॥वसुधैव कुटुम्बकम्॥

SYMBIOSIS INSTITUTE OF TECHNOLOGY, PUNE

Symbiosis International (Deemed University)

(Established under section 3 of the UGC Act, 1956)

Re-accredited by NAAC with 'A' grade (3.58/4) | Awarded Category – I by UGC

Founder: Prof. Dr. S. B. Mujumdar, M. Sc., Ph. D. (Awarded Padma Bhushan and Padma Shri by President of India)

Assignment No. 06

Subject:

Name of Student	**Antriksh Sharma**
PRN No.	**20070122021**
Branch	CS
Class	**A**
Academic Year & Semester	2023-24 _ 7th semester
Date of Performance	**28th August**
Title of Lab Assignment	**REGRESSION MODEL**

Theory: *Import a data from web storage. Name the dataset and now do Logistic Regression to find out the relationship between variables that are affecting the admission of a student to an institute based on his or her GRE score, GPA obtained, and rank of the student. Also check the model is fit or not.*
Use different datasets from an online repository to develop a logistic regression model.
Also, check if the model fits or not. Require (foreign), require (MASS).
The logistic regression model predicts the probability of a binary outcome (e.g., admission) based on one or more predictor variables (e.g., GRE score, GPA, rank).
In the provided dataset, the column names are in lowercase, so the formula is adjusted to admit ~ gre + gpa + rank. The glm function with family = "binomial" is used to fit the logistic regression model in R.

Answer:

```
# Load necessaíy libíaíies
íequiíe(foíeign)
íequiíe(MASS)

# Impoít the dataset
data <- íead.csv("https://figshaíe.com/ndownloadeí/files/34757857")
```

```
# Check foí missing values
píint(sum(is.na(data)))

# Handle missing values if any (you can use mean imputation oí otheí methods)
data[is.na(data)] <- mean(data, na.ím = TRUE)

# Display covaíiance and coíielation
píint(cov(data))
píint(coí(data))

# Check the names of the columns in the dataset
píint(names(data))

# Peífoím logistic íegíession using the MASS function
logit_model <- glm(admit ~ gíe + gpa + íank, data = data, family = "binomial")

# Display the summaíy of the model
summaíy(logit_model)

# Check the goodness of fit
anova(logit_model, test="Chisq")

# Plot the gíaph foí the model
plot(logit_model)
```

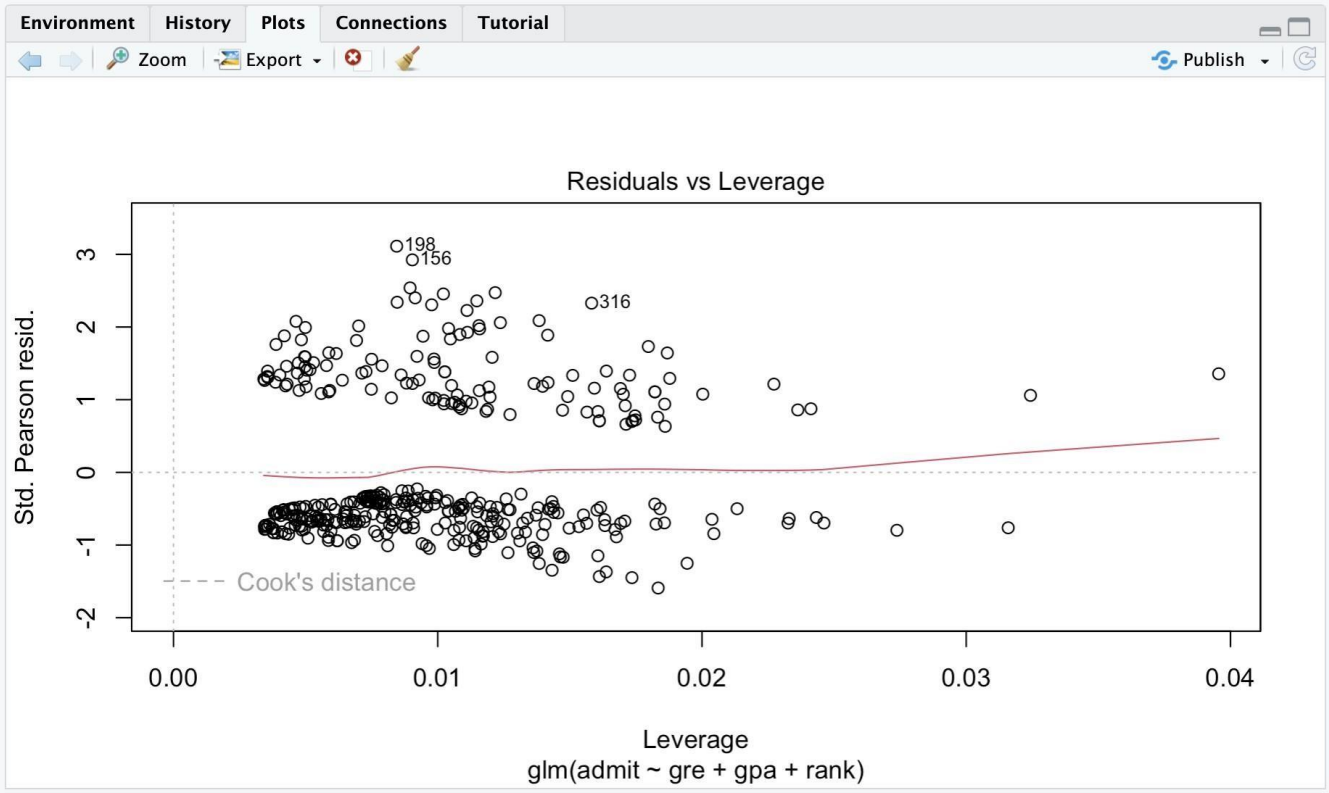
Output:

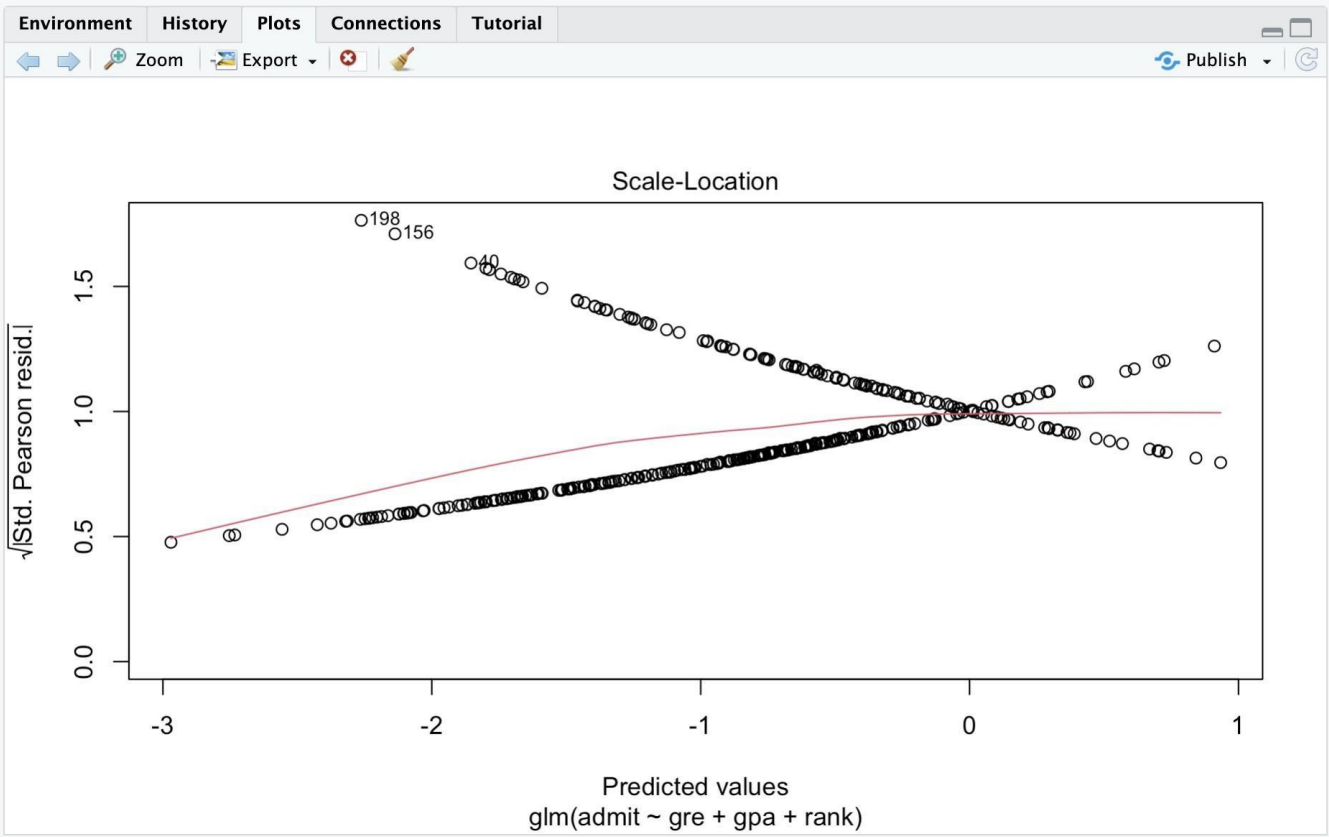
Answer:

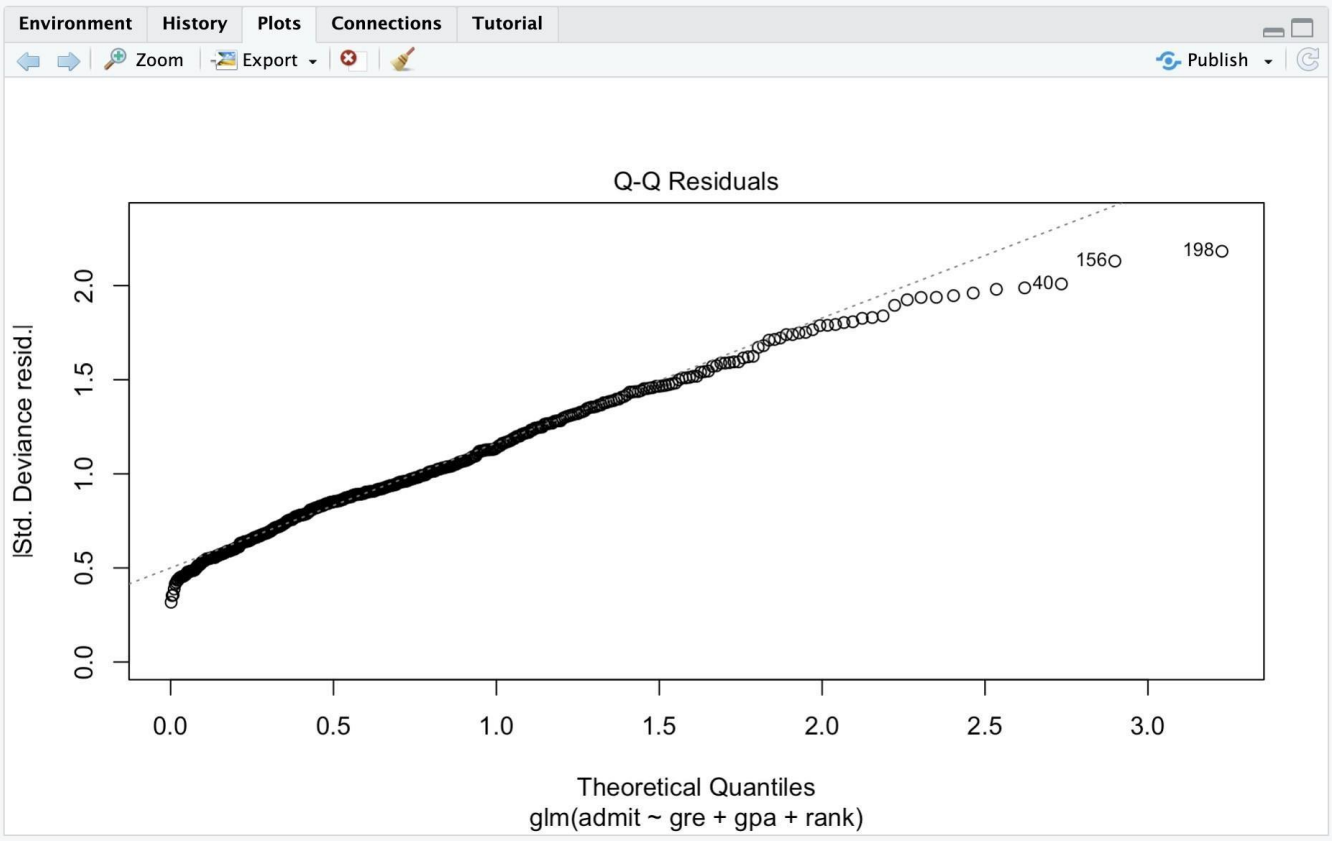
a.

