**Input:**

# Load necessary libraries

library(readr) # For reading data

library(dplyr) # For data manipulation

library(ggplot2) # For visualization

data <- read\_csv("C:/Users/rucha/OneDrive/Desktop/Ruchi/eclipse data.csv")

# Exploratory Data Analysis (EDA)

summary(data) # Summary statistics

str(data) # Structure of the dataset

head(data) # First few rows of the dataset

# Check for non-numeric variables

non\_numeric\_vars <- sapply(data, function(x) !is.numeric(x))

non\_numeric\_vars <- names(non\_numeric\_vars[non\_numeric\_vars])

# Remove non-numeric variables or convert them to appropriate types

data <- data %>% select(-one\_of(non\_numeric\_vars))

# Visualize relationships between variables

pairs(data)

model\_formula <- magnitude ~ Luna\_Num + Saros\_Num + Gamma

# Fit the multiple linear regression model

model <- lm(model\_formula, data)

# Summary of the regression model

summary(model)

# Diagnostic plots for checking model assumptions

# Residuals vs Fitted values plot

plot(model, which = 1)

# Normal Q-Q plot

plot(model, which = 2)

# Scale-Location plot

plot(model, which = 3)

# Residuals vs Leverage plot

plot(model, which = 5)

# Shapiro-Wilk test for normality of residuals

shapiro.test(residuals(model))

**Output:**

> # Load necessary libraries

> library(readr) # For reading data

> library(dplyr) # For data manipulation

> library(ggplot2) # For visualization

>

>

> # Load your solar eclipse dataset

> # Replace 'your\_dataset.csv' with the actual filename/path

> data <- read\_csv("C:/Users/rucha/OneDrive/Desktop/Ruchi/eclipse data.csv")

**Rows:** 251 **Columns:** 17

── **Column specification** ────────────────────────────────────────────────────────────────────────

**Delimiter:** ","

chr (6): Calendar Month, Ecl. Type, QLE, Lat °, Long °, Path Width km

dbl (10): Catalog Number, Calendar Year, Calendar Day, ΔT s, Luna\_Num, Saros\_Num, Gamma, ma...

time (1): Delta

ℹ Use `spec()` to retrieve the full column specification for this data.

ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

>

> # Exploratory Data Analysis (EDA)

> summary(data) # Summary statistics

Catalog Number Calendar Year Calendar Month Calendar Day Delta

Min. :4524 Min. :-99.0 Length:251 Min. : 1.00 Length:251

1st Qu.:4586 1st Qu.:-74.0 Class :character 1st Qu.: 9.00 Class1:hms

Median :4649 Median :-48.0 Mode :character Median :17.00 Class2:difftime

Mean :4649 Mean :-48.8 Mean :16.31 Mode :numeric

3rd Qu.:4712 3rd Qu.:-24.0 3rd Qu.:23.00

Max. :4774 Max. : 0.0 Max. :31.00

ΔT s Luna\_Num Saros\_Num Ecl. Type QLE

Min. :10525 Min. :-25957 Min. :46.00 Length:251 Length:251

1st Qu.:10772 1st Qu.:-25643 1st Qu.:58.00 Class :character Class :character

Median :11026 Median :-25329 Median :69.00 Mode :character Mode :character

Mean :11034 Mean :-25334 Mean :69.22

3rd Qu.:11294 3rd Qu.:-25027 3rd Qu.:80.00

Max. :11568 Max. :-24725 Max. :94.00

Gamma magnitude Lat ° Long ° Sun Alt °

Min. :-1.53140 Min. :0.0121 Length:251 Length:251 Min. : 0.00

1st Qu.:-0.80255 1st Qu.:0.4285 Class :character Class :character 1st Qu.: 0.00

Median :-0.02520 Median :0.9456 Mode :character Mode :character Median :39.00

Mean :-0.05111 Mean :0.7577 Mean :35.99

3rd Qu.: 0.74445 3rd Qu.:1.0146 3rd Qu.:57.00

Max. : 1.55130 Max. :1.0753 Max. :90.00

Path Width km Central Dur.

Length:251 Min. :0.000

Class :character 1st Qu.:0.000

Mode :character Median :1.190

Mean :2.267

3rd Qu.:4.500

Max. :8.510

> str(data) # Structure of the dataset

spc\_tbl\_ [251 × 17] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)

$ Catalog Number: num [1:251] 4524 4525 4526 4527 4528 ...

