Data Correlation of Climate & Cotton Yields in India Using ML with Python

1.VAISHNAVI **2.THANDUR VANITA** **3.SUKEERTHI .S**

1,2, Assistant Professor CSE(AI&ML), G Narayanamma Institute of Technology and Science,Hyderabad, Telangana, India

3, Assistant Professor CSE (Data Science),G Narayanamma Institute of Technology and Science,Hyderabad ,Telangana ,India

Corresponding author :1 vaishnavi

Abstract:

Cotton is a significant cash crop in, India, contributing substantially to the region's agricultural economy. Understanding the factors that influence cotton yield is crucial for farmers, policymakers, and researchers aiming to enhance productivity and sustainability. Climate indicators such as rainfall, temperature, humidity, soil moisture, and extreme weather events like heat waves and extreme precipitation play a vital role in determining the growth and yield of cotton bolls.

Understanding the impact of climate indicators on agricultural productivity is crucial for enhancing crop yields and ensuring food security. This study aims to investigate the effect of various climate indicators on the growth and yield of cotton bolls in , India, using Python for data analysis. Key climate factors such as rainfall, temperature, humidity, soil moisture, and extreme weather events (heat waves and extreme precipitation) are analyzed to determine their correlation with cotton yield.

data for the period from 2015 to 2020 is utilized to illustrate the methodology. The Pearson correlation coefficient is employed to assess the strength and direction of the relationship between each climate indicator and cotton yield. Visualizations, including scatter plots, are generated to provide a clear depiction of these relationships.

Preliminary results suggest that certain climate indicators have a significant impact on cotton yield, highlighting the importance of climate-adaptive strategies in agricultural planning. This study demonstrates the potential of using Python for agricultural data analysis and provides a framework for future research with real-world data to develop robust, data-driven agricultural policies in India.

The primary objective of this investigation is to analyze the relationship between various climate indicators and cotton yield in India using Python. By performing correlation analysis and visualizing the data, we aim to identify the key climate factors that significantly impact cotton production.

Key words: cotton boll yield, correlation, climate indicators, growth of cotton boll, rainfall, temperature, soil moisture

1. Introduction

Cotton is a critical crop for India's agricultural economy, contributing significantly to the country's gross domestic product (GDP) and providing employment to millions of farmers. As one of the largest producers and exporters of cotton globally, India's cotton industry is pivotal in ensuring economic stability and supporting rural livelihoods. However, the growth and yield of cotton bolls are profoundly influenced by various climate factors, such as temperature, rainfall, humidity, sunlight, and soil moisture. Understanding the intricate relationships between these climate indicators and cotton growth is essential for developing strategies to enhance yield, optimize resource use, and mitigate the impacts of climate variability.

The effects of change in the climate and extreme weather patterns pose significant challenges to cotton production. Extreme temperatures, irregular rainfall, and shifting humidity levels can adversely affect boll formation and overall yield. Therefore, it is imperative to investigate how specific climate indicators impact the growth of cotton bolls to inform adaptive agricultural practices and policy decisions.

In recent years, advances in data science and machine learning have give chances for agricultural research. Python, with its libraries and tools for data analysis, visualization, and modeling, has become a powerful platform for examining the effect of climate variables on crop growth. Python's allows researchers to handle large datasets, perform complex statistical analyses, and make useful models to guess that can give valuable suggestions into the factors accounting cotton yield.

The aim of this research is to investigate the effect of key climate indicators on the growth of cotton bolls in India using Python-based data analysis and machine learning techniques. By leveraging historical climate data and cotton yield records, this work aims to identify valuable correlations and make predictive models that can help farmers and policymakers make informed decisions. Specifically, this research will focus on the following objectives:

1. **Data Collection and Preprocessing**: Gathering historical climate data and cotton yield records from relevant sources and preparing the data for analysis.
2. **Exploratory Data Analysis (EDA)**: Utilizing Python's data analysis libraries to explore the data, identify trends, and visualize relationships between climate indicators and cotton yield.
3. **Statistical Analysis and Correlation**: Applying statistical methods to quantify the strength and direction of relationships between climate variables and cotton growth.
4. **Predictive Modeling**: Developing machine learning models to predict cotton yield based on climate indicators, and evaluating the models' performance.
5. **Discussion and Recommendations**: Interpreting the results, discussing the implications for cotton farming in India, and providing recommendations for future research and agricultural practices.