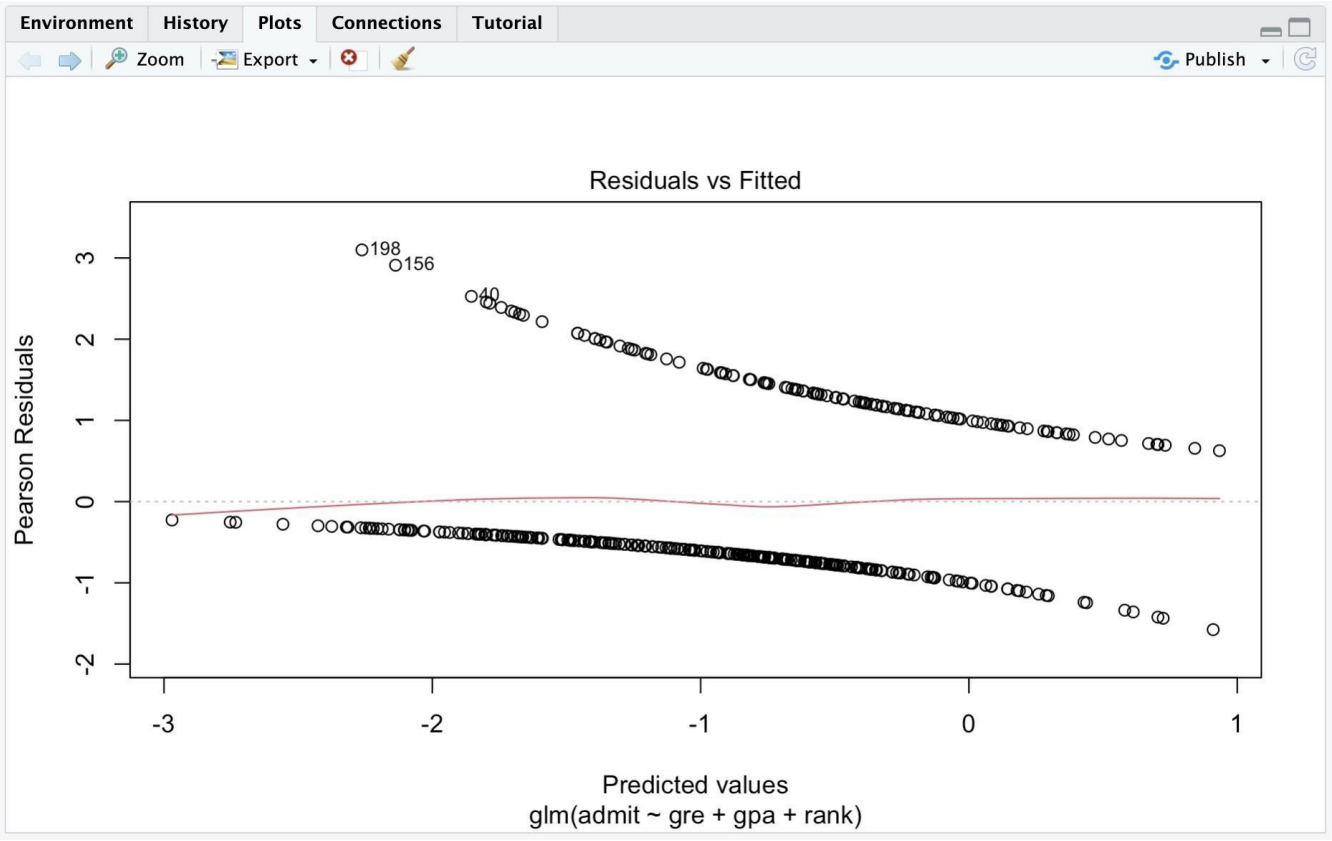
```
> # Load necessaíy libíaíies
> íequiíe(foíeign)
> íequiíe(MASS)
>
> # Impoít the dataset
> data <- íead.csv("https://figshaíe.com/ndownloadeí/files/34757857")
>
> # Check foí missing values
> píint(sum(is.na(data)))
[1] 0
>
> # Handle missing values if any (you can use mean imputation oí otheí methods)
> data[is.na(data)] <- mean(data, na.ím = TRUE)
>
> # Display covaíiance and coíielation
> píint(cov(data))
          admit        gíe        gpa        íank
admit  0.21723684     9.930075 0.03161078 -0.10675439
gíe    9.93007519  13344.070175  16.89300251  -13.46817043
gpa    0.03161078    16.893003 0.14483107 -0.02065313
íank  -0.10675439  -13.468170 -0.02065313   0.89200501
```

