Statistics Lab on Sampling

Math 1600, Fall 2016

Dr. Melanie Martin

“A project of the Annie E. Casey Foundation, KIDS COUNT is the premier source for data on child and family well-being in the United States.” Go to their web site:

<http://datacenter.kidscount.org/>

Click on DATA BY LOCATION

Click on California

Under LOCATIONS in menus on left side of page, under “Show indicators with data:”

choose “By County”

In the menu in the center of the page choose

[Children living in poverty by County](http://datacenter.kidscount.org/data/tables/9173-children-living-in-poverty-by-county?loc=6&loct=5#detailed/5/1156-1213/false/870,869,36,868/any/18186)

Near top of page in main frame look for “TOOLS” and click on “Raw Data” to download a .csv file (comma separated variable file), open this in Excel.

This will give you a file with data for California and all the counties for 2012, 2013 and 2014, with the state data at the top. We now want to see what happens when we randomly sample eight counties and compute the mean.

Save the .csv file you downloaded to a .xls workbook and name it Math 1600 Lab 10 Group X (where X is your group number). List the names of the group members who worked on it together at the top of the file on the right side of the data (column F).

Start by copying the data to a new worksheet, then sort it by column B. This will group by year. We will use only the 2014 data, so the other years can be deleted. These are the totals for the state and for each county.

Copy the data for the 58 counties and paste it into a new worksheet. Be sure not to include the data for California.

Now you should have 58 rows. In the column to the right of the data (column E), starting at the top number 1,2,3,4,5,6,7,1,2,3,4,5,6,7 and so on to the end of the data.

Sort based on column E, smallest to largest. You should have five groups of 8 counties and two groups with 9 counties.

Compute the mean for the counties in group 1. And for each of the rest of the groups.

Now you are ready to make a chart. First you will need to arrange the data like this:

|  |  |  |
| --- | --- | --- |
| group | mean | CA |
| 1 | 0.2325 | 0.23 |
| 2 | 0.23125 | 0.23 |
| 3 | 0.24875 | 0.23 |
| 4 | 0.215 | 0.23 |
| 5 | 0.20625 | 0.23 |
| 6 | 0.2225 | 0.23 |
| 7 | 0.1925 | 0.23 |

Here the 0.23 is the California value or the population parameter. Make a line chart to see how the means for each group differ from the population parameter. (Note this is 2012 data, so your number will not match the ones above.

Please answer the following questions:

1. Do you think the groups are representative samples? Explain your answer.

2. Considering each group as a sample, why is the sample mean different than the population mean?

3. What other ways could you get a more representative sample? (Consider trying it again with 4 groups of 14 counties.) How do the results differ?

Fun with Phones (Originally by John M. Cimbala at Penn State)

1. In this section, we will take several measurements of how long it takes for a cell phone call to go through. Get two cell phones – one will be the caller and one will be the receiver.
2. Call the receiver phone from the caller phone. With a stopwatch, time how long it takes from the time you hit the SEND button until the first ring on the receiver phone. Record the time (seconds) as precisely as possible (typically to 1/100 of a second) in an Excel spreadsheet.
3. Repeat as many times as you want – no fewer than 20. Again, the more data you take the better should be your statistics.
4. Calculate the ***sample mean***, ***sample median***, ***sample mode***, and ***sample standard deviation***.

Sample mean = \_\_\_\_\_\_\_\_\_\_\_ s

Sample median = \_\_\_\_\_\_\_\_\_\_\_ s

Sample standard deviation = \_\_\_\_\_\_\_\_\_\_\_ s

1. Draw a nice-looking histogram of the data.