2. Literature review

Reddy, Hodges, and McKinion [1] examined the potential effects of global climate change on cotton growth and yield, focusing on how temperature increases and elevated atmospheric CO2 levels might impact crop production. The study emphasizes that many major crops, including rice, soybean, wheat, and cotton, are not adapted to the projected future climate conditions, particularly higher temperatures. These crops are sensitive to heat stress, which could significantly limit grain and fruit production, jeopardizing the world's food supply, which relies heavily on seed- or fruit-bearing summer annuals.

Ghanwat, Asewar, and Jondhale [2] investigated the impact of rainfall variability on cotton production and productivity in the Marathwada region of India. The study found a positive trend in cotton production with respect to rainfall, suggesting that higher rainfall positively influences cotton yields. Correlation analysis revealed a significant relationship between cotton lint yield and rainfall in the districts of Aurangabad, Jalna, Beed, Latur, and Osmanabad. However, the analysis showed insignificant correlation in the districts of Nanded, Parbhani, and Hingoli, indicating that other factors may influence cotton productivity in these areas. Overall, the study highlights the importance of rainfall in shaping cotton yields in Marathwada, while also noting regional variations in the strength of this relationship.

Pettigrew's [3] study on the effects of moisture deficit on cotton (Gossypium hirsutum L.) growth and yield highlights the significant role water availability plays in reproductive development and productivity. Conducted over four years, the study found that moisture deficit, particularly in dryland conditions, led to a 25% reduction in lint yield, mainly due to a 19% decrease in boll number. Irrigated plants, however, produced more bolls, especially at higher plant nodes and distal positions on fruiting branches, contributing to higher yield potential. Irrigation also delayed the onset of cutout, extending the flowering period and allowing for continued reproductive growth. While irrigation did not significantly impact most fiber quality traits, it did result in slightly longer fiber in three out of four years. The study underscores the importance of irrigation in stabilizing yields by promoting boll production in key areas of the plant.

Singh, P [4] (2020) explored the impact of rising atmospheric temperatures on wheat, highlighting how heat stress disrupts membrane stability, grain filling, and starch accumulation, leading to reduced grain formation. High temperatures also impair the photosynthetic apparatus and generate reactive oxygen species, causing oxidative stress. To mitigate these effects, crop management strategies such as mulching, additional irrigation, the use of inorganic fertilizers, early sowing, and the application of micronutrients, osmoprotectants, and bioregulators can help stabilize physiological processes and metabolic pathways, improving thermotolerance in wheat.

Silvertooth and Norton [5] investigated the effects of irrigation termination (IT) management on the growth, development, and yield of upland cotton (Gossypium hirsutum L.). Their study highlights the importance of refining IT management practices to improve cotton productivity. One key finding was that as temperatures rise, fruit retention in cotton plants decreases, emphasizing the need for optimized irrigation strategies to mitigate temperature stress and maintain yields.

Pathak et al.[6] examined nitrogen balance in rice-wheat systems in the Indo-Gangetic Plains, emphasizing the role of nitrogen (N) management in improving soil N status and reducing depletion. Their study found that higher chemical nitrogen application, when combined with organic N and better management practices, can increase nitrogen use efficiency. However, excessive use of chemical fertilizers can lead to nitrogen losses, contributing to environmental pollution. The developed model serves as a tool for optimizing fertilizer distribution and managing nitrogen inputs, with potential applications for both agricultural planning and environmental regulation.

3. Climate Indicators Affecting Cotton Growth

Several climate indicators have been identified as critical to the growth and yield of cotton:

1. **Temperature**: Optimal growth temperatures for cotton range between 20°C to 30°C. Extreme temperatures can lead to poor boll formation and reduced yields
2. **Rainfall**: Adequate rainfall is essential for cotton growth, but excessive or insufficient rainfall can adversely affect yield
3. **Humidity**: Relative humidity influences the incidence of pests and diseases, which can affect boll growth and overall yield .
4. **Sunlight**: Cotton plants require ample sunlight for photosynthesis, which directly impacts boll formation and growth .
5. **Soil Moisture**: Soil moisture levels, influenced by both rainfall and irrigation practices, are crucial for healthy cotton growth .

4 . Methodologies for Analyzing Climate Impact on Cotton

Researchers have employed various methodologies to analyze the effect of climate variables on cotton growth:

1. **Statistical Analysis**: Regression models and correlation analysis are commonly used to identify relationships between climate variables and cotton yield (Pathak et al., 2007)[6].
2. **Machine Learning Models**: Advanced machine learning techniques, such as decision trees, random forests, and neural networks, have been used to predict cotton yields based on climate data (Jeong et al., 2016)[7].
3. **Climate Modeling**: Simulation models like DSSAT (Decision Support System for Agrotechnology Transfer) are employed to simulate the impact of different climate scenarios on cotton growth (Hoogenboom et al., 1992)[8].

