Data Analysis Report

for

The Effect of Substrate Bulk Stiffness on Focal and Fibrillar Adhesion Formation in Human Abdominal Aortic Endothelial Cells

Submitted by :

Reetika Sethi

(203114009)

Under the Guidance of

Prof Hina Amol Gokhale

Department of Metallurgical Engineering & Materials Science

INDIAN INSTITUTE OF TECHNOLOGY BOMBAY

May 2020

Table of Contents and Figure

1. Introduction

Young's modulus is a measure of the ability of a material to withstand changes in length when under lengthwise tension or compression. Sometimes referred to as the modulus of elasticity, Young's modulus is equal to the longitudinal stress divided by the strain.In this analysis we try to relate Force as a regressor variable and length as a response variable and establish a linear relationship model between them.

2. Description of the data

Data Description

971 obs. of 2 variables:

Length: num 0 0 0 0 0 0 0 0 0 0 ...

Force : num 0.00635 0.00635 0.00638 0.00638 0.00636 ...

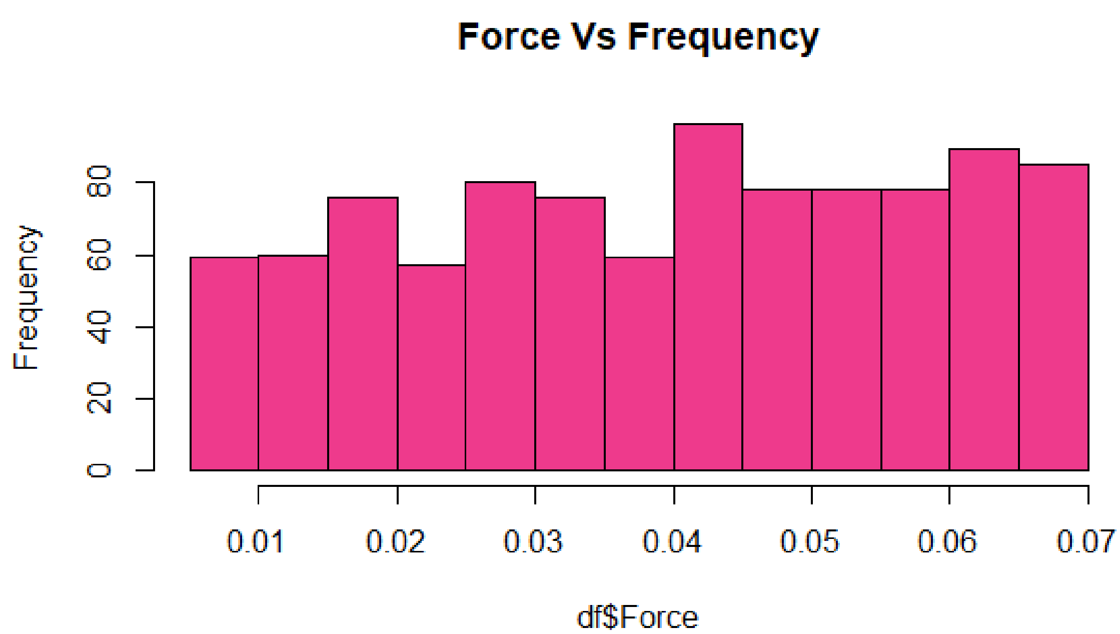

Fig 1: Uniform distribution of Force Variable

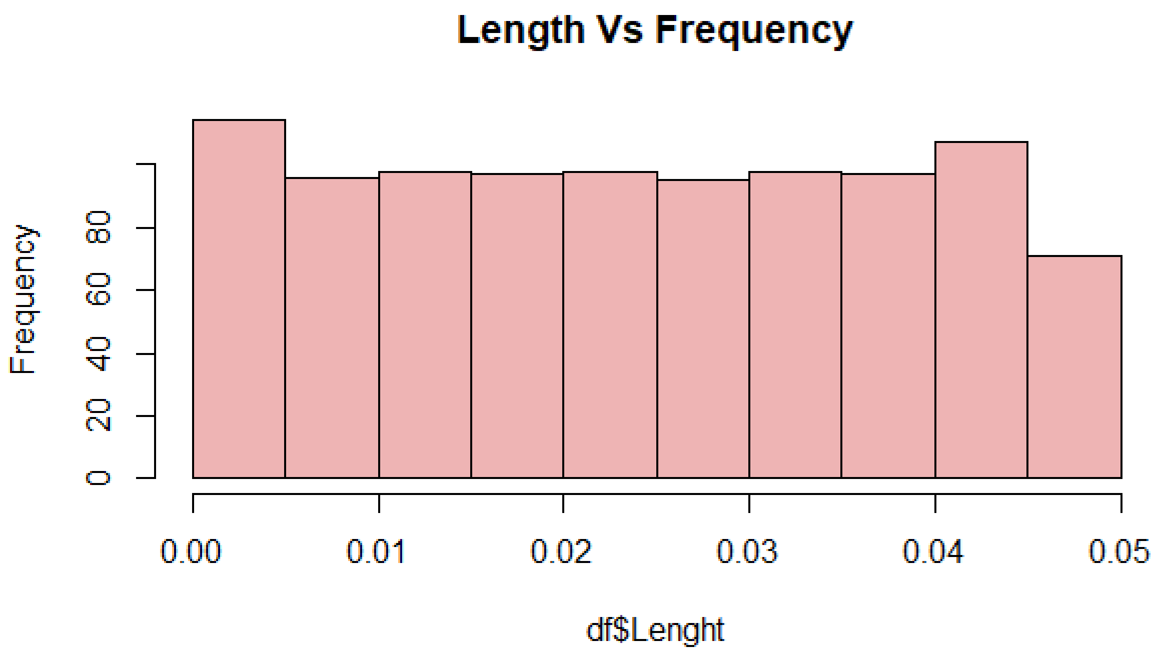

Fig 2: Uniform distribution of Length Variable

3. Distributional Analysis

In the given data, there are two variables- Length and Force
Length : Numerical Variable
Force : Numerical Variable
Length is an independent variable and Force is a dependent variable.
Number of times observation taken = 971.

Correlation Matrix

```
> cor(df[1:2])
          Lenght     Force
Lenght 1.0000000 0.9988924
Force  0.9988924 1.0000000
```

Fig 3 : Correlation Matrix

1. From Correlation Matrix shows that correlation between length and Force is very high.
2. Consequently, the Linear Regression model between them will be the best Fit to predict the output from the model.

