STATISTICAL REGRESSION ANALYSIS ON COMBINED CYCLE POWER PLANT DATA TO PREDICT THE NET HOURLY ELECTRICAL ENERGY OUTPUT OF THE PLANT

Group U

192055 – G.A.M.K. Jayathilaka,

192059 – T.N.D. Kodippily,

192126 – W.L.T. Sankalpani,

192135 – M.A.S. Shavindi

Department of Mathematical Sciences

Faculty of Applied Sciences

Wayamba University of Sri Lanka

Final Project

STAT 3232: Data Analysis and Preparation of Statistical Reports

Ms.P.M.O.P. Panahatipola and Ms.M.V.N. Thejani

November 06, 2023

ABSTRACT

This statistical regression analysis report explores the intricate relationship between Ambient Temperature (AT), Ambient Pressure (AP), Relative Humidity (RH), and Exhaust Vacuum (V) with respect to cycle power plant performance. Cycle power plants represent a diverse array of energy generation technologies, each with unique operational considerations. This study aims to unravel the complex interplay of these variables and their influence on power plant electrical energy output, efficiency, and emissions. Through a rigorous analysis, we have identified significant findings and correlations.

The primary research problem is to understand how Ambient Temperature (AT), Ambient Pressure (AP), Relative Humidity (RH), and Exhaust Vacuum (V) impact the performance of cycle power plants, with a focus on electrical energy output, efficiency, and emissions.

Key Objectives of this report are to identify the correlation and utilize regression analysis to identify and quantify the statistical correlations between Ambient Temperature (AT), Ambient Pressure (AP), Relative Humidity (RH), and Exhaust Vacuum (V). Development of robust regression models to predict power plant performance based on Ambient Temperature (AT), Ambient Pressure (AP), Relative Humidity (RH), and Exhaust Vacuum (V) providing a valuable tool for optimization and prediction is also another objective of this analysis.

Offering data-driven information to empower researchers, engineers, and policymakers in the energy sector to make informed decisions that enhance energy efficiency and reduce environmental impact. By improving the efficiency of the power plant, operators can realize significant cost savings over time. This not only benefits the organization but can also lead to more competitive pricing for consumers.

This study employs advanced regression analysis techniques to model the relationships between the aforementioned variables and power plant performance metrics. We utilize a comprehensive dataset to derive statistical insights and develop predictive models that serve as a foundation for understanding and optimizing cycle power plant operations. By applying robust statistical methodologies, this research provides valuable guidance for the energy industry's pursuit of cleaner, more efficient power generation.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

In an era where energy efficiency, environmental sustainability, and reliable power generation are of paramount importance, the statistical regression analysis of combined cycle power plant has emerged as a key area of study. Cycle power plants, which encompass a variety of energy generation technologies, are subject to a multitude of variables that influence their performance, operational efficiency and environmental impact of these power plants. The dataset contains data collected from a combined cycle power plant when the power plant was set to work with full load. Features consists of hourly ambient variables Ambient Temperature (AT) in Celsius(ºC), Ambient Pressure (AP) in milibar, Relative Humidity (RH) percentage, and Exhaust Vacuum (V) in cmHg to predict the net hourly electrical energy output of the plant (EP) in MW. A combined cycle power plant is composed of gas turbines, steam turbines and heat recovery steam generators. In a combined cycle power plant, the electricity is generated by gas and steam turbines, which are combined in one cycle, and is transferred from one turbine to another. While the vacuum is collected from and has effect on the steam turbine, the other three ambient variables effect the gas turbine performance.

This statistical report embarks on an exploration of the complex relationship between combined cycle power plant performance and the influential variables of Ambient Temperature (AT), Ambient Pressure (AP), Relative Humidity (RH), and Exhaust Vacuum (V) under ninety-five percent confidence level throughout the analysis. These factors collectively serve as crucial determinants of energy output, operational costs, and emissions. By delving into the statistical interplay of these elements, we aim to provide a comprehensive understanding of how each variable impacts power plant efficiency and environmental outcomes. The insights gained from this analysis will not only facilitate the optimization of energy production but also contribute to the development of greener, more sustainable power generation practices.

Amidst the ongoing global drive for cleaner energy and the ever-growing demand for electricity, it is imperative to scrutinize the performance of cycle power plants in great detail. This report endeavours to uncover vital statistical trends and findings that can guide the decisions of researchers, engineers, and policymakers within the energy sector. As we delve into the intricate web of correlations and dependencies between Ambient Temperature (AT), Ambient Pressure (AP), Relative Humidity (RH), Exhaust Vacuum (V), and power plant performance, we aim to provide actionable insights that will drive innovation and progress in the dynamic realm of cycle power generation.

2 STATISTICAL THEORY

2.1 What Is Regression Analysis?

Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables. It can be utilized to assess the strength of the relationship between variables and for modelling the future relationship between them.

Regression analysis includes several variations, such as linear, multiple linear, and nonlinear. The most common models are simple linear and multiple linear. Nonlinear regression analysis is commonly used for more complicated data sets in which the dependent and independent variables show a nonlinear relationship.

Regression analysis offers numerous applications in various disciplines, including finance.

2.2 Linear Regression Analysis

Linear regression analysis is based on six fundamental assumptions:

The dependent and independent variables show a linear relationship between the slope and the intercept.

- The independent variable is not random.
- The value of the residual (error) is zero.
- The value of the residual (error) is constant across all observations.
- The value of the residual (error) is not correlated across all observations.
- The residual (error) values follow the normal distribution.

2.3 Simple Linear Regression

Simple linear regression is a model that assesses the relationship between a dependent variable and an independent variable. The simple linear model is expressed using the following equation:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad \textit{Equation 1}$$

Where:

Y – Dependent variable

X – Independent (explanatory) variable

β_0 – Intercept

β_1 – Slope

ε – Random error term

In order to properly interpret the output of a regression model, the following main assumptions about the underlying data process of what you analyzing must hold:

- The relationship between variables is linear
- Homoscedasticity, or that the variance of the variables and error term must remain constant
- All explanatory variables are independent of one another
- All variables are normally-distributed

2.4 Multiple Linear Regression

Multiple linear regression analysis is essentially similar to the simple linear model, with the exception that multiple independent variables are used in the model. The mathematical representation of multiple linear regression is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon$$ *Equation 2*

Where:

Y – Dependent variable

$X_1, X_2, \ldots, X_n$ – Independent (explanatory) variables

β_0 – Intercept

$\beta_1, \beta_2, \ldots, \beta_n$ – Slopes of each explanatory variables

ε – Random error term

Multiple linear regression follows the same conditions as the simple linear model. However, since there are several independent variables in multiple linear analysis, there is another mandatory condition for the model:

Non-collinearity: Independent variables should show a minimum correlation with each other. If the independent variables are highly correlated with each other, it will be difficult to assess the true relationships between the dependent and independent variables. The multiple regression model is based on the following assumptions:

- There is a linear relationship between the dependent variables and the independent variables
- The independent variables are not too highly correlated with each other
- y_i observations are selected independently and randomly from the population
- Residuals should be normally distributed with a mean of 0 and variance (Taylor, 2022)

2.5 Scatter Plot Matrix

Scatter plot matrix shows the relationships between all the variables in the dataset. An increasing or decreasing pattern between two variables indicate a linear relationship or correlation between those two variables. Following plot shows the scatter plot matrix obtained for a dataset containing variables age, heart_rate, weight and indicator.

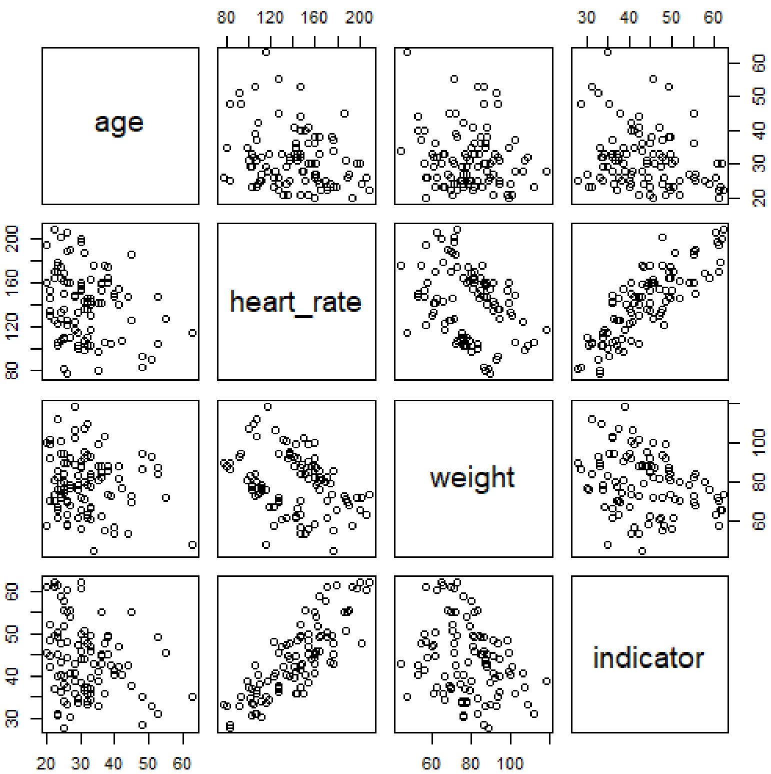

Figure 1 : Sample Scatter Plot Matrix

According to above plot indicator and heart rate has a linear relationship with weight and indicator as both shows decreasing and increasing pattern respectively.

