THE GEORGE WASHINGTON UNIVERSITY
6210,2018 SPRING

Report of Homework6

Jiaying Liu

April 3, 2018

1 Data set

The data set contain one response variable y and one explanatory variable x. We are going to fit this data set by the model

$$y = a + b * x/(c + x)$$

This is a nonlinear regression model and if we need to fit the data better, we need to set the initial parameters. The following step show the process of finding the best initial parameter.

Table 1: Original Data set.

Obs	**x**	**y**
1	0.28	1.27661
2	0.44	1.57264
3	0.88	1.55105
4	1.06	1.72785
5	1.32	1.79946
6	2.29	1.99903
7	2.69	1.81608
8	2.82	1.95772
9	3.86	2.29628
10	4.10	2.04079
...	...	...
27	9.25	2.05435
28	9.46	2.39711
29	9.73	2.82162
30	9.87	2.74975

2 Initial Parameters

To find a initial parameters, we can use following scenarios. (1) first, as x goes to zero; (2) secondly, as x goes to infinite; (3) firth, as x equal 1;

2.1 Find initial a

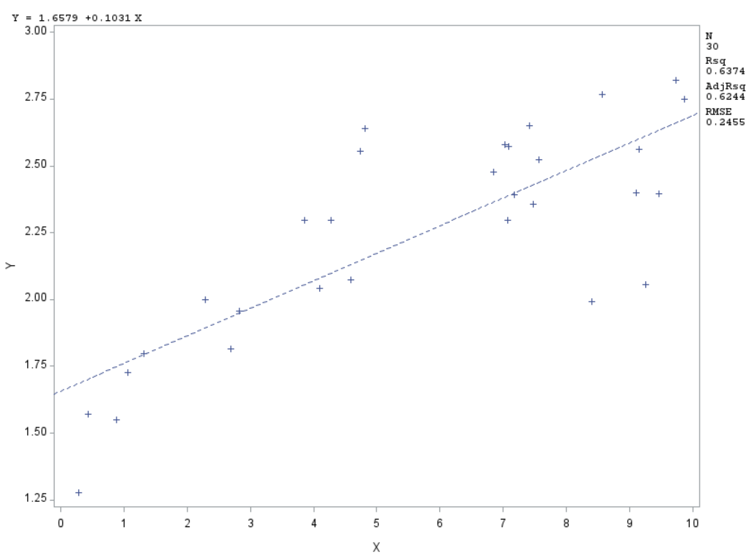
We are going to draw the scatter plot between x and y and find the value when x goes to zero. Note that when x goes to zero, the model about goes to y equal to parameter a.

$$\lim_{x \to 0} a + b * x/(c + x)$$

$$\lim_{x \to 0} y = a$$

The figure one show the result of scatter plot, we can see when x goes to zero, y hit the value of 1.2, so we can first set the initial value of a to be 1.2.

Figure 1: Scatter Plot of x vs. y



2.2 Find initial b

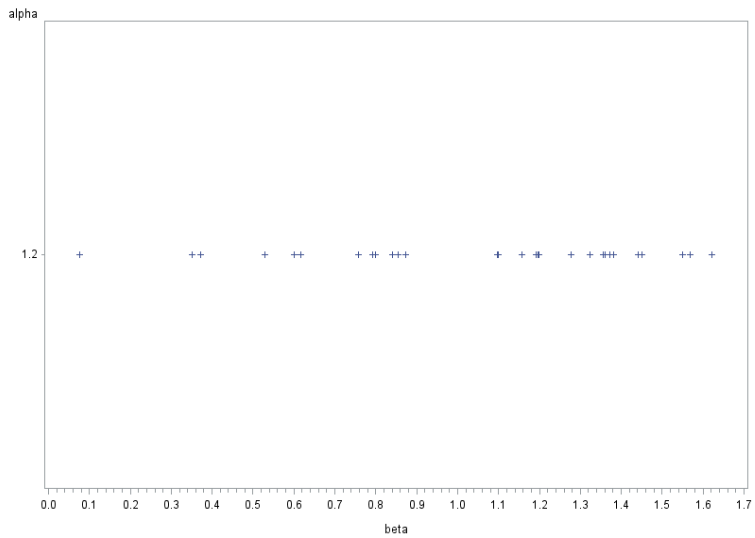
Secondly, let's find the value of parameter b. we can find the value by x goes to infinitely. By using the L'Hospital's rule, y will equal to $a + b$

$$\lim_{x \to \infty} a + b * x/(c + x) = (a * c + a * x + b * x)/(c + x)$$

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

transform the parameter $b = y - a$ and plot the scatter plot between b and a.From the figure 2 We find the value near to 1.7 is the value of parameter b.

