



Simple Linear Regression Models for Peanut Production Prediction in North Sumatra 2022

Rohana Labora Munthe^{1*}, Putri Angel Lipipian², Sardo Pardingotan Sipayung³

^{1,2,3}Universitas Katolik Santo Thomas, Medan

rohanamunthe29@gmail.com^{1*}, putriangel.lipipian163@gmail.com², pinarsiphom@gmail.com³

Abstract

Peanuts (*Arachis hypogaea* L.) are a palm oil commodity that functions strategically in supporting food security and the regional economy. North Sumatra Province has great potential in peanut development, but increasing production is still faced with a number of obstacles. This study aims to examine the relationship between harvest area and peanut production as well as assess the ability of harvest area as a predictor variable of production in North Sumatra Province. This study applies a quantitative approach using a simple linear regression analysis method. The data analyzed is secondary data obtained from the Central Statistics Agency (BPS) of North Sumatra Province through the publication of *North Sumatra in Figures 2022*, which contains data on harvest area and peanut production at the district/city level. Of the 27 districts/cities, as many as 26 districts/cities were analyzed after going through the data filtering process.

The results of data processing produced a regression equation $Y=12.356+0.000592X$, which showed that the effect of harvest area on peanut production was positive but very weak. A determination coefficient value (R^2) of 3.09% indicates that the harvest area is only able to explain a small part of the variation in production, while most of it is influenced by other factors outside the model. Significance testing through the F-test and t-test showed that the regression model was not significant at a 95% confidence level. These findings confirm that increasing peanut production cannot rely solely on land expansion, but rather requires an intensification strategy through increased productivity and input efficiency.

Keywords: *Simple Linear Regression; peanuts; Production Prediction; North Sumatra; Statistical Analysis*

1. Introduction

Peanuts (*Arachis hypogaea* L.) including palawija commodities with high economic value and strategic functions in supporting the national food system. In North Sumatra Province, the opportunity for peanut development is still relatively large, but farmers often face real problems such as market price instability and attacks of plant-disrupting organisms that have the potential to significantly reduce farming income [1]. In addition to these risk factors, production achievement is also highly determined by climatic conditions and the level of technical experience of farmers in the field [2]. Therefore, the right development strategy, including the application of SWOT analysis, is needed to identify the strengths and weaknesses of peanut agribusiness in order to be able to be competitive on a local and national scale [3].

The main factor that sustainably plays a role in determining the amount of production is the availability of land. Based on various research results, land area has a positive and significant influence on the amount of production output; The larger the land cultivated, the greater the chance of increasing the total crop yield [4]. However, land area is not optimal enough without the support of efficient input management, such as determining the appropriate number of seeds and optimal utilization of labor [5]. From an economic point of view, feasibility analysis such as the value of the \$R/C\$ ratio is also an important indicator to assess the extent to which peanut farming provides benefits compared to other commodities [1].

From the technical aspect of agronomy, increased productivity can be achieved through planting density regulation. Implementation of a double-line system (*Twin-Row*) proven to be able to optimize land use and increase crop yield potential compared to conventional single-row systems [6]. In addition, optimal planting spacing settings, such as \$40 \times 40\$ cm, combined with the provision of additional nutrients in the form of Liquid Organic Fertilizer (POC), have been shown to significantly increase stem diameter and number of pods per plant [7]. The use of various other types of organic fertilizers also has an important role in improving the physical and chemical properties of the soil that support sustainable plant growth [8].

Another challenge in peanut cultivation has to do with environmental conditions, particularly light intensity. Plants that grow in low light or shaded conditions require certain physiological adaptability through tolerant quantitative character selection [9]. The morphological character of the various strains exhibits different responses to shade stress, so the identification of superior genotypes is a very important

early stage in plant breeding programs [10]. Mapping the relationship between these various factors of production, ranging from genetic character to land area, can be accurately modeled using statistical tools [11].

