Analytical Report

Udeshika Sewwandi*, Wasana Wickramasinghe†, Thisun Hemakumara‡ and Kathusha Ignatius§
Department of Computer Science & Engineering
Faculty of Engineering
University of Moratuwa
Sri Lanka
Email: *sewwandikaus.13@cse.mrt.ac.lk, †wasanawvrd@gmail.com, ‡thisun.hemakumara@gmail.com, §kathusha1@gmail.com

Abstract—**In paddy cultivation due to loss of harvest for several seasons there are suicidal incidents happen in Sri Lanka due to farmers not able to pay debts and not able to survive with a family. Since unskilled workers who readily live with paddy harvesting related work may tend to leave the country due to not being able to survive. A hypothesis test was carried out to find if there's a negative linear correlation between total paddy harvest in previous year and current year and departure of unskilled workers for foreign employment. Further carried out a hypothesis test to find if there's a negative linear correlation between average real wages and departure of unskilled workers. Pearson correlation coefficient was used to find the correlation coefficients and t distribution was used to infer the results.**

Index Terms—**Paddy cultivation, real wages, Pearson correlation coefficient, hypothesis testing, t distribution, test statistic, p-value.**

I. Introduction

Paddy is the foremost crop in Sri Lanka from ancient times. The history, economy and culture of the country has a tight relationship with paddy cultivation. Still half of the population consisting a lot of unskilled people and people who inherits paddy fields engage in paddy cultivation. There are two seasons namely Yala season from May to August and Maha season from September to March to cultivate paddy. This tradition was established considering the two monsoons occur in the country to properly get water required. But due to adverse effects, the climate has changed and the time that monsoons occur has changed affecting the paddy cultivation. After spending a huge amount of money and ending up with poor harvest and less market opportunities, farmers loose their capital and many of them go through several hardships to repay their debts obtained from government and private institutes or from other sources. Most of the labourers engaged with cultivation are unskilled workers and solely depend on paddy cultivation and they become unemployed work due to lost of harvest. The aggregation of these reasons creates or increases poverty, starvation and depression among farmers and among their families.

Let us take a look at some important statistics of rice cultivation presented by the Department of Agriculture in Sri Lanka:

- Rice occupies 34% of total cultivation in Sri Lanka.
- Approximately 1.8 million families island - wide are engaged in paddy cultivation.
- Current production of rice is about 2.5 million tons annually to satisfy 95% of the domestic requirement.
- For an average Sri Lankan, rice provides 45% total calorie intake and 40% total protein requirement.
- The per capita consumption of rice ranges at about 100kg per year.
- The annual demand for rice is expected to grow at 1.1%
- Annual production should grow at 2.9% to meet this demand.
- Sri Lanka's current cost of production for rice is LKR 8.57 per kg, with labor occupying 55% of this cost.

Being a staple of Sri Lankan cuisine, rice is an important crop. So much so that rice paddies across the island are protected from being developed into something else.[1]

In recent years viability of the paddy production sector in Sri lanaka has been a question because of lack of net profit to the paddy-producing farmers. Even though the country could able to reach the self-sufficiency due to increased average yield and extent of cultivation it failed to satisfy the poor farmers.[2]

During the last decade there were several tragic suicides by small farmers in north-central province of Sri Lanka who were crushed by mounting debt, sky high rising of prices of commodities and insecure income has been reported [3]. The civil war which lasted for several decades and the natural hazards like droughts, floods and storms were few of the key causes for these tragedies [4]. There are several reasons for a person to leave the native land. Political instability, economic condition, infrastructure facilities given by the government and development of the country are some factors that a person would consider before leaving the country for employment purposes. Condition of living and financial status are the most important factors among them. When considering the factors mentioned above with respect to the issues in paddy cultivation and harvest, this would lead these unskilled farmers and their families to leave the country for foreign employment. In order to find the correlation between total annual paddy harvest and departures of unskilled workers for foreign employment we carried out a hypothesis testing based on correlation.

During the analysis we identified that the hurdles to earn money is the main reason for annual paddy harvest may have a correlation with departures for foreign employ-

ment. Therefore we carried-out another hypothesis testing to find the correlation between real average wages of unskilled workers under agriculture sector and departures for foreign employment.

The problem trying to address in this research is to find if there's a linear correlation between total annual paddy harvest and departures of unskilled workers for foreign employment and further find if there's a linear correlation between real average wages of unskilled workers and departures of unskilled workers for foreign employment. It would be useful for authorities and social workers who can provide support to uplift the lives of farmers.

