

Day 1: Effective Dose

INS 099
Data Analysis

Summer 2007

Data Analysis

Your spreadsheet lists data for a sample of 100 test subjects who were administered two drugs: Progen and Exineprefine. Most of the columns describe traits of each subject (age, gender, weight), but two of them describe the results of an *effective dose* experiment. In each experiment, the subject was given increasing amounts of a drug. As the dosage was increased over the course of a month, a doctor tested the subject to see if the drug had any effect (*e.g.*, reduced blood cholesterol, killed malaria parasites, *etc.*). The numbers in the last two columns list the *doses required to have an effect on the subject*. For each patient, this number is called the *effective dose*.

After a quick look at the data, you will realize that the effective dose can vary significantly. A pharmacologist uses data like these (in addition to many other types) in order to recommend doses for different types of patients. In the table below, you will see data for four patients. Please use the data to answer the following questions and complete the following tasks:

1. What is the correct dose of Progen for each patient?
2. For physicians who doesn't have access to these data, write some instructions that help them decide on the proper dose for any patient, given the patient's age, gender, and weight.

| Name | Age (years) | Gender | Mass (kg) | Progen Dose (mg) |
|-----------|-------------|--------|-----------|------------------|
| Imelda T. | 26 | Female | 65.7 | |
| Linda A. | 45 | Female | 72.1 | |
| Michel G. | 37 | Male | 83.0 | |
| Kelvin C. | 28 | Male | 61.1 | |

Exercises

1. As you work through the tutorial, figure out from the instructions what conventions are expected in scientific plots. When you are finished with the tutorial, please summarize these conventions in a list, discuss them with your group, and put together a combined list. **Please follow these conventions in all homework assignments and exams.**
2. Prepare a plot of Progen effective dose versus mass with two series: (a) for all subjects under 30-years-old, and (b) for all subjects 30-years-old and older.

Homework

(Due at the beginning of class on Day 2.)

1. For physicians who doesn't have access to these data, write some instructions to help them determine the proper dose for an arbitrary patient, given the patient's age, gender, and weight. Use text, equations, and graphs to help the doctor understand the relationships between these variables.
2. What is the correct dose of Exineprefine for each patient below?

| Name | Age (years) | Gender | Weight (lbs) | Exineprefine |
|------------|-------------|--------|--------------|--------------|
| Ellen C. | 27 | Female | 124 | |
| Shu-Wei L. | 52 | Male | 158 | |
| Miriam R. | 41 | Female | 98 | |
| Dani G. | 36 | Male | 145 | |

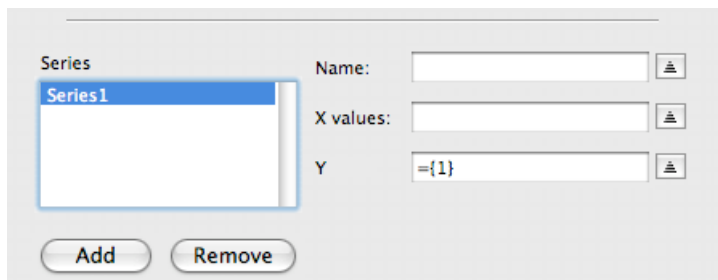
Tutorial

Graphing is an indispensable tool because plotting data allows you — and other people — to see relationships between two variables very easily and quickly. For this reason, you can use plots to:

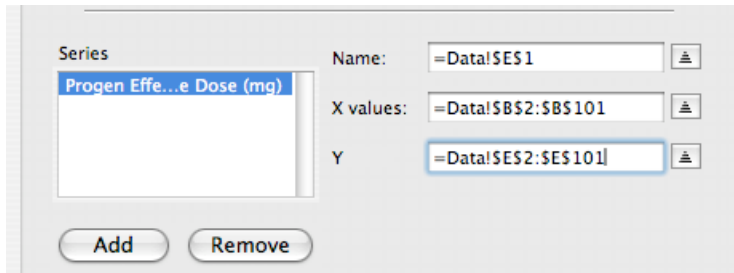
1. Decide whether two variables are related
2. Guess what kind of relationship connects two variables
3. Convince collaborators, bosses, policymakers, etc., of the relationship between variables
4. Convey the nature of a relationship to people who live in mortal fear of equations

