ECMT 461 Term Project: U.S. County Income Growth

William N. Duong

Department of Economics, Texas A&M University

ECMT 461-902: Economic Data Analysis

Dr. Craig Schulman

May 3, 2023

Introduction

Section A: Criterion and Predictor Variables

For this project, I opted to examine the U.S. County Income Growth dataset in order to analyze the potential existence of a relationship between Per Capita Income Growth Rate (1970-2020) and Population Growth Rate (1970-2020). As a result, I have selected Per Capita Income Growth Rate to be my criterion variable and Population Growth Rate to be my predictor variable. The chosen criterion variable, Per Capita Income Growth Rate, possesses economic significance because it is a measure of the percent change in per capita income over a period of time. Ideally, this measure would be a Median Income Growth Rate, which would paint a more accurate picture of the income levels across time within each county listed. This is because per capita information is unable to account for statistical outliers and important income-related factors, such as inflation, wealth, and poverty. Compared to a per capita metric, a median metric yields a more accurate depiction of a given dataset due to its ability to account for outlier values. However, the Median Income Growth Rate is not included within this dataset. Only information concerning the Per Capita Income Growth Rate is present, which is why I am using it to conduct my analysis. While not a perfect measure of income growth by any means, the Per Capita Income Growth Rate is still an extremely useful metric for comparing the income growth of different U.S. counties because of its granularity and level of detail relative to other metrics. This, in turn, gives the criterion variable economic significance.

Regarding expectations of association between the criterion and predictor variables, one would expect to observe a negative correlation between the Per Capita Income Growth Rate and Population Growth Rate because per capita income is calculated by dividing the total income of a specific group by the total population of that specific group. Therefore, as population size increases, per capita income should be expected to decrease.

Section B: Subgroup Identification

As for my category variable, I selected Region and divided it into eight subgroups: Far West, Great Lakes, Mideast, New England, Plains, Rocky Mountain, Southeast, and Southwest. I wanted to observe if the potential correlation between counties' Per Capita Income Growth Rates and Population Growth Rates held true for all U.S. counties regardless of region or if region played a role in determining the potential correlation. The basis for expecting differences in the measures of central tendency, dispersion, and degrees of association of the criterion variable among each of the subgroups is that each region of the U.S. has different characteristics which can affect per capita income and population growth. For example, regions have different push and pull factors, such as climate and government policies, which influence population migration. In addition, each region of the U.S. has different economic preferences and industry viability, which results in differences in job opportunity and income. Essentially, no two regions have the exact same economic makeup. Therefore, using Region as the category variable would allow for a more accurate and detailed analysis of any potential relationship between U.S. county Per Capita Income Growth Rate and Population Growth Rate.

Section C: Statement of Formal Hypothesis

Due to preexisting knowledge of U.S. regional economics, I expect the Far West Region to have the strongest correlation with Per Capita Income Growth Rate and Population Growth Rate from 1970 to 2020.

Literature Review

Looking at past analyses of per capita economic growth and population growth reveals experts disagree with presumed expectations of correlation between Per Capita Income Growth Rate and Population Growth Rate. In an accredited 2021 study published by the Center for Immigration Studies, research director Steven A. Camarota (2021) firmly posits there is no evidence that population growth is a driver of per capita economic growth, arguing instead that "more rapid population growth tends to slow per capita economic growth in developed countries." His findings, presented using various tables, plots, and figures, indicate "population growth is associated with lower rates of per capita economic growth and slower rates of population growth are associated with more per capita economic growth" (Camarota, 2021). This assertion directly contrasts the presumed existence of a negative correlation between the criterion variable and predictor variable expressed in the introduction. However, one should note that Camarota used per capita gross domestic product, not per capita income, as the metric for economic growth in his study. Therefore, based solely on preexisting findings, one could expect to find no correlation between per capita income growth rate and population growth rate for U.S. counties, but further analysis specifically observing per capita income is required to provide a more accurate answer.