More rural suicides in Sri Lanka

By G.G. Senaratna
21 November 2005

Four tragic suicides by farmers in the north-central province of Sri Lanka during the lead-up to the presidential election have highlighted the plight of many small farmers, who face mounting debt, rising costs and insecure incomes. Despite efforts by the candidates of the two major parties to win rural votes with various promises, hostility towards them on the part of the rural poor is growing.

Fig. 1. Suicidal incident1

Debt drives women to suicide in North and East

Over 150 debt related deaths in 2018 alone   17 March, 2019   💬 0

BY AANYA WIPULASENA   News Features

Fig. 2. Suicidal incident2

II. OVERVIEW OF THE DATASETS USED

Basically data relevant to Sri Lanka was considered over the past two decades. Four datasets used were as follows.

1) Paddy production in Yala season [5]
2) Paddy production in Maha season [6]
3) Departures for foreign employment [7]
4) Average daily wages in informal sector [8]

Fig. 3. Data Preprocessing Steps

The data preprocessing was carried out in three steps as represented in figure 3. The missing values in departures for employment dataset [7] was replaced with average of the particular attribute. Then data integration was performed by integrating data of several sources to

create the required dataset which is discussed later. Data transformation is the accumulation of data normalization and data aggregation which will also be discussed later in this section.

The departures for foreign employment dataset only contains data from year 1994 to year 2012. So found latest data related to departures for foreign employment from year 2013 to year 2017 from a statistical report of Sri Lanka Bureau of Foreign employment [9]. The dataset for paddy production in Yala and Maha seasons contains data from year 1952 to year 2012. But in order to perform the hypothesis testing got a sample from year 1994 to 2012 and from external sources published by the Sri Lankan government found recent data from year 2013 to year 2017 [10]. The average daily wages dataset only contains data from year 2006 to year 2012. The latest data from year 2013 to year 2017 were also found from external sources like statistical reports from Central of Sri Lanka. The average daily wages mentioned in the wages dataset are nominal wages. They are not indicative of the purchasing power of labourers of the informal sector. Therefore real wage values were calculated using wages dataset and the annual inflation rates of Sri Lanka provided by the International Monetary Fund (IMF). The real wages were calculated by dividing the nominal wage of the year from the Consumer Price Index (CPI) value of the year and multiplying it by 100 as shown in figure 4.

$$Real = \frac{Nominal}{Price\ Index} \times 100$$

* The price index can be the CPI, PPI, or the GDP Deflator.

Fig. 4. Formula to calculate real wage from nominal wage

Consumer Price Index (CPI) values were calculated using Consumer Price Index (CPI) of the previous year multiplied by (100% + annual inflation rate of the current year) [11].

We have used proportion transformation to proceed with the analysis [12]. Proportion transformation is a normalization which is based on the proportion each Attribute value has on the complete Attribute(i.e each value is divided by the total sum of that Attribute values). The total is calculated only from finite values, neglecting missing values and also positive and negative infinity. Since the ranges of numerical attributes in the datasets varied from thousands to metric tons normalization was done using proportion transformation which brought ranges of all attributes to 0 and 1.

Tables in figure 5 and figure 6 contain the data we have aggregated and normalized from the datasets which we have considered for analysis and hypothesis testing. In figure 5, columns Average Total Production and Total Labourers are the average of total paddy production

gained from both Yala and Maha seasons in previous year and current year and Sum of the departures of unskilled employees for foreign employment respectively. In figure 6, we have considered by Total Labourers is the total number of departures of unskilled employees for foreign employment in a particular year and the calculation details of the column average real wages has been described above(figure 4).