2.6 Correlation Matrix

Correlation matrix shows the correlations of each variable with other variables. Correlation between two variables is higher than 0.7 implies that there is a high correlation between those two variables.

2.7 Significance of A Regression Model

Significance of a regression model is determined by the p value of the model.

Hypothesis

H_0: Regression model is not significant

H_1: Regression model is significant

P value of the model is less than considered level of significance indicates the significance of the model.

2.8 Significances of A Regression Model Coefficients

Significances of a regression model coefficients are determined by the p value of the model.

Hypothesis

H_0: Coefficient is not significant

H_1: Coefficient is significant

P value of the corresponding coefficient is less than considered level of significance indicates the significance of the coefficient.

2.9 Coefficient of Determination

The coefficient of determination (R^2 or r-squared) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other words, the coefficient of determination tells one how well the data fits the model (the goodness of fit).

Although the coefficient of determination provides some useful insights regarding the regression model, one should not rely solely on the measure in the assessment of a statistical model. It does not disclose information about the causation relationship between the independent and dependent variables, and it does not indicate the correctness of the regression model. Therefore, the user should always draw conclusions about the model by analyzing the coefficient of determination together with other variables in a statistical model.

Mathematically, the coefficient of determination can be found using the following formula:

$$Coefficient\ of\ Determination(R^2) = 1 - \frac{SS_{Regression}}{SS_{Total}} \quad \text{Equation 3}$$

Where:

$SS_{Regression}$ – The sum of squares due to regression (explained sum of squares)

SS_{Total} – The total sum of squares

The coefficient of determination can take any values between 0 to 1. In addition, the statistical metric is frequently expressed in percentages. (Corporate Finance Institute Team, n.d.)

2.10 Diagnostic Checking

The model fitting is just the first part of the story for regression analysis since this is all based on certain assumptions. Regression diagnostics are used to evaluate the model assumptions and investigate whether or not there are observations with a large, undue influence on the analysis. Again, the assumptions for linear regression are:

- Linearity: The relationship between dependent and independent variables are linear
- Homoscedasticity: The variances of residual are the same for all data
- Independence: Observations are independent of each other
- Normality: For any fixed value of model residuals are normally distributed
- Multicolinearity: There should not be any relationship between independent variables

2.10.1 Diagnostic Plots

The diagnostic plots show residuals in four different ways:

- **Residuals vs Fitted**: Used to check the linear relationship assumptions. A horizontal line, without distinct patterns is an indication for a linear relationship.

- **Normal Q-Q**: Used to examine whether the residuals are normally distributed. It's better if residual points follow the straight dashed line.

- **Scale-Location** (or Spread-Location). Used to check the homogeneity of variance of the residuals (homoscedasticity): Horizontal line with equally spread points is a good indication of homoscedasticity.

- **Residuals vs Leverage:** Used to identify influential cases, that is extreme values that might influence the regression results when included or excluded from the analysis. This plot will be described further in the next sections.

2.10.2 Linearity of The Data

The linearity assumption can be checked by inspecting the Residuals Vs Fitted plot. Following plot is the residuals vs fitted plot of a simple linear regression model.

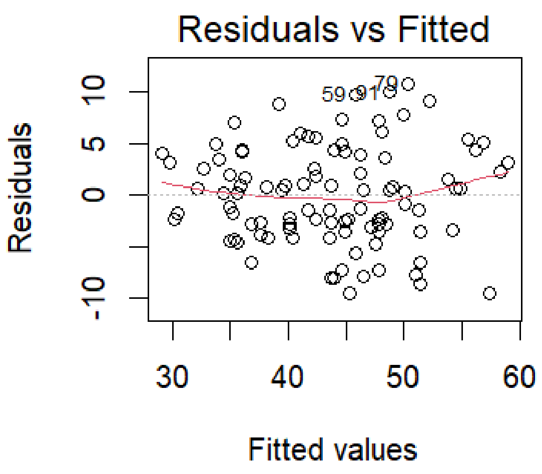

Figure 2 : Sample Residuals Vs Fitted Plot

The residual plot shows no fitted pattern. That is, the relationship between two variables is linear.

2.10.3 Homogeneity of Variance

This assumption can be checked by examining the *scale-location plot*, also known as the *spread-location plot*.

Following plot is the scale-location plot of a simple linear regression model.

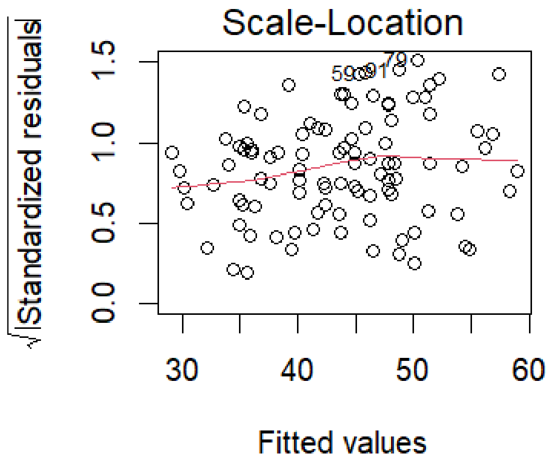

Figure 3 : Sample Scale - Location Plot

This plot shows residuals are spread equally along the ranges of predictors which indicates homoscedasticity of variances in the model.

If it can be seen that the variability (variances) of the residual points increases with the value of the fitted outcome variable, suggesting non-constant variances in the residuals errors (or *heteroscedasticity*), a possible solution to reduce the heteroscedasticity problem is to use a log or square root transformation of the outcome variable.

2.10.4 Normality of Residuals

The QQ plot of residuals can be used to visually check the normality assumption. The normal probability plot of residuals should approximately follow a straight line.

Following plot is the normal Q-Q plot of a simple linear regression model which was built to study the relationship between dependent variable sales with independent variable YouTube.

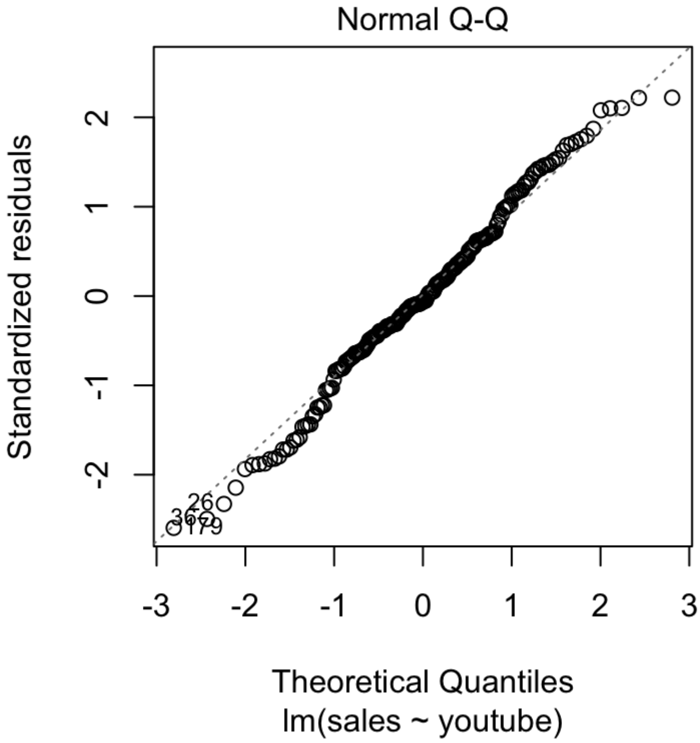

Figure 4 : Sample Normal Q-Q Plot

In the plot data points are almost scattered in the straight line which indicates residuals are normally distributed. (kassambara, 2018)

2.10.5 Independence of Residuals

The easiest way to check the assumption of independence is using the Durbin-Watson test. We can conduct this test using R's built-in function called *durbinWatsonTest* on our model. Running this test will give you an output with a p-value, which will help you determine whether the assumption is met or not.

The null hypothesis states that the errors are not auto-correlated with themselves (they are independent). Thus, if we achieve a p-value > 0.05, we would fail to reject the null hypothesis. This would give us enough evidence to state that our independence assumption is met. (Khan, 2021)

2.10.6 Multicolinearity

In regression analysis, multicollinearity occurs when two or more predictor variables are highly correlated with each other, such that they do not provide unique or independent information in the regression model.

If the degree of correlation is high enough between predictor variables, it can cause problems when fitting and interpreting the regression model.

The most straightforward way to detect multicollinearity in a regression model is by calculating a metric known as the variance inflation factor, often abbreviated VIF.VIF measures the strength of correlation between predictor variables in a model. It takes on a value between 1 and positive infinity.

We use the following rules of thumb for interpreting VIF values:

VIF = 1: There is no correlation between a given predictor variable and any other predictor variables in the model.

VIF between 1 and 5: There is moderate correlation between a given predictor variable and other predictor variables in the model.

