Statistics Software Lab Report - 9 (Outputs file)

Name of the Student: Shatansh Patnaik
Roll No: 20MA20067

IIT Kharagpur
Statistics Software Lab

Problem-1

To determine the maximum stopping ability of cars when their brakes are fully applied, 10 cars are to be driven each at a specified speed and the distance each requires come to a complete stop is to be measured. The various initial speeds (X) selected for each of the 10 cars and the stopping distances (Y) recorded are given below.

X	20	20	30	30	30	40	40	50	50	60
Y	16.3	26.7	39.2	63.5	51.3	98.4	65.7	104.1	155.6	217.2

a) Draw scatter plots of (X, Y) and $(X, \sqrt{Y})$.
b) Fit a simple linear regression line between $U = \sqrt{Y}$ and X.
c) Find 95% confidence intervals for the slope, intercept, and σ^2 in the fitted model.
d) Test the significance of the slope and the intercept.
e) Find a 95% confidence interval for the expected stopping distance when the initial speed is 35.
f) Find a 95% prediction interval for stopping distance when the initial speed is 35.
g) Carry out a lack of fit analysis for the fitted model and check whether the model is adequate.

Solution for Problem-1

a)

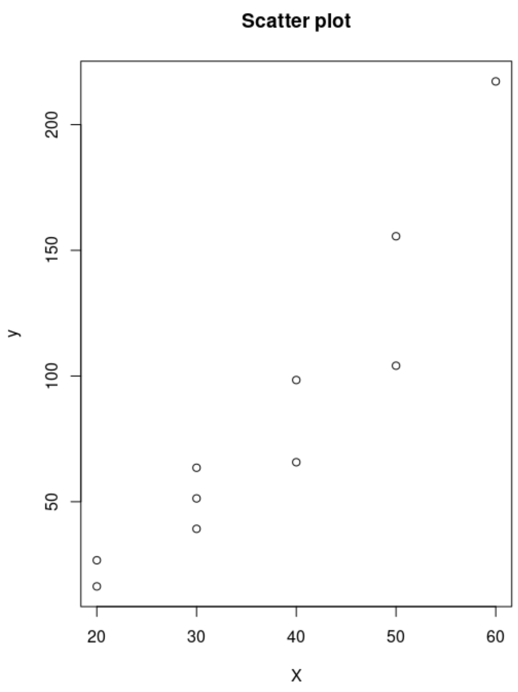

Figure 1: Scatter Plot for Y vs X

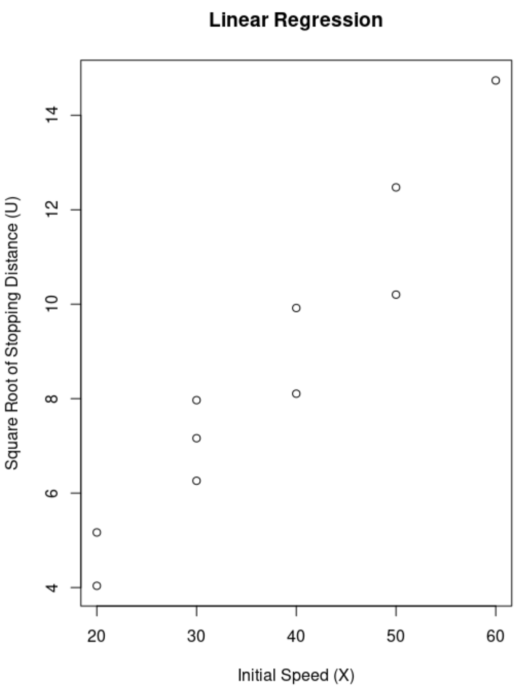

Figure 2: Scatter Plot for $\sqrt{Y}$ vs X

b)	Intercept (beta0): **-0.1665315** Slope (beta1): **0.2370318**

c)	95% Confidence Interval for Slope (beta1): **0.1820716 0.2919921** 95% Confidence Interval for Intercept (beta0): **-2.316314 1.983251** 95% Confidence Interval for σ^2 : **0.4172523 3.356529**

d)	Accept the null hypothesis: Beta0 is **NOT significant**. Reject null hypothesis: Beta1 is **significant**.

e)	95% Confidence Interval for the expected stopping distance when the initial speed is 35: **55.10992 78.06712**

f)	95% Prediction Interval for the stopping distance when the initial speed is 35: **33.80336 109.1001**

g)	The model **doesn't not lack in fitting the data**

Problem-2

An instrument which measures lactic acid concentration in the blood is to be calibrated. The investigator uses $n = 20$ samples of known concentration X (in mM) and then computes the concentration Y (in mM) determined by the instrument. The data obtained are given below. Fit a simple linear regression line of Y on X. Does the scatter diagram exhibit a strong linear relationship? Also calculate the correlation coefficient. Test for the significance of coefficients of the model. Also find 95% confidence intervals for the parameters of the model.