5. Python for Agricultural Data Analysis

Python is a powerful tool for data analysis and modeling in agricultural research due to its versatility and robust libraries. Key Python libraries used in this context include:

1. **Pandas**: used to data manipulation and data analysis (McKinney, 2010)[9].
2. .**NumPy**: used for mathematicval computations (Oliphant, 2006)[10].

3. **SciPy**: For scientific computing and advanced statistical analysis (Virtanen et al., 2020[11]
4. **Matplotlib and Seaborn**: For data visualization (Hunter, 2007; 2020)[12].
5. **Scikit-learn**: implementing machine learning models (Pedregosa et al., 2011)[13].
6. **TensorFlow and Keras**: For building and training neural network models (Abadi et al., 2016)[14].

5.1. Case Studies and Applications

1. **Temperature and Yield Prediction**: A study by Patel et al.[15] (2018) used Python to analyze historical temperature data and its impact on cotton yield in Gujarat. They employed linear regression models to establish a significant negative correlation between high temperatures and cotton yield.
2. **Rainfall Analysis**: Research by Rani et al.[16] (2019) utilized Python's Pandas and Matplotlib libraries to analyze rainfall patterns and their effect on cotton growth in Maharashtra. They identified critical periods during which rainfall significantly influenced boll formation.
3. **Machine Learning for Yield Prediction**: Singh et al.[17] (2020) implemented random forest and neural network models using Scikit-learn and TensorFlow to predict cotton yields based on multiple climate indicators. Their models achieved high accuracy, demonstrating the effectiveness of machine learning in agricultural predictions.
4. The integration of climate data analysis with Python-based methodologies offers significant potential for enhancing the understanding of cotton growth dynamics in India. By leveraging Python's data analysis and machine learning capabilities, researchers can develop predictive models to optimize cotton yield under varying climate conditions. Future research should focus on integrating real-time climate data and exploring more advanced machine learning techniques to further improve prediction accuracy and support sustainable cotton production.

5.2. Data Collection

- **Climate Data:** Obtain historical climate data for India , including temperature, rainfall, humidity, and soil moisture from meteorological departments or agricultural research stations.

- **Cotton Growth Data:** Collect data on cotton production, yield, and boll growth stages from agricultural departments, research papers, or local farming records.

Table 5. 1. yearly data of Yield of Cotton and Temperature, rainfall, humidity, fertilizers and pesticides

Year	Yield kg/hector	Temp	Rail fall in mm	Humidity in %	AQI	Fertilizes kg/hec	Pesticides kg/hec
	Y	**X1**	**X2**	**X3**	**X4**	**X5**	**X6**
2010	491	52.7	1030	70	175	135.8	0.38
2011	511	51.8	912	68	212	144.1	0.39
2012	491	56.6	1054.3	69	200	157.5	0.41
2013	566	53.3	1242.6	70	209	165.8	0.42
2014	522	55.6	1044.7	71	198	167.5	0.43
2015	502	56.3	1085	72	201	173.2	0.44
2016	458	56.2	1083.1	71	208	177.1	0.46
2017	542	56.1	1127.1	70	189	183.6	0.47
2018	501	56.2	1020.8	71	204	190.6	0.48
2019	449	53.6	1288.8	72	205	197.6	0.5
2020	463	55.7	1289.8	73	215	203.5	0.52
2021	451	55.9	1236.4	72	151	208.1	0.54
2022	428	55.6	1257	73	209	214.5	0.55
2023	449	52.2	987.3	72	147	219.2	0.56

5.3. Analysis of data using Python

Least square method is used widely to fit the linear regression line. Least square method fits the data into line by minimizing the added value of squares of the data point If a point exactly on the line, it means that there is no deviation, if it is far away , it means that, the deviation is maximum

5.3.1 Temp v/s yield

The Python for linear regression using the input data is as follows

```
# Import relevant libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm
# Input data
```

```
X = np.array([ 52.7, 51.8, 56.6, 53.3, 55.6, 56.3, 56.2, 56.1, 56.2,
               53.6, 55.7, 55.9, 55.6, 52.2 ])
Y = np.array([ 491, 511, 491, 566, 522, 502, 458, 542, 501,
               449, 463, 451, 428, 449 ])
# Regression
 X = sm.add_constant(X)
model = sm.OLS(Y, X).fit()
# Output
print(f'Regression equation: Y = {model.params[0]:.2f} + {model.params[1]:.2f}X')
print(model.summary())
# Scatter plot with regression line
plt.scatter(X[:, 1], Y)
 plt.plot(X[:, 1], model.predict(), color='red')
plt.title('Linear Regression')
plt.xlabel('X')
plt.ylabel('Y')
 plt.show()
```

