2. LECTURE 2

Objectives

- ☐ I understand the distinction between independent variable(s) and the corresponding dependent variable as well as why that distinction was chosen for the situation.
- ☐ I can define a linear equation, recognize when two measurements may have a linear relationship, and (with a calculator) I can find the best fit line to model that relationship.

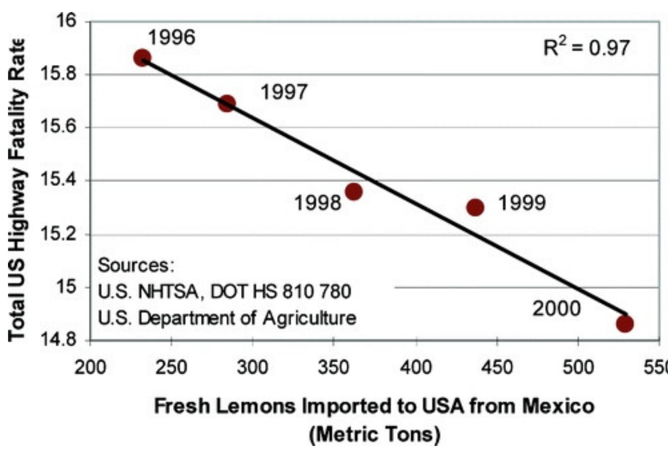
Last time, we defined how we will approach mathematics—namely, as a subject that attempts to describe and understand relationships between two or more measurements. We then discussed the two kinds of measurements: stocks and flows. In particular, flows can come in two forms. They can be inflows, rates that grow a stock, or outflows, rates that shrink a stock.

We describe and understand the relationships between measurements using functions. When we write a function, however, we are implying that some measurement (or set of measurements) is influencing one in particular. The measurement being influenced is called dependent. All other variables are called independent.

We now pick up where we left off by pointing out that although a function will indicate a direction of influence, that influence is not always *causal*.

Sometimes a causal relationship cannot be determined either because we do not know enough or it simply is not there. For example, there is a correlation between the average temperature of a day and the number of violent incidents a city experiences. Since we know violence between people will not change the temperature of a city, we treat violence as the dependent variable and temperature as the independent variable. While the temperature *influences* the number of violent incidents, we cannot say it *causes* them because we do not know a mechanism for it. It is possible that temperature causes some third factor (like encouraging people to go outside) which has a causal relationship with violence.

The image below is a more humorous correlation where causation is unlikely.



When no causal relationship can be justified, we have an obligation to write something along the lines of this: *While our function describes a relationship between two or more measurements,* *this relationship may not be causal or even meaningful.* If such a problem were to arise in the homework or exam, I expect you to write a statement to this effect.

2.1. **Determining the Dependent Variable.** If we don't know a causal mechanism, how can we determine which variable is the dependent variable? The answer is: look at the context. In many cases, there is one variable which does not have the capacity to influence the others. Below are some examples.

Example 2.1: Suppose you want to relate the amount of time a student spends drinking during the semester with their GPA at the end of the semester. Their GPA must be the dependent variable because it comes *after* the drinking was done. Since no one can go back in time, a student's GPA can't influence their amount of drinking.

Example 2.2: There exists a correlation between a person's age and their likelihood of dying by suicide. Suicide risk can, in no way, influence age. Therefore, your age must influence suicide risk.

In some cases, however, no variable is clearly the dependent variable. For example, homeless rates and crime are often correlated, but neither one is clearly dependent on the other. In those cases, your choice of dependent variable will depend on the question you are asking. For example, if you ask, "How will reducing homelessness impact crime rates?" then you will treat crime rates as dependent on homelessness.

2.2. **Linear Relationship.** If two measurements have a **linear relationship**, that means data plot (the graph of the data) is described by approximated by line. A somewhat imprecise way of determining a linear relationship is to graph the data. If the graph appears to cluster in the pattern of a line, then the measurements may have a linear relationship. We now go over a historical example to explain how to graph these points with a calculator.

During summer evenings, it is common to hear the crickets chirping. In the late 1800s, physicist and inventor Amos Dolbear conjectured that there was a relationship between the frequency of cricket chirps (the number of chirps per second) and the ambient temperature. He collected data like the chart below.