VIF > 5: There is severe correlation between a given predictor variable and other predictor variables in the model

(Zach, 2022)

3 METHODOLOGY AND RESULTS

3.1 Content of The Data Set

Dependent variable: Net hourly electrical energy output (EP)

Independent variables:

- Ambient Temperature (AT)
- Ambient Pressure (AP)
- Relative Humidity (RH)
- Exhaust Vacuum (V)

3.2 Descriptive Statistics

Table 1 : Descriptive Statistics of Dataset

Statistic	AT	V	AP	RH	EP
Minimum	4.11	35.57	994.2	32.97	425.2
First Quartile	13.77	41.45	1008.5	62.68	438.9
Median	20.57	50.83	1012.5	74.00	451.3
Mean	19.74	54.09	1012.9	72.64	454.1
Third Quartile	25.91	66.55	1017.6	84.55	468.5
Maximum	34.33	79.74	1033.0	100.15	495.2
Variance	58.50337	166.3153	37.97586	224.9981	294.1408
Standard Deviation	7.64875	12.89633	6.162456	14.99994	17.15053
Range	4.11 to 34.33	35.57 to 79.74	994.17 to 1032.98	32.97 to 100.15	425.18 to 495.23
Interquartile Range	12.15	25.09	9.0825	21.8675	29.6825

3.3 Scatter Plot Matrix

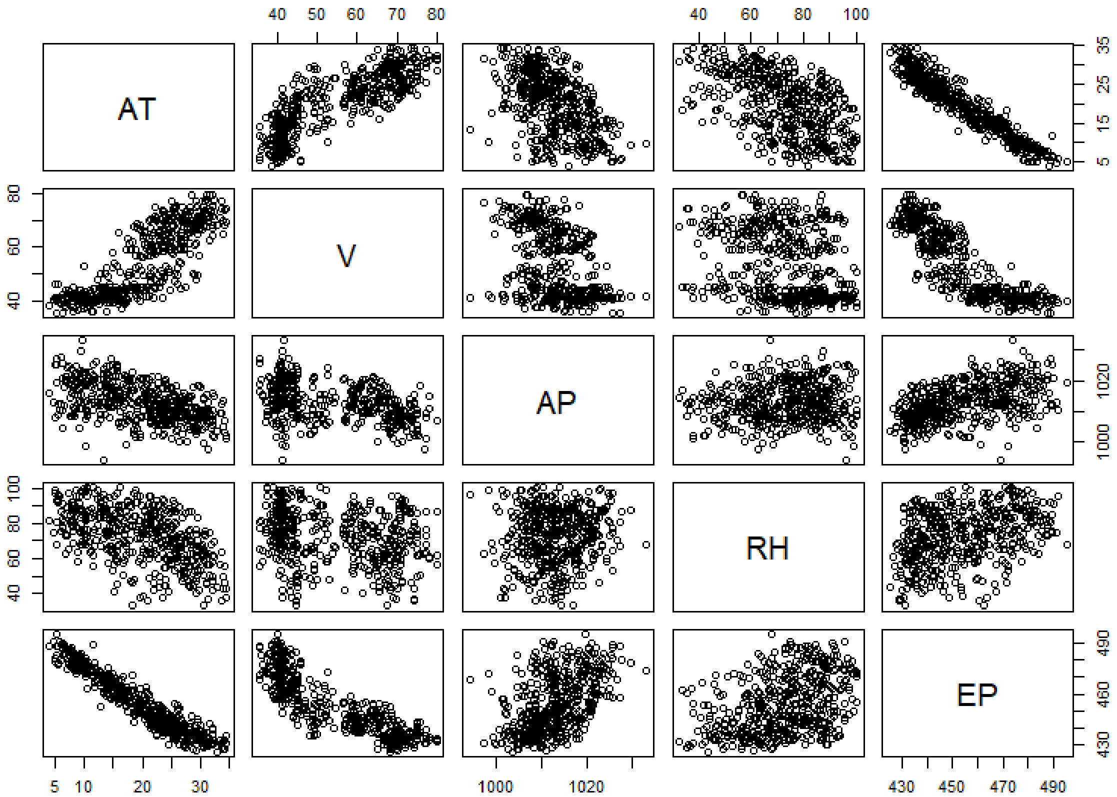

Figure 5 : Scatter Plot Matrix of Data

Scatter plot matrix indicates strong negative linear relationship between EP Vs AT and EP Vs V and weak positive linear relationship between EP Vs AP and EP Vs RH.

3.4 Correlation Matrix

Table 2 : Correlation Matrix of Data

Variable	AT	V	AP	RH	EP
AT	1.0000000	0.8325873	-0.51192125	-0.52708682	-0.9533148
V	0.8325873	1.0000000	-0.39764411	-0.26619260	-0.8611522
AP	-0.5119212	-0.3976441	1.00000000	0.08551756	0.5110355
RH	-0.5270868	-0.2661926	0.08551756	1.00000000	0.3866161
EP	-0.9533148	-0.8611522	0.51103553	0.38661609	1.0000000

Correlation between EP and AT, EP and V have high negative correlations -0.9533148 and -0.8611522 respectively. Correlations between EP and AP, EP and RH has low positive values 0.510355 and 0.3866161 respectively. This indicates AT, V are highly negatively correlated with EP and AP, RH are moderately correlated with EP.

3.5 Regression Model with All Explanatory Variables

Formula:

$$EP = 496.38392 - 1.99244 \times AT - 0.20135 \times V + 0.01789 \times AP - 0.14018 \times RH$$

Equation 4

Table 3 : Important Statistics of Regression Model with All Explanatory Variables

Important Statistics of the Model	Value of the Statistic
p-value	< 2.2e-16
Multiple R-squared	0.9334

P-Value of the model is less than 0.05. That is obtained model is significant at five present level of significance. Multiple R-squared value implies that 93.34 percent of variation in net hourly electrical energy output can be explained by the model obtained.

Table 4 : Estimates of Coefficients of Regression Model with All Explanatory Variables

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	496.38392	40.62265	12.219	< 2e-16
AT	-1.99244	0.06412	-31.072	< 2e-16
V	-0.20135	0.03038	-6.627	8.96e-11
AP	0.01789	0.03942	0.454	0.65
RH	-0.14018	0.01756	-7.981	1.02e-14

P-Value of variable AP is greater than 0.05 while that of others are less than 0.05. Therefore, the variables AT, V and RH are significant and AP is insignificant at five percent level of significance.

Therefore, step-wise regression modelling technique is used to find the best fit model for the data.

3.6 Step-Wise Regression

Table 5 : AIC Values of All the Models after Step-Wise Regression

Model Formula	AIC Value
EP ~ 1	2843.03
EP ~ AT	1647.63
EP ~ AT + RH	1535.65
EP ~ AT + RH + V	1494.92

Fourth model has the lowest AIC value. Therefore, the best fitted model is the fourth model in the table.

3.7 Best Fit Model of Step-Wise Regression

Formula:

$$EP = 514.80849 - 2.00554 \times AT - 0.19901 \times V - 0.14253 \times RH \qquad \text{Equation 5}$$

Table 6 : Important Statistics of Best Fitted Regression Model after Step-Wise Regression

Important Statistics of the Model	Value of the Statistic
p-value	< 2.2e-16
Multiple R-squared	0.9333

P-Value of the model is less than 0.05. That is obtained model is significant at five present level of significance. Multiple R-squared value implies that 93.33 percent of variation in net hourly electrical energy output can be explained by the model obtained.

Table 7 : Estimates of Coefficients of Best Fitted Regression Model after Step-Wise Regression

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	514.80849	1.56717	328.496	< 2e-16
AT	-2.00554	0.05722	-35.051	< 2e-16
RH	-0.14253	0.01677	-8.501	< 2e-16
V	-0.19901	0.02992	-6.652	7.67e-11

P-Values of all explanatory variables are less than 0.05. Therefore, the variables AT, V and RH are significant at five percent level of significance.

Hence the obtained model (EP=514.80849-2.00554×AT-0.19901×V-0.14253×RH) is a significant model.

