Ramakrishnan M
Department of Mathematics,
RKM Vivekananda College,
Chennai, Tamil Nadu, India

Viswanathan N
Department of Statistics,
Presidency College, Chennai,
Tamil Nadu, India

A study of cooperative societies using regression approach

Ramakrishnan M and Viswanathan N

Abstract
Kanchipuram is the city not only famous for Temples, but also famous for Silk industry. Silk Handloom in this city maintains the culture of Tamil Nadu, the southernmost State of Indian Peninsula that attracts peoples all over the World. Most part of the population of this city depend on the silk industry for their livelihood. The present study compares the five Handloom Co-operative societies using Multiple Linear Regression Model. The data involves the major parameters of the production process namely, Net Profit, Number of Active Looms, Number of Ideal Looms and Production value. Ten years data on these variables are used for model development and diagnosis. Issues of Multicollinearity, Serial Correlation, Homoscedasticity and Normality were discussed using appropriate Statistical tests. The Model identifies the major contributors to the Net profit and the same can be used for future predictions in the silk handloom cooperative societies.

Keywords: Multiple Regression, Serial Correlation, Homoscedasticity, Fit index and Normality

1. Introduction
The Silk weaving industry of Kanchipuram can be broadly divided in to two segments. One is the Master Weaver in the private sector and other is Weaver co-operative societies. While the Master weaver use their own financial resources to run the business, the societies which are under the administrative control of the Department of Handlooms and Textiles are given cash credit by the co- operative banks at the concessional rate of interest to run the business.

Many small societies once in the production of real zari sarees could not afford real zari production and had switched over to production of silk varieties using Half-Fine Zari. Currently, the following 6 societies in Kanchipuram are involved in the production of silk varieties using Real Zari.
1. Kanchipuram Aringnar Anna Silk Weavers Co-operative Society
2. Kanchipuram Murugan Silk Weavers Co-operative Society
3. Kanchipuram Kamakshi Amman Silk Weavers Co-operative Society
4. Kanchipuram Thiruvalluvar Silk Weavers Co-operative Society
5. Kanchipuram Silk Weavers Co-operative Society
6. Kanchipuram Kalaignar Silk Weavers Co-operative Society

The Societies are selling Kanchipuram sarees, whereas, the private traders sell most of the sarees procured from outside Kanchipuram area and selling them as Kanchipuram sarees.

The co-operative societies are under the serious condition for sustaining in to the market. Hence this paper has discussed about the performance and forecasting of sales on Silk Handloom Weavers Co-operative Societies in Kanchipuram District. The Performance reports that contains information on the variables such as Number of active looms, number of ideal looms, production units, production value, stock units, stock value, sales units, sales value, Gross profit, Net Profit, Accumulated loss, wages paid, Bonus and Dividend are used in the analysis.

Correspondence
Ramakrishnan M
Department of Mathematics,
RKM Vivekananda College,
Chennai, Tamil Nadu, India

Table 1: Profile of the societies for the year 2016-17

Particulars	Anna	Kamakshi Amman	Thiruvalluvar	KPM	Dr. Kalaignar
Date of Regn.	22/05/1971	3/3/1955	24/12/1962	19/04/1942	7/3/1973
Commencement of business	22/05/1971	6/3/1955	6/3/1963	27/04/1942	6/4/1973
No. of looms	2856	1236	2369	3736	6105
Active looms	1047	995	800	365	415
Products manufactured	Pure Silk Sarees / Half Fine Sarees	Pure Silk Sarees / Half Fine Sarees	Pure Silk Sarees / Half Fine Sarees	Pure Silk Sarees / Half Fine Sarees	Pure Silk Sarees / Half Fine Sarees
Production (Rs. lakhs)	2815.15	710.68	722.62	468.88	338.36
Sales (Rs. lakhs)	4414.35	1178.91	909.1	694.93	526.82

2. Materials and Methods

Past ten year (2007-2017) performance reports are collected from the above societies as Secondary data. Analyses of working performance of Five Co-Operative Societies in Kanchipuram are discussed. Multiple Linear Regression technique is used to identify the influencing factors among the selected independent variable. Model diagnostics on residuals are used to identify the most appropriate model. Graphical representations are used to present the results of the analysis, with p-values substantiating the Statistical significance. Comparison among the five Co-operatives has been finally carried out and the better forecasting model has been suggested.

