times), and for each of these bootstrap samples we compute its mean (each of these are called bootstrap estimates). We now have a histogram of bootstrap means. This provides an estimate of the shape of the distribution of the mean from which we can answer questions about how much the mean varies.

The idea behind bootstrap is to use the data of a sample study at hand as a "surrogate population", for the purpose of approximating the sampling distribution of a statistic; i.e. to resample (with replacement) from the sample data at hand and create a large number of "phantom samples" known as bootstrap samples. The sample summary is then computed on each of the bootstrap samples. A histogram of the set of these computed values is referred to as the bootstrap distribution of the statistic (see: Varian, 2005).

A common application of the bootstrap is to assess the accuracy of an estimate based on a sample of data from a larger population (Varian, 2005). Dekking et al. (2005) discussed the bootstrapping technique and confidence interval of bootstrapping statistics and showing parametric bootstrap. Fang and Wang (2012) discussed the problem of selection the number of clusters via the bootstrap method. They developed an estimation scheme for clustering instability based on the bootstrap, and then the number of clusters is selected so that the corresponding estimated clustering instability was minimized. Kuhn and Johnson (2018) shown that the bootstrap can be used to quantify the uncertainty associated with a given estimator or statistical learning method.

Bootstrapping techniques can be used in many fields. It is used in applied machine learning to estimate the skill of machine learning models when making predictions on data not included in the training data. It can be used when the data, or the errors in a model, are correlated. Künsch (1989) introduced the moving block bootstrap using the Jackknife for general stationary observations. Politis and Romano (1994) discussed the stationary of bootstrap. Vinod (2006) presented a method that bootstraps time series data using maximum entropy principles satisfying the Ergodic theorem with mean-preserving and mass-preserving constraints. Cameron et al. (2008) discusses the Bootstrap technique for clustered errors in linear regression.
There are some papers discussed the bootstrapping methods in regression. Stine (1989) provided two forms of bootstrap resembling in regression, illustrated their differences in a series of examples that include outliers and heteroscedasticity. Baisariyev et al. (2021) described the real-world implementation of the bootstrap method and the assessment of its performance with actual data from aviation logistics. James (2021) discussed the Efficient of computational algorithms for bootstrapping linear regression models with clustered data.

Inflation is an increase in the general price level of goods and services. When there is inflation in an economy, the value of money decreases because a given amount will buy fewer goods and services than before and it can be measured by Consumer Price Index (CPI). The inflation rate is the percentage change in the price index for a given period compared to that recorded in a previous period. It is usually calculated on a year-on-year or annual basis.

The form of the article is as follows. The next section presents Advantages of bootstrap and discusses the Re-Sampling Methods. Section 3 provides simple Linear Regression using Bootstrap with some examples for using different methods of bootstrap. Bootstrap confidence interval Methods discusses in Section 4. Finally, Section 5 provides application to real data. The paper ends with summary of this work.

2. Advantages of Bootstrap and Bootstrapping Methods

In this section, the advantages of Bootstrap and Bootstrapping Methods are discussed.

2.1 Advantages of Bootstrap

Bootstrap technique has different advantages. It can assess the variability of virtually any statistic.

DiCiccio and Efron (1996) discussed the most important advantages and it can be seen as follows:

A. A great advantage of bootstrap is its simplicity.
B. It is a straightforward way to derive estimates of standard errors and confidence intervals for complex estimators of complex parameters of the distribution, such as percentile points, proportions, odds ratio, and correlation coefficients.
C. Bootstrapping does not make assumptions about the distribution of the data.
D. Bootstrapping techniques can be used if the data containing an outlier (to detect outlier in data).
E. Bootstrap is also an appropriate way to control and check the stability of the results.
F. Can getting more than one sample with the same cost of getting one sample.
G. Bootstrap is asymptotically more accurate than the standard intervals obtained using sample variance and assumptions of normality, although for most problems it is impossible to know the true confidence interval.

