

Lab 7

The Effectiveness of Captopril

In the United States, pharmaceutical manufacturers go through a very rigorous process in order to get their drugs approved for sale. This process is designed to determine whether or not the drug works. There are a variety of factors that make this more difficult than it might seem at first blush. One factor is that different people have different reactions to the same drug. So it's not true that a drug "works" or doesn't work. In truth, it will have a different effect on different people and therefore, drug manufacturers have to convince the government that it works "on average," loosely speaking.

In this lab we'll examine a study designed to examine the effectiveness of the drug Captopril to lower blood pressure. We will focus only on Captopril's effect on systolic blood pressure.

The study, reported in the British Medicine Journal in 1979, examined 15 patients. Each patient had his or her blood pressure measured, was given the drug, and then had their blood pressure taken again several minutes later.

In earlier labs, we used data that were already *Stata* friendly. We recognize those files since their filenames end in .dta. What if we want to use our own data? In this lab, we will enter the data by hand.

A Brief Description of the Captopril Experiment

In order to test the effectiveness of the drug Captopril as a blood pressure treatment, fifteen patients were given equal doses of the drug. For each patient, the first variable represents the diastolic blood pressure before Captopril was administered, while the second variable is the diastolic blood pressure after the patient took Captopril.

before	130	122	124	104	112	101	121	124	115	102	98
after	125	121	121	106	101	85	98	105	103	98	90

before	119	106	107	100
after	98	110	103	82

To begin, we type

```
. edit
```

in the Command window and press enter. At this point, *Stata* gives the variables the generic names “var1” and “var2.” Type the first value, 130, and hit “return” to move to the next “var1” value. To move to the “var2” variable, click in the second column.

When you are done, hit the “Preserve” button on the editor window, and then close the editor by clicking the red button in the upper left corner.

To rename the variables, type

```
. rename var1 before  
. rename var2 after
```

*Question 1: Using whatever tools you think appropriate (e.g. the **histogram** command, or the **summarize** command, or whatever else you want to try), describe the distribution of blood pressures of the sample before receiving the drug.*

Question 2: Without looking, do you think the spread of the distribution of

blood pressures after taking the drug will be greater than, less than, or about the same, as it was before the drug was given? Explain.

Question 3: Again without looking, describe what you think the distribution of blood pressures will look like after taking the drug.

Question 4: Use the graph command to see what the distribution is for the 15 blood pressures after taking Captopril. How well did your expectations meet reality? If your predictions were off, can you explain why?

Now let's examine numerical summaries:

```
. summarize before after, detail
```

Question 5: According to these summaries, how do the distributions compare? What, if anything, does this say about the effectiveness of Captopril?

What we really need to do is examine both samples simultaneously. One tool for this is the boxplot:

```
. graph box before after
```

Question 6: What does this boxplot tell you that you didn't already know from our previous plots and summaries?

Question 7: It is apparent that the before group is different from the after group. Do you think all of the participants changed by the same amount? If so, explain. If not, which patients do you think changed the most? Which

changed the least? Do you think it's possible that anyone's blood pressure increased?

Can you answer these questions with any of the graphs or summaries we've seen so far? The answer is no. So far, we've looked at summaries of the blood pressures before taking the drug separately from the blood pressures after. Although we know the group changed, we have no way of knowing who in the group changed, and in which fashion.

But, of course, these questions can be answered with a little more effort. These data are what we called "paired." This means that all of the values in one variable are NOT independent of those in the second. To be more precise, every individual who contributes a value in the first variable also contributes an observation in the second variable. In fact, what we really want to focus our investigation on is how the drug changed *individuals'* scores, not how it changed the entire group.

The easiest way to examine this is to create a new variable that contains the change in blood pressure for each of the 15 patients:

```
. generate diff = after - before
```

Question 8: Examine this new variable. What was the greatest change? What was the least? What was a "typical" amount of change?

You might find it useful to examine a stem-and-leaf plot:

```
. stem diff
```

Question 9: How many people saw their blood pressure fall by 20 or more points? Did anyone show no change? Did anyone's blood pressure increase?

Question 10: If Captopril were ineffective, about where would you expect the

center of this distribution to be? What does the center of this distribution tell us, if anything, about the effectiveness of Captopril?