3. Results and Discussions

The Mean Number of Active looms of Anna, Kamakshi Amman, Thiruvalluvar, KPM and Kalaignar are 1337, 1056, 1018, 476 and 467 respectively. The Mean Number of Ideal Looms of these societies are 3074, 1970, 1384, 996 and 249 respectively. From this, it is understood that maximum number of ideal looms are in KPM Society and minimum number in Anna Society. Most of the looms in Anna society are active in the past ten years. More than 50% of the total looms are ideal in Thiruvalluvar, KPM and Kalaignar Societies.

3.1 Performance Reports analyzed by using Ordinary Least Square Method (OLS)

To apply Multiple Regression Analysis, we consider Model 1 that consists of Net Profit as dependent variable and Number of Active looms, Number of Ideal looms and Production value are taken as predictors.

Based on the summary of the regression output, the regression coefficient for purchase value is 0.25195, suggesting that an increase of 1 unit of purchase value is associated with 0.25195 unit increase in the Net profit, controlling for Number of Active loom and Ideal loom. The coefficient is significantly different from zero at the $p<0.01$ level. The coefficient for ideal loom is significantly different from zero at 5% level of significance. On the other hand, the coefficient for Active loom isn't significantly different from zero (p=0.09982) suggesting that acloom isn't linearly related when controlling for the other predictor variables. The multiple R-squared (0.8937) indicates that the model accounts

for 89.37 % of the variance in net profit. The multiple R-squared is also the squared correlation between the actual and predicted value. The residual standard error (43.51) can be thought of as the average error in predicting net profit form explanatory variables using this model. Interactions between explanatory variables are not significant.

After removing acloom variable in Model 1, it is observed that the variables idloom and pvalues are significant at 5% level of significance. But Multiple R-squared value is decreased from 0.8937 to 0.8267 and also Adjusted R-squared value is also decreased from 0.8406 to 0.7772. But Residual standard error is increased.

Model 2 is nested within model 1

Comparing nested models using the ANOVA provides a simultaneous test that active loom add to linear prediction above and beyond ideal loom and pvalues. Because the test is non-significant (p=0.09982). It is conclude that the variable active loom may be dropped from the model.

The Akaike Information Criterion (AIC) provides another method of comparing models. The index takes into account a model's statistical fit and the number of parameters needed to achieve this fit. Models with smaller AIC values – indicating adequate fit with fewer parameters-are preferred.

The AIC values for model 1 and Model 2 are 108.728 and 111.615 respectively. These AIC values suggest that inclusion of active loom variable is also better model. So it is advisable to retain Model 1. The 95% confident interval suggest that the interval [0.029, 0.173] for idloom and [0.089, 0.414] for pvalues contains the true change in Net profit for a 1 unit change in idloom and pvalues. Additionally, because the confidence interval for acloom contains zero, it can be concluded that a change in active loom is unrelated to Net profit, holding the other variables constant.