To provide an objective predictive picture in food policy formulation in North Sumatra, the application of regression analysis is very relevant. Simple Linear Regression Analysis was used in this study to explain mathematically the extent to which land area variables affect production variables [12]. By integrating technical aspects of cultivation and economic factors, this research is expected to be able to produce accurate peanut production estimates to support farmers' welfare and regional food security.

2. Research Methods

This study uses a quantitative approach with a simple linear regression analysis method to analyze the relationship between harvest area and peanut production in North Sumatra Province in 2022. The quantitative approach was chosen because it is able to process numerical data objectively and produce mathematical models that can be used to explain and predict the relationships between variables based on empirical data.

Simple linear regression is applied as an estimation method in data mining, which aims to form prediction equations based on the pattern of relationships between one independent variable and one bound variable.

2.1 Types and Approaches to Research

This research is included in quantitative research with an explanatory approach, which is research that aims to explain the cause-and-effect relationship between independent variables and bound variables through inferential statistical analysis. The explanatory approach is used to determine the extent to which the crop area variable affects peanut production and assess its feasibility as a predictor variable [13].

2.2 Data Collection

The data used in this study is secondary data obtained from the Central Statistics Agency (BPS) of North Sumatra Province through publication *North Sumatra in 2022 Figures*. The data collected included peanut harvest area (hectares) and average peanut production (quintals per hectare) at the district/city level [14]. The initial number of data consisted of 27 districts/cities, but Tebing Tinggi City was excluded from the analysis because it had a harvest area value of zero, so it did not meet the prerequisites for regression analysis. Thus, the number of data analyzed in this study is 26 districts/cities.

For example, some of the data used in the analysis are presented in Table 1.

Table 1. Sample Data on Harvest Area and Peanut Production in 2022

Regency/City	Harvest Area (Ha)	Production (Kw/Ha)
Nias	6	12,59
Christmas Decorations	183	11,39
South Tapanuli	177	15,41
Simalungun	279	12,77
Düsseldorf	259	14,10

Source: BPS North Sumatra Province, 2022

2.3 Research Variables

The variables used in this study consist of two main components, namely the independent variable (X) in the form of peanut harvest area measured in hectares, and the dependent variable, (Y) = expressing peanut productivity (quintals per hectare)

2.4 Data Analysis Methods

Data analysis was carried out by applying simple linear regression using the *Ordinary Least Squares* (OLS). This method aims to obtain an optimal regression equation through the process of minimizing the number of squares of the difference between the observation value and the estimated result value [15].

The simple linear regression model used in this study is formulated as follows:

$$Y = a + bX \quad (1)$$

with the caption: Y = expresses peanut production (quintal), X = represents the peanut harvest area (hectares), a = is a constant, and b = is the regression coefficient.

2.4.1 Stages of Regression Parameter Calculation

To obtain the value of constant (a) and regression coefficient (b), the following formula is used:

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2} \tag{3}$$

$$a = \frac{\sum Y - b\sum X}{n} \tag{4}$$

where: n = sum of observation data
 $\sum X$ = sum of the total value of the harvest area
 $\sum Y$ = sum of the total value of production
 $\sum XY$ = sum of multiplication results X and Y
 $\sum X^2$ = sum squared of value X

3. Results and Discussion

3.1 Descriptive Statistics of Research Data

The initial stage of analysis is carried out by compiling descriptive statistics to describe the characteristics of the research data. This statistic is the basis for the calculation of simple linear regression parameters using the OLS method. A summary of these statistics is presented in Table 2.

Table 2. Statistical Summary for Regression Analysis

Components	Harvest Area (X)	Production (Y)
Number of data (n)	26	26
$\sum X / \sum Y$	4.110	326,31
$\sum XY$	106.123,92	–
$\sum X^2$	20.073.269	–
Average	158,12	12,55

Based on Table 2, the total peanut harvest area in North Sumatra Province reached 4,110 hectares, with an average production of 12.55 quintals per hectare. The values $\sum X^2$ and $\sum XY$ are further used in the calculation process of simple linear regression parameters.