$ Calendar Year : num [1:251] -99 -99 -98 -98 -98 -98 -97 -97 -96 -96 ...

$ Calendar Month: chr [1:251] "May" "Oct" "Mar" "Apr" ...

$ Calendar Day : num [1:251] 7 31 28 26 20 20 17 10 6 29 ...

$ Delta : 'hms' num [1:251] 00:22:02 08:01:17 09:55:44 17:26:44 ...

..- attr(\*, "units")= chr "secs"

$ ΔT s : num [1:251] 11568 11563 11558 11558 11553 ...

$ Luna\_Num : num [1:251] -25957 -25951 -25946 -25945 -25940 ...

$ Saros\_Num : num [1:251] 75 80 47 85 52 90 57 62 67 72 ...

$ Ecl. Type : chr [1:251] "T" "A" "P" "P" ...

$ QLE : chr [1:251] "p-" "p-" "t-" "t-" ...

$ Gamma : num [1:251] 0.571 -0.835 -1.249 1.287 1.433 ...

$ magnitude : num [1:251] 1.063 0.913 0.541 0.467 0.217 ...

$ Lat ° : chr [1:251] "47N" "60S" "61S" "62N" ...

$ Long ° : chr [1:251] "158W" "59E" "168E" "149W" ...

$ Sun Alt ° : num [1:251] 55 33 0 0 0 0 56 49 79 86 ...

$ Path Width km : chr [1:251] "251" "594" "0" "0" ...

$ Central Dur. : num [1:251] 4.28 7.11 0 0 0 0 0 1.34 5.38 5.18 ...

- attr(\*, "spec")=

.. cols(

.. `Catalog Number` = col\_double(),

.. `Calendar Year` = col\_double(),

.. `Calendar Month` = col\_character(),

.. `Calendar Day` = col\_double(),

.. Delta = col\_time(format = ""),

.. `ΔT s` = col\_double(),

.. Luna\_Num = col\_double(),

.. Saros\_Num = col\_double(),

.. `Ecl. Type` = col\_character(),

.. QLE = col\_character(),

.. Gamma = col\_double(),

.. magnitude = col\_double(),

.. `Lat °` = col\_character(),

.. `Long °` = col\_character(),

.. `Sun Alt °` = col\_double(),

.. `Path Width km` = col\_character(),

.. `Central Dur.` = col\_double()

.. )

- attr(\*, "problems")=<externalptr>

> head(data) # First few rows of the dataset

# A tibble: 6 × 17

`Catalog Number` `Calendar Year` `Calendar Month` `Calendar Day` Delta `ΔT s` Luna\_Num

*<dbl>* *<dbl>* *<chr>* *<dbl>* *<time>* *<dbl>* *<dbl>*

1 4524 -99 May 7 00:22:02 11568 -25957

2 4525 -99 Oct 31 08:01:17 11563 -25951

3 4526 -98 Mar 28 09:55:44 11558 -25946

4 4527 -98 Apr 26 17:26:44 11558 -25945

5 4528 -98 Sep 20 15:21:31 11553 -25940

6 4529 -98 Oct 20 07:32:27 11552 -25939

# ℹ 10 more variables: Saros\_Num <dbl>, `Ecl. Type` <chr>, QLE <chr>, Gamma <dbl>,

# magnitude <dbl>, `Lat °` <chr>, `Long °` <chr>, `Sun Alt °` <dbl>, `Path Width km` <chr>,

# `Central Dur.` <dbl>

> # Check for non-numeric variables

> non\_numeric\_vars <- sapply(data, function(x) !is.numeric(x))

> non\_numeric\_vars <- names(non\_numeric\_vars[non\_numeric\_vars])

>

> # Remove non-numeric variables or convert them to appropriate types

> data <- data %>% select(-one\_of(non\_numeric\_vars))