Descriptive/Graphical Analysis

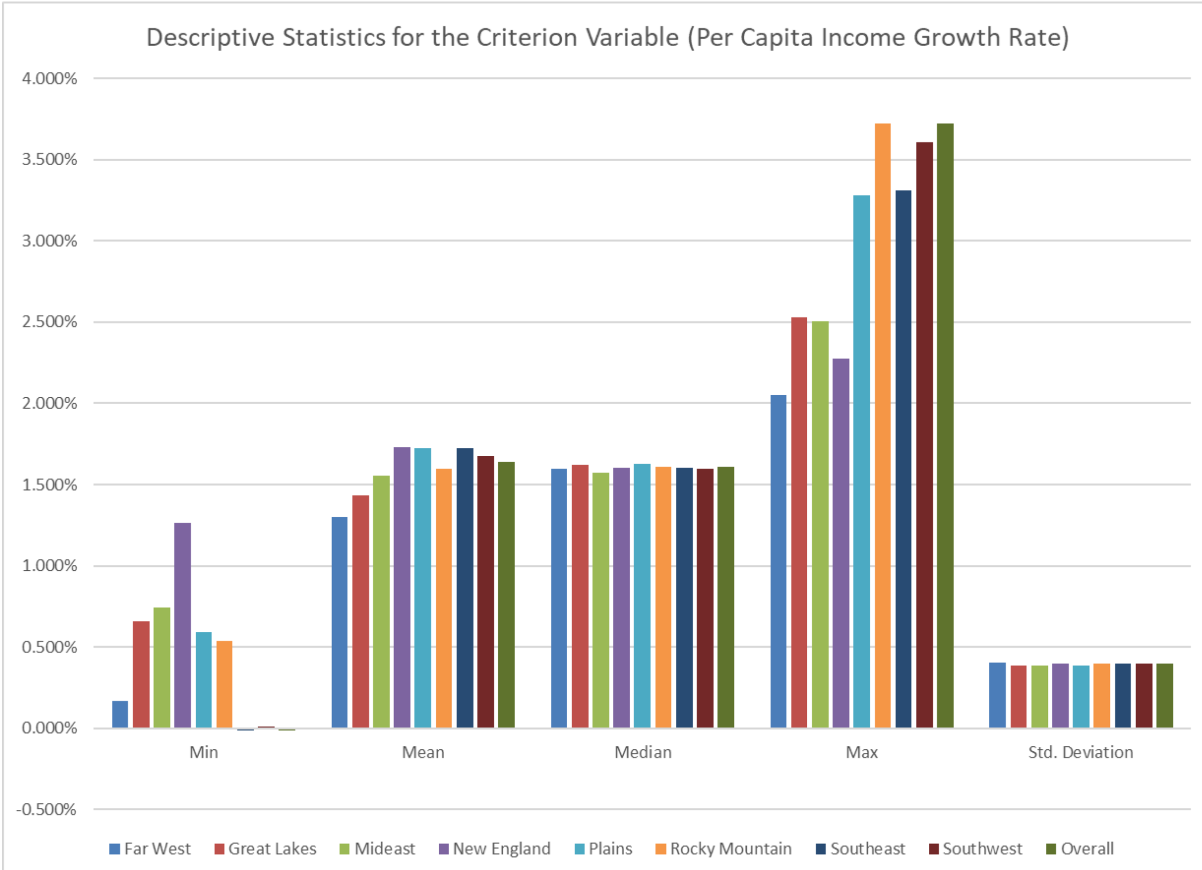
Section A: Descriptive Statistics

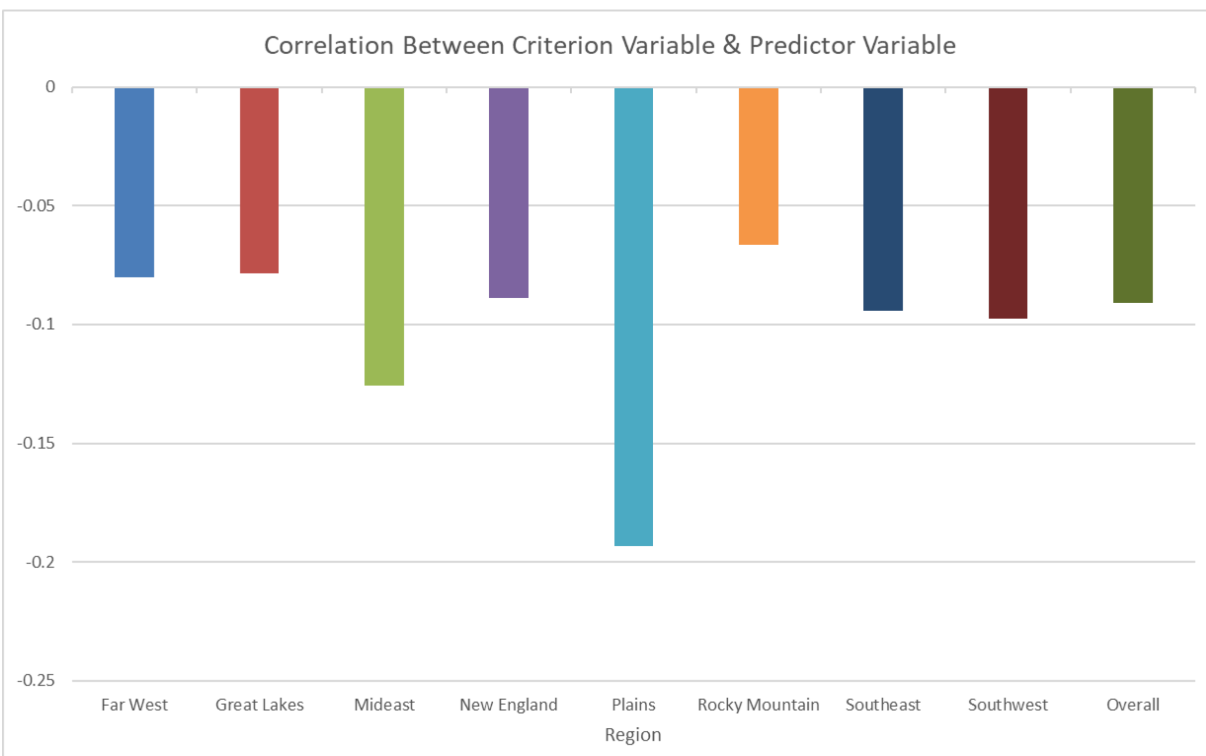
As mentioned above in the statement of formal hypothesis section, I anticipated the Far West Region to have the strongest correlation with Per Capita Income Growth Rate and Population Growth Rate from 1970 to 2020. In order to sufficiently test this hypothesis, I found basic sample statistics for the subgroups and the correlation coefficients of all eight subgroups, as well as the correlation coefficient of the overall population, and displayed them in both tabular and graphical forms. Correlation coefficients range from -1 to 1, with -1 indicating a perfect negative correlation between the two variables and 1 indicating a perfect positive correlation between the two variables. With this in mind, the strongest correlation between the criterion variable and predictor variable exists for the Plains Region at a coefficient of about -0.19, while the weakest correlation occurs for the Rocky Mountain Region at a coefficient of about -.07. In addition, each region exemplified a negative, albeit weak, correlation value, which indicates that as Per Capita Income Growth Rate increases, Population Growth Rate decreases and vice versa.

Regarding my hypothesis, when comparing the Far West Region to the other seven subgroups, the Far West Region ranks third to last in terms of strength of correlation. Therefore, my hypothesis is incorrect since the Plains Region, not the Far West Region, displays the strongest correlation between Per Capita Income Growth Rate and Population Growth Rate.

Descriptive Statistics for the Criterion Variable

Sub-Groups	Sample Size	Min	Mean	Median	Max	Std. Deviation	Correlation
Far West	150	0.167%	1.298%	1.600%	2.050%	0.004025281	-0.08017452
Great Lakes	435	0.659%	1.432%	1.619%	2.531%	0.003869364	-0.07854368
Mideast	178	0.742%	1.555%	1.572%	2.503%	0.003832442	-0.12556227
New England	67	1.263%	1.733%	1.604%	2.278%	0.003961873	-0.08898586
Plains	618	0.589%	1.723%	1.627%	3.283%	0.003834202	-0.19340288
Rocky Mountain	215	0.534%	1.594%	1.607%	3.721%	0.003962339	-0.06650343
Southeast	1035	-0.013%	1.726%	1.605%	3.313%	0.003990840	-0.09431338
Southwest	377	0.008%	1.676%	1.599%	3.606%	0.003994482	-0.09739831
Overall	3075	-0.013%	1.640%	1.606%	3.721%	0.003969783	-0.09107244


Descriptive Statistics for the Criterion Variable (Per Capita Income Growth Rate)
4.000%
3.500%
3.000%
2.500%
2.000%
1.500%
1.000%
0.500%
0.000%
-0.500%
Min
Mean
Median
Max
Std. Deviation
Far West
Great Lakes
Mideast
New England
Plains
Rocky Mountain
Southeast
Southwest
Overall


Correlation Between Criterion Variable & Predictor Variable
0
-0.05
-0.1
-0.15
-0.2
-0.25
Far West
Great Lakes
Mideast
New England
Plains
Rocky Mountain
Southeast
Southwest
Overall
Region

Section B: Frequency Distributions

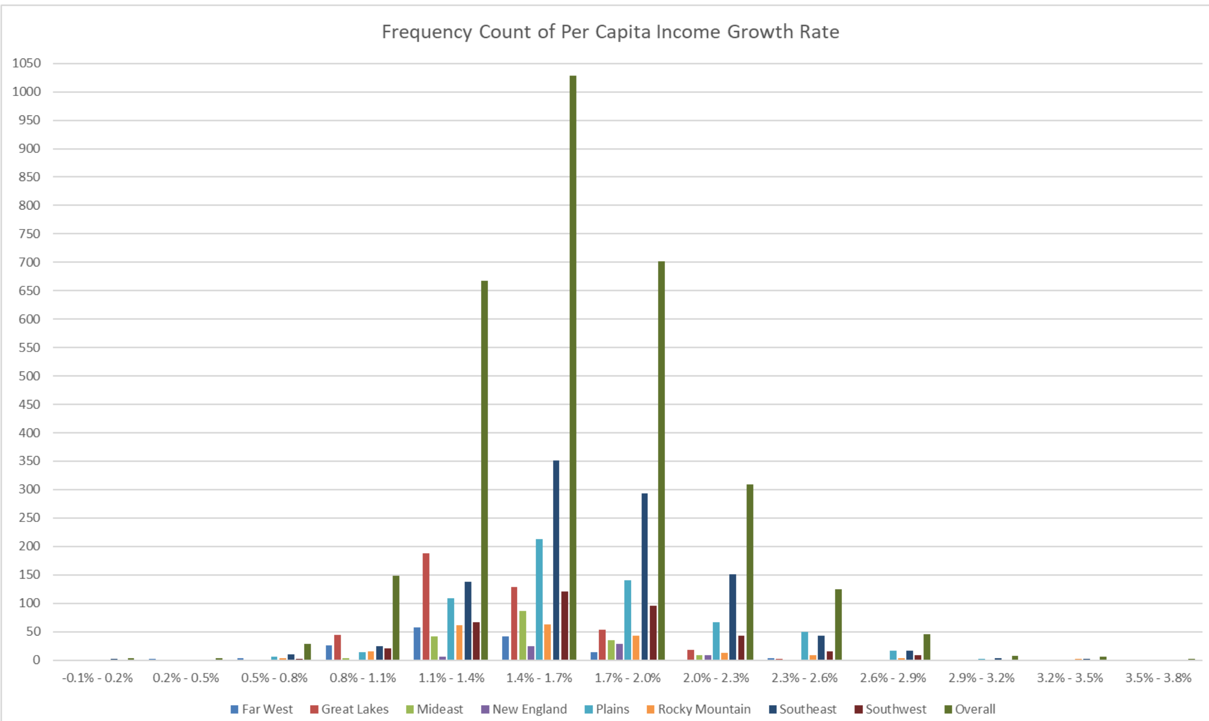
Concerning the frequency distributions of my chosen criterion variable for both the overall sample and subgroup samples, I opted to use thirteen classes of equal width, the width being 0.3%. I came to this decision because the overall minimum value was -0.013% and the overall maximum value was 3.721%, which allowed for thirteen classes of even width ranging from a value of -0.1% to a value of 3.8%. The most common range, as seen in the percentage frequency distribution table, was the 1.4% - 1.7% class, which contained 33.46% of the criterion variable data points within the given dataset. Overall, both the frequency count and percentage frequency distribution graphs demonstrated a slight left-skew, indicating that relative to the normal distribution, the U.S. County Per Capita Income Growth Rates from 1970-2020 analyzed within this dataset were ever-so-slightly greater than average.