3.8 Model Diagnostic Checking

3.8.1 Linearity

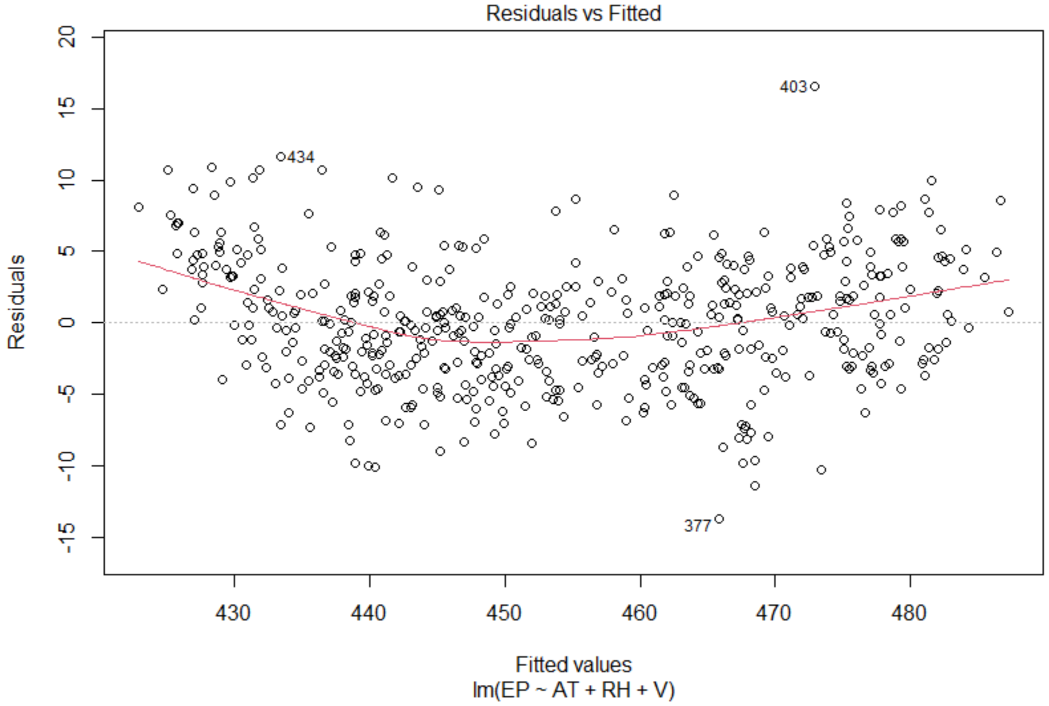

Figure 6 : Residuals Vs Fitted Values Plot of Best Fitted Model after Step-Wise Regression

Data are approximately scattered in a horizontal band without distinct patterns around the red line. Therefore, it can be assumed to have linear relationship between predictor variable and explanatory variables.

3.8.2 Normality

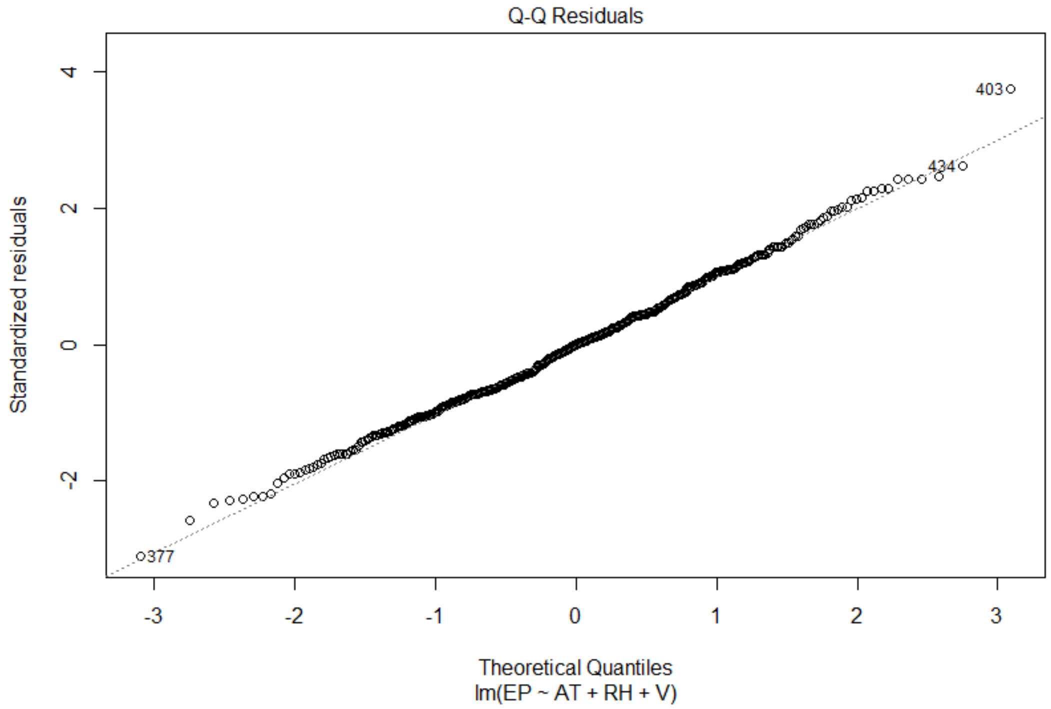

Figure 7 : Normal Q-Q Plot of Best Fitted Model after Step-Wise Regression

Residuals data are almost scattered in the straight line with few outliers. However, residuals can be assumed to be normally distributed.

3.8.3 Homoscedasticity

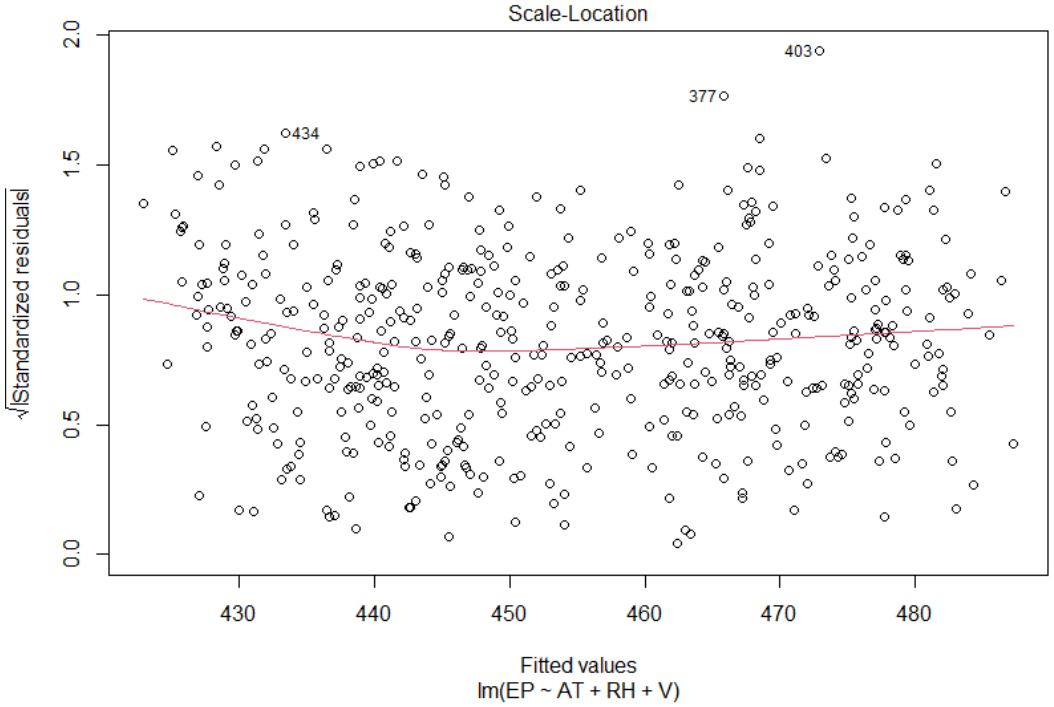

Figure 8 : Scale-Location Plot of Best Fitted Model after Step-Wise Regression

All data are spread evenly in the plot. This indicates that homoscedasticity exist in this model.

3.8.4 Independence of Residuals

Table 8 : Durbin-Watson Statistics of the Best Fitted Model after Step-Wise Regression

lag	Autocorrelation	D-W Statistic	p-value
1	-0.03175895	2.062147	0.524

P-Value of Durbin Watson test is greater than 0.05. That is residuals are independent at five percent level of significance.

3.8.5 Multicollinearity

Table 9 : VIF Values of Best Fitted Model after Step-Wise Regression

variable	VIF
AT	4.845477
RH	1.599955
V	3.766169

All the VIF values are less than 5 implies that there is no any multicollinearity in this model.

Since all the assumptions of multiple linear regression model are satisfied by the obtained model, the model (EP=514.80849-2.00554×AT-0.19901×V-0.14253×RH) is adequate for predictions.

3.9 Predictions

Table 10 : Predictions of Best Fitted Regression Model

Expected Ambient Temperature(°C)	Expected Relative Humidity(%)	Expected Exhaust Vacuum(cmHg)	Expected Net Hourly Electrical Energy Output/Prediction(MW)
18.00	71.00	53.00	458.0413
18.50	71.60	53.30	456.8933
19.30	72.00	53.70	455.1522
19.50	72.64	54.09	454.5823
19.60	72.64	54.09	454.3817
19.74	72.64	53.80	454.1587
19.74	72.64	53.90	454.1388
19.74	72.64	54.09	454.1009
19.74	72.64	54.20	454.0790
19.74	72.64	54.30	454.0591
19.74	72.80	54.09	454.0781
19.74	72.90	54.09	454.0639
19.74	72.40	54.09	454.1351
19.74	72.30	54.09	454.1494
19.80	72.80	54.50	453.8762
19.90	72.64	54.09	453.7801
19.90	72.90	54.80	453.6017
20.00	72.64	54.09	453.5795
20.00	80.00	55.00	452.3494

According to the predicted values of energy output, when decreasing AT, RH and V provide high EP values comparative to the mean level of collected data (454.1) whereas when increasing AT, RH and V provide low EP values comparative to mean level of collected data (454.1). Therefore, predicted net hourly electrical energy output values indicate that the lower the ambient temperature, relative humidity and exhaust vacuum are the higher the net hourly electrical energy output is.

