David Youngberg

Acct/Busa/Econ 222—Bethany College

**Lecture 01: Of Data and Displays**

1. What is econometrics?
   1. Econometrics is the application of statistics to economic data. Yes, the techniques that we will use in this class can be applied to other kinds of data. But due to the unique constraints in the social sciences, some of the things we will explore this semester were developed for the social sciences.
   2. What are these unique constraints? We can see them in several different jokes about economists.
      1. Put two economists in a room and you’ll get three opinions.
      2. If you lay all economists end-to-end they would not reach a conclusion.
      3. Harry Truman once wanted a one-handed economist because all of his kept saying “on the other hand…”
   3. Why is there so much disagreement among economists? Even if you consider the things economists agree on (and there is a lot of that) it pales in comparison to what physicists, chemists, and biologists agree on. Why?
   4. Consider one of economics most important questions: How do turn poor countries into wealthy ones?
      1. Economists have lots of ideas, but we don’t know what will actually work. For example, does giving the country’s government a bunch of money help?
      2. Ideally the World Bank or IMF would randomly select half of the poor countries and then give that half some large amount of money (adjusted by population). Then we can look at the results.
      3. But that’s not an option and not just because each country is so different. There are ethical and legal constraints. The struggling countries that didn’t get anything would wonder why they are left out. And, by chance, some of that money would go to countries that we know are corrupt. Even if we learn a lot, it would be a short-term disaster.
   5. Economists (and other social scientists) are prisoners of reality. Any controlled experiments that are done (and they happen, though they are a small part of the discipline) must first be approved by an ethics board. If you were confused why running a controlled experiment is hard in the social sciences, consider the Stanford Prison Experiment.
2. The Stanford Prison Experiment
   1. Back in the early 1970s, psychologist Phillip Zimbardo became curious how people’s behavior changes when they become a prisoner or a in charge of prisoners. So he set up an experiment with 24 young men. By the flip of a coin he divided them into groups: prisoners and guards with 12 of each.
   2. Zimbardo also set up a makeshift prison in the psychology department. There were cells complete with bars on the doors. Prisoners had to sleep onsite. Guards worked in eight hour shifts and were allowed to leave the site after their shift. This would continue for two weeks.
   3. The guards were given near absolute control over the prisoners and were encouraged to punish the prisoners to maintain order. Prisoners were stripped naked, forced to do push-ups, withheld food, woken in the middle of the night, etc.
   4. After the first day, the prisoners rebelled, which was put down. Two days later, one prisoner started acting “crazy:” screaming, cursing, entering rages.
   5. As the experiment continued, guards started acting increasingly cruel. About one in three demonstrated sadistic tendencies as they demanded more and more from their prisoners. Physical punishment increased, harassment became arbitrary, prisoners were forced to clean toilet bowls with their bare hands, and other behavior led to the experiment being shut down after just six days. Stricter rules for ethics in experiments followed.
3. Data Descriptions
   1. When examining data, one of the first steps is to familiarize yourself with its statistics. Recall that the central tendency can be described in three different ways: mean, median, and mode.
      1. *Mean* (or the arithmetic mean) is the average. Sum all the values and divide by the number of observations.
         1. Problem: influenced by large outliers (e.g. income).
      2. *Median* is the middle value. Half of the observations are below and half are above (if an even number of observations, take the arithmetic mean of the two middle observations).
         1. Problem: not influenced by some important changes (e.g. the very poor lose half income).
      3. *Mode* is the most common value. It’s often used for discrete data with few alternatives
         1. Problem: Difficulty with continuous variables (e.g. income, though can make into range); May also mask important changes (e.g. many poor immigrants enter country).
4. Data Displays
   1. Another way is to visually see where it is, or display the data. There are several ways to display a variable.
   2. *Pie Chart*—a circular chart divided into sections, or wedges, describing a percent of total each group is. This is one of the most widely used charts out there but it’s not perfect (as I will show you).
      1. Advantage: it is widely used and easy to understand. Despite its disadvantages, it is nice to know how to make because so many people are used to it, they expect to see one.
   3. *Histogram*—a histogram divides data into groups and displays the number of observations per group
      1. Advantage: easily organizes lots of data, especially when there are many possible groups (e.g. income or other continuous variable)
   4. *Dot Plot*—like a histogram, a dot plot shows the number of observations, but doesn’t divide them into groups
      1. Advantage: don’t lose information through generalizing (e.g. in my grade distribution diagram from last lecture, you can’t tell how many B+’s or C-‘s there were since they are grouped under Bs and Cs, respectively)