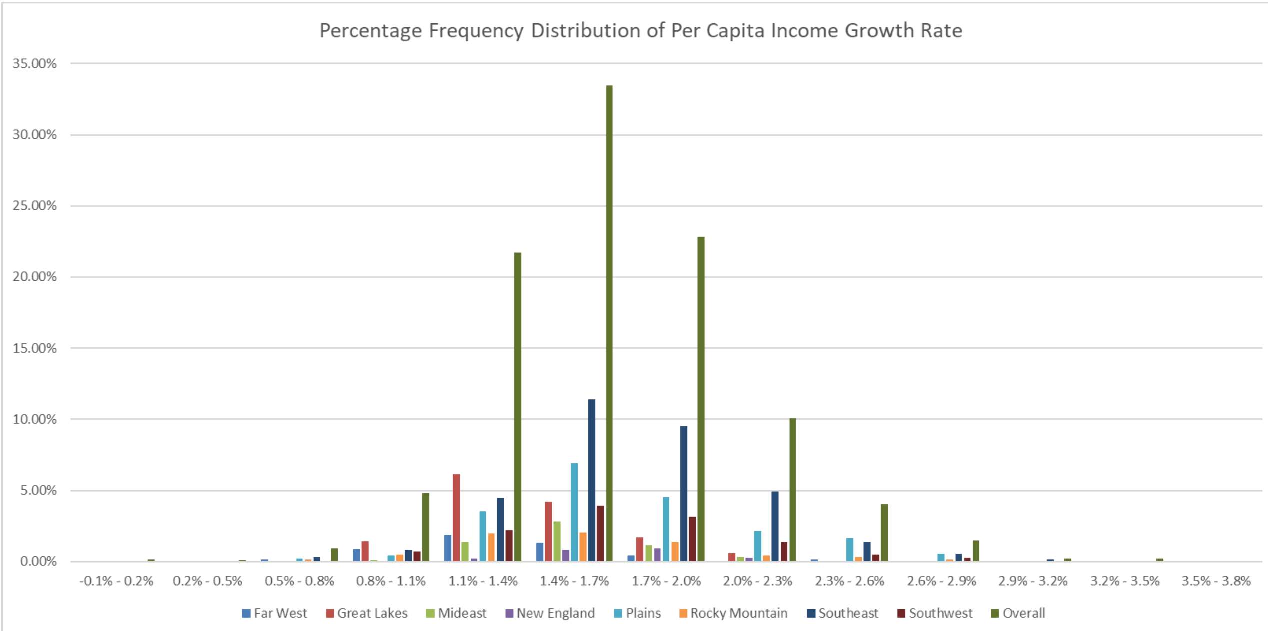
Regarding my hypothesis, the most common class present in the Far West Region subgroup was the 1.1% - 1.4% range, which included 38.0% of the 150 Far West Region data points and 1.85% of the 3075 total sample data points. This result indicates that the Far West Region subgroup was below average compared to the overall sample as far as Per Capita Income Growth Rate is concerned, further disproving my initial hypothesis.

Frequency Counts

Class	Far West	Great Lakes	Mideast	New England	Plains	Rocky Mountain	Southeast	Southwest	Overall
-0.1% - 0.2%	1	0	0	0	0	0	2	1	4
0.2% - 0.5%	2	0	0	0	0	0	0	1	3
0.5% - 0.8%	4	1	1	0	6	4	10	2	28
0.8% - 1.1%	26	44	3	0	14	15	25	21	148
1.1% - 1.4%	57	188	42	6	109	61	138	67	668
1.4% - 1.7%	41	129	87	25	213	63	351	120	1029
1.7% - 2.0%	14	53	35	28	140	43	293	96	702
2.0% - 2.3%	1	18	9	8	66	13	151	43	309
2.3% - 2.6%	4	2	1	0	50	9	43	15	124
2.6% - 2.9%	0	0	0	0	17	4	16	8	45
2.9% - 3.2%	0	0	0	0	2	0	4	1	7
3.2% - 3.5%	0	0	0	0	1	2	2	1	6
3.5% - 3.8%	0	0	0	0	0	1	0	1	2
Total	150	435	178	67	618	215	1035	377	3075

Percentage Frequency									
Class	Far West	Great Lakes	Mideast	New England	Plains	Rocky Mountain	Southeast	Southwest	Overall
-0.1% - 0.2%	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.07%	0.03%	0.13%
0.2% - 0.5%	0.07%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.10%
0.5% - 0.8%	0.13%	0.03%	0.03%	0.00%	0.20%	0.13%	0.33%	0.07%	0.91%
0.8% - 1.1%	0.85%	1.43%	0.10%	0.00%	0.46%	0.49%	0.81%	0.68%	4.81%
1.1% - 1.4%	1.85%	6.11%	1.37%	0.20%	3.54%	1.98%	4.49%	2.18%	21.72%
1.4% - 1.7%	1.33%	4.20%	2.83%	0.81%	6.93%	2.05%	11.41%	3.90%	33.46%
1.7% - 2.0%	0.46%	1.72%	1.14%	0.91%	4.55%	1.40%	9.53%	3.12%	22.83%
2.0% - 2.3%	0.03%	0.59%	0.29%	0.26%	2.15%	0.42%	4.91%	1.40%	10.05%
2.3% - 2.6%	0.13%	0.07%	0.03%	0.00%	1.63%	0.29%	1.40%	0.49%	4.03%
2.6% - 2.9%	0.00%	0.00%	0.00%	0.00%	0.55%	0.13%	0.52%	0.26%	1.46%
2.9% - 3.2%	0.00%	0.00%	0.00%	0.00%	0.07%	0.00%	0.13%	0.03%	0.23%
3.2% - 3.5%	0.00%	0.00%	0.00%	0.00%	0.03%	0.07%	0.07%	0.03%	0.20%
3.5% - 3.8%	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.00%	0.03%	0.07%
Total	4.88%	14.15%	5.79%	2.18%	20.10%	6.99%	33.66%	12.26%	100.00%





Section C: Box & Whisker Plots

Section D: Discussion of Observed Differences

When scanning for differences, sample size was one of the characteristics that stood out the most among the eight subgroups. In particular, the New England Region had the smallest sample size and least amount of data points, only 67, while the Southeast Region had the largest sample size and greatest amount of data points, 1035 in total. This is important to keep in mind because conclusions and results obtained from analyzing subgroups of smaller sample sizes tend to be weaker and less reliable than those gathered from subgroups with larger sample sizes.