Force Variable

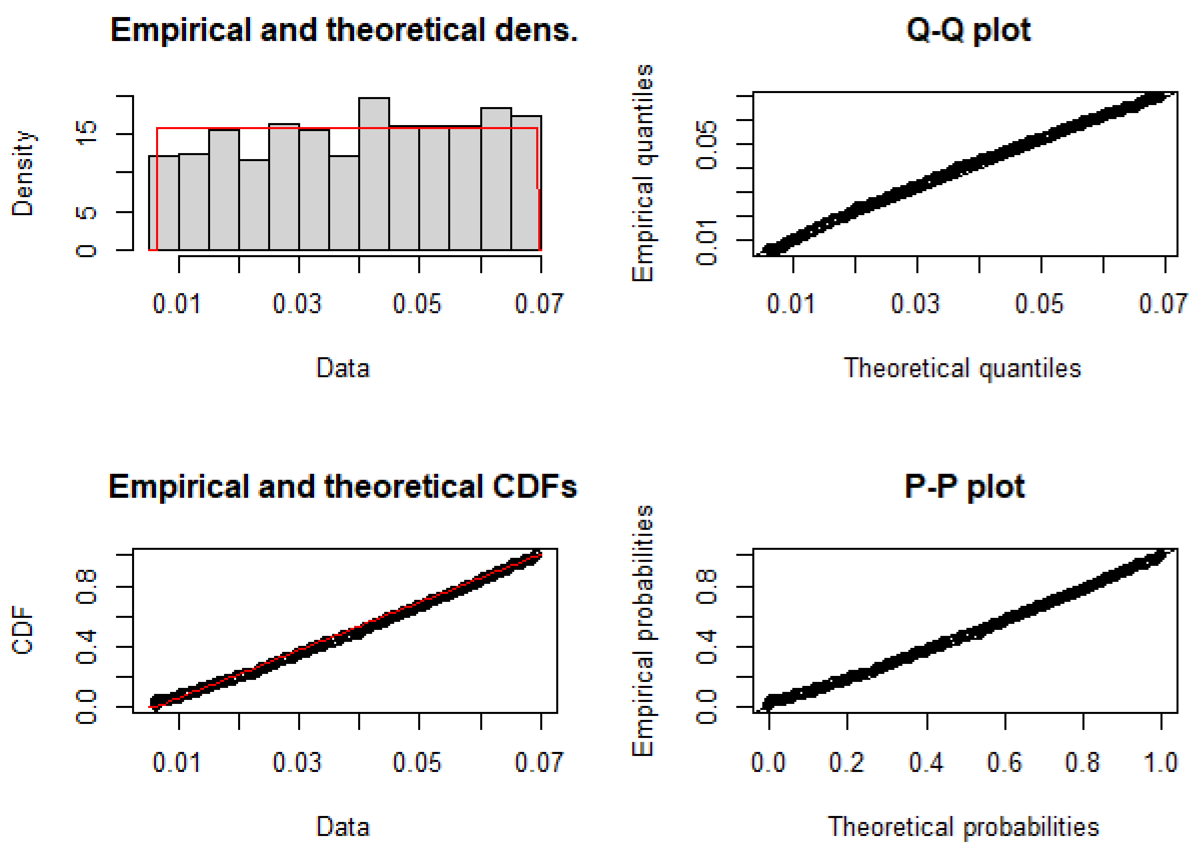

Fig 4: Distributional Analysis for Force

Length Variable

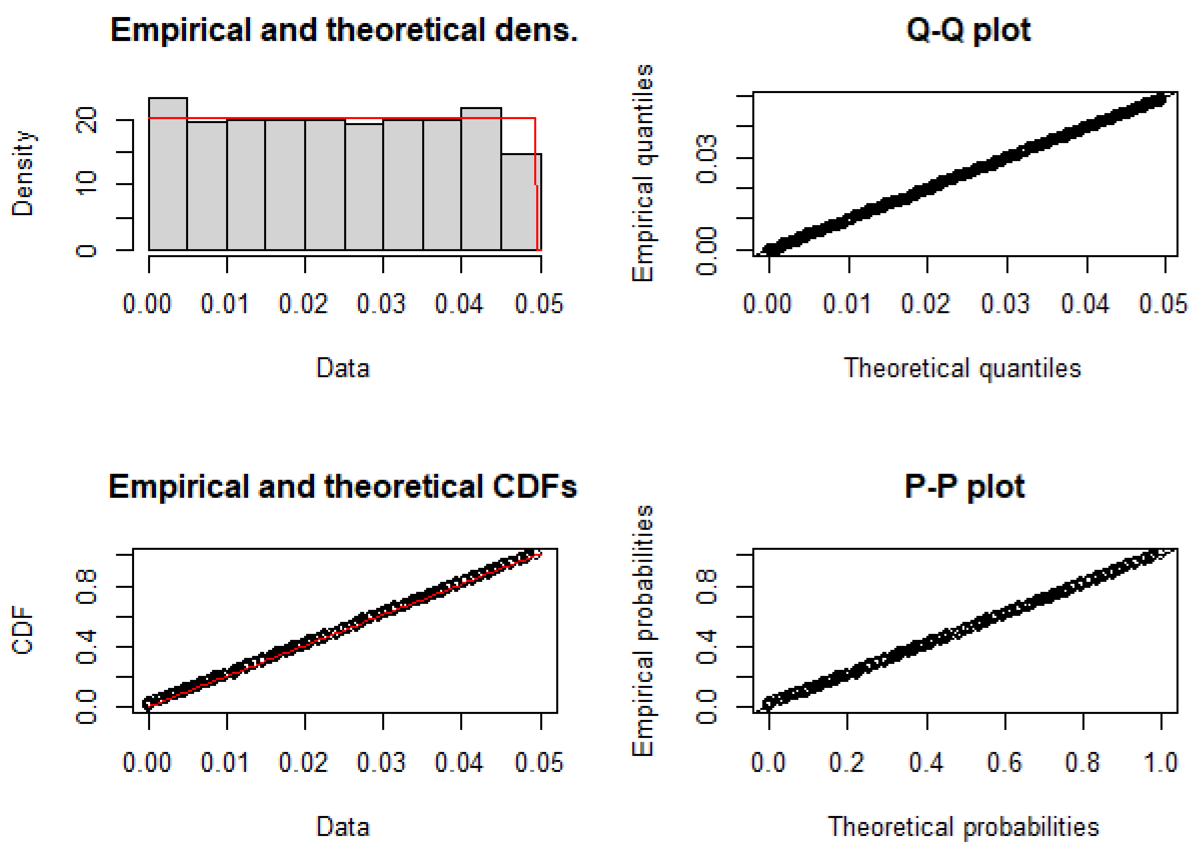

Fig 5: Distributional Analysis for Length

Predicted Distribution of Variable

Variable Name	Variable Type	Predicted Distribution
Force	Numerical	Uniform Distribution
Length	Numerical	Uniform Distribution

4. Analysis

Linear regression model Analysis

```
lm(formula = Length ~ Force, data = df)

Residuals:
      Min         1Q     Median         3Q        Max 
-0.0012875 -0.0005586 -0.0001027  0.0004576  0.0025560 

Coefficients:
              Estimate  Std. Error t value            Pr(>|t|)    
(Intercept) -0.00613296  0.00005072  -120.9 <0.0000000000000002
Force        0.76896105  0.00116361   660.8 <0.0000000000000002

Residual standard error: 0.0006741 on 969 degrees of freedom
Multiple R-squared:  0.9978,Adjusted R-squared:  0.9978 
F-statistic: 4.367e+05 on 1 and 969 DF,  p-value: < 0.00000000000000022
```

Hypothesis Testing for linear regression model

As Multiple R-squared: 0.9978 and Adjusted R-squared: 0.9978 values are approximately equal to 1, it means this linear regression model is having a good relationship between response and regressor variable.
Observed P values are close to zero it means they are rejecting the Null hypothesis and there will be linear relationship between regressor and response variable.