For this tutorial, we will use Microsoft Excel to plot data. Although Excel makes it quite easy to set up a plot using data in a spreadsheet, it also tends to produce unprofessional graphs. This tutorial will help you understand how Excel generates plots, but also how to generate *scientifically acceptable* plots. As you work through the tutorial, figure out from the instructions what conventions are expected in plots. **When you are finished with the tutorial, please summarize these conventions in a list, discuss them with your group, and put together a combined list. I expect you to follow these conventions in all homework assignments and exams.**

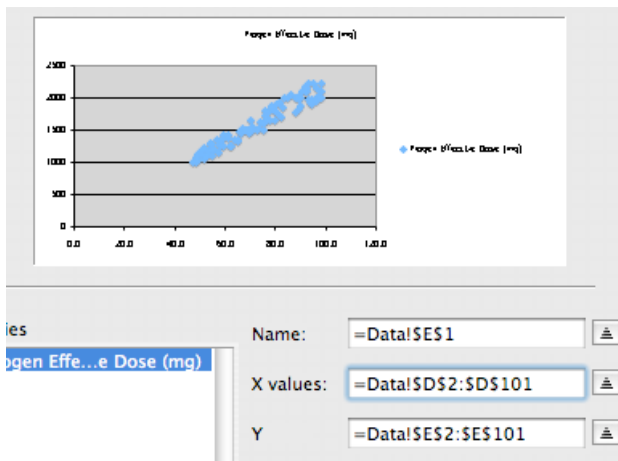
1. Make sure you have the spreadsheet open and the data is displayed. The examples in the figures start with unsorted data.
2. Select the menu command “Insert > Chart...” to bring up the “Chart Wizard” dialog.
3. The type of graph we are making is a plot, which Excel calls “XY (Scatter).” Select this type of graph. For the sub-type, select “Scatter” (*i.e.*, no connecting lines) and click “Next.”
4. Click the “Series” tab at the top of dialog. If Excel has started plotting stuff right away, click “Remove” a few times to get rid of all series. To define a new series, click “Add,” and you will see three boxes appear, each one with a little triangle next to it.



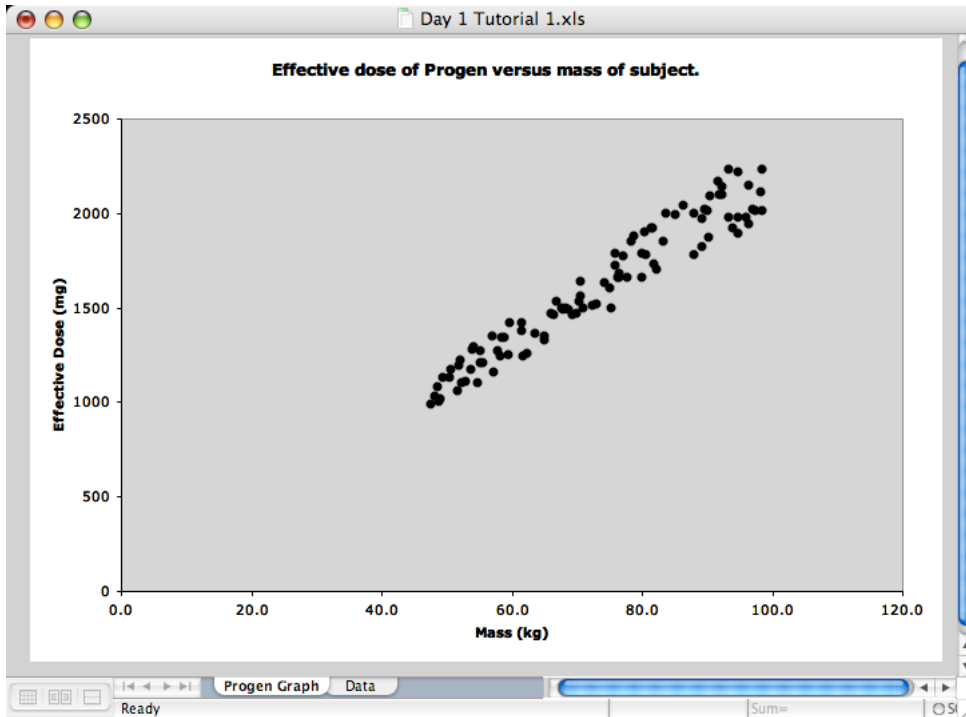
5. A *series* in Excel is a list of ordered pairs. Each series has a name, a range of x values, and a range of y values. We define each of these things using cells in the spreadsheet.
6. Let's plot the relationship between Progen Dose and Age. Click in the “Name” box, and then click on the cell E1. The box will now contain the text “=Data!\$E\$1”. This line tells Excel to look in the box E1 of the worksheet “Data” to define the name of the series. (The dollar signs have a special meaning, but for now you can just ignore them.)
7. Click on the “X values:” box. Using your mouse cursor, select the cells B2 through B101. In the box, you will see “=Data!\$B\$2:\$B\$101”. Then do the same thing for the y values, using cells E2–E101. (Delete the text “={1}” first, because otherwise Excel might give you an error due to a bug.)