Another outlier would be the maximum values, minimum values, and respective ranges of each of the subgroups. The Southeast and Southwest Regions had minimum Per Capita Income Growth Rates much smaller in magnitude than the six other regions, -0.013% and 0.008% respectively, yet both possessed maximum values within the top three subgroups, indicating that the Per Capita Income Growth Rate is much more varied and less uniform depending on location within these regions compared to the rest.

Single Sample Confidence Intervals and Hypothesis Tests

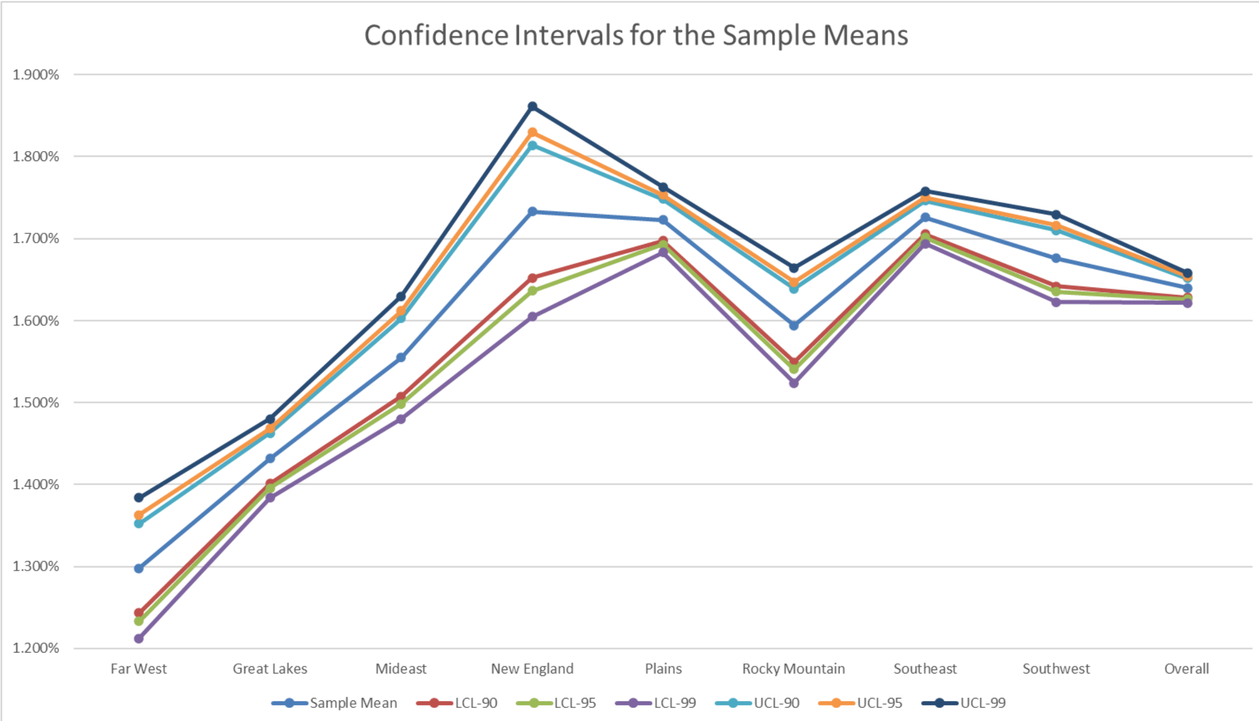
Section A: Confidence Intervals

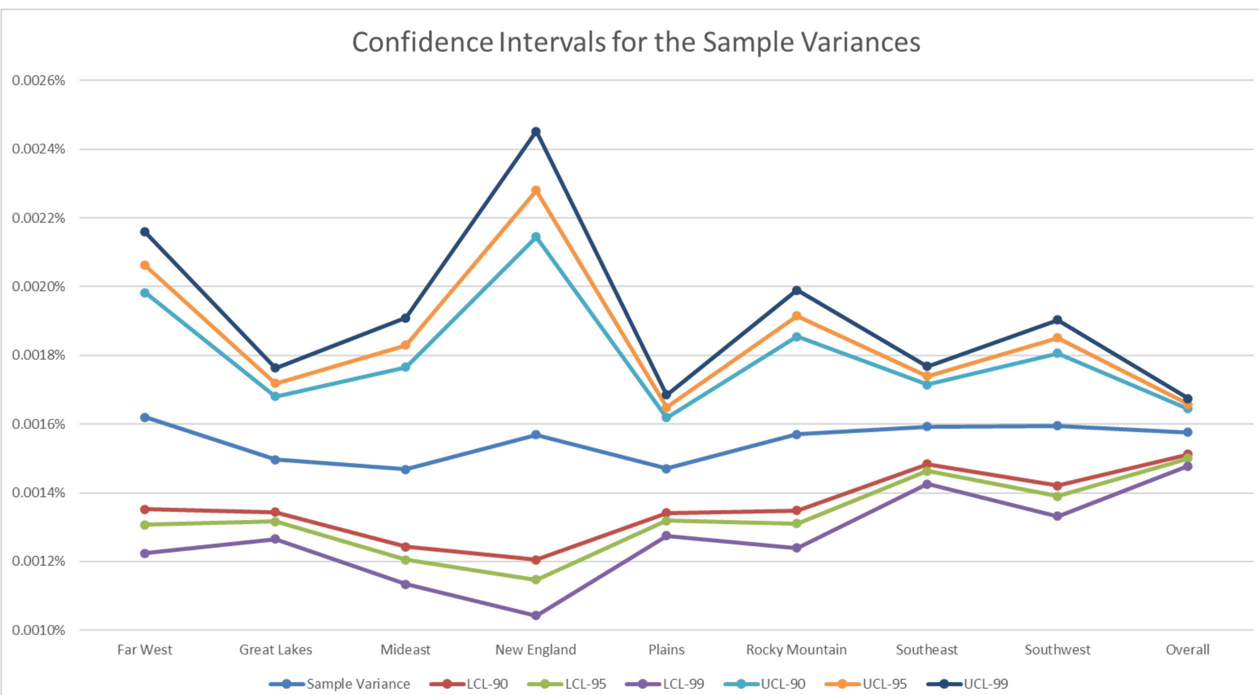
Confidence Intervals for the Sample Means