X	Y	X	Y	X	Y
11.1	33.0	13.0	41.5	10.0	12.0
10.7	31.4	57.3	10.3	13.1	15.1
11.8	34.9	58.2	10.2	12.6	15.7
10.4	34.4	56.2	10.2	13.2	15.1

Solution for Problem-2

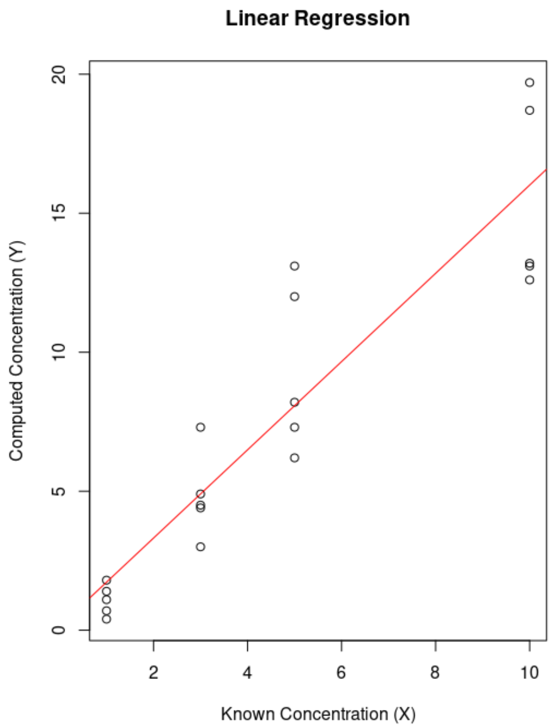

Figure 3: Scatter Plot for Y vs X

Intercept (beta0): **0.1415642**
Slope (beta1): **1.587039**

The model **doesnt not lack in fitting the data**

Correlation Coefficient:**0.916932**

Accept the null hypothesis: **Beta0 is not significant**.
Reject null hypothesis: **Beta1 is significant**.

95% Confidence Interval for Slope (beta1): **1.245025 1.929054**
95% Confidence Interval for Intercept (beta0): **-1.84536 2.128488**
95% Confidence Interval for Sigma^2 : **3.385559 12.96777**

Problem-3

The following data is available on the independent variable X and the dependent variable Y.

X	1.00	1.05	1.10	1.15	1.21	1.32	1.44	1.53
Y	3.71	3.81	3.86	3.93	3.96	4.20	4.34	4.51
X	1.63	1.79	1.91	2.03	2.12	2.26	2.37	2.51
Y	4.73	5.35	5.74	6.14	6.51	6.98	7.44	7.76

a) Fit a second-order polynomial regression model to the above data.

b) Find 95% confidence intervals for the coefficients of the model and σ^2.

c) Test for the significance for the coefficients of the model.

d) Find R^2.

Solution for Problem-3

a)

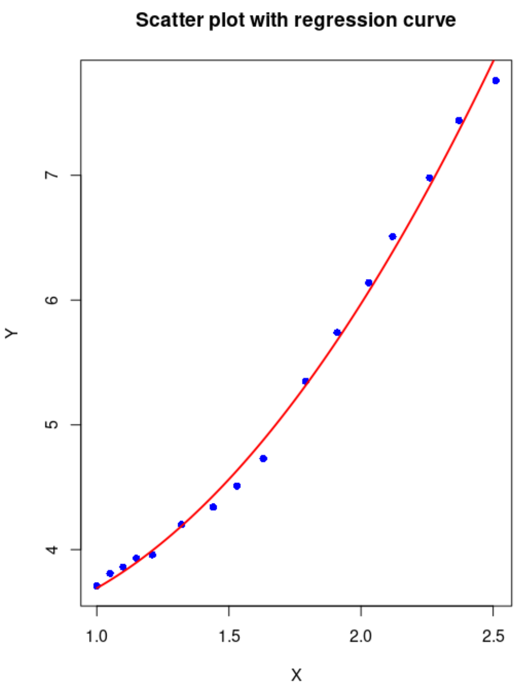

Figure 4: Scatter Plot for Y vs X

Coefficients:
β_0 : **3.55099**
β_1 : **-0.9291236**
β_2 : **1.069541**

b)

95% Confidence Intervals:
β_0 : [**2.772395 , 4.329585**]
β_1 : [**-1.894104 , 0.03585686**]
β_2 : [**0.790148 , 1.348934**]

c) β_0 : **is Significant (p ¡ 0.05)**
β_1 : **is Not Significant (p ¿= 0.05)**
β_2 : **is Significant (p ¡ 0.05)**

d) Coefficient of determination (R^2) : **0.9956444**

Problem 4

Fit a nonlinear relationship $y = ae^{bx}$ for the following data:

x	y
1	1.60
2	4.50
3	13.80
4	40.20
5	125.00
6	300.00

Draw the scatter diagram also. Does the fitted model adequately represent the data? Can you test for the significance of coefficients in the model? Give your conclusions on the tests.