3.2 Regression diagnostics

A set of techniques called regression diagnostics provides the researcher with the necessary tools for evaluating the appropriateness of the regression model and can help the researcher to uncover and correct problems. To properly interpret the coefficients of the OLS model check number of statistical assumptions on Normality, Independence, Linearity and Homoscedasticity by several methods. These assumptions are verified graphically in Figure 1.
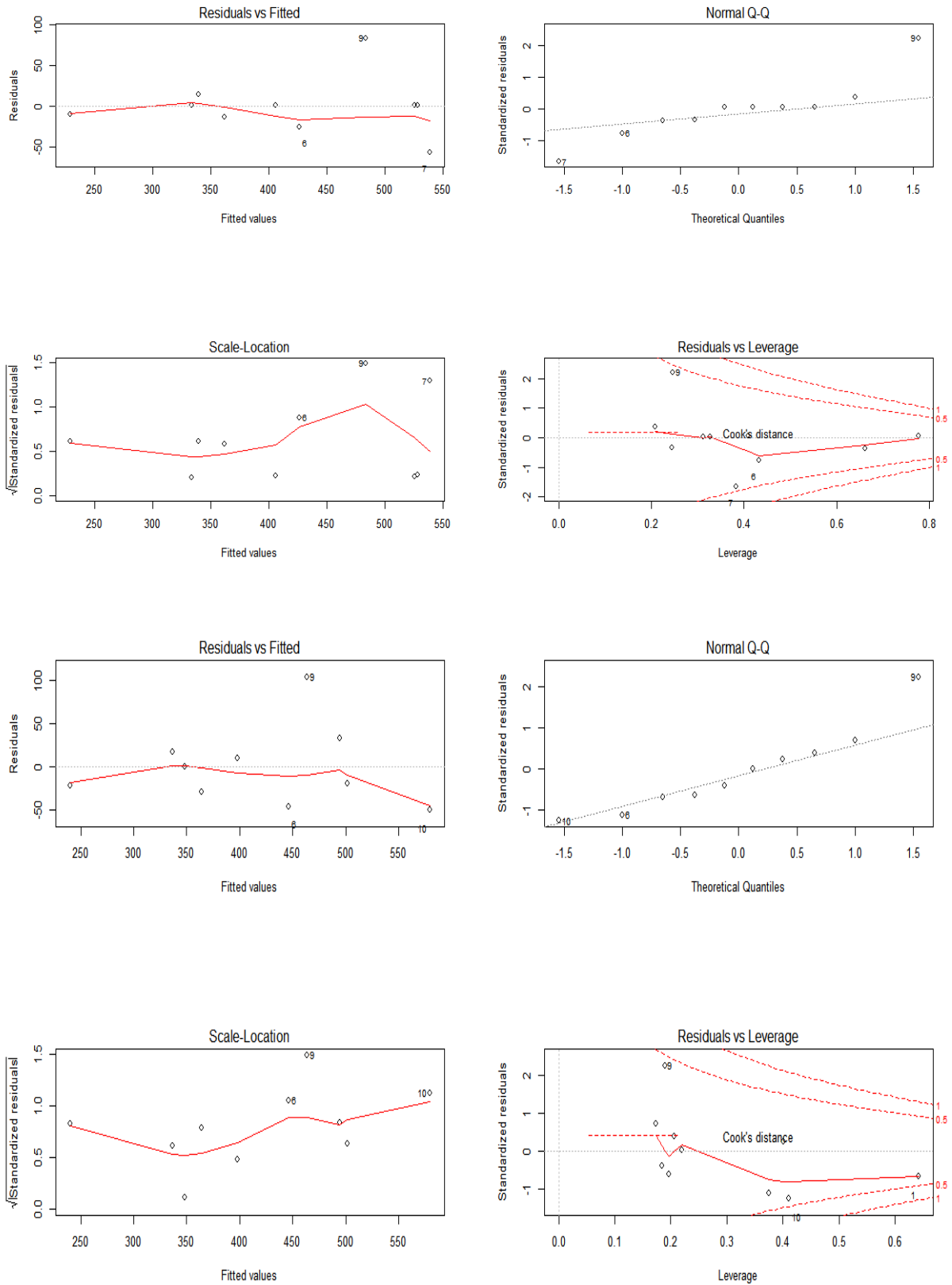
Fig 1: Diagnostic plots for Model 1 and Model 2

In both models the point 9 does not lie on Q-Q line. It has largest positive residual. To check the normality for Model 2 Use the plot which the Studentized residuals against a t distribution with n-p-1 degrees of freedom, where n is the sample size and p is the number of regression parameters (including the intercept).
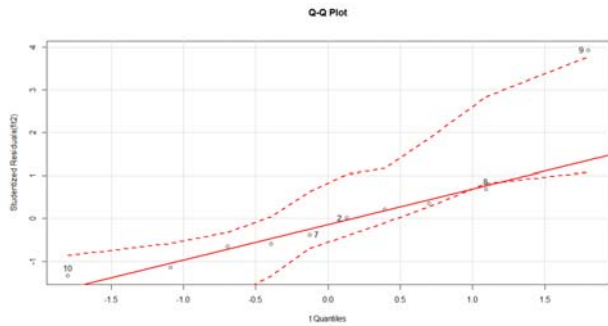