	Sample Mean	Lower CL	Upper CL
Far West - 90% CI	1.298%	1.244%	1.352%
Far West - 95% CI	1.298%	1.212%	1.363%
Far West - 99% CI	1.298%	1.352%	1.384%
Great Lakes - 90% CI	1.432%	1.401%	1.463%
Great Lakes - 95% CI	1.432%	1.396%	1.468%
Great Lakes - 99% CI	1.432%	1.384%	1.480%
Mideast - 90% CI	1.555%	1.508%	1.602%
Mideast - 95% CI	1.555%	1.498%	1.612%
Mideast - 99% CI	1.555%	1.480%	1.630%
New England - 90% CI	1.733%	1.652%	1.814%
New England - 95% CI	1.733%	1.636%	1.830%
New England - 99% CI	1.733%	1.605%	1.861%
Plains - 90% CI	1.723%	1.698%	1.748%
Plains - 95% CI	1.723%	1.693%	1.753%
Plains - 99% CI	1.723%	1.683%	1.763%
Rocky Mountain - 90% CI	1.594%	1.549%	1.639%
Rocky Mountain - 95% CI	1.594%	1.541%	1.647%
Rocky Mountain - 99% CI	1.594%	1.524%	1.664%
Southeast - 90% CI	1.726%	1.706%	1.746%
Southeast - 95% CI	1.726%	1.702%	1.750%
Southeast - 99% CI	1.726%	1.694%	1.758%
Southwest - 90% CI	1.676%	1.642%	1.710%
Southwest - 95% CI	1.676%	1.636%	1.716%
Southwest - 99% CI	1.676%	1.623%	1.729%
Overall - 90% CI	1.640%	1.628%	1.652%
Overall - 95% CI	1.640%	1.626%	1.654%
Overall - 99% CI	1.640%	1.622%	1.658%

Confidence Intervals for the Sample Variances

	Sample Variance	Lower CL	Upper CL
Far West - 90% CI	0.00153%	0.00135%	0.00198%
Far West - 95% CI	0.00153%	0.00131%	0.00206%
Far West - 99% CI	0.00153%	0.00122%	0.00216%
Great Lakes - 90% CI	0.00150%	0.00134%	0.00168%
Great Lakes - 95% CI	0.00150%	0.00132%	0.00172%
Great Lakes - 99% CI	0.00150%	0.00127%	0.00176%
Mideast - 90% CI	0.00148%	0.00124%	0.00177%
Mideast - 95% CI	0.00148%	0.00121%	0.00183%
Mideast - 99% CI	0.00148%	0.00113%	0.00191%
New England - 90% CI	0.00155%	0.00121%	0.00214%
New England - 95% CI	0.00155%	0.00115%	0.00228%
New England - 99% CI	0.00155%	0.00104%	0.00245%
Plains - 90% CI	0.00151%	0.00134%	0.00162%
Plains - 95% CI	0.00151%	0.00132%	0.00165%
Plains - 99% CI	0.00151%	0.00128%	0.00169%
Rocky Mountain - 90% CI	0.00152%	0.00135%	0.00185%
Rocky Mountain - 95% CI	0.00152%	0.00131%	0.00192%
Rocky Mountain - 99% CI	0.00152%	0.00124%	0.00199%
Southeast - 90% CI	0.00159%	0.00148%	0.00171%
Southeast - 95% CI	0.00159%	0.00146%	0.00174%
Southeast - 99% CI	0.00159%	0.00143%	0.00177%
Southwest - 90% CI	0.00161%	0.00142%	0.00181%
Southwest - 95% CI	0.00161%	0.00139%	0.00185%
Southwest - 99% CI	0.00161%	0.00133%	0.00190%
Overall - 90% CI	0.00158%	0.00151%	0.00164%
Overall - 95% CI	0.00158%	0.00150%	0.00166%
Overall - 99% CI	0.00158%	0.00148%	0.00167%


Confidence Intervals for the Sample Means
1.900%
1.800%
1.700%
1.600%
1.500%
1.400%
1.300%
1.200%
Far West
Great Lakes
Mideast
New England
Plains
Rocky Mountain
Southeast
Southwest
Overall
Sample Mean
LCL-90
LCL-95
LCL-99
UCL-90
UCL-95
UCL-99


Confidence Intervals for the Sample Variances
0.0026%
0.0024%
0.0022%
0.0020%
0.0018%
0.0016%
0.0014%
0.0012%
0.0010%
Far West
Great Lakes
Mideast
New England
Plains
Rocky Mountain
Southeast
Southwest
Overall
Sample Variance
LCL-90
LCL-95
LCL-99
UCL-90
UCL-95
UCL-99

Section B: Discussion of Confidence Interval Results

When looking at the constructed confidence intervals for the sample means of the overall sample and individual subgroup samples at each of the three confidence levels (90%, 95%, 99%), the Far West Region was observed to have the smallest sample mean, an average Per Capita Income Growth Rate of 1.298%, while the New England Region yielded the largest sample mean, an average Per Capita Income Growth Rate of 1.733%. As for the confidence intervals for the sample variances of the overall sample and individual subgroup samples, the Mideast Region possessed the smallest sample variance, a value of 0.00148% while the Southwest Region yielded the largest sample variance, a value of 0.00161%. Additionally, the subgroup with the sample mean closest to the overall sample mean was the Southwest Region, while the subgroup with the sample variance closest to the overall sample variance was the Southeast Region, indicating that on average, the Per Capita Income Growth Rate of the Southwest is on par with the national Per Capita Income Growth Rate.

Section C: Single Sample Hypothesis Tests

Single Sample Hypothesis Tests for the Sample Means: Two-Tailed Tests

Sub-Groups	Sample Mean	t-Stat	t-Crit 90	Conclusion 90	t-Crit 95	Conclusion 95	t-Crit 99	Conclusion 99
Far West	1.298%	-10.406	1.655	Reject H0	1.976	Reject H0	2.609	Reject H0
Great Lakes	1.432%	-11.212	1.648	Reject H0	1.965	Reject H0	2.587	Reject H0
Mideast	1.555%	-2.959	1.654	Reject H0	1.973	Reject H0	2.604	Reject H0
New England	1.733%	1.921	1.668	Reject H0	1.997	Fail to Reject H0	2.652	Fail to Reject H0
Plains	1.723%	5.381	1.647	Reject H0	1.964	Reject H0	2.584	Reject H0
Rocky Mountain	1.594%	-1.702	1.652	Reject H0	1.971	Fail to Reject H0	2.599	Fail to Reject H0
Southeast	1.726%	6.933	1.646	Reject H0	1.962	Reject H0	2.581	Reject H0
Southwest	1.676%	1.750	1.649	Reject H0	1.966	Fail to Reject H0	2.589	Fail to Reject H0

Single Sample Hypothesis Tests for the Sample Variances: Two-Tailed Tests

Sub-Groups	Sample Variance	Chi-Stat	Chi-Low 90	Chi-Upp 90	Concl. 90	Chi-Low 95	Chi-Upp 95	Concl. 95	Chi-Low 99	Chi-Upp 99	Concl. 99
Far West	0.002%	153.195	121.787	178.485	Fail to reject H0	117.098	184.687	Fail to reject H0	108.291	197.211	Fail to reject H0
Great Lakes	0.002%	412.321	386.703	483.571	Fail to reject H0	378.173	493.614	Fail to reject H0	361.870	513.640	Fail to reject H0
Mideast	0.001%	164.965	147.230	209.042	Fail to reject H0	142.053	215.733	Fail to reject H0	132.293	229.211	Fail to reject H0
New England	0.002%	65.737	48.305	85.965	Fail to reject H0	45.431	90.349	Fail to reject H0	40.158	99.330	Fail to reject H0
Plains	0.001%	575.575	560.378	675.895	Fail to reject H0	550.063	687.725	Fail to reject H0	530.274	711.237	Fail to reject H0
Rocky Mountain	0.002%	213.198	181.145	249.128	Fail to reject H0	175.378	256.408	Fail to reject H0	164.469	271.037	Fail to reject H0
Southeast	0.002%	1044.998	960.354	1109.920	Fail to reject H0	946.779	1125.009	Fail to reject H0	920.622	1154.890	Fail to reject H0
Southwest	0.002%	380.693	332.059	422.214	Fail to reject H0	324.171	431.616	Fail to reject H0	309.122	450.387	Fail to reject H0

Section D: Discussion of Hypothesis Test Results

Single sample hypothesis testing for the sample means was conducted by using the student's t-distribution to calculate t-statistic values for each of the tested subgroups. The null hypothesis being tested was each subgroup's sample mean being equal to the mean of the overall sample, which yielded an alternative hypothesis that each subgroup's sample mean was not equal to the mean of the overall sample and resulted in a two-tailed test. At a confidence level of 90%, the null hypothesis was rejected for every region. However, at a confidence level of 95% and 99%, the null hypothesis failed to be rejected for the New England, Rocky Mountain, and Southwest Regions.