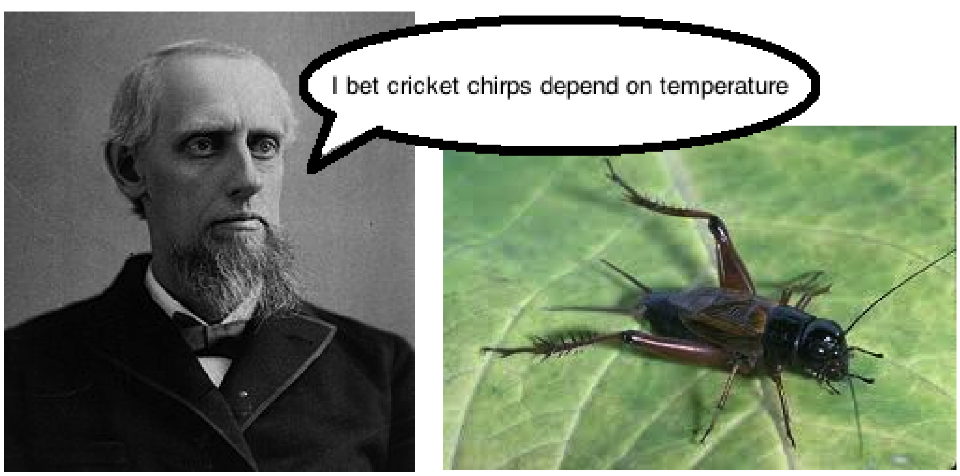


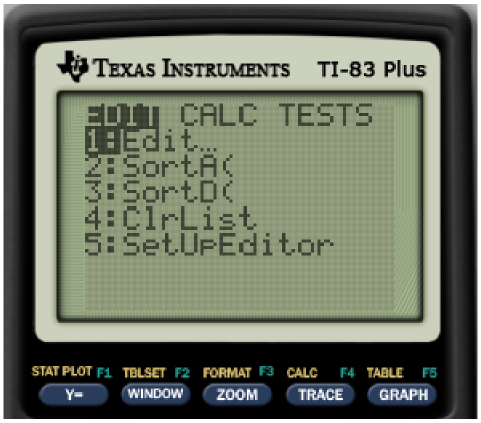
Table 2.3: Temperature versus Chirping Frequency

Temperature	Chirps per Second
20.0	88.6
16.0	71.6
19.8	93.3
18.4	84.3
17.1	80.6
15.5	75.2
14.7	69.7
15.7	71.6
15.4	69.4
16.3	83.3
15.0	79.6
17.2	82.6
16.0	80.6
17.0	83.5
14.4	76.3

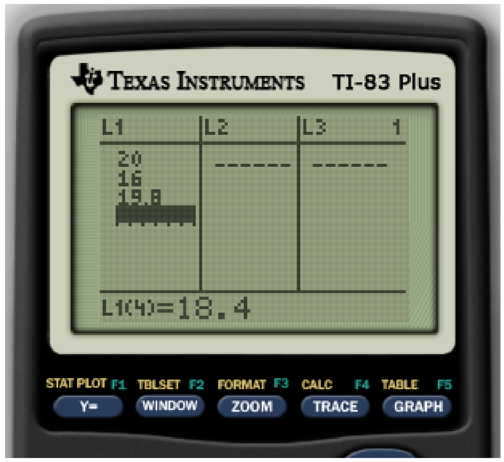
First, we need to determine which is the dependent variable. It is more sensible that the chirps should depend on temperature rather than temperature depending on the chirps (chirps don't generate a significant amount of heat). Therefore,

independent variable = time
dependent variable = chirps per second

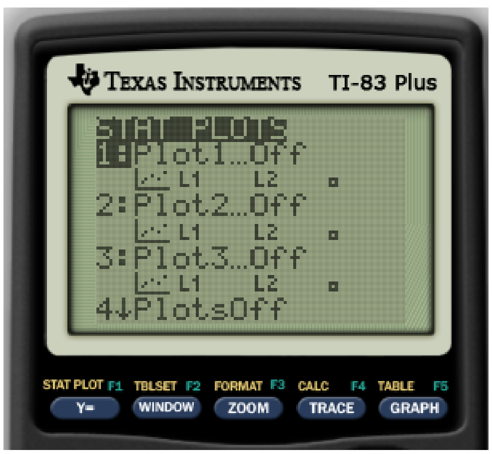
To input the above data in a TI-83 or TI-84, click `STAT` and select the `EDIT` menu at the top. Then select the first entry `1:Edit...` by pressing `ENTER`.



