FACTORS INFLUENCING IN PINEAPPLE PRODUCTION OF MANIPUR- A STATISTICAL PERSPECTIVE.

[1]Changamayum Girija Devi, [2]S. Loidang Devi
[1]Research Scholar, department of Statistics, Manipur University, India.
[2]Assistant Professor, department of Statistics, D.M. College of Science, Dhanamanjuri University, Imphal, India.
Corresponding author: loidang023@gmail.com

ABSTRACT

In Manipur the horticultural crop that offers the biggest chance for a living for the highland agricultural community is the pineapple, which has considerable popularity and production. Manipur is one of the top states for producing pineapples in north-eastern India due to its suitable temperature and soil. The state is fortunate to have an endowment of favorable agro climatic conditions, which may support a variety of horticultural crops with economic yields. It has the second largest production next to rice and it is the backbone of a sizeable section of farmers as their major source of income. Hence it is necessary to discuss the growth of the pineapple so as to enable the state to formulate appropriate policy and enhancing the production. The main objective of this study is to identify the factors influencing the production of pineapple in Manipur. Primary data is used for this analysis. Multiple linear regression model is fitted for the study. From the study, the value of R^2 for the model was found to be 0.934 which indicates that out of the total variation 93.4% of the variation in the production is explain by the independent variables included in the model.

Keywords: Pineapple, Manipur, Multiple linear regression models, Production.

INTRODUCTION

Pineapple (Ananas comosus) is a tropical plant with an edible fruit and the most economically significant plant in the family Bromeliaceae. The cultivation of pineapple is confined to high rainfall and humid coastal regions in the peninsular India and hilly areas of north-eastern region of the country. It can also be grown commercially in the interior plains with medium rainfall and supplementary irrigations. Manipur is a one of the suitable place for the development of horticulture. Pineapple is one of the main horticultural crops which are largely produced in Manipur. It has the second largest production next to rice and it is the backbone of a sizeable section of farmers as their major source of income. Manipur with its favorable climate and soil type is suitable for cultivation of pineapple. The prevailing wide agro climatic conditions of Manipur make it possible to cultivate pineapple throughout the year. Due to the absence of the reliable industry pineapple production are only mean for the human consumption. Manipur is recognized as an agricultural state for its production of large scale of agricultural products. In Manipur, pineapple is quite popular and highly produced among horticultural crops

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed(Group –I) Journal Volume 11, Iss 04, April 2022

that provide the highest livelihood opportunity to highland farming community. Hence it is necessary to discuss the growth of the pineapple so as to enable the state to formulate appropriate policy and enhancing the production. The main objective of this study is to identify the factors influencing the production of pineapple in Manipur

MATERIALS AND METHODS

Primary data were used for this analysis. In order to analyze the data, we used the SPSS-software packages. Multiple linear regression model is fitted for the study. For this a survey was planed. The procedure of the said survey may be given as follows

SAMPLING FRAME AND SAMPLING DESIGN

The sampling design for this survey is stratified three stage sampling scheme of unequal size. The sampling size is decided at 450 for this consideration.

- With the 9 districts of Manipur as strata(according to 2011 census)
- Sub division in the districts as the primary sampling unit (or the first stage sampling unit)
- Villages in the selected subdivision as the second stage sampling unit
- Farmers in the selected villages as the third stage sampling unit.

By preparing Questionnaire/schedule 450 samples were collected for fitting the model.

SELECTION PROCEDURE

According to 2011 census report, there are 9 districts in Manipur namely Imphal East, Imphal West, Senapati, Churachanpur, Bishnupur, Tamenglong, Chandel, Ukhrul and Thoubal. The 9 districts of Manipur is taken as strata. Sub division in the districts is taken as the primary sampling unit (or the first stage sampling unit). After getting the sampling frame, the subdivision with pineapple cultivation is taken into consideration otherwise excluded. The village with no cultivated area are excluded and the village having pineapple farm with area at least 0.25 Ha and above are considered as an eligible case otherwise left out as out of consideration. Under these criteria, one subdivision from each district is selected as primary sampling unit with the help of simple random sampling. From this selected subdivision villages are selected (unequal size) as second stage sampling unit. From this selected 31 villages' farmers (unequal size) are taken from each village as third stage sampling unit.

