Investigation: *How faithful is Old Faithful?*

Introduction: Now you know about geysers and how they work. In this investigation, we'll look at eruption data from Old Faithful geyser, in Yellowstone, to see how predictable its eruptions actually are. Once again, this investigation is intended to build your skills and experience analyzing and interpreting real geological data.

Part I: Make an initial prediction.
I'll give each of you a smallish set of data, consisting of several measurements of the time interval (to the nearest minute) between eruptions. In this case, the time interval between eruptions is defined as the time from the beginning of one eruption to the beginning of the next eruption.

If you arrived at Old Faithful just as the last eruption in your data set was ending, how long should you expect to wait for the next eruption? Make an estimate based on your data, and explain your prediction.

Also, make a graph of your data. (Make whatever kind of graph seems appropriate to represent the data. We will actually learn the most from these data if students come up with several different types of graph, so do not feel your graph should be the same kind as your neighbors are constructing.)

Part II: Making a preliminary, informal estimate of predictability.
Form groups of four students. Trade data sets with each other, and repeat part 1, for **each** new data set. That is, use **each** individual data set to make a new estimate and a new graph. (When you are done with this, you will have 4 estimates and 4 corresponding graphs.) You should each make your own estimates and your own graphs. You may each have a different method; that's okay. Use whatever method you used the first time.

Does Old Faithful seem faithful? That is, how consistent are your four estimates?

Part III: Comparing methodology.
In your group of four, discuss **each data set** in turn. What are your (4) estimates of wait times between eruptions? What do your four graphs look like? ***Do not be dismayed if the four of you each have a different estimate, or a different graphing method.*** On the contrary – discuss your different methodologies, and try to determine the advantages and disadvantages of each method. There is not one "right way" to analyze data. In fact, it is often appropriate to subject data to multiple methods of analysis; that's how we get the most information out of the data. If the four of you all used the same methods, brainstorm what other methods you could have used. Come up with at least three different methods of predicting wait time and three different methods of graphing the data. I will ask your group to present one of these to the class. You should be ready to present any of them, in case an earlier group presents your "favorite" method.

Part IV: Class discussion.
I will ask each group in turn to present one graph of one data set. Choose a graph and a data set that you think shows something interesting about the data, and tell us what it shows. At the end of the discussion, I will ask you to summarize what we (collectively) know about the data.

Part V: Further analysis.
What kind of data would you need in order to develop a better method of prediction of eruption intervals? You may discuss this either in your small group or as a whole class. I will provide additional data, when you present me with a request and a rationale for that request. It may help at this stage to think about the process of geyser eruption, and what physical factors could affect the length of time between eruptions. (You must ask for a specific type of data; I will not provide any additional data if you request "whatever else you have.") Then we will proceed to the geology computer lab to graph the additional data.

Investigation: *How faithful is Old Faithful?*

Part V: Further analysis, continued.
Launch Minitab (open the P: drive, open Class Programs, open "Minitab 13 for Windows," open Minitab). Open the data file (***ask me for the location!!***) as a worksheet. Make a graph of the time interval between eruptions as a function of the previous eruption duration. To do this, choose "plot" under the "Graph" menu on the toolbar. (Make time interval between eruptions be on the vertical axis.) Then have Minitab create a best-fit line using the "least squares" method. For this, go to the "Stat" menu, choose "regression," and choose "fitted line plot." Wicked cool, isn't it? Print out a copy of your graph, including the best-fit line.

After an eruption of four minutes duration, how long would you expect to wait for the beginning of the next eruption, ***on average***?

Next, let's think about whether Old Faithful really lives up to its name. The live webcam of Old Faithful includes a prediction of the next eruption time, $\pm$ 10 minutes. Draw two lines on your graph, parallel to the least squares best-fit line, above and below it by ten minutes. (That is, if you draw a vertical line segment connecting your upper line and the best-fit line, the line segment should measure ten minutes along the vertical axis of your graph. I can help with this.) What percentage of the data points on your graph fall into this "$\pm$ 10 minutes" error zone? This is a measure of the accuracy or reliability of the webcam's prediction!

Part VI: Interpretation
Now it is time to answer the question, "What does it mean?" You have observed a statistical relationship between two eruption-related variables (I hope). How is that relationship related to the physical processes that govern geyser eruptions? In other words, ***why*** does the eruption interval depend on eruption duration?

Writing your report:
As with earlier reports, you may write this report either individually or as a group. The usual requirements apply.

Your report should be a written summary of ***the investigation question***, ***how*** you went about answering it, and ***your answer*** to the question. **For the sake of brevity, please focus on what you learned from parts 4, 5, and 6 of this exercise.** Be sure to include your raw data and graphs, along with any other diagrams that would help your reader. Also, be sure to answer the questions in parts 5 & 6, and explain your answers. The quality of answers to these questions is what will distinguish decent reports from good or outstanding reports. For a reminder of my grading criteria, refer to the general grading rubric.

Finally, the most important thing: keep the written portion of your report (that is, not counting data and figures) to **less than 2 type-written pages**. This will require you to be concise, and will allow me to read all of the reports (from my 3 sections of GL 110) and give you feedback on them within a week. **This report is due one week from today.**