Now use the arrows to navigate the through the columns. In the `L1` column, type in the entries of the independent variable. In the `L2` column, type in the entries of the dependent variable. Once you're done, you can hit `2nd` then `MODE` to `QUIT`.

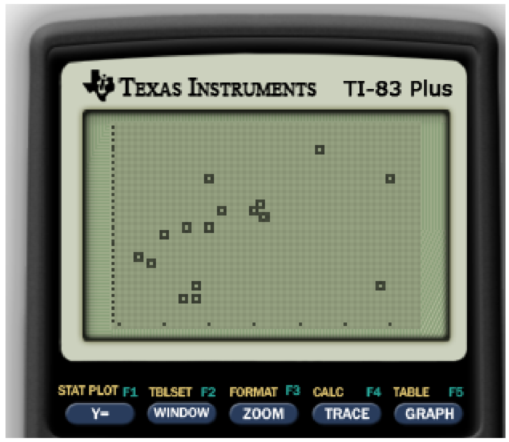


Now we would like to graph the data. To this, hit 2nd then Y= to access the STAT PLOT menu. Select the first menu and highlight the following settings according to the picture below. Once you're done, you can hit 2nd then MODE to QUIT.



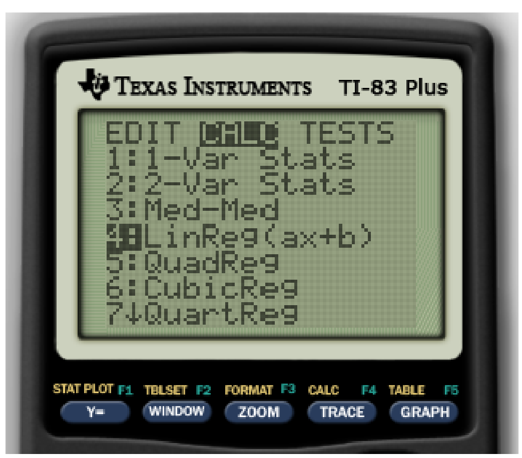
Before you hit GRAPH, I suggest you hit ZOOM and select 9: ZoomStat to adjust.

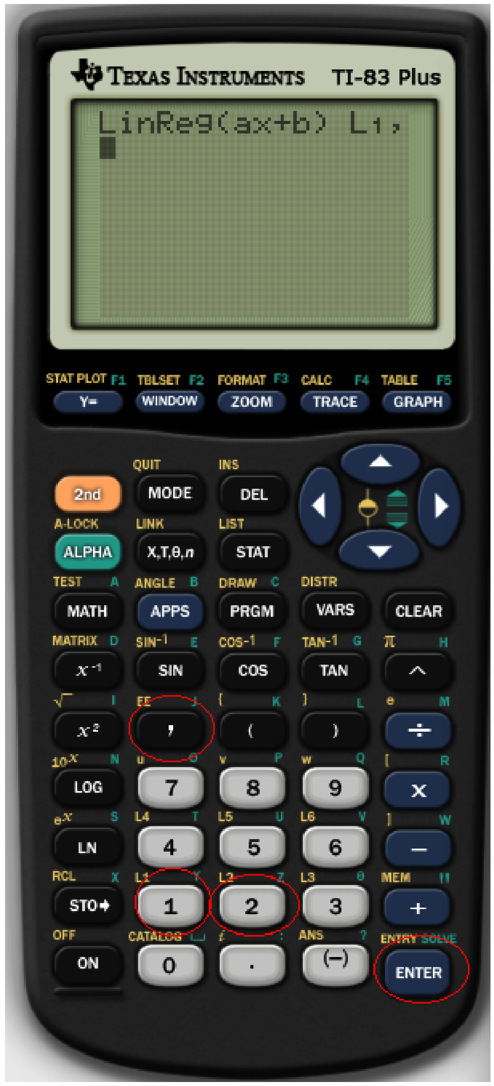
Below is the resulting scatter plot once you hit ENTER.