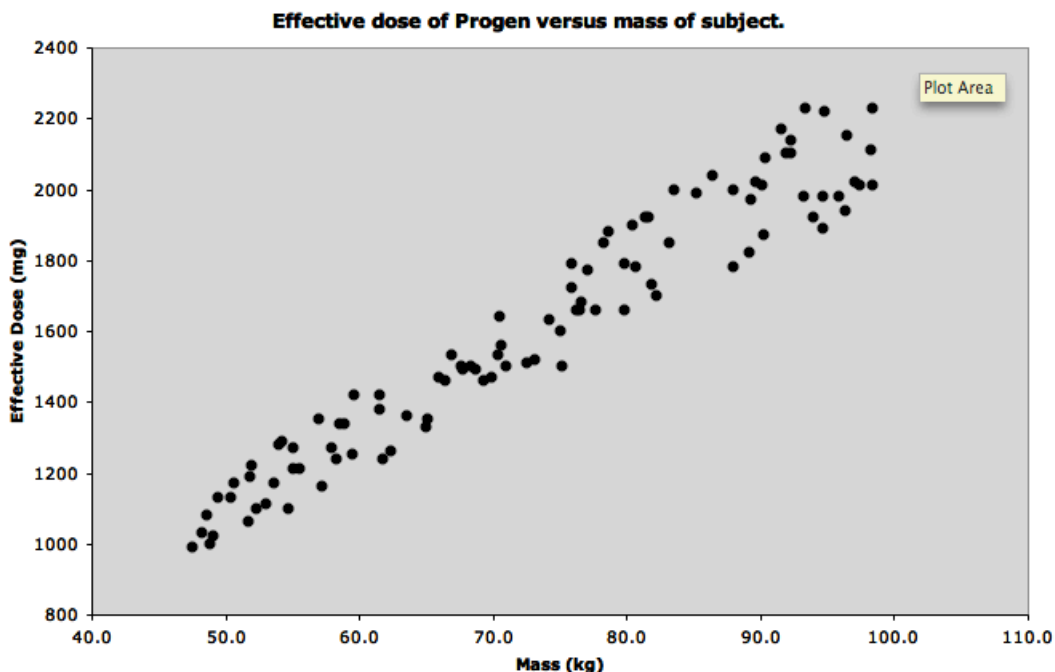
8. Looking at the three entries above, you can see that we are using the cell E1 to define the series' name, cells B2–B101 to define the x values, and cells E2–E101 to define the y values. Excel automatically pairs each x value with the corresponding y value in the same row (e.g., B2 with E2, B3 with E3, etc.).
9. Let's change our minds and plot the relationship between Progen dose and *Mass* instead of Age. You can redefine the series easily by specifying new cells in the "X values:" box. Select all of the text in this box and delete it. Then select the appropriate cells, and the miniature graph now looks like the figure below. Click "Next."