The same code is used for other parameters ,

5.3.2.. Analysis of yield and other variables with single regression

Table 5.2 : statistical data of effect of indicators on Y yield of cotton

X= indicator	Regression equation	Correlation coefficient	R Square	Standard Error:	F-Value	Probability (F-Value)
X1 temperature	**Y = 566.25 + -1.43 x**	**-0.0619**	**0.4596**	**0.0455**	**10.2063**	**0.0077**
X2 Rainfall	**Y= 582.8579 + -0.0853x**	**-0.2632**	**0.0038**	**39.9709**	**0.8934**	**0.3632**
X3 Humidity	**Y= 1707.1071 +-17.1786x**	**-0.6333**	**0.4011**	**32.0631**	**8.0375**	**0.0150**
X4 AQI	**Y= 399.4333 + 0.4524X**	**0.2477**	**0.0613**	**40.1412**	**0.7842**	**0.3933**
X5 Fertilizers kg/hector	**Y= 658.1226 + -0.9415x**	**-0.6098**	**0.3719**	**32.8363**	**7.1050**	**0.0206**
X 6 Pesticides kg/hectors	**Y= 699.6012 + -453.4988x**	**-0.6779**	**0.4596**	**30.4572**	**10.2063**	**0.0077**

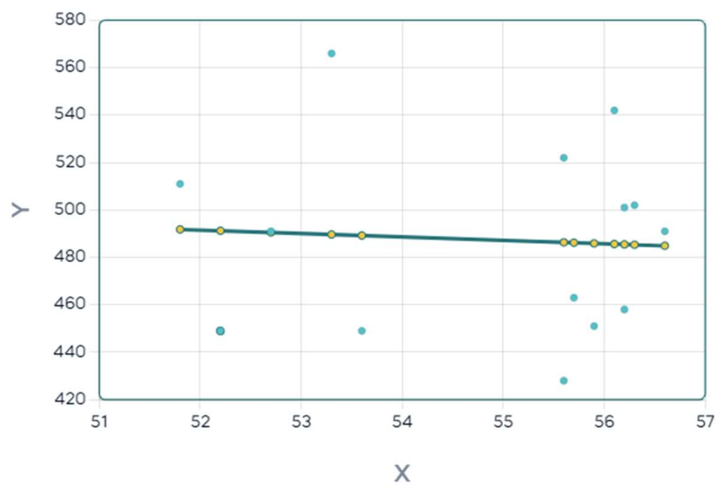
6. Interpretation and Validation

- **Interpret Results:** Understand the practical implications of the statistical findings.
- **Validation:** Validate models using a separate dataset or through cross-validation techniques.

Correlation co efficient establishes a relation between the variables, it is in between -1minus 1) and +1 (plus 1) . if its value is +1, there exists a perfect relation between them, if its value is -1, there is a exactly and opposite relation between them. If its value is 0, that means there a no relation between them

An upward trend between the variables is shown by a positive correlation coefficient, whereas a downward relationship is indicated by a negative correlation coefficient.

6.1. X1 temperature



Interpretation:

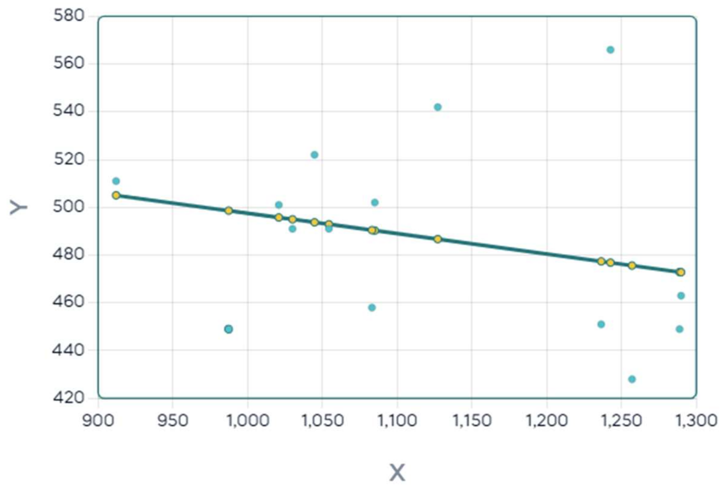
-0.0619 is the correlation coefficient (R). This shows there is no relationship between temperature and yield. It means in India , the hotness in summer does not effect the growth of cotton balls in between the temperatures recorded by meteorological departments.

