

36-220 Lab #10

Linear Models

Please write your name below, tear off this front page and give it to a teaching assistant as you leave the lab. It will be a record of your participation in the lab. Please remember to include whether you are in Section A, B or C. Keep the rest of your lab write-up as a reference for doing homework and studying for exams.

Name:

Section:

- The symbol ♣ at the beginning of a question means that, after you answer that question, you should raise your hand and have either the TA or lab assistant review your answer. Once they have reviewed your work they will place a check in the appropriate space in the table below. The purpose of this check is to be sure you have answered the question correctly.
- **You should try to complete as much of the lab exercise as possible. We understand that students work at different paces and have tried to structure the exercise so that it can be completed in the allotted time. If you work systematically through the handout and still don't complete every question don't worry. The important thing is that you understand what you are doing. Nonetheless, you are encouraged to complete the lab on your own.**

Check-Problem ♣	Instructor's Initials
-----------------	-----------------------

Question 2	
------------	--

Question 8	
------------	--

36-220 Lab #10

Linear Models

1 Cost of Nuclear Power Plants

The Minitab worksheet Nuclear.MTW, on the course webpage, contains data from a study of the cost of constructing nuclear power plants. For each of 32 “light water” nuclear power plants, the researchers measured the cost (in \$100,000, adjusted for inflation), the net capacity (in MegaWatts) and when it was constructed (in years after 1900).

We want to fit the following simple linear regression model:

$$\text{Net Capacity} = \beta_0 + \beta_1(\text{Cost}) + \text{random error}$$

and then check to see whether this is a reasonable model.

1. Construct a scatter plot of **Net Capacity** versus **Cost**.

Question #1: Describe the scatter plot. Does there appear to be a linear relationship between these two variables?

2. Here's how to do a simple regression in Minitab. Select the **Stat > Regression > Regression...** option. Choose **Net Capacity** as the **Response** variable and **Cost** as the **Predictor**. Press the button marked **Storage...** and check the boxes marked **Residuals** and **Fits**. Click **OK**. Click **OK** to perform the analysis.

♣ **Question #2:** What are the least squares estimates of β_0 and β_1 ? (These are usually denoted $\hat{\beta}_0$ and $\hat{\beta}_1$).

3. In your worksheet there should now be two new columns named **RESI1** and **FITS1**. The name **FITS1** refers to the **fitted values** from the model, i.e. for the first row the fitted value is

$$\hat{Y}_1 = \hat{\beta}_0 + \hat{\beta}_1(345.39) = 764.37$$

(as is customary, I am letting \hat{Y} denote the fitted values).

Question #3: What do you expect to see in a plot of fitted values versus **Cost**?

Question #4: Check your answer to the previous question by making the plot.

4. The column **RESI1** stores the residuals, which equal the difference between the actual observed **Net Capacity** and the fitted value. Residuals act as “approximations” to the random error term in the model, and are thus useful for testing assumptions about the model error. For example, in some cases we assume that the random error is normally distributed. Create a probability plot of the residuals to see if there is reason to conclude that our model error is not normal. Comment on what you see.

5. We also assume that the random error in our model is not related with **Cost**. In other words, our model is that **Net Capacity** equals a linear

function of **Cost** plus some random component that is independent of **Cost**. It is standard to create a scatter plot of residuals versus fitted values; do that now. A pattern in this plot would suggest that there is relationship between the random error and **Cost** or that the model is not a good fit to this data.

Question #5: Is there any noticeable pattern?

6. Create a plot of residuals versus **Date**

Question #6: Does there seem to be a relationship between these two? What does that suggest about the variable **Date**?

2 Plot of Residuals versus Fitted Values

The point about inspecting the plot of residuals versus fitted values bears reinforcing. Figure 1, is a scatter plot comparing two variables X and Y . Also on the plot is the best fitting line to the scatter plot.

Question #7: Judging only from the scatter plot, does the linear fit seem appropriate?

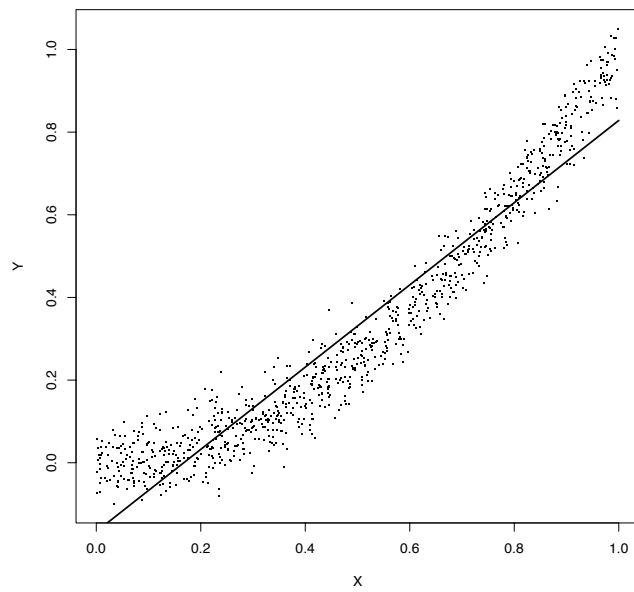


Figure 1: Scatter plot for Section 3.

♣ **Question #8:** Sketch how the plot of residuals versus fitted values will appear. Comment on how this plot would allow you to see failings of the linear fit.