2.2 Bootstrap Techniques

Bootstrap is generally useful for estimating the distribution of a statistic (e.g. mean, variance) without using normal theory (e.g. z-statistic, t-statistic). Bootstrap comes in handy when there is no analytical form or normal theory to help estimate the distribution of the statistics of interest, since bootstrap method can apply to most random quantities, e.g., the ratio of variance and mean.

There are at least two ways of performing case re-sampling:

A. The Monte Carlo algorithm for case re-sampling is quite simple. First, we resample the data with replacement, and the size of the resample must be equal to the size of the original data set. Then the statistic of interest is computed from the resample from the first step. We repeat this routine many times to get a more precise estimate of the Bootstrap distribution of the statistic.

B. The 'exact' version for case re-sampling is similar, but we exhaustively enumerate every possible resample of the data set. This can be computationally expensive as there are a total of \( C_n^2 \) different resamples, where \( n \) is the size of the data set. For example; if the size of sample equal 10, there are 92378, different resample must be use for estimating inference statistics for original sample (population).

3. Simple Linear Regression Using Bootstrap

In regression problems, case re-sampling refers to the simple scheme of re-sampling individual cases. In regression problems, the explanatory variables are often fixed, or at least observed with more control than the response variable.
Also, the range of the explanatory variables defines the information available from them. Therefore, to resample cases means that each bootstrap sample will lose some information (see: Liu and Singh (1992)).

There are two ways for using bootstrap in regression, first By using residual and second by using pairs (explanatory variables and response variable).

3.1 Re-sampling Using residual

The steps of this method proceed as follows.

A. Fit the model and calculate the fitted values $\hat{y}_i$ and the residuals $\hat{e}_i = y_i - \hat{y}_i$ where $i=1, 2… n$.
B. Create synthetic response variables $\hat{y}_{i*} = e_j + \hat{y}_i$ where $j$ is selected randomly from the list (1… n) for every $i$.
C. Refit the model using the fictitious response variables $\hat{y}_{i*}$.
D. Repeat steps (B) and (C) large number of times.

The main disadvantage for using this method is to select the residual that selected randomly from the list of residuals, i.e. this method depends on residual instead of main date.

Practical Example: Re-sampling Using Residual Method:

Consider this practical data as follow, and we want to estimate the regression model between Y and X.

<table>
<thead>
<tr>
<th>Serial</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>232</td>
<td>96</td>
<td>158</td>
<td>194</td>
<td>89</td>
<td>64</td>
<td>25</td>
<td>23</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Y</td>
<td>104</td>
<td>50</td>
<td>43</td>
<td>41</td>
<td>33</td>
<td>25</td>
<td>16</td>
<td>8</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

So, the estimate of $\beta_{01} = 2.80261$ and $\beta_{11} = 0.333679$

And after first iteration:

<table>
<thead>
<tr>
<th>X</th>
<th>232</th>
<th>96</th>
<th>158</th>
<th>194</th>
<th>89</th>
<th>64</th>
<th>25</th>
<th>23</th>
<th>4</th>
<th>2</th>
</tr>
</thead>
</table>

The estimate of $\beta_{02} = 14.47880292$ and $\beta_{12} = 0.27841738$

And after second iteration:

<table>
<thead>
<tr>
<th>X</th>
<th>232</th>
<th>96</th>
<th>158</th>
<th>194</th>
<th>89</th>
<th>64</th>
<th>25</th>
<th>23</th>
<th>4</th>
<th>2</th>
</tr>
</thead>
</table>
The estimate of $\beta_{03} = 12.97323234$ and $\beta_{13} = 0.278574521$.

And after third iteration:

<table>
<thead>
<tr>
<th>X</th>
<th>232</th>
<th>96</th>
<th>158</th>
<th>194</th>
<th>89</th>
<th>64</th>
<th>25</th>
<th>23</th>
<th>4</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>73.579</td>
<td>46.061</td>
<td>52.961</td>
<td>85.544</td>
<td>33.739</td>
<td>26.775</td>
<td>15.910</td>
<td>24.580</td>
<td>32.615</td>
<td>15.007</td>
</tr>
</tbody>
</table>

The estimate of $\beta_{04} = 16.7675328$ and $\beta_{14} = 0.269555338$.