4 CONCLUSION AND DISCUSSION

The statistical regression analysis was conducted to predict the net hourly electrical energy output based on Ambient Temperature (AT), Exhaust Vacuum (V), Ambient Pressure (AP), Relative Humidity (RH). The initial model which was included all the explanatory variables was not a good fit model as one explanatory variable was not significant. Hence step-wise regression was carried out to determine the best fit model for the data. The final regression model, obtained after performing stepwise regression, is given by the formula $EP=514.80849-2.00554\times AT-0.19901\times V-0.14253\times RH$ was a significant model with significant coefficients. All the linear regression assumptions were satisfied by the above mentioned model. The R-squared value showed an excellent fit, suggesting that the model explained approximately 93.33 percent of the variability in the net hourly electrical energy output. Specifically, as the Ambient Temperature (AT) increases by one unit, the power output decreases by approximately 2.00554 units. As Relative Humidity (RH) and Exhaust Vacuum (V) increase, the power output decreases by approximately 0.14253 and 0.19901 units, respectively. Furthermore, the predictions of the final model suggest that decreasing Ambient Temperature (AT), Relative Humidity (RH) and Exhaust Vacuum (V) increase the power output out of the current average power output observed by the observed data.

The final model obtained through stepwise regression is recommended for practical applications. It offers more predictive power comparing to the initial model with all explanatories while being simpler to understand, making it easier for decision-makers to interpret and utilize.

Since the fitted model assumes a linear relationship between the variables, which may not fully capture the complexity of the power generation process. Future research could explore nonlinear relationships also. Changing the confidence level might affect the final result of this analyse as it is taken as ninety-five throughout the analysis. The dataset may not account for all potential variables influencing power output. External factors which are not included in the dataset might affect the accuracy of the model obtained in this analysis.

Further research can explore more advanced modelling techniques, including machine learning methods, to assess if additional predictive power can be achieved. Using different confidence levels also suggested for better accurate results for enhancing quality of production and for increasing the profit by decreasing the cost of production. Investigating potential interactions and nonlinear effects between variables could enhance the models' accuracy. Consideration of time-series data could provide insights into temporal trends in power output and environmental conditions.

In practical terms, this regression model provides valuable insights for controlling and optimizing power generation in a Combined Cycle Power Plant. By adjusting the environmental factors represented by Ambient Temperature, Relative Humidity, and Exhaust Vacuum, operators can make informed decisions to enhance the plant's efficiency and reduce energy costs. However, further analysis of the model's performance and potential refinements may be necessary for comprehensive decision-making in a real-world operational setting.

5 BIBLIOGRAPHY

Corporate Finance Institute Team. (n.d.). *Coefficient of Determination*. Retrieved from Corporate Finance Institute.: https://corporatefinanceinstitute.com/resources/data-science/coefficient-of-determination/

kassambara. (2018, March 11). *Linear Regression Assumptions and Diagnostics in R: Essentials - articles - STHDA*. Retrieved from Sthda.com: http://www.sthda.com/english/articles/39-regression-model-diagnostics/161-linear-regression-assumptions-and-diagnostics-in-r-essentials/

Khan, M. (2021, April 07). *A Basic Guide to Testing the Assumptions of Linear Regression in R*. Retrieved from Godatadrive.com: https://godatadrive.com/blog/basic-guide-to-test-assumptions-of-linear-regression-in-r#:~:text=The%20easiest%20way%20to%20check

Taylor, S. (2022, November 24). *Regression Analysis*. Retrieved from Corporate Finance Institute: https://corporatefinanceinstitute.com/resources/data-science/regression-analysis/

Zach. (2022, October 24). *How to Test for Multicollinearity in R*. Retrieved from Statology: https://www.statology.org/multicollinearity-in-r/

6 APPENDIX

6.1 Dataset:

Table 11 : Dataset of the Analyze

AT	V	AP	RH	EP
8.34	40.77	1010.84	90.01	480.48
23.64	58.49	1011.4	74.2	445.75
29.74	56.9	1007.15	41.91	438.76
19.07	49.69	1007.22	76.79	453.09
11.8	40.66	1017.13	97.2	464.43
13.97	39.16	1016.05	84.6	470.96
22.1	71.29	1008.2	75.38	442.35
14.47	41.76	1021.98	78.41	464
31.25	69.51	1010.25	36.83	428.77
6.77	38.18	1017.8	81.13	484.31
28.28	68.67	1006.36	69.9	435.29
22.99	46.93	1014.15	49.42	451.41
29.3	70.04	1010.95	61.23	426.25
8.14	37.49	1009.04	80.33	480.66
16.92	44.6	1017.34	58.75	460.17
22.72	64.15	1021.14	60.34	453.13
18.14	43.56	1012.83	47.1	461.71
11.49	44.63	1020.44	86.04	471.08
9.94	40.46	1018.9	68.51	473.74
23.54	41.1	1002.05	38.05	448.56
14.9	52.05	1015.11	77.33	464.82
33.8	64.96	1004.88	49.37	427.28
25.37	68.31	1011.12	70.99	441.76
7.29	41.04	1024.06	89.19	474.71
13.55	40.71	1019.13	75.44	467.21
6.39	35.57	1025.53	77.23	487.69
26.64	62.44	1011.81	72.46	438.67
7.84	41.39	1018.21	91.92	485.66
21.82	58.66	1011.71	64.37	452.16
27.17	67.45	1015.67	49.03	429.87
13.42	41.23	994.17	95.79	468.82
20.77	56.85	1012.4	83.63	442.85
8.29	36.08	1020.38	81.53	483.26
30.98	67.45	1015.18	45.4	433.59
31.96	71.29	1008.39	47.51	433.04
15.83	52.75	1024.3	58.34	458.6
22.56	70.79	1005.85	93.09	435.14
25.91	75.6	1018.23	62.65	443.2
8.24	39.61	1017.99	78.42	477.9
24.66	60.29	1018	59.56	445.26
29.31	68.67	1006.18	63.38	435.57

21.48	66.91	1008.58	84.49	447.42
18.28	44.71	1016.99	33.71	462.28
26.96	65.34	1015.05	46.93	441.81
16.01	65.46	1014	87.68	454.16
27.37	63.73	1009.79	65.25	437.24
16.3	39.63	1004.64	85.61	464.11
23.8	48.6	1002.43	67.32	440.89
8.19	41.66	1016.57	75.38	485.2
25.28	67.69	1009.05	68.54	445.34
21.47	70.32	1011.72	88.36	440
30.54	67.45	1014.37	32.97	431.35
18.3	44.06	1017.95	63.24	456.32
25.82	72.39	1003.4	86.33	432.98
31.12	69.13	1009.61	56.54	429.41
15.99	39.63	1006.09	89.92	464.95
8.42	42.28	1008.4	87.78	481.91
23.7	69.23	1011.73	73.18	437.15
15.71	40.12	1012.81	97.46	462.6
29.11	74.9	1003.31	74.01	432.43
23.73	61.02	1009.71	78.7	442.22
28.26	65.34	1014.56	43.4	441.03
15.92	41.2	1016.04	73.37	468.91
33.4	70.8	1009.17	57.01	432.46
31.92	68.3	1015.58	41.55	430.07
26.87	74.99	1002.48	71.91	437.78
5.23	40.78	1025.13	92.74	483.12
15.72	41.76	1023.07	66.95	462.19
17.74	50.88	1015.56	89.78	457.71
27.13	59.54	1004.33	73.27	438.32
25.82	66.49	1013.2	60.32	434.35
34.2	68.14	1002.7	56.09	427.05
19.13	68.61	1010.65	94.92	448.69
11.77	41.06	1021.45	76.63	475.88
10.25	41.46	1018.67	84.41	479.28
23.82	48.92	1010.48	44.45	446.85
30.1	66.25	1009.38	65.41	434.99
29.92	69.75	1009.12	67.15	441.5
29.63	66.75	1018.08	58.46	433.26
30.61	77.17	1009.49	73.19	430.46
16.18	50.88	1014.06	100.09	455.14
31.66	73.91	1004.49	68.13	431.26
29.14	61.5	1009.64	39.95	437.76
18.4	43.96	1012.4	77.81	464.63
15.86	38.62	1016.65	67.51	462.58
6.31	42.07	1004.61	77.74	483.27
23.4	72.99	1009.48	88.82	438.51

33.62	77.17	1009.38	63.71	431.03
6.89	43.65	1019.87	72.77	484
27.41	69.98	1008.51	65.49	431.64
26.37	54.5	1014.48	66.31	451.78
26.58	59.39	1014.49	65.87	439.46
15.25	43.02	1012.04	41.43	467.23
21.24	41.67	1012.6	49.27	459.81
24.98	58.05	1011.69	69.97	447.15
26.63	64.44	1012.66	61.19	442
18.87	52.08	1005.25	99.19	449.61
5.12	40.78	1025.45	96.88	481.28
31.1	71.32	1008.09	45.87	437.54
7.91	39.96	1023.57	88.44	475.52
14.97	58.2	1019.52	81.13	458.68
28.92	61.02	1011.19	56.13	436.76
22.92	61.9	1013.27	78.32	446.08
12.55	38.91	1012.58	75.44	474.35
9.38	41.44	1015.09	86.8	481.96
10.28	41.46	1018.44	85.23	479.66
30.31	77.24	1008.02	67.84	435.28
25.34	71.06	1007.74	92.52	434.86
30.47	70.02	1009.65	62.03	435.72
20.18	56.53	1020.3	77.1	454.14
21.31	46.97	1014.08	62.57	455.76
22.1	74.87	1009.4	85.78	442.83
20.09	63.94	1020.82	74.16	449.47
22.36	58.79	1016.69	73.6	445.76
9.43	37.14	1013.03	74.99	473.57
28.86	49.3	1003.93	52.85	435.75
23.25	70.32	1009.28	86.52	437.96
22.7	60.84	1019.15	71.48	438.59
21.3	59.39	1013.83	84.53	444.21
19.94	44.9	1008.52	74.69	459.47
14.97	43.67	1011.13	87.33	465.54
13.83	40.78	1024.85	73.08	458.85
23.39	71.25	1005.22	93.7	435.68
21.48	47.43	1006.88	63.54	449.2
22.63	58.41	1013.69	90.32	445.43
20.01	45.09	1014.21	38.19	453.96
20.52	43.79	1015.9	68.1	463.93
13.58	49.83	1008.9	86.8	465.82
14.38	44.84	1024.59	81.68	471.6
28.53	69.84	1004.46	68.03	427.72
13.89	35.71	1016.4	65.49	468.54
25.93	69.34	1008.91	79.43	434.08
17.84	61.27	1020.1	70.68	454.57