METHODOLOGY

MULTIPLE LINEAR REGRESSION MODEL

The proposed models are of the form

$$\underline{Y} = \mathbf{X}\underline{\theta} + \underline{\varepsilon} \tag{1}$$

Where, $\underline{Y} = (Y_1, Y_2, \dots, Y_i, \dots, Y_n)'$, i.e., (n x 1)vetor of observations,

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research Paper UGC CARE Listed(Group -I) Journal Volume 11, Iss 04, April 2022

$$\mathbf{X} = (X_{ij})_{nxp}, \text{i.e.}, (n \times p) \text{matrix of co} - \text{efficients},$$

$\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_i, \ldots, \theta_p)'$, i.e., a (p x 1) vector of parameters, and

$$\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_i, \ldots, \varepsilon_n)', \text{i.e.}, \text{an } (n \times 1) \text{ vector of error components}$$

With $E(\underline{\varepsilon}) = \underline{0}$ and dispersion matrix, $\sigma^2 I$,

I is the n x n identity matrix.

The least squares estimates are given by minimizing the scalar sum of squares

$$S = (\underline{Y} - \mathbf{X}\underline{\theta})'(\underline{Y} - \mathbf{X}\underline{\theta}) \qquad (2)$$

A necessary condition that equation (2) be minimized is that

$\frac{\partial S}{\partial \underline{\theta}} = \underline{0}$ [by maxima and minima theory of calculus]

Differentiating, we have

$$2\mathbf{X}'(\underline{Y} - \mathbf{X}\,\underline{\theta}) = \underline{0}$$
$$\Rightarrow \mathbf{X}'\,\underline{Y} = \mathbf{X}'\mathbf{X}\,\underline{\theta}$$

which are the usual normal equation.

Then,

$$\hat{\underline{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\underline{Y}$$

Where we assume that ($\mathbf{X}'\mathbf{X}$), the matrix of sums of squares and products of elements of column-vectors composing **X**, is non-singular and can therefore, be inverted.

TESTING PROBLEMS

(a) A hypothesis, whether the regression hyper-plane is at all helpful in predicting the values of Y may be tested, the null hypothesis under consideration is

$$H_0: \theta_{i1} = \theta_{i2} = \cdots = \theta_{ip} = 0, \quad i = 1,2, \ldots, n$$
$$\Rightarrow H_0: \underline{\lambda}_i'\,\underline{\theta} = \underline{0}, \qquad i = 1,2, \ldots, n;$$

Where, $\underline{\lambda}_1, \underline{\lambda}_2, \ldots, \underline{\lambda}_n$ are linearly independent and the functions $\underline{\lambda}_i'\underline{\theta}$ is estimable.

The above null hypothesis is equivalent to

$$H_0: \underline{\rho} = \underline{0}$$

i.e., the mean of $\underline{Y}$ is as accurate in predicting $\underline{Y}$ as the regression plane.

Where $\underline{\rho}$ is the multiple correlation coefficient of Y_i on X_{ij}'s.

In order to test the null hypothesis, we apply the method of ANOVA.