Finally, the 4th iteration:

<table>
<thead>
<tr>
<th>X</th>
<th>232</th>
<th>96</th>
<th>158</th>
<th>194</th>
<th>89</th>
<th>64</th>
<th>25</th>
<th>23</th>
<th>4</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>95.787</td>
<td>40.3450</td>
<td>60.9699</td>
<td>62.0425</td>
<td>33.1620</td>
<td>37.4352</td>
<td>15.9104</td>
<td>37.7366</td>
<td>10.6015</td>
<td>10.9102</td>
</tr>
</tbody>
</table>

The estimate of $\beta_{05} = 13.17000269$ and $\beta_{15} = 0.308005202$.

So, the estimate of $\beta_0 = \frac{1}{5} \sum_{i=1}^{5} \beta_{0i} = 12.03843615$.

And, the estimate of $\beta_1 = \frac{1}{5} \sum_{i=1}^{5} \beta_{1i} = 0.293646288$.

3.2 Re-sampling Using pairs (explanatory variables and response variable)

The steps of this method proceeds as follows.

A. Fit the model (estimate the parameters).
B. Select the Pairs of explanatory variables and response variable ($x_i$ and $y_i$), and refitting the model using the selected sample ($\hat{\beta}_i$).
C. Repeat step B large number of times (let n times).
D. The estimate of parameters ($\hat{\beta}_i$) is $\frac{1}{n} \sum_{j=1}^{n} \beta_{ij}$.

Practical Example: Re-sampling Using Pairs Method:

Consider we have the following data; and want to estimate the regression model between Y and X.

<table>
<thead>
<tr>
<th>Serial</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>232</td>
<td>96</td>
<td>158</td>
<td>194</td>
<td>89</td>
<td>64</td>
<td>25</td>
<td>23</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Y</td>
<td>104</td>
<td>50</td>
<td>43</td>
<td>41</td>
<td>33</td>
<td>25</td>
<td>16</td>
<td>8</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

The estimate of $\beta_{01} = 2.80261$ and $\beta_{11} = 0.333679$.

And after first iteration:

<table>
<thead>
<tr>
<th>Serial</th>
<th>5</th>
<th>7</th>
<th>5</th>
<th>9</th>
<th>9</th>
<th>8</th>
<th>3</th>
<th>10</th>
<th>9</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>89</td>
<td>25</td>
<td>89</td>
<td>4</td>
<td>4</td>
<td>23</td>
<td>158</td>
<td>2</td>
<td>4</td>
<td>89</td>
</tr>
</tbody>
</table>
The estimate of $\beta_{02} = 3.378302$ and $\beta_{12} = 0.292027$

And after second iteration:

<table>
<thead>
<tr>
<th>serial</th>
<th>4</th>
<th>5</th>
<th>7</th>
<th>4</th>
<th>4</th>
<th>1</th>
<th>5</th>
<th>6</th>
<th>4</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>194</td>
<td>89</td>
<td>25</td>
<td>194</td>
<td>194</td>
<td>232</td>
<td>89</td>
<td>64</td>
<td>194</td>
<td>232</td>
</tr>
<tr>
<td>Y</td>
<td>41</td>
<td>33</td>
<td>16</td>
<td>41</td>
<td>41</td>
<td>104</td>
<td>33</td>
<td>25</td>
<td>41</td>
<td>104</td>
</tr>
</tbody>
</table>

The estimate of $\beta_{03} = 1.914852$ and $\beta_{13} = 0.305144$

And after third iteration:

<table>
<thead>
<tr>
<th>serial</th>
<th>7</th>
<th>3</th>
<th>3</th>
<th>7</th>
<th>5</th>
<th>3</th>
<th>2</th>
<th>8</th>
<th>9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>25</td>
<td>158</td>
<td>158</td>
<td>25</td>
<td>89</td>
<td>158</td>
<td>96</td>
<td>23</td>
<td>4</td>
<td>232</td>
</tr>
<tr>
<td>Y</td>
<td>16</td>
<td>43</td>
<td>43</td>
<td>16</td>
<td>33</td>
<td>43</td>
<td>50</td>
<td>8</td>
<td>3</td>
<td>104</td>
</tr>
</tbody>
</table>

The estimate of $\beta_{04} = 2.718309$ and $\beta_{14} = 0.342786$

And after 4\textsuperscript{th} iteration:

<table>
<thead>
<tr>
<th>s</th>
<th>7</th>
<th>3</th>
<th>9</th>
<th>10</th>
<th>7</th>
<th>1</th>
<th>4</th>
<th>1</th>
<th>10</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>25</td>
<td>158</td>
<td>4</td>
<td>2</td>
<td>25</td>
<td>232</td>
<td>194</td>
<td>232</td>
<td>2</td>
<td>64</td>
</tr>
<tr>
<td>Y</td>
<td>16</td>
<td>43</td>
<td>3</td>
<td>1</td>
<td>16</td>
<td>104</td>
<td>41</td>
<td>104</td>
<td>1</td>
<td>25</td>
</tr>
</tbody>
</table>

The estimate of $\beta_{05} = 0.918383$ and $\beta_{15} = 0.367608$

So, the estimate of $\beta_0 = \frac{1}{5} \sum_{i=1}^{5} \beta_{0i} = 2.346492$

And, the estimate of $\beta_1 = \frac{1}{5} \sum_{i=1}^{5} \beta_{1i} = 0.328249$

4. Bootstrap confidence interval

Drawing resample with replacement from the observed data, we record the means found in a large number of resample. Looking over this set of means, we can read the values that bound 90% or 95% of the entries, this method called a bootstrap confidence interval.

Confidence Interval of parameter $\theta$, is the range that $\theta$ can be lie within its bounds with a high probability $\alpha$ (significant Level), The two mostly used levels of confidence are 95% and 99%.

Wang et al. (2022) provided confidence interval localization of pipeline leakage via bootstrap method.
There are many types of bootstrap confidence interval; we discuss some of them such as Bootstrap Percentile method, Centered Bootstrap Percentile Method, Bootstrap-t Methods

4.1 Bootstrap Percentile method

In bootstrap’s most elementary application, one produces a large number of “copies” of a sample statistic, computed from these phantom bootstrap samples. Then, a small percentage, say 100(\(\alpha/2\))% (usually \(\alpha = 0.05\)), is trimmed off from the lower as well as from the upper end of these numbers. The range of remaining 100(1-\(\alpha\))% values is declared as the confidence limits of the corresponding unknown population summary number of interest, with level of confidence 100(1-\(\alpha\))%. The above method is referred to as bootstrap percentile method.

Suppose the estimation of parameter \(\theta\) at 1000 bootstrap replications of \(\hat{\theta}\), denoted by \((\theta_1^*, \theta_2^*, \ldots, \theta_{1000}^*)\). After ranking from bottom to top, let us denote these bootstrap values as \((\theta_{(1)}^*, \theta_{(2)}^*, \ldots, \theta_{(1000)}^*)\). Then the bootstrap percentile confidence interval at 95% level of confidence would be \((\theta_{(25)}^*, \theta_{(975)}^*)\) (see: Hall (1988)).

4.2 Centered Bootstrap Percentile Method

Suppose the estimation of parameter \(\theta\) at 1000 bootstrap replications of \(\hat{\theta}\), denoted by \((\theta_1^*, \theta_2^*, \ldots, \theta_{1000}^*)\). After ranking from bottom to top, let us denote these bootstrap values as \((\theta_{(1)}^*, \theta_{(2)}^*, \ldots, \theta_{(1000)}^*)\). Then the Centered bootstrap percentile confidence interval at 95% level of confidence would be \((2\theta - \theta_{(975)}^*)\) and \((2\theta - \theta_{(25)}^*)\).