Fig 2: Studentized residual plot for Model 2

With the exception of point 9 (2015-16), all the points fall close to the line and are within the confidence envelope. This model met the normality assumption fairly well. Box-cox transformation is used to generate a maximum-likelihood estimation of the power λ most likely to normalize the variable X^λ. This results suggest that you can normalize the variable Net profit by replacing it with Nprofit[1,2]. But in this case the hypothesis that $\lambda = 1$ can't be rejected (p=0.8210703), so there's no strong evidence that a transformation is actually needed in this case. This is consistent with the results of the Q-Q plot.

To find the independence of errors use Durbin-Watson test to detect such serially correlated errors. The non-significant p-value (p=0.482) suggests a lack of autocorrelation, and conversely an independence of errors. The lag value (1 in this case) indicates that each observation is being compared with the one next to it in the dataset.

If the dependent variable is linearly related to the independent variables, there should be no systematic relationship between the residuals and the predicted values. In other words, the model should capture all the systematic variance present in the data, leaving nothing but random noise. In the Residuals versus Fitted graph (Upper left) suggest that it satisfies linearity assumption. To check nonlinearity in the relationship between the dependent variable and the independent variables, component plus residual plots were used. To create a component plot residual plot for the variable X_j, plot the points $\varepsilon_i + (\beta_j * X_{ji})\ versus\ X_{ji}$ where the residuals are based on the full model, and i=1,2,…,n
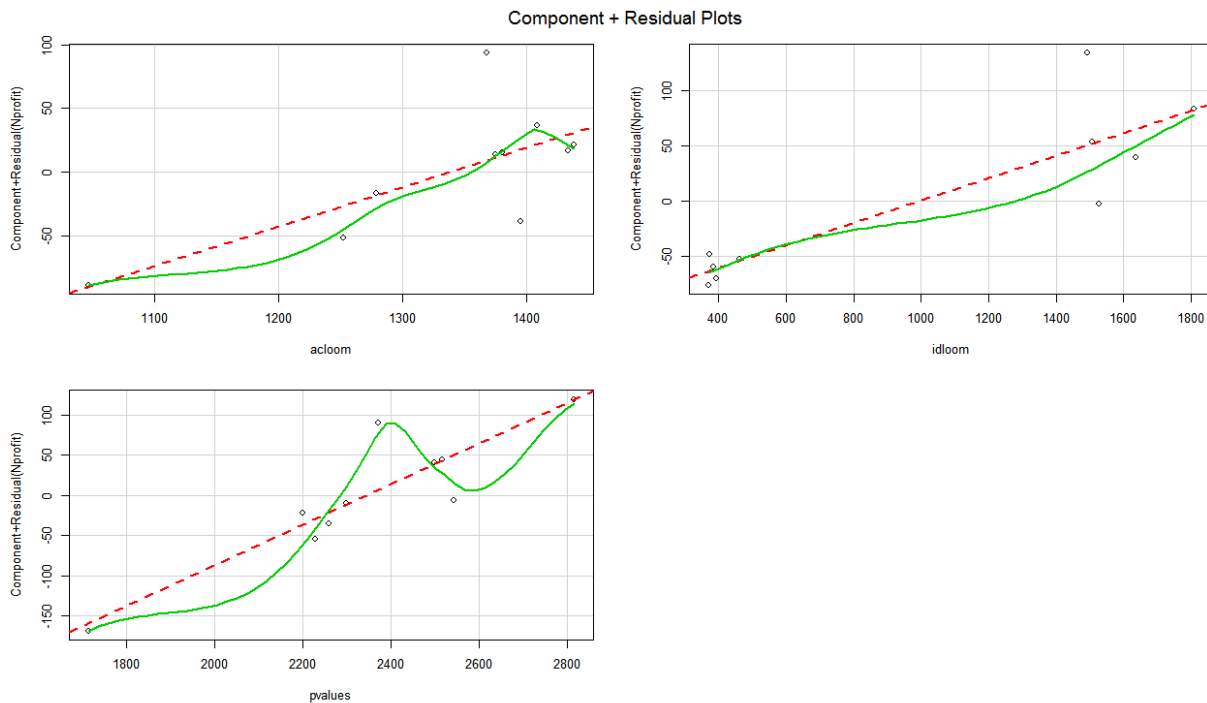