ANOVA TABLE

Sources of variation	Degrees of freedom	Sum of squares	Mean sum of squares	F

Regression	p	$SS_{reg} = \underline{\hat{\theta}}'\boldsymbol{X}'\underline{Y} = \sum_{i-1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2$	$\frac{SS_{reg}}{p} = MS_{reg}$	MS_{reg}
Residual(or error)	n-p-1	$SS_{res} = \underline{Y}'\underline{Y} - \underline{\hat{\theta}}'\boldsymbol{X}'\underline{Y} = \sum_{i-1}^{n}\left(Y_i - \hat{Y}_i\right)^2$	$\frac{SS_{res}}{n-p-1} = MS_{res}$	MS_{res}
Total	n-1	$\underline{Y}'\underline{Y} = \sum_{i-1}^{n}\left(Y_i - \overline{Y}\right)^2$		

Alternating, it may be tested by the statistic

$F = \frac{n-p-1}{p} \cdot \frac{R^2}{1-R^2}$,

which has a Snedecor F distribution with p and n-p-1 degrees of freedom whenever the null hypothesis is true. Here R denotes the sample multiple correlation co efficient. From the above result, we can take the reasonable conclusion by comparing the calculated values with the corresponding table values.

b) Further testing problems

Test of hypothesis can be used to assess whether variables are contributing significantly to the regression equation.

For any specific variable, X_{ij},we can test the null hypothesis, as

$$H_0: \theta_{ij} = 0, \qquad j = 1,2, \ldots, p,$$

By comparing

$$t = \frac{\hat{\theta}_{ij}}{SE(\hat{\theta}_{ij})}$$

with the table value of student's t on n-p-1 degrees offreedom at a commonly assigned level of significance.

The above testing problems is equivalent to testing

$$H_0: \underline{\rho}_{ij} = 0, \qquad j = 1,2, \ldots, p,$$

i.e., the simple correlation co-efficient between Y_i and X_{ij}, for each j, is zero.

Under this null hypothesis, the test statistic to be used is

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

Where r is the simple correlation co-efficient between Y_i and X_{ij}, and we can conclude accordingly. The static t so defined follows student's t-distribution on n-2 degrees of freedom whenever the hypothesis is true.

METHODS OF DIAGNOSTIC CHECK

In addition to the above methodologies, we shall use the corresponding diagnostic methods for testing goodness of fit for each. To be précised, the methods used have been meant for

1) Detecting outliers
2) Checking linearity or homogeneity of variance of residuals
3) Checking independence and normality of residuals
4) Checking multicollinearity or independence of random variables and
5) Checking whether the data were drawn from a normal population

ANALYSIS AND RESULTS

The statistical model developed for the production is defined as

$$Y_i = C\prod_{j=1}^{p} X_{ij}^{a_{ij}}\, e^{\varepsilon_i} \quad ; \qquad i = 1,2,\dots\dots,n;$$
$$j = 1,2,\dots\dots,p$$

Which by taking logarithm, may equivalently given as

$$\log_e Y_i = b_{i0} + \sum_{j=1}^{p} b_{ij}\log_e X_{ij} + \varepsilon_i\,;\ i = 1,2,\dots,450;\ j = 1,2,\dots,13$$

Where,

Y_i= production of the pineapple in the i^{th}farm (kg),
X_{i1}= age of the i^{th} farmers,
X_{i2}=household size of the i^{th} farmer
X_{i3}= number of cultivators in the i^{th} farm
X_{i4}= total area of the i^{th} farm in Ha
X_{i5}=quantity of chemical fertilizer used by the i^{th} farmer, in kg
X_{i6}= experience in years of the i^{th} farmer
X_{i7}= market distance from the i^{th} farm in km
X_{i8}= amount of credit in rupees, if any
X_{i9}= number of times of weeding per year in the i^{th} farm
X_{i10}= average expenditure by the i^{th} farmer for pineapple cultivation per year
X_{i11}=average income of the i^{th} farmer per year
X_{i12}= Years of schooling of the i^{th} farmer
X_{i13}= 1 if soil of the i^{th}farmer has been tested
=0, otherwise
b_{i0}= constant
b_{ij}= regression coefficient
ε_i= error components which are assumed to be independently and identically distributed $N(0,\sigma^2)$
The following are the results and discussions

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed(Group –I) Journal Volume 11, Iss 04, April 2022