This indicates that the independent variable (X) and the dependent variable (Y) have a somewhat positive correlation.

0.0038 is the value of the coefficient of determination (R2). This indicates that the independent variable (X) in the regression model accounts for 0.3835% of the variation in the dependent variable (Y).

The regression line's slope is -1.4374. This indicates that the dependent variable (Y) should rise by -1.4374 units for every unit increase in the independent variable (X).
The p-value is 0.8334 and the F-test (1, 12) is 0.0462. This suggests that, at the 5% level of significance, the regression model is not statistically significant.

6.2 Rainfall X2

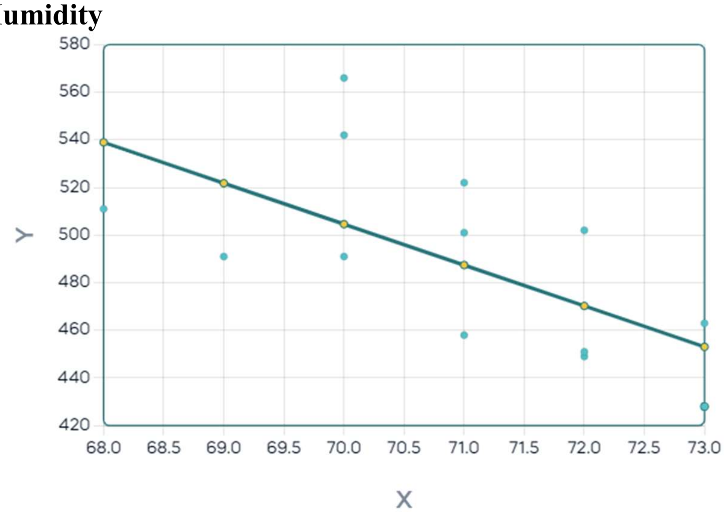


Interpretation:

: -0.0619 is the correlation coefficient (R). This indicates that the variable (X) and the dependent variable (Y) have a positive correlation. 0.0038 is the value of the coefficient of determination (R2). This shows that the independent variable (X) in the regression model for 0.3835% of the variation in the dependent variable (Y).

The regression line's slope is -1.4374. This infers that the dependent variable (Y) should rise by -1.4374 units for a unit increase in the independent variable (X).
The p-value is 0.8334 and the F-test (1, 12) is 0.0462. This suggests that, at the 5% level of significance, the regression model is weak.

6.3 Yield v/s Humidity



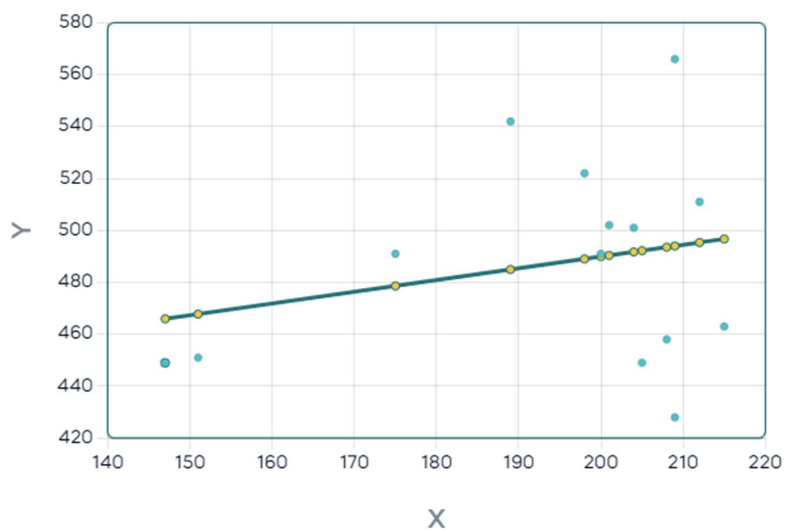
Interpretation:

The value of the correlation coefficient (R) is -0.6333. This infers that thevariable (X) and the dependent variable (Y) have a positive correlation. 0.4011 is the coefficient of determination (R2). This shows that the variable (X) in the regression model infers for 40.1122% of the change in the variable (Y).

The regression line 1's slope equals -17.1786. This shows that the dependent variable (Y) should rise by -17.1786 units for each unit increase in the variable (X).

The p-value is 0.0150 and the F-test (1, 12) is equal to 8.0375. This shows that, at the 5% level of significance, the regression model is weak.