6.28	41.06	1020.96	90.91	489.79
8.16	39.64	1011.21	85.56	481.78
18.61	67.71	1004.07	84.49	443.57
16.23	43.72	1009.96	83.01	457.9
8.66	41.01	1022.51	98.55	472.16
19.53	66.54	1013.72	81.46	451.19
10.17	40.46	1018.54	68.21	470.33
6.48	40.27	1010.55	82.12	486.68
12.69	40.75	1016.05	82.92	472.68
23.77	61.87	1009.8	54.83	445.72
14.43	43.5	1021.86	83.14	466.58
28.25	57.19	1007.99	67.9	436.99
12.26	39.96	1011.79	79.06	470.87
14.55	42.74	1027.29	66.64	457.77
24.57	51.95	1005.8	70.07	448.15
10.25	41.26	1007.44	98.08	476.03
21.96	51.43	1007.28	89.96	442.87
9.82	39.64	1010.79	69.22	477.93
18.58	44.6	1016.61	48.15	459.06
20.57	44.9	1008.39	62.69	456.18
8.61	41.04	1024.81	95.6	473.62
9.99	43.52	1022.95	99.67	473.64
13.97	52.05	1012.05	85.36	463.64
16.76	43.67	1011.27	70.55	462.44
25.58	47.01	1014.5	63.53	454.39
17.34	46.48	1007.18	83.15	452.06
19.13	66.86	1012.95	71.24	453.31
12.75	40.64	1020.44	83.17	466.93
29.46	49.3	1003.88	54.1	436.37
15.59	52.75	1024.3	59.93	458.87
14.2	43.02	1012.18	57.07	470.67
14.05	40.35	1011.52	66.21	475.54
32.35	77.95	1014.76	60.88	432.72
21.41	56.9	1007.03	79.41	441.41
10.61	37.5	1008.8	100.13	472.9
19.13	40.79	1004.47	91.7	457.83
17.28	44.06	1018.15	66.93	462.05
24.79	50.78	1008.78	61.88	445.3
28.66	50.05	1006.08	47.74	435.94
12.91	41.16	1011.9	89.19	472.71
17.32	43.7	1015.13	61.66	464.5
18.74	59.21	1018.3	91.55	449.77
26.69	70.36	1006.82	68.29	435.04
12.28	40.55	1005.72	98.56	472.47
14.46	44.78	1007.64	67.03	459.31
30.54	70.17	999.42	64.32	434.62

27.58	66.48	1004.82	51.23	435.34
32.17	73.21	1001.43	69.86	430.64
26.23	59.87	1012.93	66.62	447.16
11.69	35.76	1019.02	55.6	471.72
14.73	39.82	1012.67	80.82	470.41
12.36	48.04	1012.8	93.59	468.37
9.52	40.96	1023.48	86.49	483.59
27.32	73.42	1012.07	86.3	432.74
23.81	73.17	1012.13	86.13	438.11
7.92	42.28	1008.55	81.68	484.75
23.58	58.82	1009.68	81.38	445.01
20.78	60.84	1018.66	80.4	445.82
19.68	68.28	1002.62	71.57	452.47
14.12	40.92	1022.46	73.28	459.76
22.97	62.4	1010.25	75.18	445.3
6.89	39.37	1020.21	74.17	486.9
19.98	56.53	1020.22	73.53	454.89
21.71	60.27	1015.74	89.76	441.19
23.02	59.21	1011.74	83.18	440.5
27.71	69.51	1009.65	58.38	435.04
10.46	37.5	1013.12	76.74	472.16
27.18	62.66	1009.89	52.83	442.16
25.56	50.05	1005.67	69.29	442.49
19.69	39.72	1001.49	60.34	456.55
31.37	76.86	997.34	62.94	431.03
14.17	41.67	1013.4	72.77	471.42
15.6	43.71	1024.88	80.37	463.34
25.3	71.58	1010.18	87.36	433.97
7.09	39.37	1020.07	73.26	487
6.07	41.14	1027.57	86.98	480.19
9.86	38.38	1018.22	85.35	479.56
14.73	40.35	1011.15	65.2	470.03
31.32	74.33	1012.92	36.48	429.57
6.14	38.5	1012.4	75.35	489.26
27.92	69.23	1013.21	44.43	435.62
25.63	58.86	1014.86	52.75	443.84
16.37	36.99	1006.37	90.11	463.76
28.31	65.71	1014.67	59.29	447.26
24.44	69.84	1005.34	80.37	430.3
11.33	41.5	1013.58	88.7	475.01
23.87	63.94	1019.02	44.28	445.11
21.14	48.78	1020.39	86.2	450.35
22.24	60.07	1015.93	62.25	446.14
23.07	70.47	1008.64	83.37	439.04
25.95	59.14	1016.42	67.03	438.46
16.44	43.96	1013.66	76.31	468.57

20.66	41.54	1014.68	79.7	461.55
15.01	39.52	1017.53	72	468.02
20.51	63.86	1020.65	75.36	447.18
24.1	52.84	1006.47	47.58	444.69
9.38	40.46	1019.29	75.77	474.1
27.49	71.94	1008.25	68.29	428.27
32.84	68.14	1003.59	43.88	425.18
20.83	44.78	1008.51	35.9	460.6
28.69	67.25	1017.71	44.6	436.21
9.35	44.03	1011.02	88.11	476.25
17.37	41.23	998.79	68.44	461.08
14.84	41.48	1017.26	63.42	460.3
27.88	68.94	1007.68	75.68	435.04
17.83	43.72	1008.64	78.72	453.86
7.64	42.28	1008.46	84.29	484.78
25.67	74.78	1010.24	68.87	440.48
23.1	51.3	1011.93	80.05	452.17
27.13	69.45	1013.9	51.16	434.47
26.95	70.8	1009.73	82.95	436.8
19.4	42.44	1013.19	58.63	459.76
13.5	44.92	1023.87	87.17	468.85
11.97	41.04	1026.42	70.75	468.81
18.11	58.95	1016.61	89.17	454.29
24.1	71.29	1007.96	67.56	442.8
20.29	65.27	1013.17	74.94	445.52
21.82	65.27	1013.86	72.81	447.43
29.45	75.6	1018.12	50.68	437.31
10.87	38.38	1020.85	69.78	477.03
17.36	39.53	1008.28	68.87	456.57
24.97	58.95	1017.19	56.24	440.03
13.72	45.09	1012.95	87.41	466.24
26.49	61.86	1012.55	74.24	440.21
8.25	41.26	1020.59	91.84	475.17
12.8	44.9	1007.58	66.6	466.95
17.08	40.12	1012.17	81.5	457.81
31.4	71.06	1008.44	74.77	433.34
8.26	44.71	1020.85	68.82	485.19
21.22	59.15	1014.68	59.71	450.96
26.1	62.91	1012.44	60.91	446.05
19.91	51.19	1008	90.23	449.23
14.56	40.69	1015.48	73.73	469.31
10.22	41.74	1022.65	83.58	479.01
29.12	67.83	1008.05	46.86	432.46
16.97	39.16	1005.7	69.13	458.59
26.19	49.3	1003.69	65.88	440.13
23.65	66.05	1019.6	78.21	442.9

