**MDS531A: Econometrics**

**LAB 1: Simple Regression Analysis**

**Name: Neelanjan Dutta**

**Register number: 2448040**

**Tool used:**

**EViews (Econometric Views)** is a powerful statistical and econometric software tool widely used for time series, cross-section, and panel data analysis. It offers an intuitive graphical user interface along with advanced features for statistical modeling, forecasting, and hypothesis testing. EViews is especially popular in economics, finance, and social sciences due to its ability to handle large datasets, perform regression analysis, and generate high-quality graphs.

The software allows users to import data from various sources, perform descriptive and inferential statistical analyses, compute correlation and covariance matrices, estimate econometric models, and interpret relationships between variables with precision. With EViews, researchers and analysts can efficiently explore datasets, test assumptions, and build models for forecasting and decision-making.

**Dataset:**

|  |  |  |
| --- | --- | --- |
| Year | Y | X |
| 1982 | 3081.5 | 4620.3 |
| 1983 | 3240.6 | 4803.7 |
| 1984 | 3407.6 | 5140.1 |
| 1985 | 3566.5 | 5323.5 |
| 1986 | 3708.7 | 5487.7 |
| 1987 | 3822.3 | 5649.5 |
| 1988 | 3972.7 | 5865.2 |
| 1989 | 4064.6 | 6062 |
| 1990 | 4132.2 | 6136.3 |
| 1991 | 4105.8 | 6079.4 |
| 1992 | 4219.8 | 6244.4 |
| 1993 | 4343.6 | 6389.6 |
| 1994 | 4486 | 6610.7 |
| 1995 | 4595.3 | 6742.1 |
| 1996 | 4714.1 | 6928.4 |

**Dataset description:**

* **Year:** The time period for each observation, ranging from 1982 to 1996.
* **Y**: The dependent variable, which might represent a measure such as output, income, or demand.
* **X**: The independent variable, possibly representing a factor like investment, expenditure, or another influencing factor.

**Objective:**

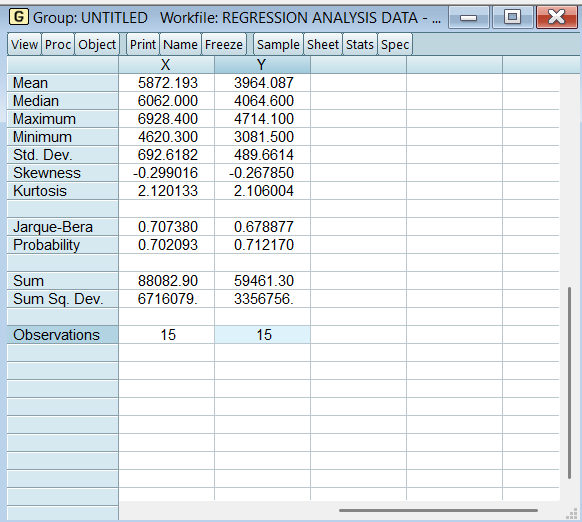
The purpose of this analysis is to examine the relationship between the variables X and Y using statistical methods such as descriptive statistics, correlation, covariance, and regression analysis in EViews.

**Results:**

1. **Dataset Uploading / Importing:**

The dataset containing yearly observations from 1982 to 1996 was imported into **EViews** from an Excel file named Regression Analysis Data-Sheet1.xlsx. The file was opened by selecting **File → Open → Foreign Data as Workfile**, and the Excel file was chosen from the system directory. The data includes three columns: Year, Y, and X. The Year column was set as the workfile’s date identifier by specifying an annual frequency, while Y and X were set as series for analysis.