4.3 Bootstrap-t Methods

Bootstrapping a statistical function of the form \(T = (\hat{\theta} - q) / SE\) where SE is a sample estimate of the standard error of \(\hat{\theta}\), brings extra accuracy. This additional accuracy is due to so called one-term Edge worth correction by the bootstrap (see: Hall (1992)).
The bootstrap counterpart of such a function \( T \) is \( (TB) = (\hat{\theta}_B - \hat{\theta})/SE_B \) where \( SE_B \) is exactly like \( SE \) but computed on a bootstrap sample. The confidence interval using Bootstrap – t Method can be given by:
\[
(\hat{\theta} - SET_{.975}, \hat{\theta} - SET_{.025})
\]
This range for \( q \) is called bootstrap-t based confidence interval for \( q \) at coverage level 95%. Such an interval is known to achieve higher accuracy than the earlier method, which is referred to as “second order accuracy” in technical literature.

5. Real Data Applications

This section provided a practical application (real data) for using bootstrapping technique to generate different samples from one sample using pair method and make a regression analysis between two variables.

5.1 The relation between Inflation Rate and Exchange Rate:

This example is explaining the relation between Inflation rate and Exchange Rate, from 2013 to 2017 (Quarterly). Data can be recorded as follows:

Table 1: Inflation Rate and Exchange Rate in Egypt (Quarterly)

<table>
<thead>
<tr>
<th>Exchange Rate</th>
<th>6.78</th>
<th>7.02</th>
<th>6.94</th>
<th>6.94</th>
<th>6.95</th>
<th>7.05</th>
<th>7.10</th>
<th>7.18</th>
<th>7.63</th>
<th>7.63</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inflation Rate</td>
<td>7.60</td>
<td>9.80</td>
<td>10.10</td>
<td>9.95</td>
<td>9.80</td>
<td>8.20</td>
<td>11.10</td>
<td>10.10</td>
<td>11.50</td>
<td>11.40</td>
</tr>
<tr>
<td>Exchange Rate</td>
<td>7.83</td>
<td>7.83</td>
<td>7.83</td>
<td>8.88</td>
<td>8.88</td>
<td>18.14</td>
<td>16.34</td>
<td>17.02</td>
<td>17.71</td>
<td>17.02</td>
</tr>
<tr>
<td>Inflation Rate</td>
<td>9.20</td>
<td>11.10</td>
<td>9.00</td>
<td>14.00</td>
<td>14.10</td>
<td>23.30</td>
<td>30.90</td>
<td>29.80</td>
<td>31.60</td>
<td>30.77</td>
</tr>
</tbody>
</table>


Figure 1 describes the scatter plot between inflation rate, as explanatory variable, and Exchange rate, as response variable.
Figure 1: Scatter plot for inflation rate and Exchange rate

The Regression equation of original data is $y = 0.489x + 2.512$

Using bootstrap pairs technique for select samples, withdraw 1000 Samples from original sample (i.e. make 1000 iterations) and estimate the parameter of linear regression model from each sample. The average of estimated slope and y-intercept are $0.531087$ and $2.2729$ respectively. So, the linear regression model is $y = 0.531087x + 2.27292$.

The Bootstrap confidence interval of sample regression parameters are $(0.041, 0.761)$, $(1.158, 4.341)$ for slope and y-intercept respectively.

Figure 2 represent the Error of estimate based on original sample and Bootstrap samples.

![Figure 2](image-url)

Figure 2: Error estimate from original sample and Bootstrap sample.

From Figure 2, we can see that error estimate from Bootstrap sample is lower than error estimate from original sample. It can be indicating that the estimation using Bootstrap method more efficient than estimation using original sample (one sample).