Row No.	Average Total Production (previous year & current year)	Total labourers	Year
1	0.033	0.014	1994
2	0.034	0.043	1995
3	0.031	0.041	1996
4	0.027	0.038	1997
5	0.031	0.037	1998
6	0.035	0.041	1999
7	0.036	0.042	2000
8	0.035	0.043	2001
9	0.035	0.045	2002
10	0.037	0.046	2003
11	0.036	0.048	2004
12	0.037	0.052	2005
13	0.041	0.044	2006
14	0.041	0.048	2007
15	0.044	0.052	2008
16	0.047	0.051	2009
17	0.050	0.054	2010
18	0.051	0.054	2011
19	0.049	0.057	2012
20	0.053	0.035	2013
21	0.050	0.033	2014
22	0.051	0.025	2015
23	0.058	0.019	2016
24	0.059	0.037	2017

Fig. 5. Total paddy production and departures for foreign employment

Row No.	Total labourers	Average Real wages	Year
1	The position of the example in the (filtered) view on the example table.		
2	0.095	0.071	2007
3	0.103	0.072	2008
4	0.101	0.076	2009
5	0.106	0.078	2010
6	0.105	0.080	2011
7	0.112	0.084	2012
8	0.068	0.087	2013
9	0.064	0.089	2014
10	0.050	0.093	2015
11	0.038	0.097	2016
12	0.072	0.101	2017

Fig. 6. Average real wages and departures for foreign employment
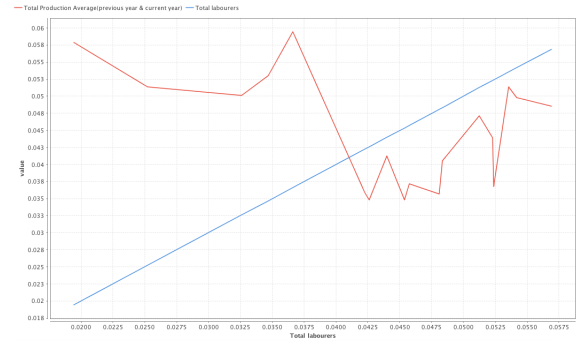

Fig. 7. Total paddy harvest vs departures for foreign employment

Figure 7 depicts the association between the Total paddy harvest and Total departures for foreign employment. We have calculated the Total paddy harvest by getting the average of previous year's and current year. According to the analysis done using the data sets it is identified that when departures for foreign employment increases, total paddy production reduces. Therefore we conclude that Total paddy harvest and departures for foreign employment has a negative correlation.
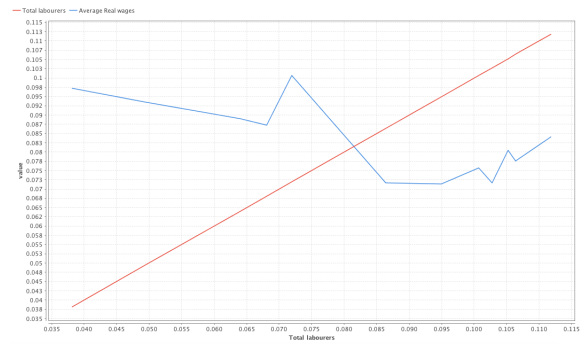

Fig. 8. Average real wages vs departures for foreign employment

Figure 8 is the result of the analysis between average real wages and foreign employment. Since departures for foreign employment increases, the average wages of the population of Sri Lanka decreases. Hence a negative correlation was identified on average real wages and the departures for foreign employment.

III. METHODOLOGY

In order to solve the problem of finding if there's a negative linear correlation between total annual paddy harvest and departures of unskilled workers for foreign employment and further finding if there's a negative linear correlation between real average wages of unskilled workers of agriculture sector and departures of unskilled workers for foreign employment, Pearson correlation was considered.

A. Pearson Correlation Coefficient

Pearson correlation coefficient is used to find the linear correlation or association between two variables of interest which gives a value between -1 and +1 based on the correlation between two variables [13].It indicates the strength of the relationship.

$$r = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n}}{\sqrt{\left[\left(\Sigma X^2 - \frac{(\Sigma X)^2}{n_x}\right)\left(\Sigma Y^2 - \frac{(\Sigma Y)^2}{n_y}\right)\right]}}$$

- ΣX This simply tells you to add up all the X scores
- ΣY This tells you to add up all the Y scores
- ΣX^2 This tells you to square each X score and then add them up
- ΣX^2 This tells you to square each Y score and then add them up
- ΣXY This tells you to multiply each X score by its associated Y score and then add the resulting products together (this is called a "cross-products")
- n   This refers to the number of "pairs" of data you have.