10. How will anyone know what this graph is showing? We need to label this graph unambiguously, so other people will know what the x values and y values are. Graphs should have either a title or a caption to briefly explain its meaning. Fill out the fields under "Titles" as follows:
 - **Chart title:** "Effective dose of Progen versus mass of subject"
 - **Value (X) axis:** "Mass (kg)"
 - **Value (Y) axis:** "Effective dose (mg)"
11. Depending on your computer's default preferences, there might be some gridlines on the graph. Let's leave them out for now. Under "Gridlines," uncheck all gridline options.
12. Legends are pointless if there's only one series of data. Uncheck the option "Show legend." You only use a legend if you are plotting multiple series of data.
13. Insert the graph as a new sheet named "Progen Graph" and click "Finish":



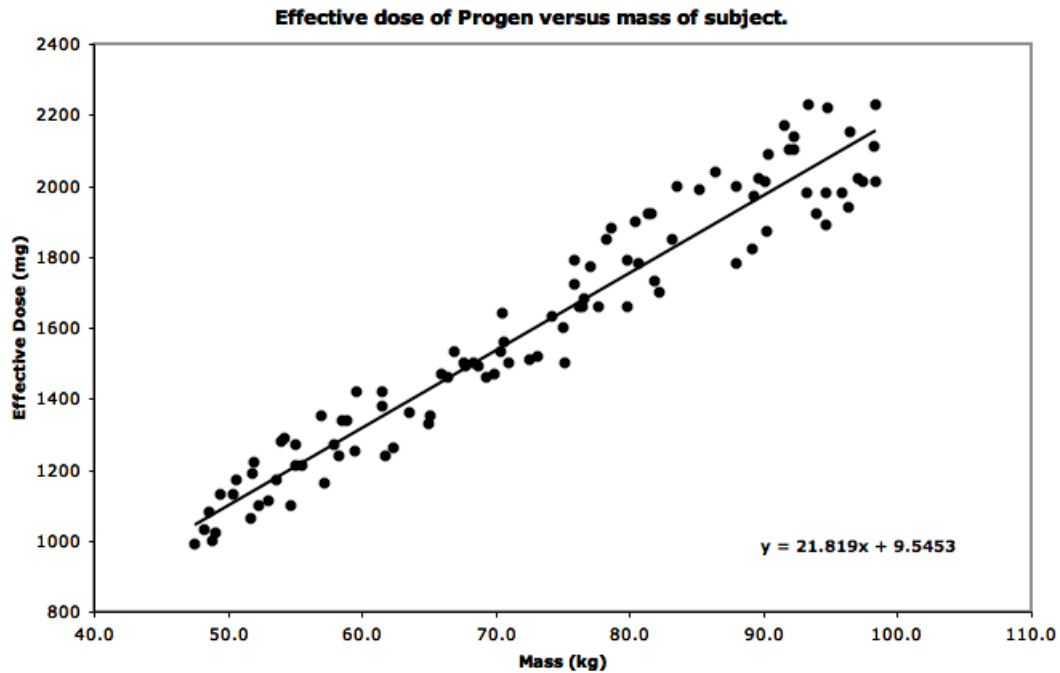
14. The mass data have no values from 0 to 40 kg. As a result, there is “dead space” on the left side of the plot. Let’s adjust the x axis scale to exclude these values. Double click on the x axis, and a dialog named “Format Axis” will appear. Under the “Scale” tab, set the “Minimum” to 40. Using a similar procedure, format the y axis to exclude values from 0 to 800 mg.



15. The graph shows that effective dose increases with mass, so we suspect the relationship is linear. To fit a linear model to these data, select the menu command “Chart > Add Trendline...” and specify a linear model. Since the Progen data are the only series on the graph, the series is selected already.

16. Linear fits are useless unless they tell you what the best parameters are. Click on the “Options” tab, select “Display equation on chart,” and click “OK.” Your chart now has a best-fit line and an equation. Write down the parameters in this equation.

(In many scientific communities, leaving the equation in the plot area is considered unprofessional. Instead, you would list the best-fit parameters in an annotation, in a caption, in a table, or in separate text.)



17. Please summarize the conventions expected of scientific graphs. In other words, what features should a good plot have?