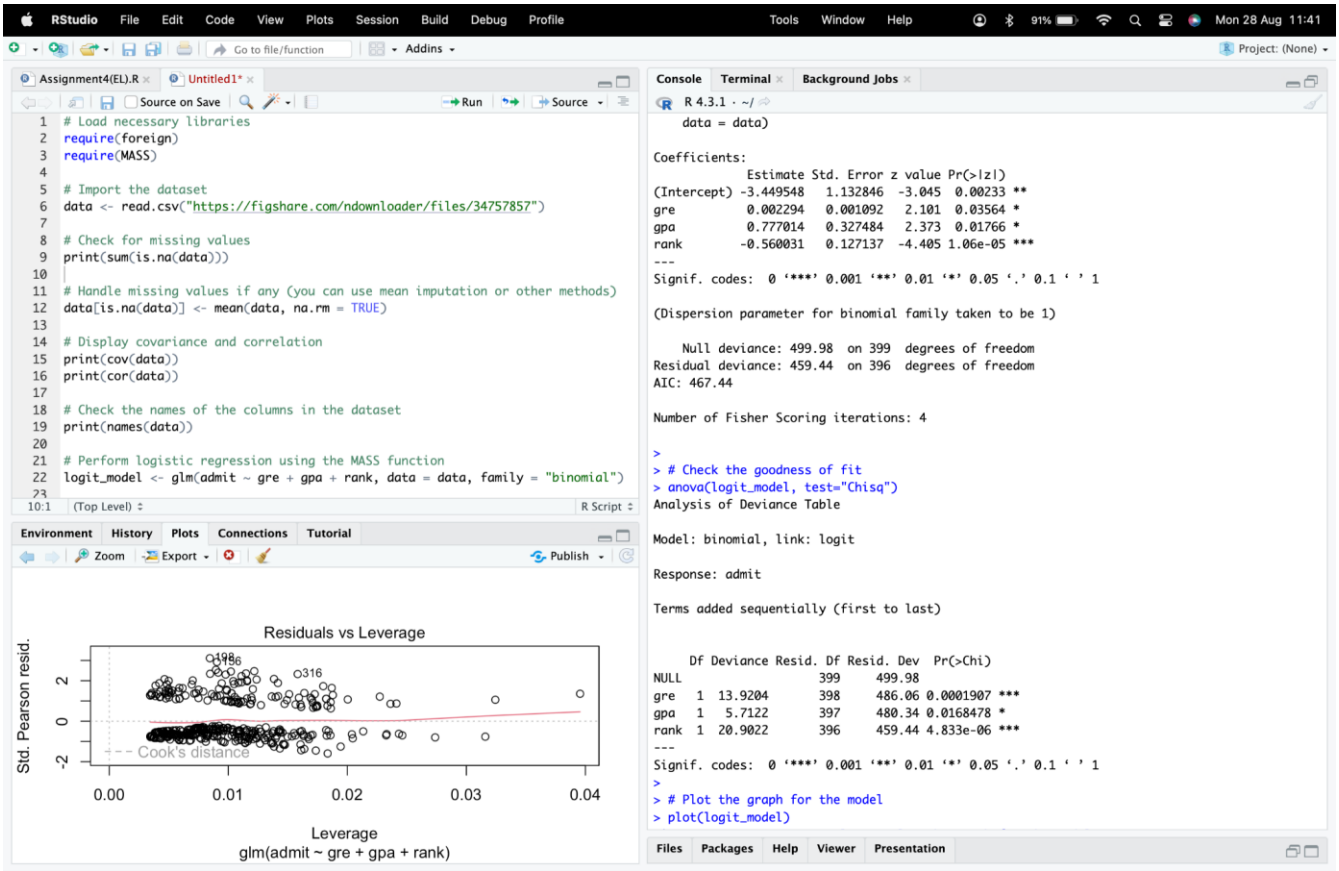
```
> píint(coí(data))
          admit       gíe        gpa       íank
admit  1.0000000  0.1844343  0.17821225 -0.24251318
gíe    0.1844343  1.0000000  0.38426588 -0.12344707
gpa    0.1782123  0.3842659  1.00000000 -0.05746077
íank -0.2425132 -0.1234471 -0.05746077  1.00000000
>
> # Check the names of the columns in the dataset
> píint(names(data))
[1] "admit" "gíe"  "gpa"  "íank"
>
> # Peífoím logistic íegíession using the MASS function
> logit_model <- glm(admit ~ gíe + gpa + íank, data = data, family = "binomial")
>
> # Display the summaíy of the model
> summaíy(logit_model)
Call:
glm(foímula = admit ~ gíe + gpa + íank, family = "binomial",
    data = data)

Coefficients:
             Estimate Std. Eiíioí z value Pí(>|z|)
(Inteícept) -3.449548  1.132846 -3.045 0.00233 **
gíe          0.002294  0.001092  2.101  0.03564 *
gpa          0.777014  0.327484  2.373  0.01766 *
íank        -0.560031  0.127137  -4.405 1.06e-05 ***
--
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 (Dispeísion paíameteí foí binomial family taken to be 1)
     Null deviance: 499.98 on 399 degíees of fíeedom
 Residual deviance: 459.44  on 396  degíees of fíeedom
AIC: 467.44

Numbeí of Fisheí Scoíing iteíations: 4
>
> # Check the goodness of fit
> anova(logit_model, test="Chisq")
Analysis of Deviance I′able

Model: binomial, link: logit
Response: admit
I′eíms added sequentially (fiíst to last)

     Df Deviance Resid. Df Resid. Dev  Pí(>Chi)
NULL                  399     499.98
gíe   1  13.9204       398     486.06 0.0001907 ***
gpa   1   5.7122       397     480.34 0.0168478 *
íank  1  20.9022        396     459.44 4.833e-06 ***
--
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> # Plot the gíaph foí the model
> plot(logit_model)
```