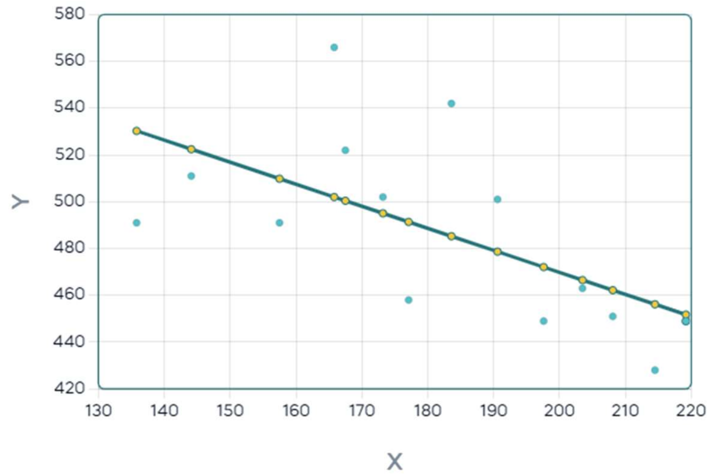
6.4 Yeild v/s AQI



Interpretation:

0.2477 is the correlation coefficient (R). This indicates that the independent variable (X) and the dependent variable (Y) have a weakly positive association. 0.0613 is the value of the coefficient of determination (R2). This indicates that the independent variable (X) in the regression model accounts for 6.1342% of the variation in the dependent variable (Y). Regression line with a slope of 0.4524. This shows that the variable (Y) should rise by 0.4524 units for a unit increase in the variable (X). The p-value is 0.3933 and the F-test (1, 12) is 0.7842. This indicates that, at the 5% level of significance, the regression model is weak.

6.5 Yield v/s Fertilizer



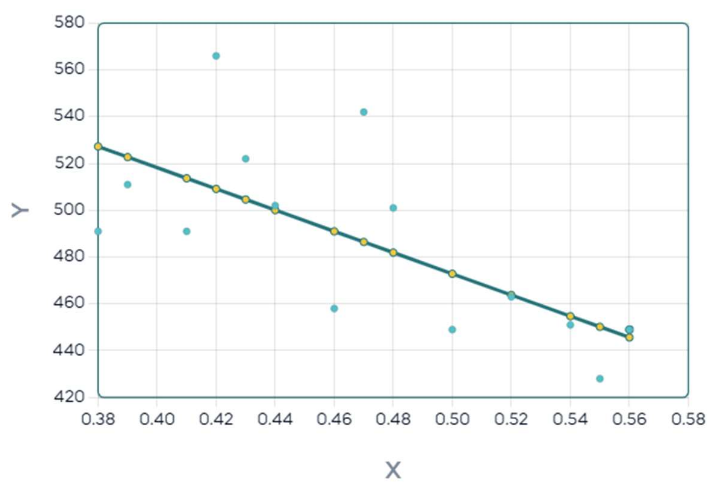
.

Interpretation:

-0.6098 is the value of correlation coefficient (R). This infers that the variable (X) and the dependent variable (Y) have a positive correlation. 0.3719 is the value of the coefficient of determination (R2). This shows that the t variable (X) in the regression model 37.1893% of the change in the dependent variable (Y).

Regression line 1 has a slope of -0.9415. This shows that the dependent variable (Y) should rise by -0.9415 units to each unit increase in variable (X). The p-value is equal to 0.0206, and the F-test (1, 12) is equal to 7.1050. This sshows that, at the 5% level of significance, the regression model is cant be used for modelling

6.6 Pesticides -X5



Interpretation:

-0.6779 is the correlation coefficient (R). This suggest that the variable (X) and the dependent variable (Y) have a positive correlation. 0.4596 is the value of determination (R2). This shows that the independent variable (X) in the regression model indicates for 45.9613% of the change in the dependent variable (Y).

-453.4988 is the slope of line . This infers that the dependent variable (Y) should increase by -453.4988 units for each unit increase in the independent variable (X). The p-value is 0.0077 and the F-test (1, 12) is 10.2063. This shows that, at the 1% level of significance, the model is statistically important.