Figure 2: Scatter Plot of a vs. b



2.3 Find initial c

From the above, we already found the initial value of a and b. Now let set x goes to 1, then y will equal to $a + b/(c + 1)$

$$\lim_{x \to 1} a + b * x/(c + x) = (a * c + a * x + b * x)/(c + x)$$

$$\lim_{x \to 1} y = a + b/(c + 1)$$

$$c = b/(y - a) - 1$$

we know that a equal to 1.2 and the b equal to 1.7, if the x goes to 1, from the data set, we will find that the y equal to 1.7. So depend on those value, we can calculate that the initial c value is 2.4.

2.4 estimate the model

Let's set the initial value of a,b and c. Then fit the nonlinear model by using this value. The table 2 show that the model is significant and the table 3 show the estimate of parameter. We find that as the sum of squares converge, the estimate of parameters go to stable. From the Gauss-Newton method, we find the exact value of each parameter.

$$a = 1.2006 \; b = 1.7148 \; c = 2.7389 \; y = 1.2006 + 1.7148 * x/(2.7389 + x)$$

Table 2: Analysis of Variance.

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

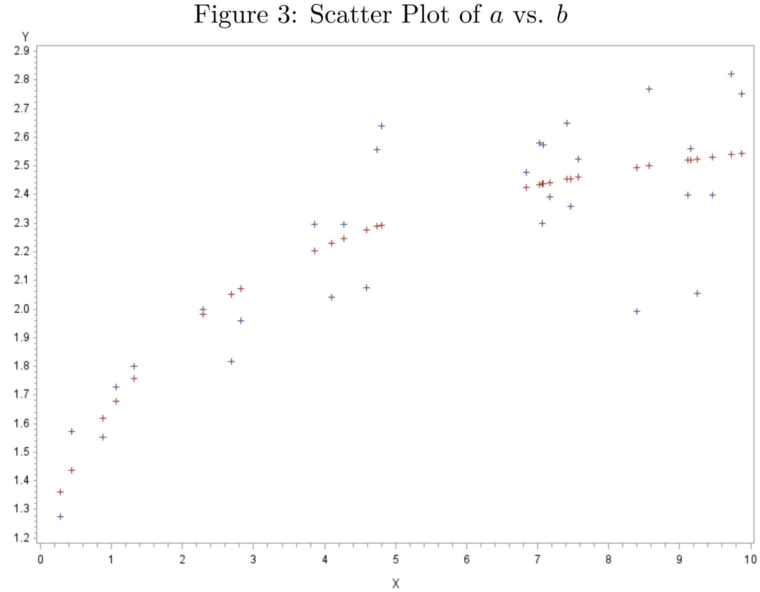
Table 3: parameter estimate.

parameter	estimate	approx std error	confident interval	
alpha	1.2006	0.2242	0.7406	1.6607
beta	1.7148	0.1995	1.3055	2.1240
theta	2.7389	1.7176	-0.7854	6.2632

2.5 goodness of fitting

The figure 3 show the result of original scatter plot and the prediction plot. The blue dot show the original observation and the red dot show the prediction value. We can find that the prediction dot approximately describe the trend of observation dot.It is a nonlinear regression.

Figure 3: Scatter Plot of a vs. b



3 Data set

The original data set contains one response variable y and two explanatory variable x and z and one time variable t. In the section we are going to fit the date by generalized linear models. We assume that y follows a Possion distribution with mean μ and μ is related with time length $\mu = \lambda * t$. So first we need to preprocesse the variable time. Set it to log value which is log(time).

Table 4: Original Data set.

Obs	**T**	**X**	**Z**	**Y**	**logtime**
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
4	1.7	0.7	1.7	1	0.53063
5	2.6	0.7	1.4	2	0.95551
...	...	...	...	...	...

3.1 estimate parameters

Base on the Possion distribution, we estimate the parameter and get the result in table 5. From the table 5, we get the model

$$log(\mu) = -3.9463 + 3.1401X + 10.135Z + log(T)$$

which $log(\mu)$ represent the expectation of y. We can see both variable have significant p value which mean they have influence on variable y.

Table 5: Analysis Of Maximum Likelihood Parameter Estimates.

Parameter	**DF**	**Estimate**	**SE**	**WALD CI**		**WALD chisquare**	**P value**
Intercept	1	-3.9463	0.7616	-5.4391	-2.4535	26.85	<.0001
X	1	3.1401	0.7800	1.6114	4.6687	16.21	<.0001
Z	1	1.0135	0.3457	0.3358	1.6911	8.59	0.0034
Scale	0	1.0000	0.0000	1.0000	1.0000		

3.2 Verify assumption

From the figure 4, displaying the qqplot of residual. Since the dot follow a string line so we can conclude that the residual is normal distribution which verify the model assumption.

Figure 4: QQplot