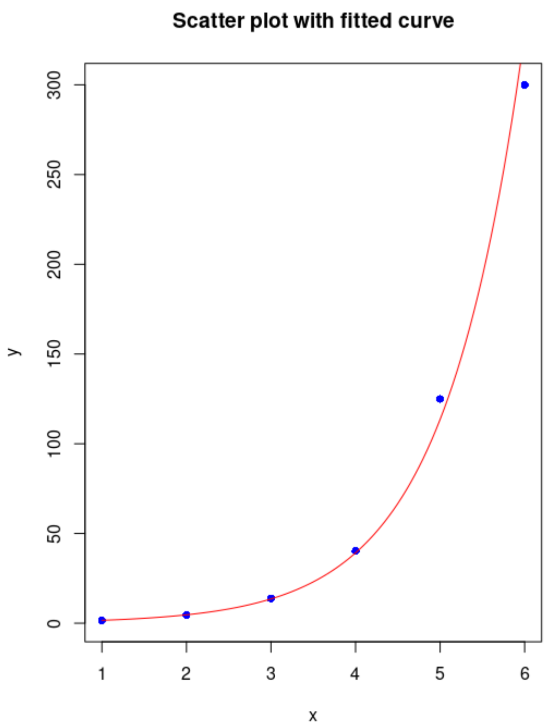

Figure 5: Scatter Plot for Y vs X

We note that:
$\beta_0 : loga$
$\beta_1 : b$
$F - statistic : 0.000764297$
Accept the null hypothesis: Lack of fit is not significant.

Reject null hypothesis: Beta0 is significant.
Reject null hypothesis: Beta1 is significant.

Based on the results obtained, we can infer the following:

- The estimated parameter β_0, representing loga, is found to be significant, indicating that the intercept term has a meaningful impact on the model.
- Similarly, the estimated parameter β_1, representing b, is also found to be significant, suggesting that the coefficient associated with the nonlinear term has a significant effect on the response variable.
- The F-statistic value of 0.000764297 suggests that the lack of fit is not significant, indicating that the model adequately fits the data.

Overall, the results support the validity of the nonlinear relationship model $y = ae^{bx}$ for the given data.

Problem 5

Fit a nonlinear relationship $y = a \cdot b^x$ for the following data:

x	y
2	144.0
3	172.8
4	207.4
5	248.5
6	298.5

Draw the scatter diagram also. Does the fitted model adequately represent the data? Can you test for the significance of coefficients in the model? Give your conclusions on the tests.

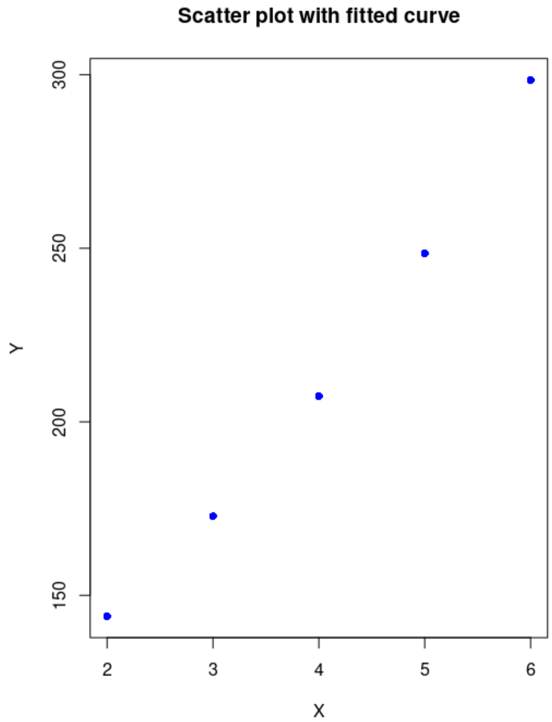

Figure 6: Scatter Plot for Y vs X

We note that:
$\beta_0 : loga$
$\beta_1 : logb$
$F - statistic : 2.205712e - 06$
$Critical value : 19.16429$
Accept the null hypothesis: Lack of fit is not significant.

Reject null hypothesis: Beta0 is significant.
Reject null hypothesis: Beta1 is significant.

Based on the results obtained, we can infer the following:

- The estimated parameter β_0, representing $\log(a)$, is found to be significant, suggesting that the intercept term has a meaningful impact on the model.
- Similarly, the estimated parameter β_1, representing $\log(b)$, is also found to be significant, indicating that the coefficient associated with the logarithm of the independent variable has a significant effect on the response variable.
- The F-statistic value of 2.205712×10^{-6} is compared to the critical value of 19.16429. Since the F-statistic is significantly lower than the critical value, we fail to reject the null hypothesis, indicating that the lack of fit is not significant. This suggests that the fitted model adequately represents the data.

Overall, the results support the validity of the nonlinear relationship model $y = ab^x$ for the given data.