3.2 Simple Linear Regression Calculation Process

Based on the values in Table 2, the regression parameters are calculated using the OLS formula. By entering the values of $\sum X$, $\sum Y$, $\sum XY$, $\sum X^2$, and the amount of data (n = 26) into the regression equation, the value of the regression coefficient (b) of 0.000592 and the value of constant (a) of 12.356 were obtained.

The results of the regression parameter estimation are presented in detail in Table 3.

Table 3. Regression Analysis Results

Parameters	Value	Interpretation
Regression coefficient (b)	0,000592	Very small
Constant (a)	12,356	Basic production
R ²	0,0309 (3,09%)	Very low
F-count	1,443	Insignificant
t-count	0,879	Insignificant

Based on Table 3, the regression equation is obtained as follows:

$$Y = 12,356 + 0,000592X \tag{2}$$

A constant value of 12.356 indicates that when the harvest area is assumed to be zero, the basic production rate of peanuts is in the range of 12.356 quintals per hectare. The regression coefficient of 0.000592 indicates that any increase in the harvest area of one hectare only increases production by 0.000592 quintals per hectare, which confirms that the effect of harvest area on production is very weak.

3.3 Results of Regression Model Statistical Test

To assess the feasibility of the regression model, statistical tests were carried out which included the determination coefficient, the F test, and the t test.

Table 4. Results of Simple Linear Regression Statistical Test

Parameters	Value
Coefficient of determination (R ²)	0,0309 (3,09%)
F-count	1,443
t-count	0,879

The value of the determination coefficient (R^2) of 3.09% indicates that the harvest area variable is only able to explain a small part of the variation in peanut production. Most of the production variation, which is 96.91%, is influenced by other factors outside the model. The results of the F-test showed that the F-calculated value was smaller than the F-table value at a significance level of 5%, so that the regression model was simultaneously insignificant. The t-test also showed that the regression coefficient was partially insignificant at a 95% confidence level.

3.4 Analysis and Interpretation of Results

The results of the analysis show that the harvest area is not the dominant factor in determining the level of peanut production in North Sumatra Province. Although theoretically an increase in the harvest area can increase the total output, in this study the effect on productivity per hectare is very small.

The low value of the determination coefficient indicates that peanut production is more influenced by other factors such as agroclimatic conditions, soil fertility, use of superior seeds, cultivation techniques, fertilization, and farm management.

3.5 Production Prediction Simulation

As an application of the regression model, production prediction simulations were carried out on several harvest scales. The results of these predictions are presented briefly in Table 5.

Table 5. Examples of Production Predictions on Various Scales of Harvest Area

Scale	Harvest Area (Ha)	Production Prediction (Kw/Ha)
Small	100	12,415
Intermediate	500	12,652
Large	1.000	12,948
Industry	2.000	13,540

The simulation results show that the increase in harvest area provides relatively small additional production, so that the strategy of increasing peanut production through land expansion alone is considered less effective without being accompanied by increased productivity through technology and input efficiency.

Conclusion

Based on the results of a simple linear regression analysis of peanut harvest area and production data in North Sumatra Province in 2022, it can be concluded that the relationship between harvest area and peanut productivity is very weak. The regression equation obtained, i.e. $Y=12.356+0.000592X$, shows that each addition of one hectare of harvest area only contributes to an increase in production of 0.000592 quintals per hectare, which practically does not have a significant impact. The determination coefficient value (R^2) of 3.09% indicates that the harvest area variable is only able to explain 3.09% of the variation in peanut production. Meanwhile, 96.91% of production variations were influenced by external factors not covered by the model, such as climatic conditions, soil fertility, cultivation technology, and farming management. The results of statistical significance testing through the F test and the insignificant t-test at a confidence level of 95% further strengthen the finding that the single harvest area variable is not sufficient to be used as an accurate predictor of peanut productivity in North Sumatra Province. The main implication of this study emphasizes the need to shift the orientation of agricultural policy from an extensification approach, namely the expansion of planting area, to an intensification strategy that focuses on increasing productivity per unit area through the application of efficient cultivation technology, the use of appropriate agricultural inputs, and strengthening the capacity and competence of farmers. For further research, it is recommended to develop a more comprehensive analysis model by integrating additional variables to obtain more accurate production estimates.