As for testing the sample variances, the chi-squared distribution was used to calculate a chi-squared statistic for each of the eight subgroups alongside a chi-squared upper and chi-squared lower value at each of the three confidence levels (90%, 95%, 99%). The null hypothesis being tested was each subgroup's sample variance being equal to the variance of the overall sample, which yielded an alternative hypothesis that each subgroup's sample variance was not equal to the variance of the overall sample and, similarly, resulted in a two-tailed test. Test results showed that for every region at every confidence level, the null hypothesis failed to be rejected.

Two Sample Confidence Intervals and Hypothesis Tests

Section A: Pair-wise Hypothesis Tests of Equal Variances

To determine whether variances across subgroups are equal or not, I needed to arrange my data into unique pairs to conduct pair-wise hypothesis tests. Since I had eight different subgroups, I ended up with twenty-eight unique pair combinations. Afterwards, I found the F-statistic and P-value of each pair using each region's sample variance and sample size.

Following these calculations, for each of the twenty-eight unique pairs, I failed to reject the null hypothesis since none of the pairs' P-values were greater than 0.95 or less than 0.05. As a result, I conducted the next round of tests, the hypothesis tests of subgroup sample means, under the assumption of equal variances.

Pair-wise Sample Variance Hypothesis Tests

	Variance	Sample Size
Far West	0.0016%	150
Great Lakes	0.0015%	435
Mideast	0.0015%	178
New England	0.0016%	67
Plains	0.0015%	618
Rocky Mountain	0.0016%	215
Southeast	0.0016%	1035
Southwest	0.0016%	377

Variance Test F-Stat

	Far West	Great Lakes	Mideast	New England	Plains	Rocky Mountain	Southeast
Great Lakes	1.0822						
Mideast	1.1032	1.0194					
New England	1.0323	0.9538	0.9357				
Plains	1.1022	1.0184	0.9991	1.0677			
Rocky Mountain	1.0320	0.9536	0.9355	0.9998	0.9364		
Southeast	1.0173	0.9400	0.9222	0.9855	0.9230	0.9858	
Southwest	1.0155	0.9383	0.9205	0.9837	0.9214	0.9840	0.9982

P-Values

	Far West	Great Lakes	Mideast	New England	Plains	Rocky Mountain	Southeast
Great Lakes	0.7302135						
Mideast	0.7747659	0.55329866					
New England	0.5492848	0.3962491	0.36029665				
Plains	0.7212797	0.55926033	0.51766444	0.658975918			
Rocky Mountain	0.6032976	0.32897391	0.26924253	0.491925893	0.271978278		
Southeast	0.5735043	0.28959119	0.23202029	0.454770538	0.234618337	0.45564563	
Southwest	0.5662481	0.22244857	0.16071177	0.418429508	0.163388957	0.419511277	0.486171593

Conclusion: Reject if P-Value is GREATER THAN 0.95 or LESS THAN 0.05

	Far West	Great Lakes	Mideast	New England	Plains	Rocky Mountain	Southeast
Great Lakes	F.T.R.						
Mideast	F.T.R.	F.T.R.					
New England	F.T.R.	F.T.R.	F.T.R.				
Plains	F.T.R.	F.T.R.	F.T.R.	F.T.R.			
Rocky Mountain	F.T.R.	F.T.R.	F.T.R.	F.T.R.	F.T.R.		
Southeast	F.T.R.	F.T.R.	F.T.R.	F.T.R.	F.T.R.	F.T.R.	
Southwest	F.T.R.	F.T.R.	F.T.R.	F.T.R.	F.T.R.	F.T.R.	F.T.R.

Section B: Pair-wise Confidence Intervals and Hypothesis Tests of Subgroup Means

Pair-wise Confidence Intervals for Differences in Sample Means

Confidence Intervals-95

		Far West	Great Lakes	Mideast	New England	Plains	Rocky Mountain	Southeast
Great Lakes	UCL	-0.0330%						
	LCL	-0.1471%						
Mideast	UCL	-0.1443%	-0.0744%					
	LCL	-0.2822%	-0.1720%					
New England	UCL	-0.2962%	-0.2302%	-0.1059%				
	LCL	-0.4871%	-0.3730%	-0.2510%				
Plains	UCL	-0.3096%	-0.2466%	-0.1042%	0.1101%			
	LCL	-0.4526%	-0.3355%	-0.2315%	-0.0890%			
Rocky Mountain	UCL	-0.1637%	-0.1045%	0.0375%	0.2537%	0.1946%		
	LCL	-0.3401%	-0.2192%	-0.1149%	0.0258%	0.0638%		
Southeast	UCL	-0.3189%	-0.2541%	-0.1123%	0.1002%	0.0362%	-0.0739%	
	LCL	-0.4490%	-0.3338%	-0.2292%	-0.0848%	-0.0419%	-0.1902%	
Southwest	UCL	-0.2566%	-0.1952%	-0.0528%	0.1624%	0.0998%	-0.1564%	0.0958%
	LCL	-0.4127%	-0.2941%	-0.1901%	-0.0485%	-0.0069%	-0.1564%	0.0027%

Pair-wise Hypothesis Tests for Differences in Sample Means

t-Statistics

	Far West	Great Lakes	Mideast	New England	Plains	Rocky Mountain	Southeast
Great Lakes	-3.098407						
Mideast	-6.083449	-4.96058989					
New England	-8.085187	-8.29713122	-4.8434981				
Plains	-10.46148	-12.848567	-5.1767466	0.208188805			
Rocky Mountain	-5.618481	-5.54457573	-0.9983859	2.41403479	3.875801129		
Southeast	-11.57721	-14.4714215	-5.7321066	0.163346909	-0.143328541	-4.456472699	
Southwest	-8.423611	-9.70614318	-3.476084	1.062167486	1.70850355	-2.207940946	2.076465656

P-Values

	Far West	Great Lakes	Mideast	New England	Plains	Rocky Mountain	Southeast
Great Lakes	0.0020393						
Mideast	3.295E-09	9.1215E-07					
New England	4.477E-14	9.9827E-16	2.2729E-06				
Plains	4.941E-24	3.3545E-35	2.8632E-07	0.835143614			
Rocky Mountain	3.841E-08	4.2932E-08	0.3187097	0.01641864	0.000114648		
Southeast	1.939E-29	1.8101E-44	1.2511E-08	0.870275307	0.886048207	9.07959E-06	
Southwest	3.513E-16	3.8167E-21	0.00054875	0.288739651	0.087855454	0.027632299	0.038031685

Conclusions: Reject if P-Value is LESS THAN 5%

	Far West	Great Lakes	Mideast	New England	Plains	Rocky Mountain	Southeast
Great Lakes	Reject						
Mideast	Reject	Reject					
New England	Reject	Reject	Reject				
Plains	Reject	Reject	Reject	F.T.R.			
Rocky Mountain	Reject	Reject	F.T.R.	Reject	Reject		
Southeast	Reject	Reject	Reject	F.T.R.	F.T.R.	Reject	
Southwest	Reject	Reject	F.T.R.	F.T.R.	F.T.R.	Reject	Reject

To find the confidence intervals of the subgroup sample means, I used the information found in the prior section, namely which pairs rejected or failed to reject the hypotheses. With this information, I calculated which pairs had assumed equal or unequal variances to conduct t-Tests, find P-Values, and form conclusions based on the acquired information. In the end, each pair rejected the null hypothesis except for the following pairs: Mideast-Rocky Mountain, Mideast-Southwest, New England-Plains, New England-Southeast, New England-Southwest, Plains-Southeast, and Plains-Southwest.