5.2 The relation between Inflation Rate and GDP growth Rate:

This example is explaining the relation between Inflation rate and GDP growth rate, from 2009 to 2020 (Annually). Data can be recorded as follows:

Table 2: Inflation Rate and GDP Growth Rate in Egypt (Quarterly)

<table>
<thead>
<tr>
<th></th>
<th>Inflation Rate</th>
<th>Growth Rate of GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>11.8</td>
<td>4.6735998</td>
</tr>
<tr>
<td>Year</td>
<td>Inflation Rate</td>
<td>GDP Growth Rate</td>
</tr>
<tr>
<td>------</td>
<td>----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>2010</td>
<td>11.1</td>
<td>5.147234859</td>
</tr>
<tr>
<td>2011</td>
<td>10.1</td>
<td>1.764571949</td>
</tr>
<tr>
<td>2012</td>
<td>7.1</td>
<td>2.226199797</td>
</tr>
<tr>
<td>2013</td>
<td>9.5</td>
<td>2.185466055</td>
</tr>
<tr>
<td>2014</td>
<td>10.1</td>
<td>2.915911879</td>
</tr>
<tr>
<td>2015</td>
<td>10.4</td>
<td>4.372019079</td>
</tr>
<tr>
<td>2016</td>
<td>13.8</td>
<td>4.346643453</td>
</tr>
<tr>
<td>2017</td>
<td>29.5</td>
<td>4.181221001</td>
</tr>
<tr>
<td>2018</td>
<td>14.4</td>
<td>5.314121037</td>
</tr>
<tr>
<td>2019</td>
<td>9.2</td>
<td>5.557683888</td>
</tr>
<tr>
<td>2020</td>
<td>5.0</td>
<td>3.569669475</td>
</tr>
</tbody>
</table>

**Source:**


Figure 3 describes the scatter plot between inflation rate, as explanatory variable, and Growth Rate of GDP, as response variable. As seen, from Figure 3, there is an outlier in original data (at year 2017, the inflation rate became 29.5). This outlier may affect on estimation of parameter for linear regression model. While in case of using Bootstrap technique to generate samples from original sample, the problem of outlier can be detected (because in Bootstrap, we calculate the average of estimated parameter).

The Regression equation of original data is $y = 0.0559x + 3.192$.

Using bootstrap pairs technique for select samples, withdraw 20 Samples from original sample (i.e. make 1000 iterations) and estimate the parameter of linear regression model from each sample. The average of estimated slope and y-intercept is 0.111099 and 2.800987 respectively. So, the linear regression model is $y = 0.111099x + 2.800987$. 
Figure 3: Scatter plot for inflation rate and Growth Rate of GDP from 2009 to 2020.

The Bootstrap confidence interval for Slope and y-intercept, using percentile method, can be seen as (0.0406, 0.3921) and (2.289, 3.134) respectively.

Figure 4 represent the comparison between Errors of estimate based on original sample and Bootstrap samples.

From Figure 4, we can see that error estimate from Bootstrap sample is lower than error estimate from original sample. It can be indicating that the estimation using Bootstrap method more efficient than estimation using original sample (one sample).

**Conclusion**
Bootstrap methods are a collection of sample re-uses techniques designed to estimate Mean, standard errors and confidence intervals. Conversely, the traditional methods often assume that the data follow the normal distribution or some other distribution. For the normal distribution, the central limit theorem might let you bypass this assumption for sample sizes that are larger than ~30. Consequently, we can use bootstrapping for a wider variety of distributions, unknown distributions, and smaller sample sizes. This paper provides the Bootstrap methods that can be used in regression models and shed light on confidence interval methods that can be used. Finally, practical and real data examples are provided to illustrate the proposed methods.

References:

5- DiCiccio T.J., Efron B., "BootstrapConfidenceIntervals (With Discussion)" , Statistical Science , 1996; 11, 189-228
8- Fang, Y. and Wang, J., "Selection of the Number of Clusters Via the BootstrapMethod", Computational Statistics and Data Analysis, 2012; 56(3), 468-477. [https://doi.org/10.1016/j.csda.2011.09.003].