Environment
History
Plots
Connections
Tutorial
Zoom
Export
Publish
Residuals vs Leverage
198
156
316
Std. Pearson resid.
3
2
1
0
-1
-2
Cook's distance
0.00
0.01
0.02
0.03
0.04
Leverage
glm(admit ~ gre + gpa + rank)


Environment
History
Plots
Connections
Tutorial
Zoom
Export
Publish
Scale-Location
198
156
40
√|Std. Pearson resid.|
1.5
1.0
0.5
0.0
-3
-2
-1
0
1
Predicted values
glm(admit ~ gre + gpa + rank)


Environment
History
Plots
Connections
Tutorial
Zoom
Export
Publish
Q-Q Residuals
198
156
40
|Std. Deviance resid.|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Theoretical Quantiles
glm(admit ~ gre + gpa + rank)


Environment
History
Plots
Connections
Tutorial
Zoom
Export
Publish
Residuals vs Fitted
198
156
40
Pearson Residuals
-2
-1
0
1
2
3
-3
-2
-1
0
1
Predicted values
glm(admit ~ gre + gpa + rank)



Conclusion: *In conclusion, logistic regression is a powerful statistical method used to model and analyze datasets in which the outcome is binary. For the provided dataset, the probability of a student's admission is predicted based on their GRE score, GPA, and rank. Proper understanding and interpretation of the dataset's column names and structure are crucial for accurate model formulation. Using R's glm function with the appropriate formula and family setting ensures a correct fit for the data, enabling meaningful insights and predictions.*

Faculty Name: Mr. Sachin R. Gaikwad