>

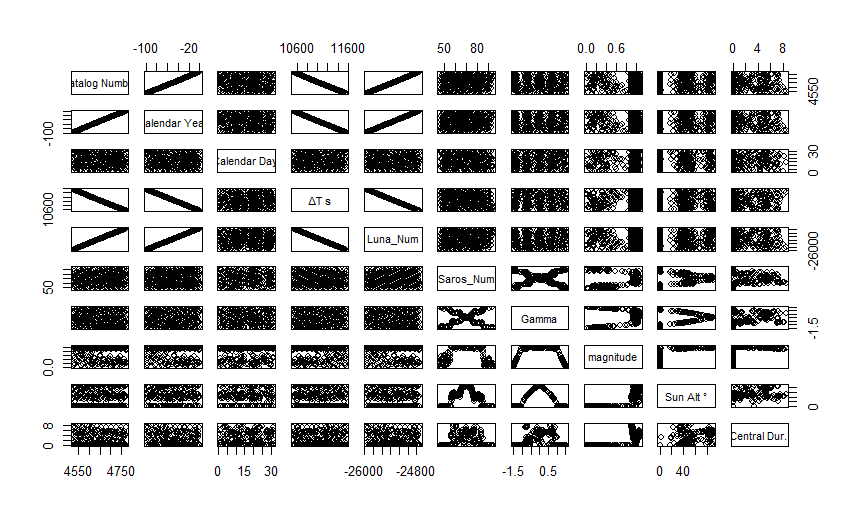
>

> # Visualize relationships between variables

> # For example:

> # Scatterplot matrix

> pairs(data)



> model\_formula <- magnitude ~ Luna\_Num + Saros\_Num + Gamma

> # Fit the multiple linear regression model

> model <- lm(model\_formula, data)

>

> # Summary of the regression model

> summary(model)

Call:

lm(formula = model\_formula, data = data)

Residuals:

Min 1Q Median 3Q Max

-0.8421 -0.3222 0.1827 0.2626 0.3496

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -4.765e-01 1.545e+00 -0.308 0.7580

Luna\_Num -4.728e-05 6.050e-05 -0.781 0.4353

Saros\_Num 5.591e-04 1.625e-03 0.344 0.7310

Gamma 4.395e-02 2.274e-02 1.933 0.0544 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.341 on 247 degrees of freedom

Multiple R-squared: 0.01756, Adjusted R-squared: 0.00563

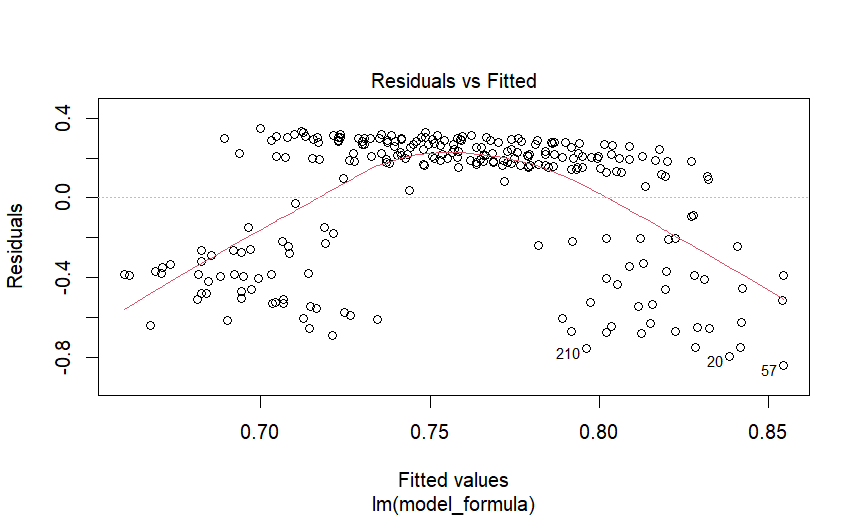
F-statistic: 1.472 on 3 and 247 DF, p-value: 0.2227

>

> # Diagnostic plots for checking model assumptions

> # Residuals vs Fitted values plot

> plot(model, which = 1)