29.53	70.94	1007.22	70.68	432.43
19.69	56.65	1020.84	72.14	455.07
33.71	69.98	1013.09	42.62	431.99
32.14	75.08	1005.35	49.92	439.22
29.28	77.17	1009.4	76.3	433.14
7.01	37.8	1021.47	62.84	483.95
30.47	68.24	1009.55	64.98	427.94
29.13	47.93	1002.5	49.1	435.61
11.86	39.85	1013	66.92	478.29
26.16	52.36	1013.62	58.52	453.08
10.56	39.13	1010.19	87.12	463.14
8.58	40.72	1022.04	87.55	480.96
14.26	40.92	1022.07	73.96	460.38
23.48	73.67	1007.07	88.49	437.19
5.54	45.87	1008.69	95.91	478.02
16.98	43.99	1021.65	81.17	456.07
30.08	67.25	1017.6	53.09	434.01
10.63	39.61	1020	72.42	476.99
30.36	74.87	1008.96	61.87	435.31
29.56	70.4	1006.49	82.95	432.85
19.94	56.53	1020.48	76.43	453.8
27.55	64.96	999.73	57.34	431.33
31.44	79.74	1006.95	56.99	431.69
25.35	74.33	1015.29	77.84	435.66
4.11	38.44	1015.9	81.79	488.05
30.59	70.04	1010.28	50.56	429.13
14.04	48.79	1017.62	75.25	464.09
15.47	44.9	1021.59	81.74	465.61
8.67	40.77	1011.81	89.4	479.23
10.53	37.5	1008.55	99.91	472.32
7.99	41.38	1021.95	78.77	487.57
22.29	45.01	1012.4	51.58	449.08
9.06	36.3	1015.14	58.71	477.38
10.05	37.14	1012.64	75.97	474.16
20.07	49.21	1012.34	56.8	454.25
30.65	71.98	1004.92	63.39	429.85
14.38	40.73	1017.48	90.64	462.98
27.76	71.97	1008.53	86.97	434.04
16.7	40.56	1019.04	50.9	470.84
5.35	35.57	1027.12	80.81	488.65
11.57	41.38	1021.71	61.49	473.94
30.57	73.03	1013.75	73.68	437.42
16	40.66	1016.12	89.23	457.12
15.6	41.04	1025.48	64.43	457.49
22.12	47.93	1005.5	84.47	443.41
21.33	58.46	1016.17	79.73	445.27

23.65	59.92	1010.19	89.53	436.74
6.8	39.37	1020.24	73.29	487.33
31.35	71.32	1005.99	62.35	434.43
23.99	68.12	1012.92	70.68	437.14
26.53	54.2	1012.9	66.73	439.96
31.59	62.7	1009.3	50.16	442.58
24.06	69.94	1005.47	60.46	436.88
27.28	44.89	1009.02	49.9	441.94
23.69	71.97	1009.62	93.03	438.62
11.46	39.1	1010.17	100.15	470.55
22.61	49.3	1003.51	83.02	441.76
21.84	45.09	1013.94	46.11	450.88
9.2	39.82	1013.19	91.25	482.89
17.75	49.25	1020.86	63.67	454.41
7.53	43.65	1018.58	64.85	479.15
16.76	45.09	1014.28	57.71	458.67
22.69	65.12	1016.35	72.02	446.91
11.91	44.45	1021.39	84.49	466.52
26.63	61.47	1008.18	69	444.16
6.72	39.85	1011.84	84.66	489.09
9.44	43.99	1020.31	98.06	475.01
18.32	45	1022.67	46.38	471.43
18.31	62.1	1020.38	79.02	455.24
26.57	44.57	1007.76	50.39	445.49
34.33	69.05	1000.89	37.9	430.56
14.85	40.8	1025.38	53.29	471.78
29.33	64.84	1010.76	53.32	443.15
8.77	40.46	1019.68	77.84	475
11.35	38.91	1017.93	83.45	474.22
21.38	44.05	1005.69	81.66	445.71
4.88	40.27	1012.66	74.51	491.29
26.66	64.27	1012.83	58.88	437.87
13.04	45.09	1013	87.68	467.44
29.47	71.32	1008.07	67	437.14
18.91	56.9	1006.98	84.07	448.24
24.86	44.05	1005.69	66.65	447.2
9.42	41.4	1029.6	87.43	478.12
12.75	42.34	1018.17	94.67	465.44
21.32	49.02	1008.81	85.81	445.65
24.26	67.69	1008.48	66.74	437.4
20.01	43.77	1012.13	65.53	455.66
21.35	59.04	1012.28	95.71	445.83
17.43	63.09	1019.87	92.77	449.36
27.86	65.18	1012.87	55.53	436.26
20.74	59.8	1015.46	76.39	452.99
27.73	67.69	1007.47	59.42	431.72

9.53	41.44	1018.01	80.09	481.83
28.5	79.74	1006.75	87.09	433.08
21.39	52.3	1009.2	79.77	446.87
14.54	46.18	1015.29	94.48	463.03
28.01	65.34	1014.68	46.86	440.77
30.51	70.94	1007.64	59.42	430.91
21.63	60.84	1018.16	77.53	445.93
26.39	71.25	999.8	89.12	430.34
13.84	41.26	1020.92	74.74	462.46
17.27	43.52	1021.37	76.73	460.08
21.89	58.05	1011.31	85.71	441.76
24.02	64.63	1020.49	55.17	449.66
31.92	75.33	1001.81	61.97	436.38
12.97	49.83	1008.69	91.49	452.07
15.89	43.96	1014.02	75.24	467.35
9.25	41.26	1020.51	91.21	473.87
15.99	43.5	1021.34	73.22	460.7
9.06	39.66	1019.64	76.98	479.53
14.48	44.06	1018.73	75.88	467.46
27.15	59.21	1013.49	51.71	440.27
21.68	64.15	1021.09	70.88	454.32
28.16	61.47	1009.19	62.53	442.48
34.33	74.34	1001.74	42.75	435.76
31.98	64.05	1010.49	58.01	439.56
15.21	43.56	1012.92	56.61	467
13.56	42.34	1017.95	93.83	462.6
10.32	42.02	998.27	99.12	472.16
22.46	58.49	1011.5	70.54	448.46
29.3	68.08	1011.23	56.23	434.11
22.3	44.57	1008.48	67.58	449.74
21.91	70.32	1009.17	90.57	439.29
10.95	43.8	1022.61	63.46	473.2
24.96	68.12	1011.27	70.22	437.62
5.91	39.33	1010.18	95.53	491.49
5.25	40.07	1019.48	67.7	495.23
21.47	60.93	1006.46	87.83	442.75
10.16	39.3	1019.71	81.21	480.74
17.29	44.06	1016.24	77.56	461.38
17.46	40.22	1006.7	91.96	461.74
11.56	40.43	1025.48	74.75	489.54
7.51	41.01	1024.61	97.41	477.61
17.08	38.58	1015.41	73.42	461.49
9.27	42.02	1004.29	95.05	474.93
21.3	48.92	1010.92	65.09	447.89
8.51	40.81	1015.54	83.16	481.02
12.77	41.5	1014.13	86.8	470.68

25.04	65.48	1017.66	57.08	440.92
21.83	41.1	1001.91	41.79	453.38
15.92	41.79	1006.24	79.42	458.68
29.97	69.14	1007.92	66.74	433.82
9.8	41.82	1032.98	67.55	473.72
9.01	38.56	1016.67	63.61	482.37
24.71	61.9	1013.78	85.82	443.39
20.57	63.78	1016.76	74	449.93
24.32	65.48	1017.55	59.02	443.31
21.41	50.32	1008.61	84.47	443.63
27.09	59.15	1013.02	55.18	438.9
23.49	47.45	1008.46	66.18	446.72
20	63.31	1017.06	63.1	449.68
13.4	41.58	1020.5	71.17	461.54
23.6	60.08	1017.51	63.02	451.97
21.03	70.02	1010.21	95.69	444.53
5.15	45.87	1008.02	98.36	479.61
17.16	58.86	1016.4	86.39	453.76
24.59	71.94	1006.89	88.52	430.28
23.05	65.94	1010.99	69.73	442.39
10.48	37.5	1009.81	95.26	474.57
15.72	43.02	1011.98	45.74	466.33
25.56	70.32	1009.07	90.63	433.72
13.78	41.16	1021.65	75.77	460.44
29.35	69.04	1008.55	61.78	445.04
32.09	79.74	1006.92	61.5	432.85
25.16	77.3	1000.73	86.4	439.37
14.63	40	1018.03	63.4	457.09
5.7	40.35	1012.18	91.57	488.8
28.15	74.9	1003.67	76.32	433.63
22.3	57.19	1006.4	75.28	445.1
27.71	72.99	1007.18	72.68	433.01
23.09	45.01	1012.25	43.95	453.45
19.66	42.18	1001.51	95.75	454.46
10.02	52.75	1021.93	73.68	473.09
23.97	68.67	1006.63	76.83	441.53
7.82	38.91	1015.58	90.11	479.15
30.21	69.4	1003.69	62.89	438.18
24.12	62.44	1013.83	88.02	443.3
21.05	63.09	1016.49	91.23	447.45
13.93	40.24	1017.88	88.35	469.25
9.68	41.06	1022.75	87.44	476.67
17.48	49.39	1021.51	84.53	460.01
28.76	66.56	1006.07	62.41	432.3
19.21	53.16	1013.18	81.64	454.01
25.63	72.99	1008.2	86.8	436.64