Fig. 9. Equation for Pearson Correlation coefficient

B. Hypothesis Formulation

Hypothesis for scenario 1

H0 : Correlation coefficient(ρ) between departure of the unskilled workers for the foreign employment and the total paddy production = 0

H1: Correlation coefficient(ρ) between departure of the unskilled workers for the foreign employment and the total paddy production $<> 0$

Hypothesis for scenario 2

H0 : Correlation coefficient(ρ) between departure of the unskilled workers for foreign Employment and average real wages of informal sector = 0

H1 : Correlation coefficient (ρ) between departure of the unskilled workers for foreign Employment and average real wages of informal sector $<> 0$

C. Population and sample statistics

The population for scenario 1 is all the annual departures of unskilled workers for foreign employment and all the annual paddy harvest in Yala and Maha seasons in the past. A sample of 24 was taken for analysis from year 1994 to 2017. The population for scenario 2 is all the annual departures of unskilled workers for foreign employment and all the annual average real wages for unskilled workers in the past. A sample of 12 was taken for analysis from year 2006 to 2017.

D. Test Statistic

The test statistic is the calculated value of t using the equation which is used to find if the null hypothesis can

be accepted or rejected. The test statistic equation used to find the correlation is represented in figure 10.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r2}}$$

t - Test Statistic

n - Sample Size

r - Sample Correlation Coefficient

Fig. 10. Test statistic equation

E. P values

The p-value is the area under the probability distribution curve used to conclude if the null hypothesis is accepted or rejected.

If p-value $> \alpha$, where is level of significance we accept the null hypothesis or fail to reject the null hypothesis. If p-value $<= \alpha$, we reject the null hypothesis and accept the alternative hypothesis.

F. Hypothesis Testing using correlation coefficient

Hypothesis testing based on correlation was carried out for the two formulated hypothesis as mentioned earlier using t distribution [14]. The hypothesis test was carried-out using 95% confidence level so the level of significance (α) is 0.05. Since our alternative hypothesis in both cases is on existence of correlation for the hypothesis made, a two-tailed test was used.

Scenario 1

	Maha Production 000 Mt.	Total Production	Total prod previous year	Total Production Average(previous year & current year)	Unskilled No.	Housemaid No.	Total labourers	Year
Maha Production 000 Mt.	1.000000	0.940654	0.677965	0.900930	0.085068	-0.622002	-0.445307	0.894182
Total Production	0.940654	1.000000	0.637792	0.914103	0.171645	-0.607172	-0.401276	0.882235
Total prod previous year	0.677965	0.637792	1.000000	0.895314	0.156964	-0.604254	-0.404573	0.861117
Total Production Average(previous year & current year)	0.900930	0.914103	0.895314	1.000000	0.181903	-0.669261	-0.445061	0.963566
Unskilled No.	0.085068	0.171645	0.156964	0.181903	1.000000	0.464240	0.733414	0.152891
Housemaid No.	-0.622002	-0.607172	-0.604254	-0.669261	0.464240	1.000000	0.942570	-0.681756
Total labourers	-0.445307	-0.401276	-0.404573	-0.445061	0.733414	0.942570	1.000000	-0.465591
Year	0.894182	0.882235	0.861117	0.963566	0.152891	-0.681756	-0.465591	1.000000

Fig. 11. Correlation coefficient matrix for scenario 1

According to the correlation matrix we can see that there's a negative correlation between Total Production Average(previous year current year) and Total labourers includes unskilled labourers and the Pearson correlation coefficient is -0.445. That means when paddy harvest has a decreasing trend there's an increasing trend in departures of unskilled workers for foreign employment. Hypothesis test was carried out as mentioned earlier using t distribution and 95 % confidence level. The degree of freedom is 23 here which is one less than the sample size(24). The critical values of t for the regions used to accept or reject the null hypothesis are -2.068 and 2.068 as depicted in figure 12.
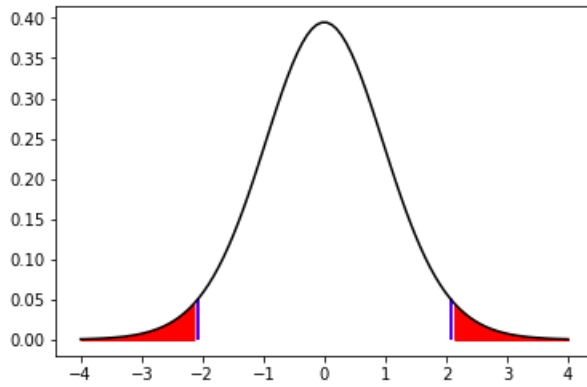
Fig. 12. t distribution for DoF 23