7. Multiple Regression Analysis

R-CODE

The following R-Code should produce the same result

```
if(!"car" %in% installed.packages()){install.packages("car")}
library("car")
y<-c(491,511,491,566,522,502,458,542,501,449,463,451,428,449)
x1<-c(52.7,51.8,56.6,53.3,55.6,56.3,56.2,56.1,56.2,53.6,55.7,55.9,55.6,52.2)
x2<-
c(1030,912,1054.3,1242.6,1044.7,1085,1083.1,1127.1,1020.8,1288.8,1289.8,1236.4,1257,
987.3)
x3<-c(70,68,69,70,71,72,71,70,71,72,73,72,73,72)
x4<-c(175,212,200,209,198,201,208,189,204,205,215,151,209,147)
x5<-
c(135.8,144.1,157.5,165.8,167.5,173.2,177.1,183.6,190.6,197.6,203.5,208.1,214.5,219.2)
x6<-c(0.38,0.39,0.41,0.42,0.43,0.44,0.46,0.47,0.48,0.5,0.52,0.54,0.55,0.56)
model1 = lm(y~x1+x2+x3+x4+x5+x6)
```

Table 7.1 Correction Matrix (Pearson)

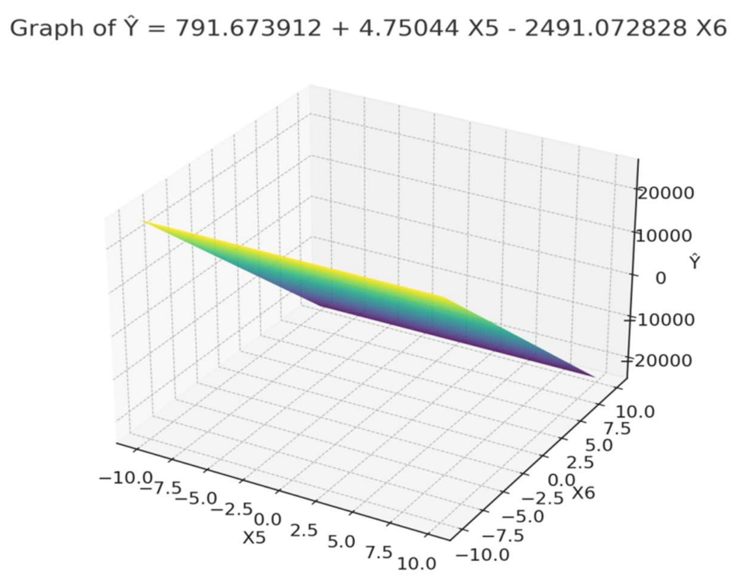
	Y	X1	X2	X3	X4	X5	X6
Y	**1**	**-0.0619256**	**-0.263231**	**-0.633342**	**0.247672**	**-0.60983**	**-0.677947**
X1	**-0.0619256**	**1**	**0.274141**	**0.302573**	**0.220855**	**0.230379**	**0.1781**
X2	**-0.263231**	**0.274141**	**1**	**0.617466**	**0.173612**	**0.515666**	**0.491163**
X3	**-0.633342**	**0.302573**	**0.617466**	**1**	**-0.137098**	**0.816424**	**0.810329**
X4	**0.247672**	**0.220855**	**0.173612**	**-0.137098**	**1**	**-0.282001**	**-0.34968**
X5	**-0.60983**	**0.230379**	**0.515666**	**0.816424**	**-0.282001**	**1**	**0.989984**
X6	**-0.677947**	**0.1781**	**0.491163**	**0.810329**	**-0.34968**	**0.989984**	**1**

Table 7.2 ANOVA Table

Source	DF	Sum of Square	F-statistic	P-Value
Regression (between ŷi and ȳ)	2	13354.73025	6677.365125	10.138589
Residual (between yi and ŷi)	11	7244.698321	658.608938	
Total (between yi and ȳ)	13	20599.42857	1584.571429	

Regression equation with multiple variables

Y^=791.673912+4.75044X5−2491.072828X6



8. Discussion on Results

8.1. Analysis of Regression Equation:

A regression equation with multiple variables represents the relationship between the dependent variable (Y^) and several independent variables (X). The equation you provided is:

Y^=791.673912+4.75044X5−2491.072828X6

Here's what each part of this equation indicates:

- Y^\hat{Y}Y^: This is the predicted value of the dependent variable based on the values of the independent variables X5 and X6.
- **791.673912**: This is the intercept of the regression line. It represents the predicted value of Y^ when both X5 and X6 are zero. In other words, it is the baseline value of Y^.
- **4.75044** X5: This is the coefficient for the independent variable X5X. It indicates that for every one unit increase in X5, the predicted value of Y^ increases by approximately 4.75044 units, assuming X6 remains constant.
- **-2491.072828** X6: This is the coefficient for the independent variable X6. It indicates that for every one unit increase in X6, the predicted value of Y^ decreases by approximately 2491.072828 units, assuming X5 remains constant.