1. **Descriptive Statistics:**

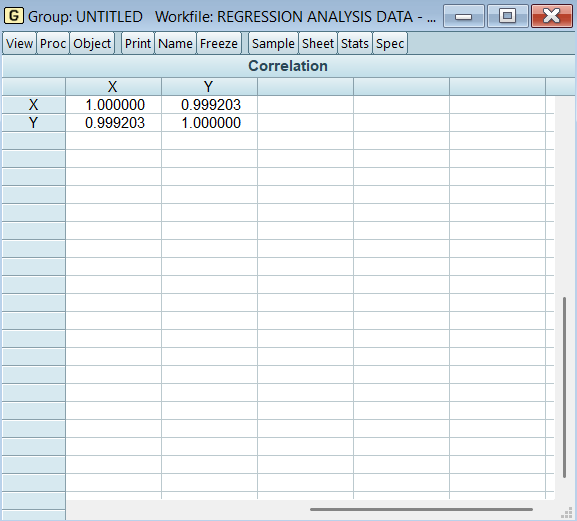
Descriptive statistics were computed by selecting **Quick→ Group Statistics → Descriptive Statistics → Individual Samples** after opening the dataset in EViews. The analysis provides summary measures such as the mean, standard deviation, minimum, and maximum values for both Y and X.

**Fig 1: Descriptive statistics**

**Interpretation:**

* Both X and Y show moderate variability around their means.
* The distributions are slightly negatively skewed but close to normal.
* The Jarque-Bera test confirms that both variables can be considered normally distributed, making them suitable for further analysis like correlation and regression.
* The data shows consistent trends without extreme outliers, providing a stable foundation for modeling relationships between X and Y.

1. **Correlation Analysis:**

The correlation between Y and X was computed using **Quick→ Group Statistics → Correlation**.

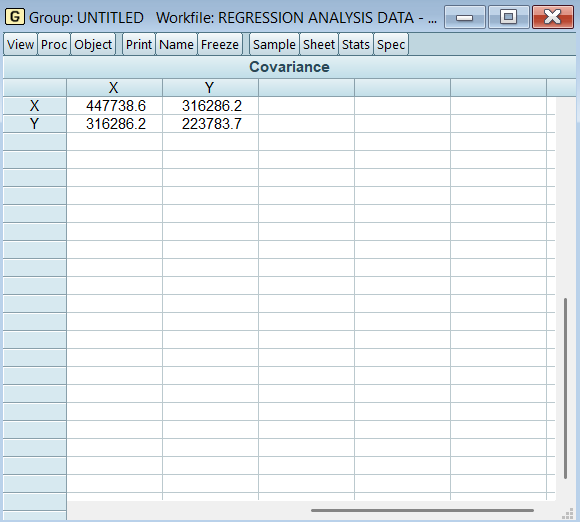
**Interpretation:**

The correlation coefficient was found to be **0.99203**, indicating an extremely **strong positive relationship** between the two variables. This suggests that as X increases, Y also tends to increase proportionally.

**Fig 2: Correlation Analysis**

1. **Covariance Analysis:**

The covariance between Y and X was computed using **Quick→ Group Statistics → Covariance**.



**Fig 3: Covariance Analysis**

**Interpretation:**

The covariance between **Y** and **X** was calculated to understand the direction of their relationship. A **positive covariance** of **316,286.2** confirms that both variables move together in the **same direction**. As **X** increases, **Y** also tends to increase, and when **X** decreases, **Y** follows the same trend. This supports the observed pattern in the dataset, indicating that higher values of **X** are associated with higher values of **Y**, reflecting a **strong positive relationship** between the two variables.

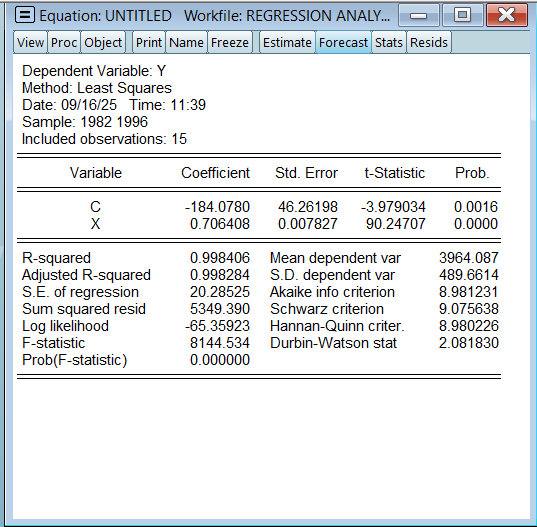
1. **Regression Analysis:**

A simple linear regression was performed using **Quick → Estimate Equation** in EViews. In the dialogue box that appears, we specified the regression equation by typing:

