STAT 6210 Homework 6 Report
Xue Ming Wang (Vivian)
G20580112

Part A

1. Data set

The given dataset comprises an explanatory variable X and a response variable Y. We intend to use a nonlinear regression model to fit this dataset, represented by the equation:

$$y = a + \frac{bx}{c + x}.$$

In order to improve the fit of the model to the data, we must determine the most appropriate initial parameters. The subsequent steps outline the procedure for identifying the optimal initial parameters.

2. Initial Parameters

In order to determine the initial parameters, we can consider the following scenarios: (1) when x approaches zero, (2) when x approaches infinity, and (3) when x equals 1.

2.1 Initial a

To determine the initial value for parameter a, we will plot a scatter-plot of the dataset and identify the value of Y when X approaches zero. It should be noted that when x approaches zero, the model approximates y to be equal to the parameter a.

$$\lim_{x \to 0} a + \frac{bx}{c + x} => \lim_{x \to 0} y = a$$

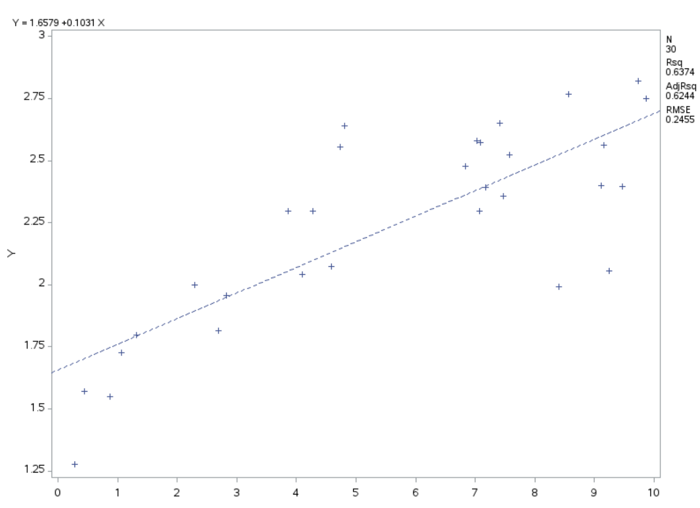

Figure 1. X vs. Y Scatter-Plot

From Figure 1, we can see that when x approaches zero, y takes a value of 1.6579. Therefore, we can use this value as the initial estimate for the parameter a.

2.2 Initial b

Next, we will determine the initial value of the parameter b. To do this, we can calculate the value of y as x approaches infinity. Using L'Hospital's rule, as x approaches infinity,

$$\lim_{x\to\infty} a + \frac{bx}{c+x} = \frac{\mathrm{ac} + \mathrm{a\,x} + \mathrm{b\,x}}{\mathrm{c} + \mathrm{x}}.$$

Taking the limit of this equation, we get:

$$\lim_{x\to\infty} y = a + b$$

as x approaches infinity. To obtain the initial estimate for b, we can transform the equation as b = y - a, and then create the b vs. a scatter-plot. From Figure 2, we can observe that the value of b is approximately 1.2, where a is 1.6579.