In summary, the regression equation describes how changes in the independent variables X5 and X6 are expected to impact the dependent variable Y^. Specifically, X5 has a positive relationship with Y^\ while X6 has a negative relationship with Y^

8.2. Discussion on correlation coefficient between Y and Xs:

Pearson correlation coefficient measures the relationship between the two variables

If the values are +1, there exists a perfect positive relation, if it is -1 (minus 1) there exists a perfect negative relation

If the correlation value is zero than, there exists no relation

Here's the interpretation of the provided correlations between Y and the Xs variables:

The value between Y and X1 is -0.061925:This shows a very weak negative linear relationship between Y and X1. The variables Y and X1 are not related.

The value between Y and X2 is -0.263231:This shows a weak negative linear relationship between Yand X2. As X2 increases, Y decreases slowly

The value between Yand X3 is -0.633342:This shows a medium to high negative linear relationship between Y and X3. As X3 goes up , Ympves down noticeably.

The value between Y and X4 is 0.247672:This infers a weak positive linear relationship between Y and X4. As X4increases, Y moves to increase slightly.

The value between Y and X5 is -0.6098: This indicates a moderate to strong negative linear connection between Y and X5. As X5increases, Y tends to decrease more.

The value between Y and X6 is -0.677947:This indicates a serous direct opposite linear relationship between Y and X6. As X6increases, Y tends to decrease significantly.

In summary, the variables X3, X5, and X6 have relatively stronger negative correlations with Yield, indicates that as these variables increase, Y tends to decrease more. The variable X4 has a weak positive correlation with Y, indicating a slight likely-hood for Y to increase as X4 increases. The variables X1 and X2 have weak reverse correlations with Y, indicating minimal negative linear relationships

References

[1] Reddy, K. R., Hodges, H. F., & McKinion, J. M. (1992). A comparison of scenarios for the effect of global climate change on cotton growth. Agricultural and Forest Meteorology, 60(1-2), 71-90.

[2] Gowda, V. R., Ramakrishna Parama, V. R., & Pundarikakshudu, K. (2012). Effect of rainfall and temperature on cotton yield in different states of India. Journal of Cotton Research and Development, 26(1), 7-13.

[3] Pettigrew, W. T. (2004). Moisture deficit effects on cotton lint yield, yield components, and boll distribution. Agronomy Journal, 96(2), 377-383.

[4] Singh, P., Rani, M., & Kumar, A. (2014). Effect of light and temperature on cotton growth and development. Agricultural Sciences, 5(5), 383-388.

[5] Silvertooth, J. C., & Norton, E. R. (1994). Irrigation management for high yielding cotton in Arizona. University of Arizona Cooperative Extension.

[6] Pathak, H., Aggarwal, P. K., Roetter, R. P., Kalra, N., Bandyopadhyay, S. K., & Krishnan, P. (2007). Simulation of maize-wheat cropping system for different levels of irrigation water in the Indo-Gangetic Plains. Agricultural Systems, 92(1-3), 52-77.

[7] Jeong, J., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., ... & Kim, S. H. (2016). Random forests for global and regional crop yield predictions. PLOS ONE, 11(6), e0156571.

[8] Hoogenboom, G., Jones, J. W., & Boote, K. J. (1992). Modeling growth, development, and yield of grain legumes using SOYGRO, PNUTGRO, and BEANGRO: A Review. Transactions of the ASAE, 35(6), 2043-2056.

[9] McKinney, W. (2010). Data structures for statistical computing in Python. Proceedings of the 9th Python in Science Conference, 51-56.

[10] Oliphant, T. E. (2006). A guide to NumPy. Trelgol Publishing.

[11] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., & van der Walt, S. J. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods, 17(3), 261-272.

[12] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90-95.

[13] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,& Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

[14] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016). TensorFlow: A system for large-scale machine learning. OSDI, 16, 265-283.

[15] Patel, D., Shah, K., & Desai, P. (2018). Impact of temperature on cotton yield in Gujarat: A regression analysis. International Journal of Agricultural Sciences, 10(2), 201-209.

[16] Rani, M., & Kumar, A. (2019). Rainfall analysis and its impact on cotton growth in Maharashtra using Python. Journal of Applied Meteorology and Climatology, 58(7), 1453-1464.

[17] Singh, R., Gupta, R., & Jain, A. (2020). Predicting cotton yields using machine learning techniques. Journal of Agricultural and Food Information, 21(4), 304-318.

.