Another method for examining paired data is with a scatterplot. Examine the graph of the “after” blood pressure versus the “before.”

. scatter after before

Question 11: What can you learn about the change in blood pressure from this graph? Do you think this is a better (or worse) graph for answering this question than any of the other techniques we’ve used so far? Explain.

Since not everyone experiences the drug in the same way, it is natural to explore why this might be. One theory is that the drug might help those who need it the most. After all, if your blood pressure is unusually high, then it has further to fall, and might respond more dramatically to treatment.

Question 12: Make a scatterplot of the difference score against the before score. Do you think this provides evidence that this hypothesis (that those with the highest initial blood pressures improved the most) is correct?

Question 13: If you were a doctor, would you prescribe Captopril to your patients with high blood pressure? (Assume no side effects!) If yes, explain. If no, what other studies would you like to see?

Commands used for Captopril Lab

Here is a list of *Stata* commands we used in this lab. Use the space next to each command to make notes on what that command does.

edit

rename

summarize

graph box

generate

stem

scatter

Assignment for Inputting Our Own Data Lab

(This dataset is taken from Bruce Trumbo, *Learning Statistics with Real Data*, Duxbury Press, 2001.)

In 1982, a group of medical researchers conducted a study to determine if factory workers who worked around lead were inadvertently exposing their children to lead. (D. Morton. Lead absorption in children of employees in a lead-related industry. *American Journal of Epidemiology*, 155:549-555, 1982.) Lead poisoning is particularly dangerous to children because it interferes with normal physical and mental development. You can get lead into your blood by ingesting it, either by eating it or by breathing in lead particles. Once in your bloodstream it stays there; the body has no natural mechanism for cleansing itself of lead. (You may remember that many commonly used house paints used to have lead in them. These were discontinued, however, because children would sometimes chip paint off of walls and eat the paint.)

Other studies have shown that children of factory workers have high lead levels, but these studies had no control group, and so it was not known how these levels compared to children whose parents did not work at the factory. This study, in contrast, uses a control group. They examined 33 children, each of whom had parents who worked at a battery factory in Oklahoma and whose daily work brought them in contact with lead. For each of these 33 children (the “exposed” group), they found a “control” child of the same age living in the same neighborhood, but whose parents did not work around lead.

The lead levels reported below are measured in micrograms per deciliter. Most doctors consider levels of 40 or higher to be dangerous, and levels higher than 60 require immediate hospitalization. Here are the data from this study:

Exposed	38	23	41	18	37	36	23	62	31	34	24	14	21	17
Control	17	18	18	24	19	11	10	15	16	18	18	13	19	10

Exposed	16	20	15	10	45	39	22	35	49	48	44	35	43	39
Control	16	16	24	13	9	14	21	19	7	18	19	12	11	22

Exposed	34	13	73	25	27
Control	25	16	13	11	13

The data are shown in pairs. That is, the first child in the exposed group, whose lead level was 38, was matched with a child of the same age and living in the same neighborhood who had a lead level of 17. The second child in the Exposed group should be matched with the second child in the Control group.

Question 14: All things being equal, who would you expect to have the higher lead level and why, an 11-year-old or a 9-year-old?

Question 15: Why were the children in the control group matched by age and where they lived?

Question 16: Consider the Exposed group and Control group separately. What evidence do you see to support the theory that working at the factory caused the higher lead levels? Support your answer with appropriate summaries and graphs.

Question 17: Treat these data as paired. What evidence do you see to support the theory that working at the factory caused the higher lead levels? Support your answer with appropriate summaries and graphs.

Question 18: No matter which approach you use (looking at the groups separately or looking at them paired), it seems pretty clear that the Exposed group has higher lead levels. However, which approach do you think makes the strongest case?

Question 19: The Morton study also reports that the children of factory workers who showered and changed clothes at the factory before coming home had the lowest lead levels in the Exposed group. A politician wants to write a law requiring factory workers at lead factories to take showers. A politician in the opposition party argues that this would be an unnecessary expense for the factory owners because this study was an observational study. Choose sides, and write a letter to the politician you disagree with explaining why you think he or she is wrong. Make sure to include appropriate graphs and/or statistical summaries to support your claim.