Scenario 2

Fig. 13. Correlation coefficient matrix for scenario 2

According to the correlation matrix we can see that there's a negative correlation between Average Real Wages and Total labourers which includes only unskilled labourers. The relevant Pearson correlation coefficient is -0.742. That means when average real wages has a decreasing trend there's an increasing trend in departures of unskilled workers for foreign employment. Hypothesis test was carried out as mentioned earlier using t distribution and 95% confidence level. The degree of freedom is 11 here which is one less than the sample size(12). The critical values of t for the regions used to accept or reject the null hypothesis are -2.201 and 2.201 as depicted in figure 14.
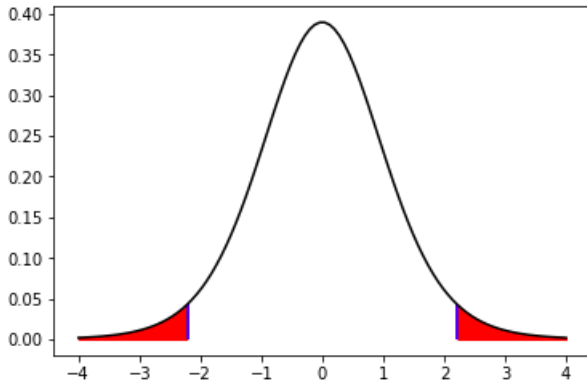

Fig. 14. t distribution for DoF 11

Scenario 1

The p-value obtained for the first hypothesis testing of finding correlation of departures of total labourers for foreign employment and total paddy harvest of current year plus previous year is 0.05 which is equal to the level of significance (α) as shown in figure 15.
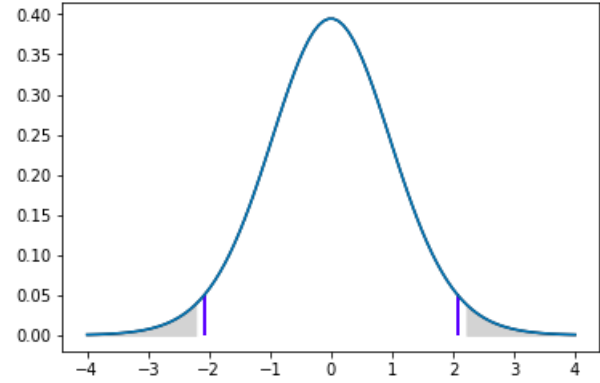

Fig. 15. Resulting p-values for scenario 1

So we reject the null hypothesis and accept the alternative hypothesis. So we can conclude that there's a correlation between departures of total labourers for foreign employment and total paddy harvest of current year plus previous year and the correlation is negative. That is when paddy harvest of current year plus previous year increases the departures of total labourers(unskilled labourers) for foreign employment decreases.

Scenario 2

The p-value obtained for the second hypothesis testing of finding correlation of departures of total labourers for foreign employment and average real wages is 0.005 which is a very small compared to the level of significance (α) as shown in figure 16.
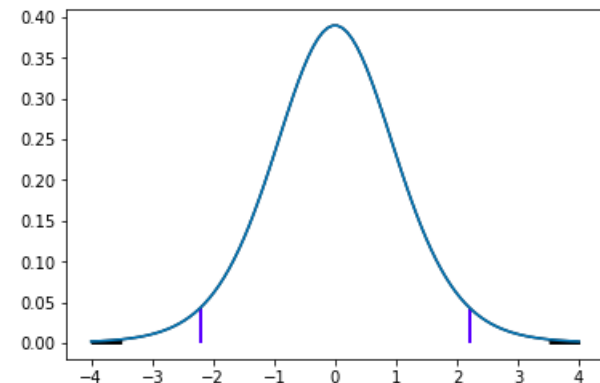

Fig. 16. Resulting p-values for scenario 2

So we reject the null hypothesis and accept the alternative hypothesis. So we can conclude that there's a strict correlation between departures of total labourers for foreign employment and average real wages and the correlation is negative. That is when average real wages increases the departures of total labourers(only unskilled labourers) for foreign employment decreases.

V. CONCLUSION

After analyzing these aforementioned datasets and performing hypothesis tests it was identified that there is a negative correlation exist between total annual paddy harvest and departures of unskilled workers for foreign employment. When further performing the analysis identified that there also exists a negative correlation among average real wages and departures of unskilled workers for foreign employment. This would be a useful analysis for the country since the related authorities can take actions to minimize the departures of unskilled labourers with less average daily wages who engage in paddy cultivation.

