

The purpose of this homework is to calculate basic summary statistics and create some data visualizations for two data samples. For this homework, you will need to download U.S. County Unemployment data from the class website (<http://people.tamu.edu/~cschulman/>) in the Excel workbook “HW1 County Unemployment Data.xlsx”. When you open the data in Excel it should look like the following:

	A	B	C	D	E
1	Southeast Region Sample			Southwest Region Sample	
		2009 SE Counties Unemployment Rate			2009 SW Counties Unemployment Rate
2	County Name			County Name	
3	Wayne County, WV	8.5		Gillespie County, TX	4.4
4	Washington County, KY	13.5		Lynn County, TX	5.9
5	Boone County, AR	7.4		Houston County, TX	9.1
6	Colonial Heights city, VA	8.0		Polk County, TX	9.0
7	Catahoula Parish, LA	10.1		Swisher County, TX	5.7
8	Glascocock County, GA	12.0		Colfax County, NM	7.1
9	Richmond County, VA	7.9		Kleberg County, TX	6.7
10	Brevard County, FL	9.6		Hunt County, TX	8.2
11	Glynn County, GA	8.9		Kenedy County, TX	5.9

The “County Name” columns simply identify the county (or ‘place’) and State. Note that the data are in no particular sort order. The Unemployment Rate columns show the 2009 annual average unemployment rate for a sample of 250 counties from the Southeast Region (in column B) and 250 counties from the Southwest (in column E). At this point, go to File, Save As, and save your file as file type “Excel Workbook (*.xlsx)” using the naming convention “Lastname Firstname Studentid HW1.xlsx.” For example:

Smith James 987654321 HW1.xlsx

Click the “+” button next to the tab name to insert a new worksheet. Right-click on the new worksheet name and rename it “Solutions.” Organize your solutions to the following questions on this “Solutions” tab.

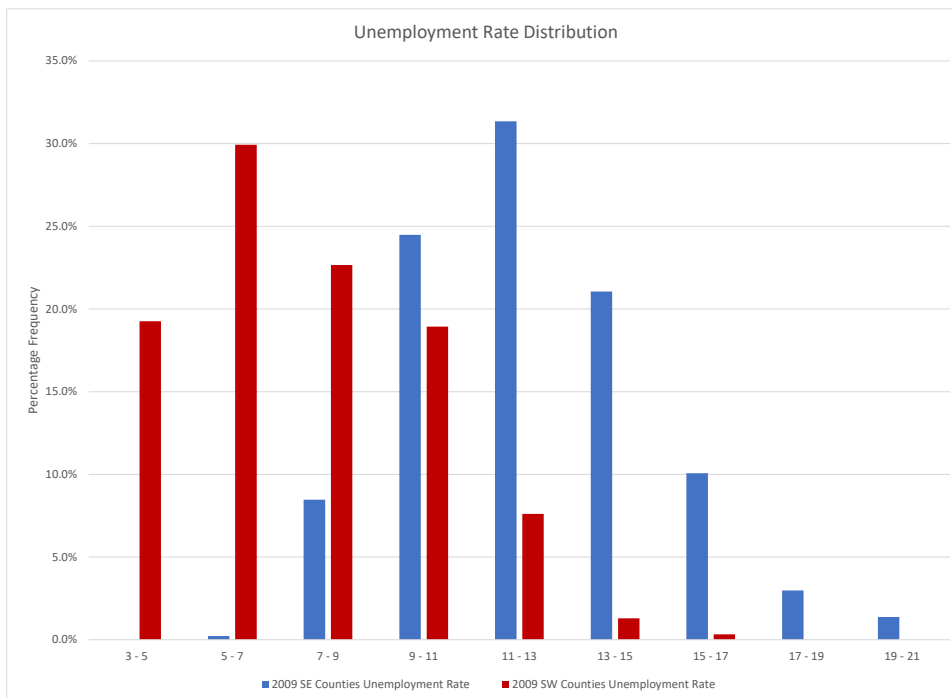
1. **Create a table of summary statistics** across the 250 counties in each of the samples of the unemployment rate variable like that at right, including the, Minimum (Excel function “Min”), Maximum (“Max”), Range (Max – Min), the total number of observations – Count, (“Count”), Mean (“Average”), Median (“Median”), the First Quartile Q1 (“Quartile.exc”) and Third Quartile Q3 (“Quartile.exc”).
 - a. How do the two samples compare regarding their Means, Medians and Interquartile Range?

	Southeast Unemployment	Southwest Unemployment
Minimum		
Maximum		
Range		
Count		
Mean		
Median		
Q1		
Q3		

2. For each of the two samples
 - a. What three counties had the highest Unemployment Rates (3 highest SE and 3 highest SW)?
 - b. What three counties had the lowest Unemployment Rates (3 lowest SE and 3 lowest SW)?
3. Follow the procedures described in the lecture notes to **create a frequency distribution table** like that below of the two unemployment rate variables using the Excel “Countifs” function. Include the frequency count and percentage frequency of the number of counties with Unemployment Rates that fall in each of **9** equally spaced class intervals from 3 to 21 (greater than or equal to 3 but less than 5, greater than or equal to 5 but less than 7, etc.).

Class	Southeast Frequency Count	Southeast Percentage Frequency	Southwest Frequency Count	Southwest Percentage Frequency
3-5				
5-7				
7-9				
9-11				
etc.				

4. Using the information from your frequency table, **create a bar chart** of the percentage frequency distributions of Unemployment Rates like that below for the two samples (**your numbers will be different**). Do the distributions appear to be **symmetric** around the mean or **skewed** to the right or the left?



5. Using the two Unemployment Rate variables create Box-and-Whisker plots for the two samples like that in the Lecture Notes (you do not need to add all the textboxes, arrows, and descriptions)
6. How do the relative values of the Q1, median, mean, and Q3 relate to your answer in (4) that the distributions of Unemployment Rates are somewhat symmetric, skewed right, or skewed left? Why?

Save your Excel file with your solutions using the naming convention “Lastname Firstname Studentid HW1.xlsx” noted above for submission in Canvas.