**IMPACT OF CLIMATIC EXTREMES ON SOYBEAN AND CORN PRODUCTION IN ONTARIO**

**Name: Sundas Ahmed**

**Student Number: 500910967**

**Data: 10th November 2023**

**CIND820: Big Data Analytics Project**

**Results:**

Regression Analysis for Corn:

OLS Regression Results

==============================================================================

Dep. Variable: Corn R-squared: 0.436

Model: OLS Adj. R-squared: 0.404

Method: Least Squares F-statistic: 13.41

Date: Tue, 07 Nov 2023 Prob (F-statistic): 1.33e-06

Time: 19:48:29 Log-Likelihood: -98.041

No. Observations: 56 AIC: 204.1

Df Residuals: 52 BIC: 212.2

Df Model: 3

Covariance Type: nonrobust

===================================================================================

coef std err t P>|t| [0.025 0.975]

-----------------------------------------------------------------------------------

const -22.2506 7.825 -2.843 0.006 -37.953 -6.548

Precipitation 0.1479 0.047 3.174 0.003 0.054 0.241

Max Temperature 2.4303 0.683 3.558 0.001 1.060 3.801

Min Temperature -1.5259 0.693 -2.202 0.032 -2.916 -0.136

==============================================================================

Omnibus: 0.316 Durbin-Watson: 0.699

Prob(Omnibus): 0.854 Jarque-Bera (JB): 0.462

Skew: 0.152 Prob(JB): 0.794

Kurtosis: 2.675 Cond. No. 2.72e+03

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.72e+03. This might indicate that there are

strong multicollinearity or other numerical problems.

The ordinary least squares model is used to understand the relationship between the dependent and independent variables. The independent variables (Precipitation, Maximum Temperature, and Minimum Temperature explain about 43% of the variation in the dependent variable i.e. Corn.

To check the overall significance of the model, the lower p-value and higher F statistics suggest that the model as a whole is statistically significant. We have used Jarque-Bera Tests to check the normality and distribution of the residuals, the results show that the data is normally distributed. The Durbin-Watson test results also confirm little to no autocorrelation.

The results of the regression model confirm our hypothesis, the results are described below:  
**Maximum Temperature**: The maximum temperature is positively associated with corn yield. A 1-degree Celsius increase in maximum temperature will lead to 2.43 metric tons increase in corn yield, keeping all other variables constant.

**Minimum temperature**: The minimum temperature is negatively associated with corn yield. A 1-degree Celsius increase in minimum temperature will lead to 1.526 metric tons decrease in corn yield, keeping all other variables constant.

**Precipitation:** Precipitation is positively associated with corn yield. A 1 millimeter increase in precipitation will lead to 0.15 metric tons increase in corn yield, keeping all other variables constant.

Regression Analysis for Soybean:

OLS Regression Results

==============================================================================

Dep. Variable: Soybean R-squared: 0.253

Model: OLS Adj. R-squared: 0.210

Method: Least Squares F-statistic: 5.860

Date: Tue, 07 Nov 2023 Prob (F-statistic): 0.00159

Time: 19:49:07 Log-Likelihood: -29.068

No. Observations: 56 AIC: 66.14

Df Residuals: 52 BIC: 74.24

Df Model: 3

Covariance Type: nonrobust

===================================================================================

coef std err t P>|t| [0.025 0.975]

-----------------------------------------------------------------------------------

const -2.7097 2.283 -1.187 0.241 -7.292 1.872

Precipitation 0.0247 0.014 1.813 0.076 -0.003 0.052

Max Temperature 0.4476 0.199 2.245 0.029 0.048 0.848

Min Temperature -0.2624 0.202 -1.298 0.200 -0.668 0.143

==============================================================================

Omnibus: 13.793 Durbin-Watson: 1.202

Prob(Omnibus): 0.001 Jarque-Bera (JB): 19.291

Skew: -0.856 Prob(JB): 6.47e-05

Kurtosis: 5.310 Cond. No. 2.72e+03

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 2.72e+03. This might indicate that there are

strong multicollinearity or other numerical problems.

The ordinary least squares model is used to understand the relationship between the dependent and independent variables. The independent variables (precipitation, Maximum Temperature, and Minimum Temperature explain about 25.3% of the variation in the dependent variable i.e. soybean.

To check the overall significance of the model, the lower p-value and higher F statistics suggest that the model as a whole is statistically significant, which in case of soybean is p-value 0.00159 and F statistics 5.860. We have used Jarque-Bera Tests to check the normality and distribution of the residuals, the results show that the data is not normally distributed. The Durbin-Watson test results confirms a positive autocorrelation.

The results of the regression model confirm our hypothesis, the results are described below:  
**Maximum Temperature**: The maximum temperature is positively associated with soybean yield. A 1-degree Celsius increase in maximum temperature will lead to increase of 0.44 metric tons in soya bean yield, keeping all other variables constant.

**Minimum temperature**: The minimum temperature is negatively associated with soybean yield. A 1-degree Celsius increase in minimum temperature will lead to 0.2624  metric tons decrease in soya bean yield, keeping all other variables constant.

**Precipitation:** Precipitation is positively associated with soybean yield. A 1 millimeter increase in precipitation will lead to 0.0247 n metric tons increase in soya bean yield, keeping all other variables constant.

Within the regression model for both Corn and Soya bean, the conditional numbers appears to be fairly large, indicating multicollinearity in the independent variables. Since, the multicollinearity indicates that the independent variables (Max temperature, min temperature and precipitation) might be highly correlated, it is difficult to interpret the individual impact of each variable. Henceforth, the natural log of the data will be taken to transform the variables to address the issue of multicollinearity.