VI. ACKNOWLEDGMENT

REFERENCES

[1] L. Tharanga, "The importance of rice in Sri Lanka", *BlueLanka Tours*, 2018. [Online]. Available. https://www.bluelankatours.com/blog/the-importance-of-rice-in-sri-lanka. [Accessed:14- March- 2019]

[2] A.R.Mohamed and A.A.Naseer, "Trends in paddy production in Sri Lanka", *E-Repository*, 2005. [Online]. Available. http://ir.lib.seu.ac.lk/handle/123456789/47. [Accessed:14- March-2019]

[3] G. G. Senaratna, "More Rural Suicides in Sri Lanka" *World Socialist Web Site*,November 2005. [Online]. Available. https://www.wsws.org/en/articles/2005/11/suic-n21.html. [Accessed:15- March- 2019]

[4] L. Kurukulasuriya, "For a few rupees more" *Sunday Times & Tech.*, 2000. [Online]. Available. http://www.sundaytimes.lk/001029/news4.html. [Accessed:15-March- 2019]

[5] Information and Communication Technology Agency of Sri Lanka, "Paddy Production in Yala season 1952 - 2012", *OPENDATA portal of Sri Lanka*, 2019. [Online]. Available. http://www.data.gov.lk/dataset/paddy-production-yala-season-1952-2012. [Accessed:14- March- 2019]

[6] D. Chathuri,W. Aanya, Traditional farmers turn to alternative crops, ventures"Paddy farmers in quagmire", *The Sunday Times*, 2014. [Online]. Available. http://www.sundaytimes.lk/140223/news/paddy-farmers-in-quagmire-86826.html. [Accessed:14- March- 2019]

[7] Information and Communication Technology Agency of Sri Lanka, "Paddy Production in Maha season 1952 - 2012", *OPENDATA portal of Sri Lanka*, 2019. [Online]. Available. http://www.data.gov.lk/dataset/paddy-production-maha-season-1952-2012. [Accessed:14- March- 2019]

[8] Information and Communication Technology Agency of Sri Lanka, "Departures for Foreign Employment by Manpower Level 1994 - 2012", *OPENDATA portal of Sri Lanka*, 2019. [Online]. Available. http://www.data.gov.lk/dataset/departures-foreign-employment-manpower-level-1994-2012. [Accessed:13-March- 2019]

[9] Information and Communication Technology Agency of Sri Lanka, "Average Daily Wages of Informal Sector 2006 - 2012", *OPENDATA portal of Sri Lanka*, 2019. [Online]. Available. http://www.data.gov.lk/dataset/average-daily-wages-informal-sector-2006-2012. [Accessed:15- March- 2019]

[10] Sri Lanka Bureau of Foreign Employment, "Corporate Plan 2017 - 2021", Sri Lankan Government, 2017. [Online]. Available. http://www.slbfe.lk/file.php?FID=386. [Accessed:16-March- 2019]

[11] Department of Census and Statistics, "Paddy Statistics", Sri Lankan Government, 2017. [Online]. Available. http://www.statistics.gov.lk/agriculture/Paddy

[12] W. Kenton, "Consumer Price Index - CPI", *Investopedia*, 2019. [Online]. Available. https://www.investopedia.com/terms/c/consumerpriceindex.asp. [Accessed:16- March- 2019]

[13] RapidMiner, "Normalize", *rapidminer Documentation*, 2019. [Online]. Available. https://docs.rapidminer.com/latest/studio/operators/cleansing/normalization/norma [Accessed:16- March- 2019]

[14] SPSS Tutorials, "Pearson Correlations - Quick Introduction", *SPSS Tutorials*, 2019. [Online]. Available. https://www.spss-tutorials.com/pearson-correlation-coefficient/. [Accessed:18-March- 2019]

[15] Texas Gateway, "Testing the Significance of the Correlation Coefficient (Optional)", *SPSS Tutorials*, 2019. [Online]. Available. https://www.texasgateway.org/resource/124-testing-significance-correlation-coefficient-optional. [Accessed:19- March- 2019]

[16] Information and Communication Technology Agency of Sri Lanka, "Open Data Initiative of Government of Sri Lanka Datasets Freely Available to the Public", *OPENDATA portal of Sri Lanka*, 2019. [Online]. Available. http://www.data.gov.lk/. [Accessed:15- March- 2019]