TABLE-1
Multiple correlation co-efficient(R) and Durbin- Watson (d) statistic
Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.966[a]	.934	.932	.18496	1.687

a. Predictors: (Constant), X_{i13}, X_{i12}, X_{i7}, X_{i9}, X_{i5}, X_{i3}, X_{i6}, X_{i8},X_{i10}, X_{i1}, X_{i2}, X_{i11}, X_{i4}
b. Dependent Variable: Y_i

TABLE-2
ANOVA[a]

Model		Sum of Squares	Df	Mean Square	F
	Regression	209.719	13	16.132	471.562
	Residual	14.916	436	.034	
	Total	224.635	449		

a. Dependent Variable: Y_i
b. Predictors: (Constant), X_{i13}, X_{i12}, X_{i7}, X_{i9}, X_{i5}, X_{i3}, X_{i6}, X_{i8}, X_{i10}, X_{i1}, X_{i2}, X_{i11}, X_{i4}

Since the calculated F value namely 471.562 is highly significant, we reject the null hypothesis that the b_{ij}'s are zero at the 0.01 probability level of significance. Hence each Y_i can be predicted by the X_{ij}'s accurately.

TABLE-3
Coefficients, t-statistics, collinearity statistics

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	6.948	.510		13.630	.000		
	X_{i1}	.043	.039	.017	1.104	.270	.680	1.470
	X_{i2}	.021	.026	.013	.820	.413	.639	1.565
	X_{i3}	-.025	.026	-.015	-.956	.340	.651	1.535
	X_{i4}	.879	.048	.800	18.322	.000	.080	12.504
	X_{i5}	.004	.008	.007	.538	.591	.937	1.067
	X_{i6}	-.012	.013	-.014	-.921	.357	.702	1.425

	X_{i7}	.011	.010	.015	1.114	.266	.891	1.123
	X_{i8}	.006	.002	.036	2.716	.007	.873	1.145
	X_{i9}	.299	.045	.086	6.669	.000	.908	1.101
	X_{i10}	-.235	.032	-.226	-7.442	.000	.165	6.043
	X_{i11}	.351	.027	.378	13.102	.000	.183	5.460
	X_{i12}	.002	.009	.003	.253	.800	.834	1.199
	X_{i13}	.009	.050	.002	.174	.862	.962	1.039
a. Dependent Variable: Y_i								

It is found that the regression coefficients, namely those of total area(X_{i4}), credit amount (X_{i8}), number of weeding (X_{i9}), expenditure(X_{i10}) and average income(X_{i11}) are significant while the remaining are found insignificant. So it confirms that the linearity condition between Y_i and X_{ij}'s$\forall$ j as embodied by the model may reasonably be assumed. The scatter plot, Figure 3 is the type of the ellipse of concentration and it confirms the absence of heteroscedasticity of variances of the residuals. It follows from the Table-1, in which the computed value of the Durbin Watson statistic (which is 1.687) is very close to 2, and this indicates that the residuals ε_i's are independent. Since the normal probability plot of standardized residuals, shown in Figure 2 is close to the diagonal of the box. It can be concluded that the residuals follow a joint multivariate normal distribution. Next on reading the Table 3 which show the tolerance, VIF, etc, it can be observed that the tolerance limits are not small , and each is not less than 0.01 and the VIF is not large and it is not greater than 100, and these result together imply that multicollinearity is absent. Furthermore, it may be observed in Fig 1. that almost all the position of histogram are superimposed by normal density curve, and this clearly indicates that the data are drawn from a normal distribution.

TABLE-4
CasewiseDiagnostics[a]

Case Number	Std. Residual	Y_i	Predicted Value	Residual
4	-4.894	7.93	8.8312	-.90517
8	-3.062	8.04	8.6033	-.56640
89	-3.239	11.54	12.1343	-.59911
268	-5.012	9.67	10.5964	-.92707

Table-4 shows that the observation 4, 8, 89 and 268 out of 450 observations are the outliers.