Fig 3: Component plus Residual Plot for Model 1

The component plus residual plots confirms that this Model met the linearity assumption. The form of the linear model seems to be appropriate for this dataset.

3.3 Homoscedasticity
Error term follows constant variance assumption. In diagnostic plot for Model 1, scale Location plot contains random band around a horizontal line. It seen that the assumptions are met. Also, the score test of the hypothesis of constant error variance against the alternative that the error variance changes with the level of the fitted values. A significant result suggests heteroscedasticity (non-constant error variance). The score test is non-significant (p=0.096), suggesting that this Model met the constant variance assumption.

Spread-Level Plot generates a scatter plot of the absolute standardized residuals versus the fitted values, and superimposes a line of best fit. The points form a random horizontal band around a horizontal line of best fit.
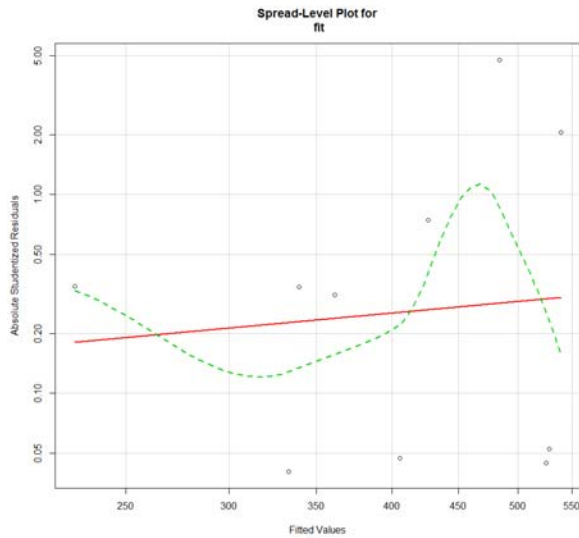
Fig 4: Spread-Level Plot for Model 1

3.4 Global validation of linear model assumption

The gvlma() function in R performs a global validation of linear model assumptions as well as separate evaluations of skewness, kurtosis, and heteroscedasticity. This is the single test of model assumptions. The data meet all the statistical assumptions that go with the OLS regression model (p=0.12278).

3.5 Multicollinearity

Multicollinearity can be detected using a statistic called Variance Inflation Factor (VIF). For any predictor variable, the square root of the VIF indicated the degree to which the confidence interval for that variable's regression parameter is expanded relative to a model with uncorrelated predictors. As a general rule, $\sqrt{VIF} > 2$ indicates a multicollinearity problem. So, there is no multicollinearity in this model because the square roots of the VIF for all predictor variables are less than 2.

3.6 Outliers

Outliers are observations that aren't predicted well by the model. They have either unusually large positive or negative residuals. Positive residuals indicate that the model is underestimating the response value, while negative residuals indicate an over estimation. Already, we have seen from the Q-Q plot, the point 9 is an outlier. To confirm this we can use outlier Test reports the Bonferroni adjusted p-value for the largest Studentized residual. Among the 10 observations, for the year 2015-16 residual value is very high. i.e., 83.78. Even though at 1% level of significance it is not significant and also the data contains only 10 observations so the best way is include this observation also for our study.

3.7 High Leverage Points

Observations that have high leverage are outliers with regard to the other predictors. In other words, they have an unusual combination of predictor values. The response value isn't involved in determining leverage. Observations with high leverage are identified through the hat statistic. For a given dataset, the average hat value is p/n, where p is the number of parameters estimated in the model (including the intercept) and n is the sample size. Roughly speaking, an observation with a hat value greater than 2 or 3 times the average hat value should be examined. In this data hat statistic value for

2015-16th year is equal to 2 times of average hat value. Therefore there is no unusual observation. But the number of active loom is very low and ideal looms are high for the year 2015-16 when compared to other year data.