Residual Analysis

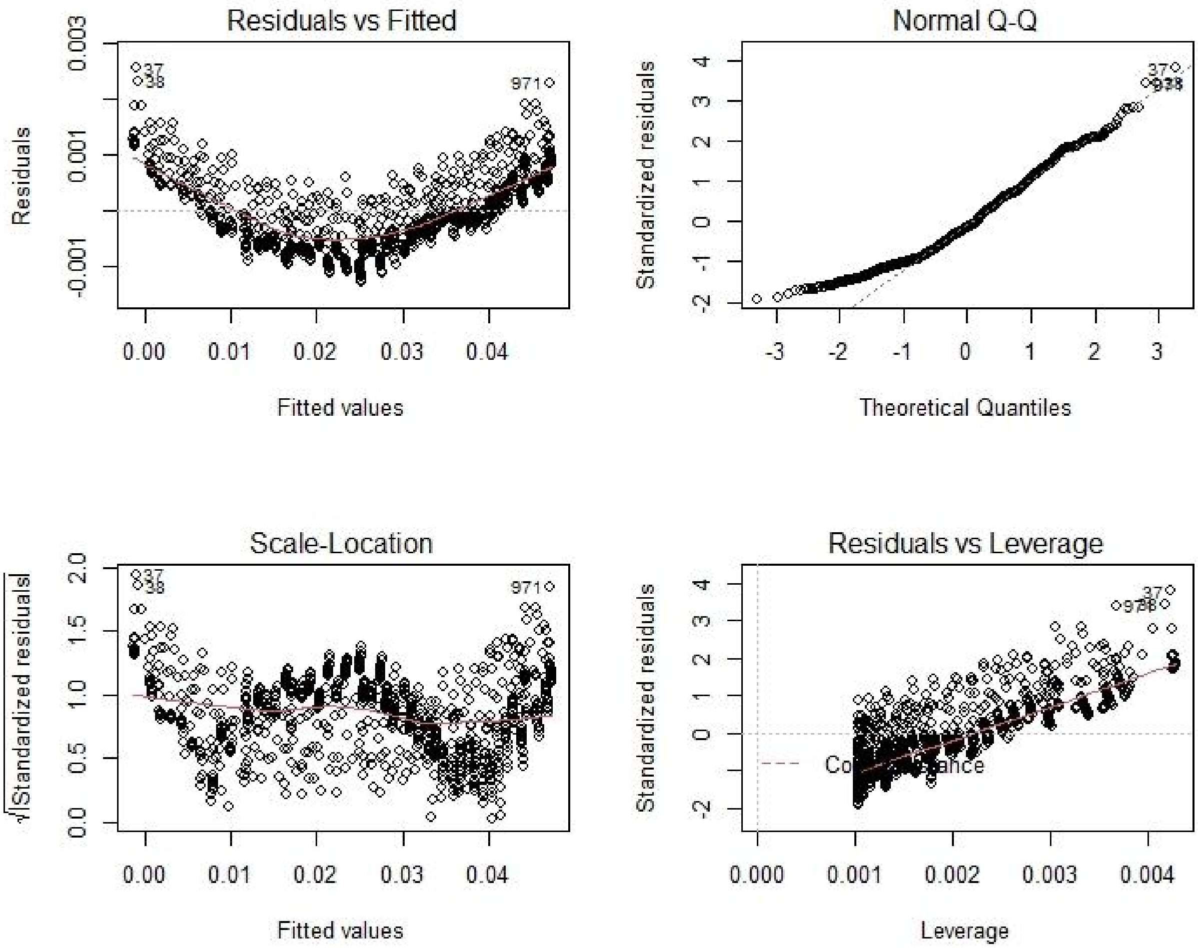

Fig 6: Residual Analysis for Fitted Model

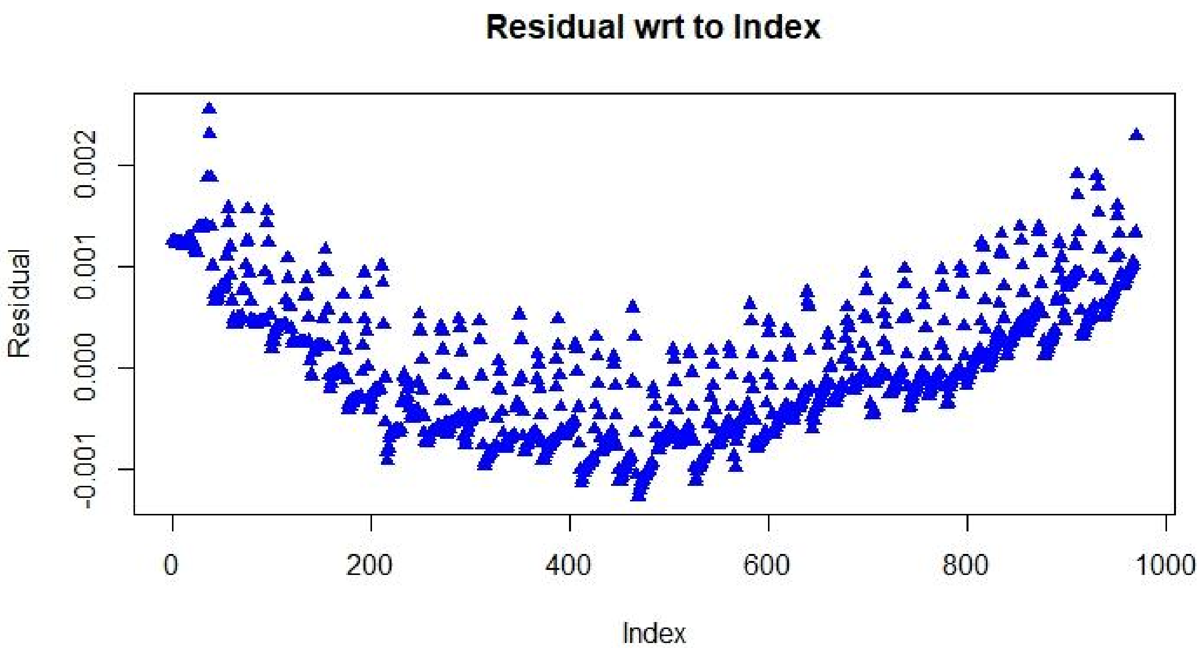

Fig 7: Residual Analysis for Index of Regressor Variable

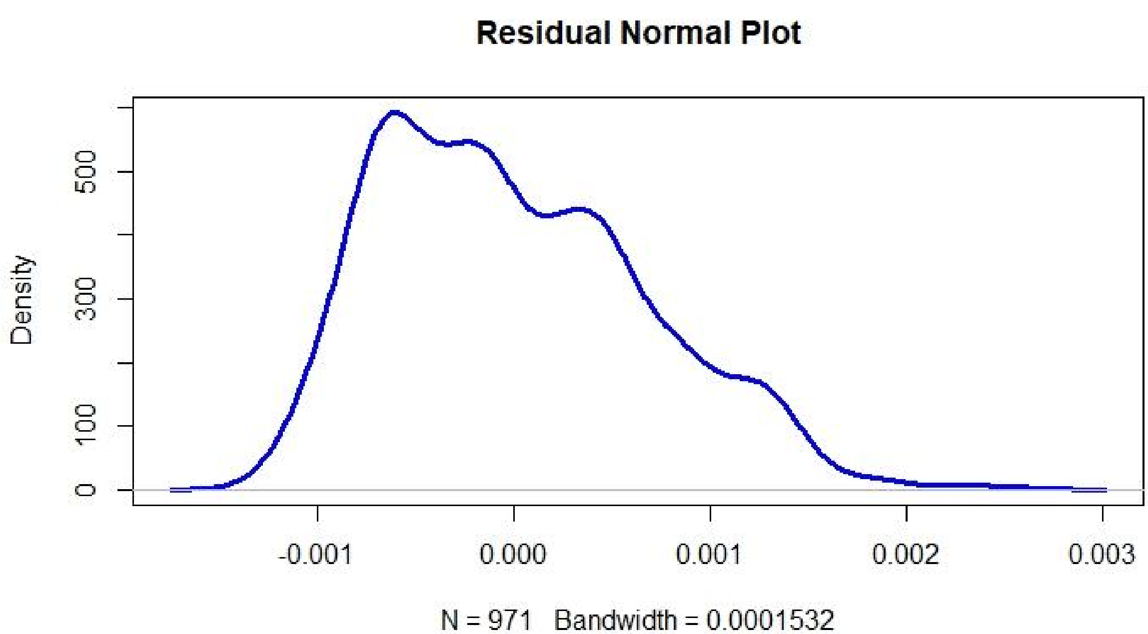

Fig 8: Residual Density Plot

5. Conclusion

1. Length is our Response variable and Regressor variable.
2. Response and Regressor both variable follow the Uniform distribution
3. Best Possible fit for Regressor and Response Variable is

$$\boxed{\textbf{Length} = -0.006133 + 0.768961 * \textbf{Force}}$$

4. Residual analysis indicates that our variance is constant throughout the fitted value and there is no problem related to heteroscedasticity.
5. With the help of QQ plot we can interpret the distribution of our data.
6. Confirmation of Distribution of data with the help of Shapiro Test and observed P_Value.
7. Residual density plots follow the normal distribution with zero mean that means our error follows the normal distribution with zero means (Its basic assumption for any linear regression model and satisfied here).
8. As Multiple R-squared 0.9978 and Adjusted R-squared 0.9978 values are approximately equal to 1, it means this linear regression model is having a good relationship between response and regressor variable.
9. Standardised residual with respect to fitted values shows that our model has enough variability within itself and there is no problem related to heteroscedasticity and its variance is constant throughout and we can apply Linear Regression in it.
10. Residual graphs with respect to the Index of Response variable show that our data is random which also satisfies the randomness assumption of linear regression model.