ANOVA Tests

Section A: Single Factor ANOVA Test for Joint Equality

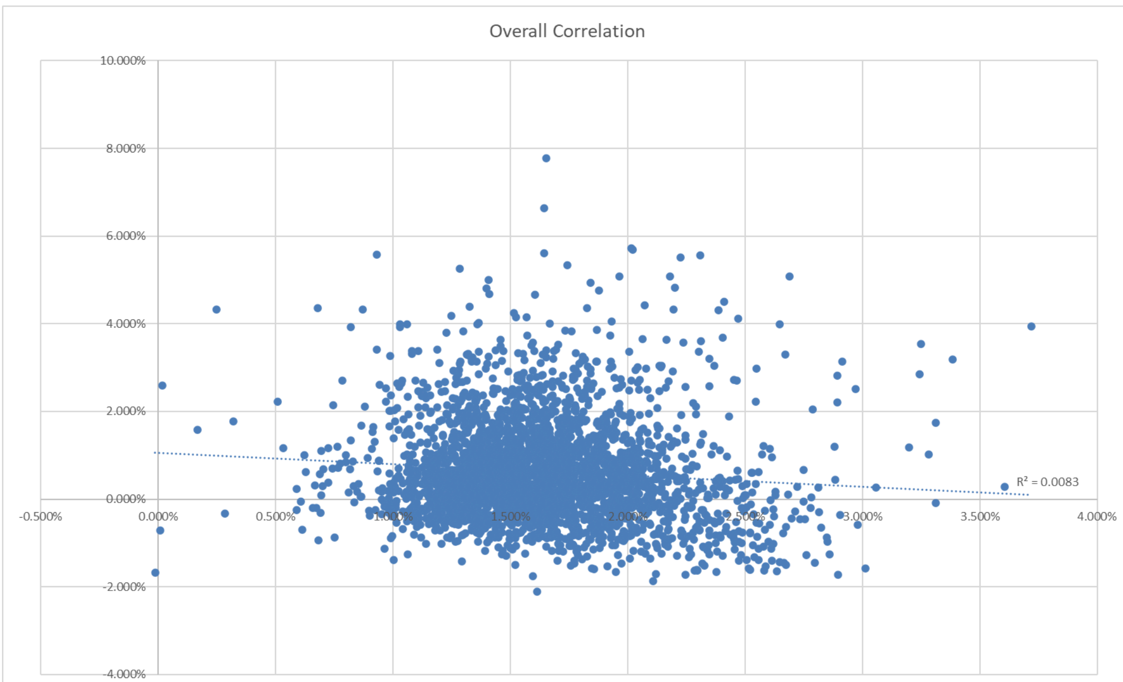
Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Far West	150	2.012492005	0.013416613	1.33949E-05		
Great Lakes	435	6.227894641	0.014316999	8.05402E-06		
Mideast	178	2.767710889	0.015548938	7.14404E-06		
New England	67	1.161326392	0.01733323	5.16329E-06		
Plains	618	10.64669634	0.017227664	1.66521E-05		
Rocky Mountain	215	3.426195261	0.015935792	2.08044E-05		
Southeast	1035	17.86016841	0.017256201	1.45567E-05		
Southwest	377	6.319792356	0.016763375	1.83428E-05		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.47%	7	0.0670%	46.9254	1.42E-63	2.0126
Within Groups	4.38%	3067	0.0014%			
Total	4.85%	3074				

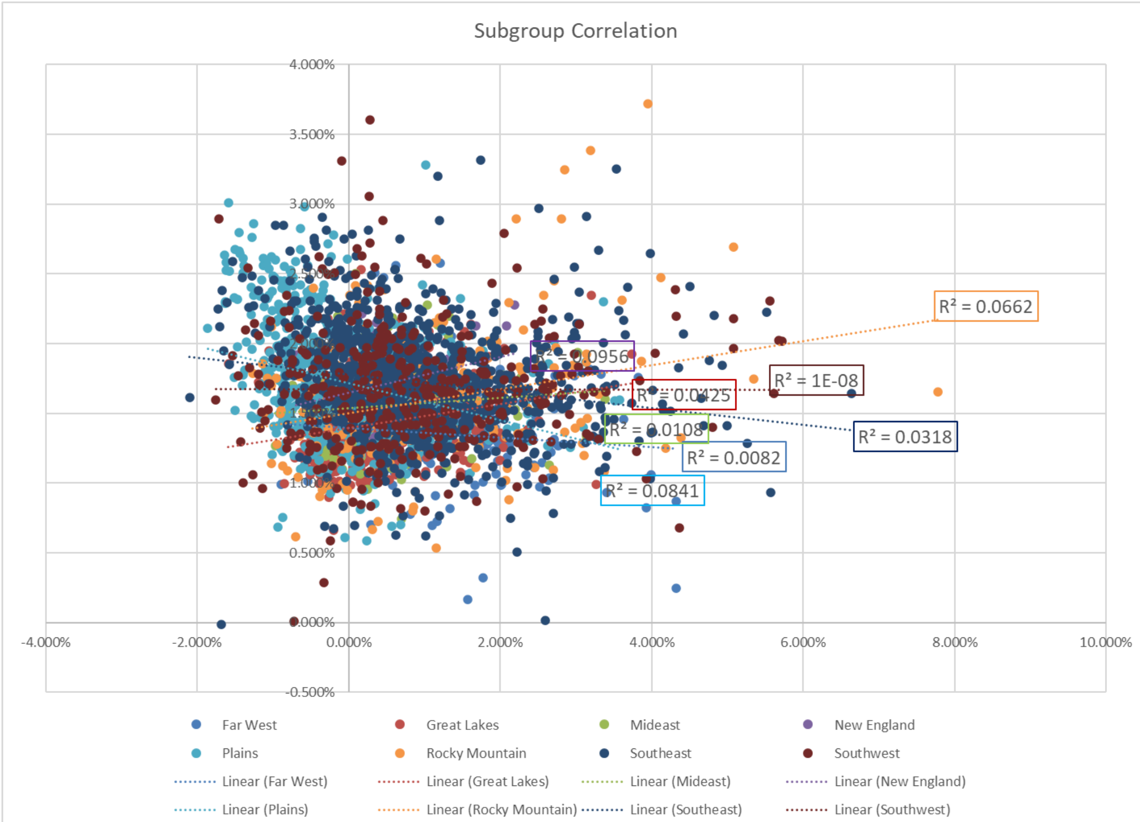
Section B: Discussion of ANOVA Results

While the other tests performed so far were done to determine the differences among all the sub-group sample means, the single factor ANOVA test differs in that it is intended to find the joint equality among subgroups. The null hypothesis being tested is the sample means across subgroups are equal. The alternative hypothesis being tested is that the subgroup sample means are unequal. In order to form this table, I used the built-in data analysis tool within Excel to conduct a single factor ANOVA test. The test results show the F-Calc value of 46. 9254 is greater than the F-Crit value of 2.0126. Therefore, I concluded that the null hypothesis is rejected. In other words, the sample means of the sub-groups are unequal.