Influence Plot combine the information from outlier, leverage, and influence plots into one high informative plot. The resulting plot indicates the year 2016 is the outlier.
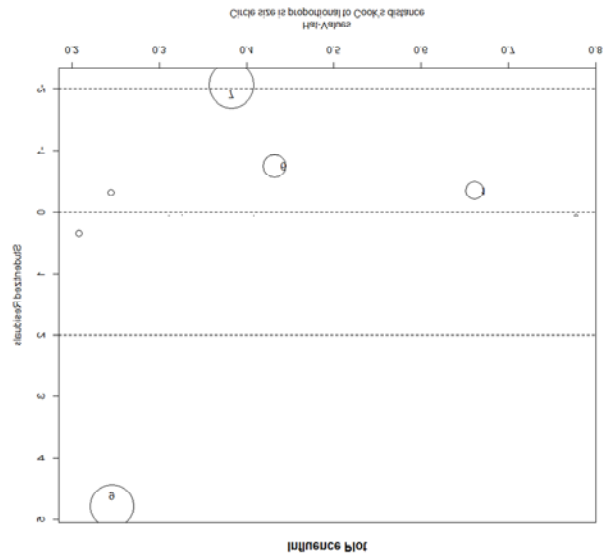

Fig 5: Influence plot, Years above +2 or below -2 on the vertical axis are considered outliers. Years above 0.8 or 1.2 on the horizontal axis have high leverage (unusual combinations of predictor values). Circle size is proportional to influence. Observations depicted by large circles may have disproportionate influence on the parameters estimates of the model.

3.8 Influential observations

Influential observations are observations that have a disproportionate impact on the values of the model parameters. There are two methods for identifying influential observations: Cook's D values greater than 4/(n-k-1), where n is the sample size and k is the number of predictor variables, indicate influential observations. Cook's D plot is given below
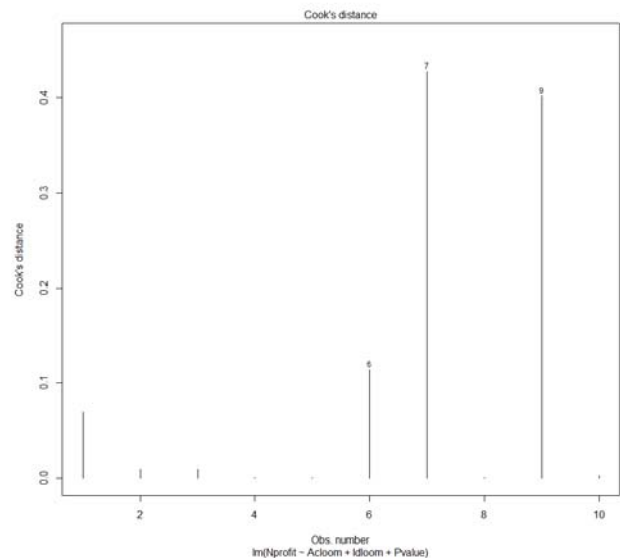

Fig 6: Cook's D Plot

This data contains only 10 values. Cook's D values are not greater than 0.6667. So the graph identifies no influential observations.

3.9 Relative Importance
There have been many attempts to develop a means for assessing the relative importance of predictors. The simplest has been to compare standardized regression coefficients. Standard regression coefficients describe the expected change in the response variable for a standard deviation change in a predictor variable, holding the other predictor variables constant.

One standard deviation increase in production value yields 0.67 standard deviation increase in Net profit rate, when controlling for Number of Active loom and Ideal loom. Using standardized regression coefficients as our guide, production value is the most important predictor and Number of Ideal loom and Number of Active loom are respective important predictors.

There have been many other attempts at quantifying relative importance. Relative importance can be thought of as the contribution each predictor makes to R-square, both alone and in combination with other predictors.