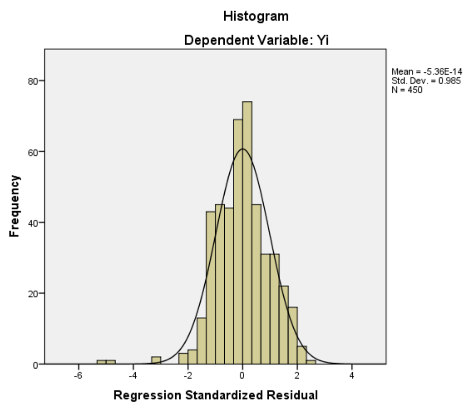

FIG.1 Histogram

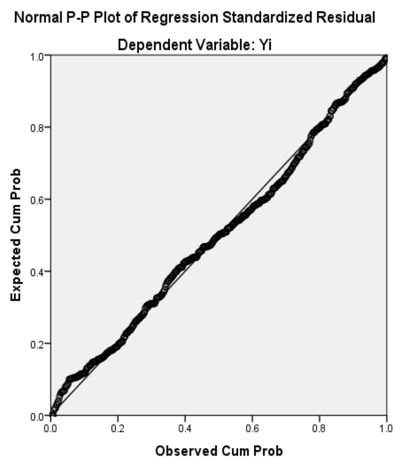

FIG.2 Normal P-P Plot of regression standardized Residual

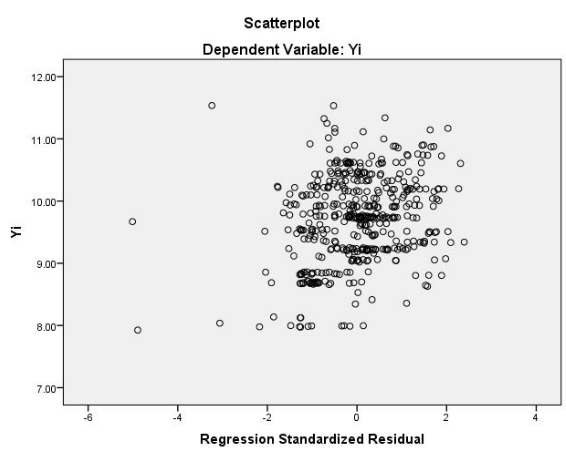

FIG. 3 Scatterplot

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed(Group –I) Journal Volume 11, Iss 04, April 2022

CONCLUSION

From the study it was found that the value of R^2 for the model is found to be 0.934 which indicates that out of the total variation 93.4% of the variation in the production is explain by the independent variables included in the model. It is found that the regression coefficients, namely those of total area(X_{i4}), credit amount (X_{i8}), number of weeding (X_{i9}), expenditure(X_{i10}) and average income(X_{i11}) are significant while the remaining are found insignificant.

REFERENCE

1. Himmati, R. (2017). The Analysis of Agricultural Product in Pineapple Commodities in Ngancar District, Kediri Regency, East Java Province.*Journal of Indonesian Applied Economics,***7**(1): 48-58.
2. Joseph, N.M.; Mbabazi, P. and Shukla, J. (2017). The Contribution of Contract Farming in the Improvement of Pineapple Production in Rwanda: Case Study of Ucocuan in Ngoma District. *International Journal of Interdisciplinary Research and Innovations.***5**(2): 62-67.
3. Olah, O.N. and Okon, E.T. (2022). Factors Affecting Pineapple Production in Central Agricultural Zone of Cross River State, Nigeria.*American Journal of Environmental and Resource Economics*.**7**(1):8-14.
4. Triyono; Kamardiani, D.R.; Prasetio, M.A,(2021). Alocative Efficiency of Honey Pineapple Farm in Pemalang Regency, Central Java, Indonesia.*IConARD 2020*, **232**:1-10.