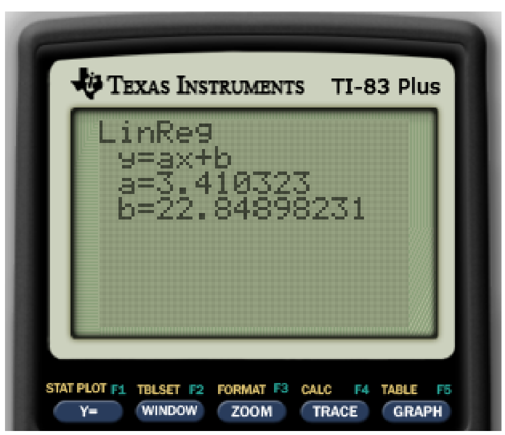
The picture appears to have a linear relationship. If we suspect a linear relationship, then the next step is to define the **best fit line** to describe the relationship. For this class, we simply covered how to use a calculator to find this line. A little later in the course, we will actually go over the mathematical process of finding the line.

To find a line, press STAT and select the CALC menu at the top. Then select the first entry 4: LinReg(ax+b). Once you do, it will display LinReg(ax+b) with the cursor blinking. Press 2nd then 1 to print L_1. Then press , and press 2nd then 2 to print L_2.



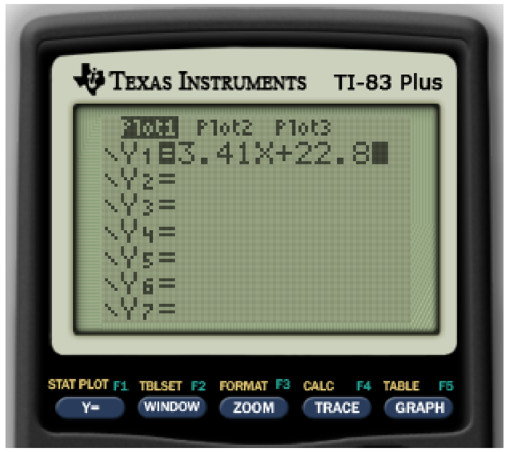


Hit ENTER. You should then get a print out of the line that best fits the data.

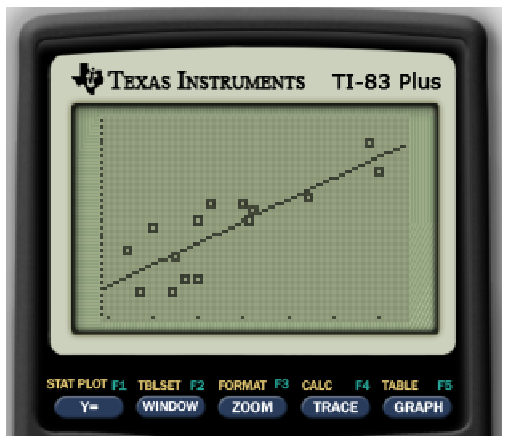


Our data went only up to the 10th decimal place, so we can treat this function as $y = 3.41x + 22.8$, where x is the temperature and y is chirps per second.

To compare the data against the line, we can press Y= and type the function of the line. To type "x", press the X,T,Θ,n key.



Press GRAPH and see how it compares.



2.2.1. *Interpretation of our Findings.* If we look at the picture, we see a line that appears to explain a significant relationship between temperature and chirping. If we take the derivative of the equation, we get

$$\frac{dy}{dx} = 3.41.$$

Remember from Math 110 (or your previous calculus course) that a derivative explains the amount y changes in relation to x at a particular point. Getting the derivative 3.41 tells us that for every 1 degree change in temperature (x), we should see an increase of 3.41 chirps per second (y).

The next question to ask is this: Is this relationship consistent with the data? That is, should we believe a real-life relationship exists between these two measurements. In general, this is a difficult question to answer. For now, we will determine this visually by comparing if the data appear to follow the line well. When we graphed the line with the data in the calculator, we could see a linear trend (if we ignored the point in the bottom right corner). The more the data agrees with the line we see, the more we should believe this relationship exists in real life. The less it agrees, the less we should believe these two measurements share a relationship.

We will make this ideal more concrete later on.

Summary of Ideas: Lecture 2

- Functions imply a causal relationship. Sometimes that is true and sometimes it is not. When we are not sure or when we believe it is not true, we have an obligation to point that out in our work.
- When modeling a relationship between two measurements, we treat one variable as depending on the other. We determine this based on the context or the question we ask.
- We can plot data with the help of a calculator. If that data appears to grow in the shape of a line, we suspect two the measurements have a linear relationship.
- We can find that line with the help of a calculator. We can then compare it with the actual data.
- We can always find a line regardless of the data. We should only suspect a relationship exists between two measurements if the data appears to follow the line well.