A new method called relative weights shows significant promise. The method closely approximates the average increase in R-square obtained by adding a predictor variable across all possible sub models.

The results show that the total amount of variance accounted for by the model (R-square=0.894) has been divided among the predictor variables. P value accounts for 48 percent of the R square, Number ideal looms accounts for 44 percent and Number of active loom account for 7 percent. Based on the method of relative weights, P value has the greatest relative importance, followed by Number of Ideal looms and Number of Active looms, in that order.

Regression equation for Model 1 is
Nprofit=-688.75162
+0.30966*acloom+0.10161*idloom+0.25195*pvalues
When applied the same model for other societies, it is observed that the above predictors are not significant. But, for Kalaignar society, Production value is the only significant predictor.

Our next aim is to analyze the mean net profit of all societies using statistical test.

From the mean Netprofit for the five societies, it appears that Anna society obtained the highest Netprofit, whereas KPM society obtained the least Netprofit, Standard deviations are also in the same level. The following diagram also reveals the same thing. ANOVA F test for Societies is significant ($p<0.001$), providing evidence that the five societies mean Netprofit aren't equal.

3.10 Multiple comparisons
The ANOVA F test for NetProfit tells us that the five societies aren't equal Netprofit, but it doesn't tell us which societies differ from one another. So use multiple comparison procedure to answer this question. Tukey multiple comparison of means 95% family-wise confidence level test is used to test multiple comparison.

Anna Society NetProfit is significantly different from other societies. Kamatchi Amman and Thiruvalluvar Net Profit aren't significantly different and KPM and Kalaignar societies are also not significant difference in NetProfit. The following diagram also reveals the same thing.
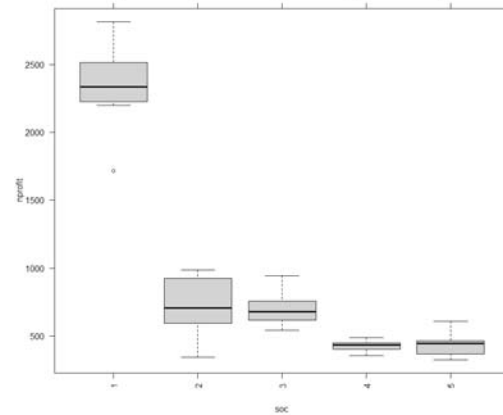

Fig 7: Box Plot for Multiple Comparison

Groups (represented by box plots) that have the same letter don't have significantly different means. It also reveals the same thing. Bartlett's test indicates that the variances in the five societies differ significantly ($p<0.001$). From the output, you can see that there's no indication of outliers in the data. Taking the Q-Q plot, Bartlett's test and outlier test together, the data appear to fit the ANOVA model quite well. This, in turn, adds to your confidence in the results.

4. Conclusion
In this study, the researcher have discussed fitting OLS regression models, using regression diagnostics to assess the data's fit to statistical assumptions, and methods for modifying the data to meet these assumptions more closely. A final regression model is selected from many possible models. In our study, Production values, Number of Ideal Looms and Number of Active looms are more important variables for predicting Netprofit and identified their importance also. Using One-Way ANOVA compared the dependent variable Net Profit means of five societies and observed significant difference exists among group means and variances.

5. References
1. Secondary sources: Records of the societies
2. Frontline: An Article on the Silk Industry of Kanchipuram, TamilNadu
3. Indian Silk Industry in the global scenario: http://zenithresearch.org.in
4. Buying Behavior and promotions of Silk retailing in Tamilnadu: https://www.worldwidejournals.com/indian-journal-of-applied-research-(IJAR)/ volume:3/issue:4/April 2013/ISSN-2249-555X
5. Robert I. Kabacoff. R in Action, Data Analysis and graphics with R. Manning Publications Co. United States of America., 2011.
6. Thilaga G, Raman GP, Ramakrishnan M. A study on performance and forecasting of Silk handloom weaver's Co-operative society limited in Kanchipuram city in Tamil Nadu, International Journal of Business and Administration Research Review, 2017; 2(19):42-51.