Econometrics 461 Spring 2025
Homework #1 Due in Canvas by midnight, Friday, January 31st

The purpose of this homework is to help you begin building your Excel skills – creating new variables, calculating basic summary statistics, and creating some data visualizations for two groups of U.S. Counties. The goal with this analysis is to calculate the Average Compound Annual Growth Rate (ACAGR) in Real Per-Capita Personal Income for counties in two U.S. Regions – the Plains States and the Rock Mountain States. For this homework, you will need to download U.S. County Personal Income and Population data from the class website (http://people.tamu.edu/~cschulman/) in the Excel workbook "ECMT 461 HW1 Data.xlsx". These data are from the U.S. Bureau of Economic Analysis. When you open the data in Excel, the HW1 Data tab should look like the following:

	A	B	C	D	E	F	G	H	I	J
1	GEOFIPS	State FIPS	County FIPS	GeoName	Region	Region Name	Personal Income 2010 (000)	Population 2010	Personal Income 2020 (000)	Population 2020
2	08001	08	001	Adams, CO*	7	Rocky Mountain	13713370	443693	25292075	520479
3	08003	08	003	Alamosa, CO	7	Rocky Mountain	448363	15519	665499	16377
4	08005	08	005	Arapahoe, CO	7	Rocky Mountain	24797601	575067	45295999	655207
5	08007	08	007	Archuleta, CO	7	Rocky Mountain	367401	12032	680737	13425

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

I strongly suggest that you create a dedicated folder on your computer to store all your homework files so that you have only one place to look to find your files.

The data definitions for the various columns are as follows:

GEOFIPS	5-digit Geographic code that uniquely identifies each U.S. County
State FIPS	First two digits of GEOFIPS, to identify each U.S. State
County FIPS	Last 3 digits of GEOFIPS to identify each county within a State
GeoName	Geographic Name
Region	Region Code: There are two regions included in these data, Region 4 – the Plains States; and Region 7 – the Rock Mountain States;
Region Name	Region Name
Personal Income 2010 (000)	Total Personal Income for the County in 2010 in thousands of nominal dollars
Population 2010	County Population in 2010
Personal Income 2020 (000)	Total Personal Income for the County in 2020 in thousands of nominal dollars
Population 2020	County Population in 2020

There should be a total of 834 counties included in these data.

The second worksheet tab "Price Index data" has annual data on the Personal Consumption Expenditures Chain-type Price Index, Index 2023=100, for the years 2010 – 2023: PCEPI.

The first step in the analysis is to calculate Per Capita Personal Income (personal income per person) for each county in both 2010 and 2020. For example:

$$PC\ PInc\ 2010 = \frac{1000 \times Personal\ Income\ 2010}{Population\ 2010}$$

And similarly for PC PInc 2020. These two variables are in *nominal* dollars – that is, they do not account for inflation. To adjust for inflation and get *real* per capita personal income in 2023 dollars, we use the Price Index data as follows:

$$Real\ PC\ PInc\ 2010 = \frac{100 \times PC\ PInc\ 2010}{PCEPI_{2010}}$$

And similarly for Real PC PInc 2020. Note that you will only need to use the 2010 and 2020 PCEPI values for these calculations.

Given the Real PC PInc variables, you can now calculate the Average Compound Annual Growth Rate for each county:

$$ACAGR = \left(\frac{Real\ PC\ PInc\ 2020}{Real\ PC\ PInc\ 2010}\right)^{\frac{1}{10}} - 1$$

In Excel, I suggest liberal use of parentheses to make sure you get the proper result. For example:

	PC PInc 2010	PC PInc 2020	Real PC PInc 2010	Real PC PInc 2020	ACAGR
9	30907	48594	40766	55339	=((N2/M2)^(1/10))-1
7	28891	40636	38107	46277	2.0%

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** for the counties in each of the two regions for the ACAGR 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").

	Plains ACAGR	Rocky Mountain ACAGR
Minimum		
Maximum		
Range		
Count		
Mean		
Median		
Q1		
Q3		

Note: There are several ways to organize the data to perform these calculations separately for the counties in each of the two regions. Perhaps the most efficient method is to use conditional *array* formulas in Excel. If you have done all the calculations to get the ACAGR variable on the HW1 Data tab, with the Region variable in column E and the ACAGR variable in column O, an example of an array formula is as follows:

	Plains ACAGR	Rocky Mountain ACAGR
Minimum	=MIN(IF('HW1 Data'!E:E=4,'HW1 Data'!O:O))	
Maximum		

For the QUARTILE.EXC function, an array formula would look like the following:

Q1	=QUARTILE.EXC(IF('HW1 Data'!E:E=4,'HW1 Data'!O:O),1)
Q3	

a. How do the two samples compare regarding their Means, Medians and Interquartile Range? (On your solutions tab, simply type your answer in a cell below your summary statistics table.)

2. For each of the two regions
   a. What three counties had the highest ACAGR (3 highest Plains and 3 highest Rocky Mountains)?
   b. What three counties had the lowest ACAGR (3 lowest Plains and 3 lowest Rocky Mountains)?
   c. (NOTE: Either sorting the data by Region and ACAGR, or using a data filter, with the Top 10 and Bottom 10 number filter may make this task easier.)
3. Follow the procedures described in the lecture notes to **create a frequency distribution table** like that below of the ACAGR variable for the two regions using the Excel "Countifs" function. Include the frequency count and percentage frequency of the number of counties with ACAGRs that fall in each of 22 equally spaced class intervals from -3.5% to 7.5% in 0.5% increments (greater than or equal to -3.5% but less than -3.0%, greater than or equal to -3.0% but less than -2.5%, etc.).

		Plains Frequency	Rocky Mountain Frequency	Plains Percent Frequency	Rocky Mountain Percent Frequency
-3.5%	-3.0%	=COUNTIFS($E$2:$E$1654,"=4",$O$2:$O$1654,">=" &$S4,$O$2:$O$1654,"<" &$T4)			
-3.0%	-2.5%				

4. Using the information from your frequency table, **create a bar chart** of the percentage frequency distributions of ACAGRs like that below for the two regions. Do the distributions appear to be **symmetric** around the mean or **skewed** to the right or the left?

Note: To create a column of text labels for the horizontal axis like that shown above, you can use the Excel TEXT function combined with the & concatenation operator like the following:

		Plains Frequency	Rocky Mountain Frequency	Plains Percent Frequency	Rocky Mountain Percent Frequency	
-3.5%	-3.0%	0	1	0.0%	0.5%	=TEXT(S4,"0.0%")&" to "&TEXT(T4,"0.0%")
-3.0%	-2.5%	1	0	0.2%	0.0%	-3.0% to -2.5%
-2.5%	-2.0%	1	0	0.2%	0.0%	-2.5% to -2.0%
-2.0%	-1.5%	0	0	0.0%	0.0%	-2.0% to -1.5%

5. Insert a new worksheet tab named "BW Chart Data." Using a data filter on the Region variable in the HW1 Data tab, copy the ACGR variable separately for the two regions into this new tab. Note that when you paste the data into this new worksheet tab, you will need to Paste-Special, "Values and number formats." Use these data to create Box-and-Whisker plots for the two regions like that in the Lecture Notes (you do not need to add all the textboxes, arrows, and descriptions).
6. Is you answer to 1.a above visually evident in the Box-and-Whisker plots

Save your Excel file with your solutions using the naming convention

Lastname Firstname Studentid HW1.xlsx

noted above for submission in Canvas.