Correlation Analysis

Section A: X-Y Scatter Plots





Section B: Hypothesis Tests of Overall and Subgroup Correlation Significance

Group	Correlation	R-Squared	Sample Size	zr
Overall	-0.0911	0.0083	3075	
Far West	-0.0802	0.0064	150	-0.0803
Great Lakes	-0.0785	0.0062	435	-0.0787
Mideast	-0.1256	0.0158	178	-0.1262
New England	-0.0890	0.0079	67	-0.0892
Plains	-0.1934	0.0374	618	-0.1959
Rocky Mountain	-0.0665	0.0044	215	-0.0666
Southeast	-0.0943	0.0089	1035	-0.0946
Southwest	-0.0974	0.0095	377	-0.0977

Tests of Individual Correlation Significance H0: ρ = 0				
Two-Tailed Tests, alpha =			0.05	
Group	t-Calc	t-Critical	Conclusion	
Overall	-5.0696	1.960736256	Reject	
Far West	-0.9785	1.976122494	Fail to reject	
Great Lakes	-1.6395	1.965457757	Fail to reject	
Mideast	-1.6791	1.973534388	Fail to reject	
New England	-0.7203	1.997137908	Fail to reject	
Plains	-4.8925	1.963822523	Reject	
Rocky Mountain	-0.9727	1.971163885	Fail to reject	
Southeast	-3.0448	1.962263119	Reject	
Southwest	-1.8951	1.966310161	Fail to reject	

Next, I conducted hypothesis tests to find which subgroup correlated to my chosen criterion and predictor variables. The null hypothesis being tested is that there is no correlation between the Per Capita Income Growth Rate and Population Growth Rate. The alternative hypothesis being tested is that there is correlation between the variables. From the test results, the overall, Plains, and Southeast regions reject the null hypothesis, while the other six regions fail to reject the null hypothesis.

Section C: Pair-wise Hypothesis Tests Among Subgroup Correlations

Group	Correlation	R-Squared	Sample Size	zr
Overall	-0.0911	0.0083	3075	
Far West	-0.0802	0.0064	150	-0.0803
Great Lakes	-0.0785	0.0062	435	-0.0787
Mideast	-0.1256	0.0158	178	-0.1262
New England	-0.0890	0.0079	67	-0.0892
Plains	-0.1934	0.0374	618	-0.1959
Rocky Mountain	-0.0665	0.0044	215	-0.0666
Southeast	-0.0943	0.0089	1035	-0.0946
Southwest	-0.0974	0.0095	377	-0.0977

Pair-wise Correlation Tests H0: pi = pj							
Z-Calc							
	Far West	Great Lakes	Mideast	New England	Plains	Rocky Mountain	Southeast
Great Lakes	-0.01718764						
Mideast	0.41009754	0.530355711					
New England	0.05926116	0.078513454	-0.253331496				
Plains	1.25830761	1.866382105	0.812850465	0.811980196			
Rocky Mountain	-0.12806533	-0.14434374	-0.583811113	-0.158598211	-1.623096046		
Southeast	0.16161502	0.277268267	-0.386953182	0.041707496	-1.988082367	0.37123051	
Southwest	0.17834157	0.269039185	-0.311403942	0.062733899	-1.496986113	0.361829223	0.051586714
Two-Tailed Test, alpha =		0.05					
Z-Critical	1.95996398						
Conclusions							
	Far West	Great Lakes	Mideast	New England	Plains	Rocky Mountain	Southeast
Great Lakes	Fail to reject						
Mideast	Fail to reject	Fail to reject					
New England	Fail to reject	Fail to reject	Fail to reject				
Plains	Fail to reject	Fail to reject	Fail to reject	Fail to reject			
Rocky Mountain	Fail to reject	Fail to reject	Fail to reject	Fail to reject	Fail to reject		
Southeast	Fail to reject	Fail to reject	Fail to reject	Fail to reject	Reject	Fail to reject	
Southwest	Fail to reject	Fail to reject	Fail to reject	Fail to reject	Fail to reject	Fail to reject	Fail to reject

To conduct the pair-wise degree of correlation tests, I calculated the zr transformation for each subgroup. Each subgroup's zr value is displayed in the table above. The null hypothesis being tested is that each pair's correlation values are equal to one another. The alternative hypothesis that they are not equal to one another. Using the zr values and respective degrees of freedom values, I calculated the Z-Calc values and found the appropriate Z-Critical value. Following these calculations, I concluded that each pair failed to reject the null hypothesis except for the Plains-Southeast pair, which rejected the null hypothesis.

Section D: Joint Equality Test for Subgroup Correlations

Joint Multi-Group Correlation Test, alpha = 0.05						
Group	Sample Size	Correlation	zr trans.	nj-3	(nj-3)*zr^2	(nj-3)*zr
Far West	150	0.0478	0.047852195	147	0.336605387	7.034272667
Great Lakes	435	0.0478	0.047794107	432	0.986807517	20.6470542
Mideast	178	0.0581	0.058168419	175	0.592123874	10.17947336
New England	67	0.0504	0.050394718	64	0.162536167	3.225261956
Plains	618	0.0614	0.061506377	615	2.326566197	37.82642213
Rocky Mountain	215	0.0497	0.049706895	212	0.523804395	10.53786182
Southeast	1035	0.0487	0.048755666	1032	2.453182606	50.3158469
Southwest	377	0.0491	0.049179362	374	0.904560024	18.39308155
Sums				3051	8.286186167	158.1592746
						25014.35614
	Chi-Sq Calc	0.087446036				
	Chi-Sq Crit	7.814727903				
	Conclusion	Fail to reject				

Lastly, I used this table to test for joint equality across the eight subgroups. The null hypothesis being tested is that the correlation between the Per Capita Income Growth Rate and Population Growth Rate is jointly equal across subgroups. The alternative hypothesis is that the correlation between the two variables is different across subgroups. From these tests, I determined the appropriate chi-squared calculation value and chi-squared critical value and determined that I was unable to reject the null hypothesis.

Conclusion

In summary, my initial hypothesis that the Far West Region would have the strongest correlation with Per Capita Income Growth Rate and Population Growth was shown to be incorrect. It was instead the Plains Region, not the Far West Region, that displayed the strongest correlation between Per Capita Income Growth Rate and Population Growth Rate.

Regarding possible limitations, the dataset analyzed was rather limited, only containing 3075 data points and breaking down the entire continental United States into eight regions. With a country as large in population as the United States, 3075 data points is not nearly enough to paint an accurate picture of the different regions within the U.S. and their respective Per Capita Income Growth Rates and Population Growth Rates. In addition, as discussed earlier in this report, there are a multitude of reasons which influence migration toward and away from various locations, including but not limited to climate, government policies, and crime rate.

As for extensions, I believe increasing the overall sample size by adding in data points for more cities, states, and regions could improve the overall accuracy of the data.

References

Camarota, S. A. (2021, June 1). *There is no evidence that population growth drives per capita economic growth in developed economies*. CIS.org. Retrieved February 25, 2023, from https://cis.org/Camarota/There-No-Evidence-Population-Growth-Drives-Capita-Economic-Growth-Developed-Economies