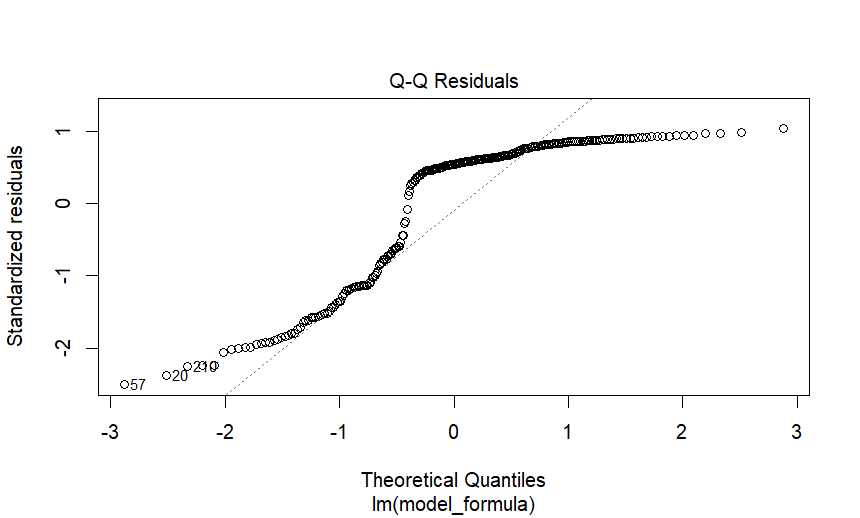
> # Normal Q-Q plot

> plot(model, which = 2)

>

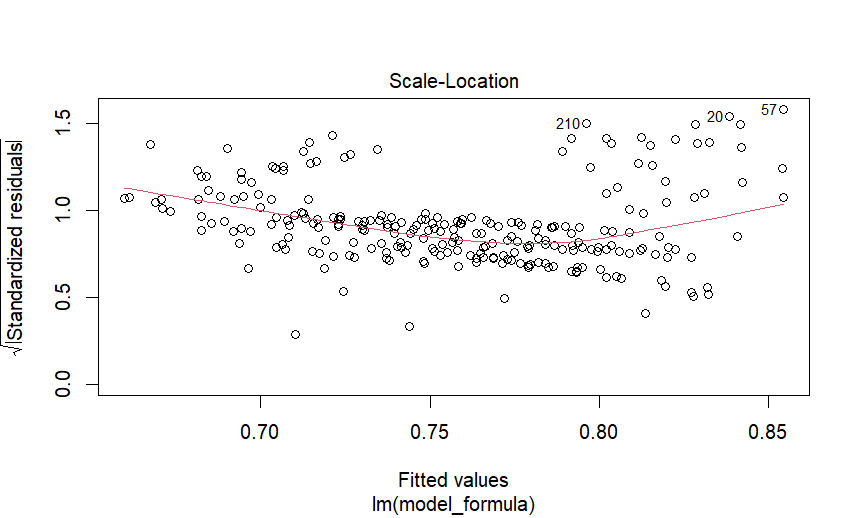
Code file:





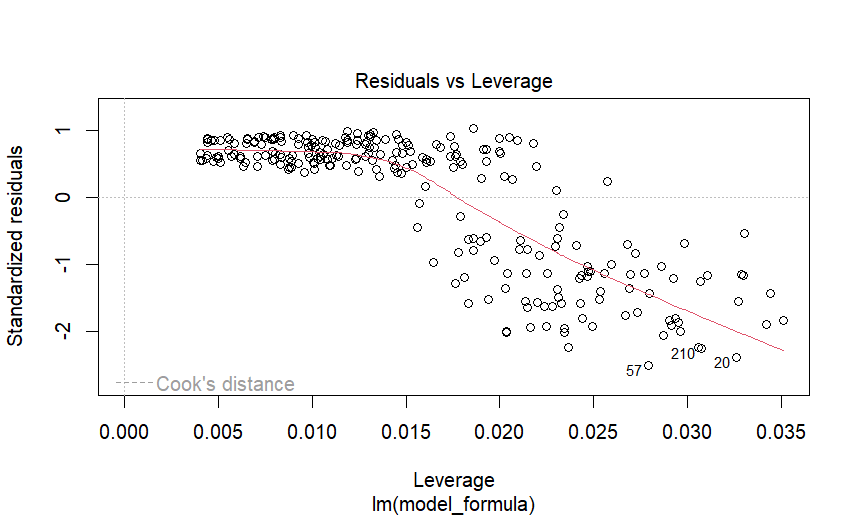
> # Scale-Location plot

> plot(model, which = 3)

> 

> # Residuals vs Leverage plot

> plot(model, which = 5)

> 

>

> # Shapiro-Wilk test for normality of residuals

> shapiro.test(residuals(model))

Shapiro-Wilk normality test

data: residuals(model)

W = 0.80263, p-value < 2.2e-16