**y c x**

* **y**: The dependent variable we want to explain or predict.
* **c**: The constant term (intercept) in the regression equation.
* **x**: The independent variable that is used to explain changes in **y**.

After entering this, we clicked **OK**, and EViews estimated the regression model using the **Ordinary Least Squares (OLS)** method.



**Fig 4: Regression Analysis**

**Interpretation:**

1. **Regression Coefficients:**

**Table 1: Regression coefficients**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient | Std. Error | t-Statistic | Probability |
| C | **-184.0780** | **46.26198** | **-3.979** | **0.0016** |
| X | **0.706408** | **0.007827** | **90.247** | **0.0000** |

1. **Intercept (C = -184.0780):**

* The constant term is approximately **-184.08**.
* It indicates that if **X = 0**, **Y** would be around **-184.08**, although this is not practically relevant since **X = 0** is outside the data range.
* The associated **t-statistic (-3.979)** and **p-value (0.0016)** show that the intercept is statistically significant.

1. **Coefficient of X (0.706408):**

* The slope is approximately **0.7064**, meaning for every 1 unit increase in **X**, **Y** increases by about **0.71 units**.
* The **t-statistic (90.247)** is extremely high and the **p-value (0.0000)** confirms that this coefficient is highly significant.

**Goodness of Fit:**

1. **R-squared = 0.998406**

* About **99.84%** of the variation in **Y** is explained by **X**.
* This is an excellent fit, suggesting that **X** is a very strong predictor of **Y**.

1. **Adjusted R-squared = 0.998284**

* Adjusted for the number of observations, it confirms the model’s robustness.

1. **Standard Error of the Regression = 20.285**

* The average distance between the actual and predicted values of **Y** is around **20.29 units**, indicating high prediction accuracy.

**Model Diagnostics:**

1. **Sum Squared Residual = 5349.39**
   * The total squared error between observed and predicted values is relatively small.
2. **Log-likelihood = -65.3592**
   * Used for comparing models and likelihood estimation.
3. **F-statistic = 8144.534** and **Prob(F-statistic) = 0.0000**
   * The overall regression is statistically significant, meaning the model explains the variation in **Y** well.
4. **Akaike, Schwarz, and Hannan-Quinn criteria**
   * These are information criteria used for model selection, indicating good fit in this case.
5. **Durbin-Watson stat = 2.081830**
   * This value is close to 2, suggesting that there is no significant autocorrelation in the residuals.

**Equation interpretation:**

The equation was specified with Y as the dependent variable and X as the independent variable. The regression output provided the following equation:

**Y = −184.08 + 0.71 X**

This implies that for **every unit increase** in **X**, the **value of Y increases** by approximately **0.71 units**. The intercept term suggests that when X = 0, Y would theoretically be around -184.08, though this is outside the range of the observed data and should be interpreted cautiously. The model shows a very strong relationship between X and Y, supported by the high significance of the coefficient.

**Conclusion:**

The analysis of the dataset revealed that both X and Y exhibit consistent growth over the period, with moderate variability as shown by their standard deviations. The correlation coefficient between the two variables was found to be extremely high, indicating a strong positive relationship. This was further supported by a positive covariance, confirming that increases in X are associated with increases in Y. The regression analysis provided a statistically significant model with a high R-squared value, explaining over 99% of the variation in Y based on X. The results clearly demonstrate that X is a key factor influencing Y, and the relationship is both robust and reliable for prediction purposes.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*