Acknowledgments

The author would like to thank the Central Statistics Agency (BPS) of North Sumatra Province for providing the data used in this study. Gratitude was also conveyed to the parties who had provided input and support during the research process.

References

- [1] R. Fitriani, M. Zubair, and Wuryantoro, "Comparative study of costs and income of corn farming and peanut farming," *Journal of Agribusiness*, vol. 18, no. 2, pp. 1–18, 2022.
- [2] M. R. Aulia, A. Putra, E. Purnama, A. Anantama, and N. Aisyah, "Qualitative Study of the Potential, Utilization, Threat and Success of Peanut Production in Rantau Panyang Timur Village, West Aceh Regency," vol. 3, no. 2, pp. 60–64, 2023.
- [3] T. Di and S. South, "SWOT-AHP ANALYSIS IN THE DEVELOPMENT OF PEANUT AGRIBUSINESS".
- [4] K. Tanah, D. I. Kecamatan, P. Raya, and K. Aceh, "Analysis of factors affecting peanut production in pasie raya sub-district, aceh jaya regency 1," vol. 25, pp. 29–37, 2024.
- [5] K. Tanah, D. I. Desa, and K. Kecamatan, "ANALYSIS OF FACTORS AFFECTING PRODUCTION," no. 2022, pp. 183–190, 2025.
- [6] A. S. Assopi, A. F. Hemon, and I. K. Ngawit, "Growth and Yield of Peanuts (*Arachis hypogaea* L.) on Several Plant Densities under a Twin-Row Planting System," vol. 4, no. 3, pp. 760–769, 2025.
- [7] D. A. N. J. Tanam, "3 1,2,3".
- [8] F. Yunanda, I. N. Soemeinaboedhy, and I. P. Silawibawa, "The Effect of Various Organic Fertilizers on Soil Physical Properties, Soil Chemistry, and Peanut Production (*Arachis hypogaea* L.) In Kediri District The Effect of Various Organic Fertilizers on Soil Physical Properties , Soil Chemistry , and Peanut (*Arachis hypogaea* L.) Production in The Kediri Regency," vol. 1, no. 3, pp. 294–303, 2022.
- [9] L. L. Intensity, "Quantitative Characteristics and Tolerance of Several Peanut Lines Grown at," vol. 2, no. 2, pp. 283–292, 2023.

-
- [10] P. S. Agroekoteknologi, U. Mataram, and K. Mataram, "Lifiani, R., Sutresna., I.W., and Hemon, A.F. 70," vol. 31, no. 1, pp. 70–83, 2021.
- [11] R. Ayumnuazmi, A. F. Hemon, and I. W. Sutresna, "Quantitative Character of Several Genotypes of Peanuts (*Arachis hypogaea* L .) on Various Density of Double-Row Plants Quantitative Characteristics of Some Genotypes of Peanut Plants (*Arachis hypogaea* L .) At Various Plant Density of Double Rows," vol. 4, no. 2, pp. 620–628, 2025.
- [12] E. Suhesti, "Analysis of the feasibility of peanut farming in Bayeman Village, Arjasa District, Situbondo Regency," vol. 1, no. 1, pp. 1–6, 2024.
- [13] K. L. Barat, M. S. Wedastra, and I. D. G. Suartha, "ANALYSIS OF THE USE OF PRODUCTION FACTORS ON PEANUT PRODUCTION LEVELS (CASE STUDY IN DOPANG VILLAGE, GUNUNG SARI DISTRICT)," pp. 241–252, 2024.
- [14] F. Economics, D. A. N. Business, U. I. State, S. Ali, and H. Ahmad, No Title.
- [15] M. Y. Nugroho, "Analysis of Factors Affecting Rice Production in Sumatra Using the Linear Regression Method," pp. 36–40, 2023.