20.36	66.05	1015.86	87.19	450.23
15.34	40.73	1018.03	86.19	458.55
9.55	40.72	1023.12	68.06	481.08
13.45	41.16	1022.04	82.96	460.57
10.8	39.58	1011.99	87.54	478.24
28.15	72.99	1007.62	70.03	431.83
15.42	40.56	1021.23	59.37	469.33
33.6	74.33	1012.31	35.66	428.62
30.92	78.05	1011	59.62	434.12
25.24	57.17	1010.44	74	442.77
13.6	41.2	1015.92	85.05	467.42
18.68	46	1002.2	97.17	447.8
22.43	69.84	1006.94	75.54	436.19
23.37	70.94	1008.01	81.28	441.56
19.62	43.56	1012.89	37.6	463.33
14.35	39.39	1012.92	82.41	470.53
24.88	63.91	1009.41	77.75	447.28
20.67	63.56	1013.27	79.39	450.7
34.11	73.56	1006.77	45.92	432.78
19.15	44.89	1009.6	74.52	451.09
24.96	70.47	1007.73	76.48	438.23
18.86	42.18	1001.16	98.58	457.07
25.1	60.37	1005.83	72.53	438.44
10.15	43.41	1018.4	82.07	473.43
26.53	65.18	1012.77	64.37	437.39
26.05	46.21	1011.31	55.25	450.86
25.45	74.33	1015.53	80.95	435.12
8.14	39.96	1023.41	88.86	477.05
31.95	67.17	1007.11	61.58	432.6
30.8	73.56	1007.39	75.78	432.43
7.69	43.02	1014.51	85.23	486.46
15.3	43.67	1011.75	95.31	468.14
27.12	70.32	1009.94	52.67	429.08
25.76	74.87	1010.18	65.15	443.7
10.28	39.64	1010.45	74.15	477.65
30.91	79.74	1006.92	56.32	433.84
24.22	61.86	1013.39	80.11	442.65
5.67	45.87	1008.91	93.29	478.44
22.96	68.61	1011.15	76.62	447.18
24.54	65.46	1014.27	40.77	446.26
11.32	36.18	1015.03	64.68	472.67
15.15	53.82	1016.34	62.59	461.6
23.53	50.05	1005.63	78.4	443.71
29.08	66.54	1004.67	64.5	430.16
20.42	60.77	1017.65	87.07	447.85

6.2 R Syntaxes and Outputs:

```
> ccpp=read.csv(file.choose())
> attach(ccpp)
> str(ccpp)
'data.frame':   500 obs. of  5 variables:
 $ AT: num  8.34 23.64 29.74 19.07 11.8 ...
 $ V : num  40.8 58.5 56.9 49.7 40.7 ...
 $ AP: num  1011 1011 1007 1007 1017 ...
 $ RH: num  90 74.2 41.9 76.8 97.2 ...
 $ EP: num  480 446 439 453 464 ...
> summary(ccpp)
       AT              V               AP               RH
 Min.   : 4.11   Min.   :35.57   Min.   : 994.2   Min.   : 32.97
 1st Qu.:13.77   1st Qu.:41.45   1st Qu.:1008.5   1st Qu.: 62.68
 Median :20.57   Median :50.83   Median :1012.5   Median : 74.00
 Mean   :19.74   Mean   :54.09   Mean   :1012.9   Mean   : 72.64
 3rd Qu.:25.91   3rd Qu.:66.55   3rd Qu.:1017.6   3rd Qu.: 84.55
 Max.   :34.33   Max.   :79.74   Max.   :1033.0   Max.   :100.15
       EP
 Min.   :425.2
 1st Qu.:438.9
 Median :451.3
 Mean   :454.1
 3rd Qu.:468.5
 Max.   :495.2
> var(AT)
[1] 58.50337
> var(V)
[1] 166.3153
> var(AP)
[1] 37.97586
> var(RH)
[1] 224.9981
> var(EP)
[1] 294.1408
> sd(AT)
[1] 7.64875
> sd(V)
[1] 12.89633
> sd(AP)
[1] 6.162456
> sd(RH)
[1] 14.99994
> sd(EP)
[1] 17.15053
> range(AT)
[1]  4.11 34.33
> range(V)
[1] 35.57 79.74
> range(AP)
[1]  994.17 1032.98
> range(RH)
```

```
[1]  32.97 100.15
> range(EP)
[1] 425.18 495.23
> IQR(AT)
[1] 12.15
> IQR(V)
[1] 25.09
> IQR(AP)
[1] 9.0825
> IQR(RH)
[1] 21.8675
> IQR(EP)
[1] 29.6825
> plot(ccpp)
> cor(ccpp)
           AT          V          AP          RH         EP
AT  1.0000000  0.8325873 -0.51192125 -0.52708682 -0.9533148
V   0.8325873  1.0000000 -0.39764411 -0.26619260 -0.8611522
AP -0.5119212 -0.3976441  1.00000000  0.08551756  0.5110355
RH -0.5270868 -0.2661926  0.08551756  1.00000000  0.3866161
EP -0.9533148 -0.8611522  0.51103553  0.38661609  1.0000000
> m=lm(EP~.,data=ccpp)
> summary(m)

Call:
lm(formula = EP ~ ., data = ccpp)

Residuals:
    Min      1Q  Median      3Q     Max
-13.660  -3.126  -0.034   2.946  16.461

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 496.38392   40.62265  12.219  < 2e-16 ***
AT           -1.99244    0.06412 -31.072  < 2e-16 ***
V            -0.20135    0.03038  -6.627 8.96e-11 ***
AP            0.01789    0.03942   0.454     0.65
RH           -0.14018    0.01756  -7.981 1.02e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.445 on 495 degrees of freedom
Multiple R-squared:  0.9334,    Adjusted R-squared:  0.9328
F-statistic:  1734 on 4 and 495 DF,  p-value: < 2.2e-16

> null=lm(EP~1,data=ccpp)
> mStep=step(null,scope=list(upper=m),direction="both")
Start:  AIC=2843.03
EP ~ 1

       Df Sum of Sq    RSS    AIC
+ AT    1    133392  13385 1647.6
+ V     1    108847  37929 2168.4
```

```
+ AP    1     38332 108445 2693.7
+ RH    1     21939 124837 2764.1
<none>              146776 2843.0

Step:  AIC=1647.63
EP ~ AT

       Df Sum of Sq    RSS    AIC
+ RH    1      2728  10656 1535.7
+ V     1      2176  11209 1560.9
+ AP    1       105  13279 1645.7
<none>               13385 1647.6
- AT    1    133392 146776 2843.0

Step:  AIC=1535.65
EP ~ AT + RH

       Df Sum of Sq    RSS    AIC
+ V     1       873   9784 1494.9
<none>               10656 1535.7
+ AP    1         9  10647 1537.2
- RH    1      2728  13385 1647.6
- AT    1    114181 124837 2764.1

Step:  AIC=1494.92
EP ~ AT + RH + V

       Df Sum of Sq   RSS    AIC
<none>               9784 1494.9
+ AP    1       4.1  9779 1496.7
- V     1     872.8 10656 1535.7
- RH    1    1425.6 11209 1560.9
- AT    1   24233.1 34017 2116.0
> summary(mStep)

Call:
lm(formula = EP ~ AT + RH + V, data = ccpp)

Residuals:
     Min       1Q   Median       3Q      Max
-13.7695  -3.0906  -0.0185   2.9138  16.6160

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 514.80849    1.56717 328.496  < 2e-16 ***
AT           -2.00554    0.05722 -35.051  < 2e-16 ***
RH           -0.14253    0.01677  -8.501  < 2e-16 ***
V            -0.19901    0.02992  -6.652 7.67e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.441 on 496 degrees of freedom
Multiple R-squared:  0.9333,    Adjusted R-squared:  0.9329
```

```
F-statistic:  2315 on 3 and 496 DF,  p-value: < 2.2e-16

> plot(mStep,1)
> plot(mStep,2)
> plot(mStep,3)
> library(car)
Loading required package: carData
> durbinWatsonTest(mStep)
 lag Autocorrelation D-W Statistic p-value
   1     -0.03175895      2.062147   0.524
 Alternative hypothesis: rho != 0
> vif(mStep)
      AT       RH        V 
4.845477 1.599955 3.766169 

# Predictions

> 
AT=c(18,18.5,19.3,19.5,19.6,19.74,19.74,19.74,19.74,19.74,19.74,19.7
4,19.74,19.74,19.8,19.9,19.9,20,20)
> 
RH=c(71,71.6,72,72.64,72.64,72.64,72.64,72.64,72.64,72.64,72.8,72.9,
72.4,72.3,72.8,72.64,72.9,72.64,80)
> 
V=c(53,53.3,53.7,54.09,54.09,53.8,53.9,54.09,54.2,54.3,54.09,54.09,5
4.09,54.09,54.5,54.09,54.8,54.09,55)
> expectedChanges=data.frame(AT,RH,V)
> expectedChanges$Predictions=predict(mStep,expectedChanges)
> expectedChanges
      AT    RH     V Predictions
1  18.00 71.00 53.00    458.0413
2  18.50 71.60 53.30    456.8933
3  19.30 72.00 53.70    455.1522
4  19.50 72.64 54.09    454.5823
5  19.60 72.64 54.09    454.3817
6  19.74 72.64 53.80    454.1587
7  19.74 72.64 53.90    454.1388
8  19.74 72.64 54.09    454.1009
9  19.74 72.64 54.20    454.0790
10 19.74 72.64 54.30    454.0591
11 19.74 72.80 54.09    454.0781
12 19.74 72.90 54.09    454.0639
13 19.74 72.40 54.09    454.1351
14 19.74 72.30 54.09    454.1494
15 19.80 72.80 54.50    453.8762
16 19.90 72.64 54.09    453.7801
17 19.90 72.90 54.80    453.6017
18 20.00 72.64 54.09    453.5795
19 20.00 80.00 55.00    452.3494
```