|  |  |
| --- | --- |
| 0 | 9 |
| 1 | 0056 |
| 2 | 12 |

* 1. *Stem-and-Leaf Display*—a table displaying data with their “tens+” digit on the left and the last digit, in numerical order, listed on the right.
     1. Example of stem-and-leaf showing data values of 15, 10, 9, 22, 10, 16, 21
     2. Advantage: Gives us all the specific values of the data while showing a distribution, though better for smaller data sets
  2. *Box Plot*—a display which shows where quartiles of data are
     1. A quartile is a part of a data set with one-fourth of the total observations. The 1st quartile is a data value which indicates when, from the minimum to that value, are the first fourth of the observations are
     2. Note you can also divide the data into other segments such as in five equal parts (quintiles), ten equal parts (deciles), one hundred equal parts (percentiles), etc.
     3. The lines on either side of the box show the range between the maximum and 3rd quartile and between the minimum and 1st quartile
     4. The box is between the 1st and 3rd quartile with a line (the median, or 2nd quartile); the box is the *interquartile range*.
     5. The larger the distance between these points, the more disperse the observations. The shorter the distance, the more concentrated
     6. Advantage: Like the steam-and-leaf diagram, it illustrates dispersion but it is able to handle virtually any number of observations. All you need to make a box plot are five numbers: maximum, minimum, 1st quartile, 3rd quartile, and median (2nd quartile).

1. Labs
   1. On occasion, I don’t like how the lab is organized in the workbook or I want to add additional material. Use the lab section, below, to help you follow along during the lab for the course, on Wednesday.

**Lab Section**

1. Chapter 1
   1. Microsoft Excel is a spreadsheet program, or a program that organizes all its data in a table.
   2. The table is composed of rows (labeled by number) and columns (labeled by letter). A combination of a number and a letter is a cell.
      1. You can select a cell with your mouse; you can select within a cell by double-clicking.
      2. If you start typing after selecting a cell, it will replace what you’ve already typed with what you’re typing now.
      3. Selecting within a cell lets you edit what’s already there.
   3. After selecting a cell or typing within a cell you can move…
      1. Down by pressing Enter
      2. Up by pressing Shift + Enter
      3. Right by pressing Tab
      4. Left by pressing Shift + Tab
      5. You can also use the arrow keys.
   4. Excel also lets you do math, using standard notation (\* for multiplication, + for addition, ^ for “to the power”.)
      1. What makes this really powerful is that you can reference other cells for values.
      2. Select a cell and press “=” and then use your mouse to select other cells separated by mathematical operators. You’ll notice the code for the cells appears: A1, B5, etc. You can also type these codes in if you don’t want to use your mouse.
      3. Press Enter to see the result. The result will update if you change the referenced cells.
   5. Copy and pasting references
      1. If you select a cell with references to other cells and you copy/paste (*not* cut/paste) the cell, you’ll notice the references will change based on where you pasted the cell. For example, if you pasted one row down from the original, the new references will also be one row down.
      2. You can “lock” a reference by selecting within a cell and placing your cursor within or adjacent to the reference (say, between the letter and number) and pressing F4. The $ that appear around the reference indicate it is locked; if you copy and paste the cell, it won’t update the locked reference.
   6. Commands
      1. Excel has hundreds of equations built into it that we can access with the “=” sign. For example, instead of typing A1+A2+A3+A4+A5, you can just type =SUM(A1:A5) or highlight A1 to A5 after typing =SUM(.
      2. Like other references, the values will update if you copy and paste the cell (but not the actual text of the cell, just the cell as a whole).
   7. Printing your homework
      1. When you print your homework, make sure all the columns you’re using fits within the width of the page. It makes it much easier to read.
      2. First, select all the cells you want printed. Then go to “Page Layout” tab and press “Print Area” and then “Set Print Area.” A dotted line will appear around your selection. If there are any dotted lines within that selection that means you’ll be spilling onto another page. If it’s a vertical dotted line, you need to make your columns fit.
      3. One way to fix this is to shrink the columns. By double-clicking between columns at the very top, where they are lettered. This will automatically adjust their width to exactly fit what you have in the columns. You can also manually drag by click and holding.
      4. Another way it so increase the margins. A few icons left of “Print Area” you’ll see “Margins.” Click it to change margins, perhaps by selecting “Narrow” margins.
      5. A third way is to switch from “Portrait” to “Landscape” in “Orientation,” the icon to the right of “Margins.”
   8. Helpful shortcuts (note that while the letter is capitalized, you do not need to press and hold “Shift” as well; it is capitalized because it is capitalized on the keyboard).
      1. CTRL + S saves.
      2. CTRL + Z undos what you just changed in the document.
      3. CTRL + X cuts what you’ve highlighted.
      4. CTRL + C copies what you’ve highlighted.
      5. CTRL + V pastes what you’ve cut or copied.
      6. CTRL + B turns **bold** on or off.
      7. CTRL + I turns *italics* on or off.
      8. CTRL + U turns underline on or off.
      9. CTRL + F activates the “find” window.
      10. CTRL + P brings up the “print” window.
2. Chapter 2
3. Homework
   1. Chapter 2: 1-4
   2. For a fifth question, do the following:
      1. Using the values on how your grade is determined for this course (described in the syllabus), construct a pie chart. When indicated in the syllabus, use totals rather than individual values (one slice of the pie chart will be all homeworks; another will be all exams).
      2. Make sure to label each slice of the chart. Include a corresponding percent for each slice.