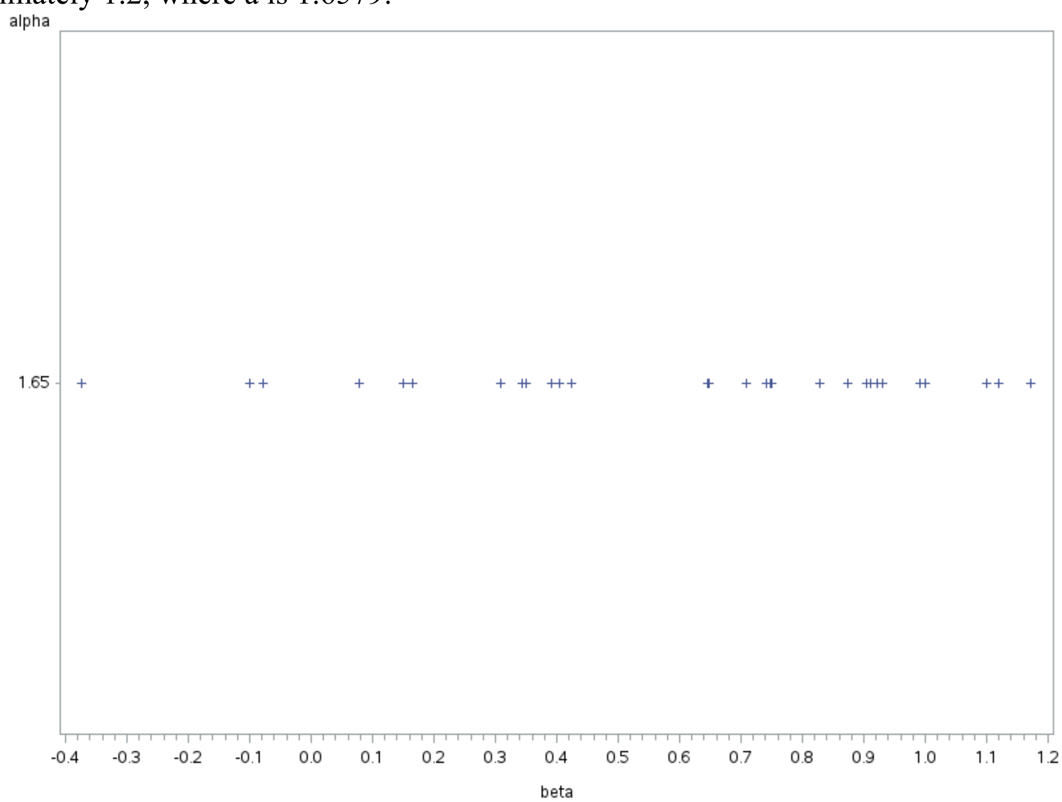

Figure 2. a vs. b Scatter-Plot

2.3 Initial c

To determine the initial value of parameter c, we can set x approaches to 1 in the equation for y.

$$\lim_{x \to 1} a + \frac{bx}{c + x} = \frac{\mathrm{ac\ +\ a\,x\ +\ b\,x}}{\mathrm{c\ +\ x}}$$

Taking the limit of this equation, we get:

$$\lim_{x \to 1} y = a + \frac{b}{c + 1}$$

To obtain the initial estimate for c, we can transform the equation as:

$$c = \frac{b}{y - a} - 1$$

Given that a is equal to 1.6579 and b is equal to 1.2, we can use the dataset to find that when x approaches 1, the corresponding value of Y is 1.7. Based on these values, we can calculate the initial value of c to be 27.50.

3. Estimate the Model

To estimate the model, we can set the initial values of a, b, and c that we calculated earlier. We can then fit the nonlinear regression model using these initial values.

Parameter	Estimate
alpha	1.2006
beta	1.7148
theta	2.7389

Table 1. Parameter Estimate

Source	DF	Sum of Squares	Mean Square	F value	Pr > F
Model	2	3.4559	1.7280	39.0	<.0001
Error	27	1.1963	0.0443		
Corrected Total	29	4.6523			

Table 2. Analysis of Variance.

Table 1 provides the estimates of the parameters, while Table 2 shows that the model is significant. As the sum of squares converges, the estimates of the parameters become stable. Using the Gauss-Newton method, we can determine the exact value of each parameter, where:

a = 1.2006
b = 1.7148
c = 2.7389

$$y = 1.2006 + \frac{1.7148x}{(2.7389 + x)}.$$

4. Goodness of Fit

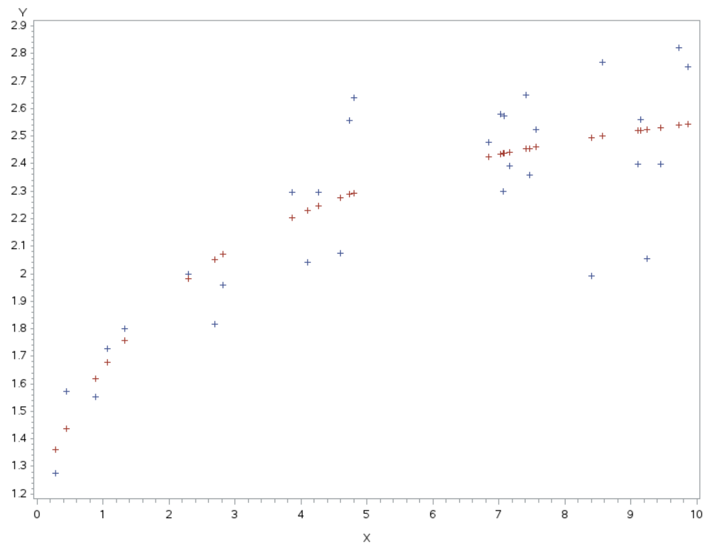

Figure 3. X vs. Y Scatter-Plot

The goodness of fit can be evaluated by comparing the original scatter-plot with the prediction plot, as shown in Figure 3. The blue dots represent the original observations, while the red dots represent the predicted values. We can see that the predicted values closely follow the trend of the observed values, indicating a good fit. As the model is a nonlinear regression, it provides a more accurate description of the relationship between the variables than a linear regression model.

5. Conclusion

The estimated parameters are a = 1.2006, b = 1.7148, and c = 2.7389, and the model is:

$$y = 1.2006 + \frac{1.7148x}{(2.7389 + x)}.$$

Through the scatter-plot, we can conclude that this nonlinear regression model is a good fit.

Part B

1. Data Set

The original dataset (Table 3) comprises one response variable (y), two explanatory variables (x and z), and one time variable (t). In this section, we aim to fit the data using generalized linear models. We assume that y follows a Poisson distribution with a mean of μ, where μ is related to the time length (t) as $\mu = \lambda * t$. To preprocess the time variable, we take the logarithm of its values and use the resulting log(time) variable in our analysis.

Obs	T	X	Z	Y	Log(Time)
1	2.6	0.9	1.4	1	0.95551
2	1.6	0.4	0.6	0	0.47000
3	1.6	0.8	1.9	1	0.47000
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

Table 3. Original Data set

2. Estimated Parameters

Based on the Poisson distribution, we have estimated the parameters, and the results are presented in Table 4.

Parameter	Estimate	SE	WALD 95% CI		WALD chi-square	Pr > ChiSq
Intercept	-3.9463	0.7616	-5.4391	-2.4535	26.85	<.0001
X	3.1401	0.7800	1.6114	4.6687	16.21	<.0001
Z	1.0135	0.3457	0.3358	1.6911	8.59	0.0034
Scale	1.0000	0.0000	1.0000	1.0000		

Table 4: Analysis Of Maximum Likelihood Parameter Estimates.

From Table 4, we have obtained the model:

$$\log(\mu) = -3.9463 + 3.1401X + 1.0135Z + \log(T)$$

Here, $\log(\mu)$ represents the expected value of y. We observe that both variables X and Z have significant p-values, which indicates that they have a considerable influence on the variable y.

3. Verify the Assumption

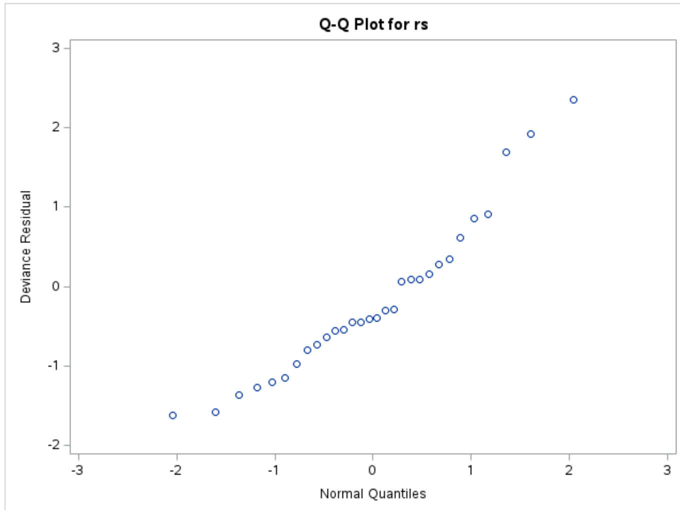

Figure 4. QQ-plot

The qqplot of residuals in figure 4 shows that the dots follow a straight line. This indicates that the do not show clear deviation from normality, So the model assumption is valid.

4. Conclusion

The model is $\log(\mu) = -3.9463 + 3.1401X + 1.0135Z + \log(T)$, and it is well fitted, and the model assumption is valid.