

**MG-212 Course Project**

From: Brandon Rank

To: Dr. Thomas Craig

Date: May 10, 2021

Subj: Analysis of Homeless and Unemployment Figures Across United States

---

**Section 1: Introduction/Overview**

Homelessness and unemployment have become a common problem in society over the past few decades, with constant evolution in our economy. With employment and the ability to afford housing a constant struggle for millions across America, events like the Covid-19 pandemic wreak havoc on families and individuals alike. It is important that we as members of society look at the trends of homelessness and unemployment over the past few decades to better understand how economic changes affect the state of these individuals' ability to find housing and employment opportunities. Living on the streets can be a dangerous time in someone's life, with rates of diseases like Tuberculosis, Hepatitis C, and HIV vastly elevated due to a lack of sanitary and safe shelter. In addition, these individuals face the elements every day. We have all heard stories of homeless sleeping on a bench during winter, only to never awaken again. As a result, it is essential that we continue to explore the data and new solutions to preventing this persistent problem across America.

**Section 2: Statement of Hypotheses**

While studying the data provided by the Bureau of Labor Statistics, the United States Interagency Council on Homelessness, and the Census Bureau, it will be interesting to look at how the unemployment rates, homeless counts and median household income figures are associated and how they have changed over the years. The homeless figures presented by the USICH are from a point-in-time count each year in each of the 50 states. While these figures represent a large portion of the homeless population, it is expected that these figures exceed those presented. The unemployment figures provided by the BLS were calculated from employment figures across the 50 states. Finally, the median household income figures for each year are provided by the Census Bureau. The following hypotheses will be tested to determine the trends and relationships among the data.

**Hypotheses:**

Null Hypothesis ( $H_0$ ): There is no difference in the homeless counts in 2019 between the geographic regions defined by the Census Bureau.

Alternative Hypothesis ( $H_1$ ): There is a difference in the homeless counts in 2019 between the geographic regions defined by the Census Bureau.

Null Hypothesis ( $H_0$ ): There is no association between homelessness counts, median household income and unemployment rates.

Alternative Hypothesis ( $H_1$ ): There is an association between homelessness counts, median household income, and unemployment rates.

Null Hypothesis ( $H_0$ ): The homeless counts in each geographic region from 2007 to 2019 have changed equally.

Alternative Hypothesis ( $H_1$ ): The homeless counts in each geographic region from 2007 to 2019 have not changed equally.

Null Hypothesis ( $H_0$ ): In 2019, the overall homeless populations across the age groups under 18, 18 – 24, and over 24 are equal.

Alternative Hypothesis ( $H_1$ ): In 2019, the overall homeless populations across the age groups under 18, 18 – 24, and over 24 are not equal.

**Section 3: Data and Analysis Methodology**

The data gathered is from the Bureau of Labor Statistics, the Census Bureau, and the USICH as mentioned before, with links provided on the last page of the memo. From the BLS, I gathered data on the unemployment rate from 2007 to 2019. In addition, from the USICH, I obtained a data set which includes the homeless figures from point-in-time counts across the United States. These figures include a breakdown of age, gender, sheltered vs. non-sheltered and overall figures from 2007 to 2019, however for the purposes of this project, I will only be using the overall, gender, and age breakdowns. Furthermore, the age and gender breakdown data are only provided for recent years, and therefore any comparisons using these specific breakdowns will only apply to those years, respectively.

The final piece of data I collected was from the Census Bureau, and contained the median household income from 2007-2019.

To test the hypotheses above, I will be using a Tukey comparison ANOVA test to compare the homeless counts across the geographic regions in 2019, a multiple regression analysis to determine if there is an association between homeless figures and the unemployment rate and median household income in 2019, a Tukey comparison ANOVA test to determine if there has been a change in the overall homeless counts across 2007 to 2019 in the geographic regions, and a final Tukey ANOVA test to determine if there is a difference in 2019 between the overall homeless counts for the age groups under 18, 18 – 24, and over 24. The Tukey ANOVA tests will be used due to their ability to determine the differences between various samples, in this case, between the different geographic regions and age groups. Due to the fact that Tukey tests cannot run with 51 levels, those being all 50 states and Washington D.C., it was necessary to perform the tests using geographic regions as defined by the Census Bureau. To perform these tests I used Minitab, with data collected and formatted in an Excel worksheet.

## **Section 4: Results**

### **One-Way ANOVA: Overall Homeless, 2019 vs. Region**

As stated above, the first hypothesis tested is that there is no difference in homeless counts across the geographic regions in 2019. From the test results of a Tukey comparison one-way ANOVA, we are 95% confident that there is no difference between the homeless counts in 2019 across the geographic regions. There is not statistically significant data to suggest that the counts across the regions are different as the regions all fall into the same cluster in the comparison, and therefore, we fail to reject the null hypothesis that there is no difference.

#### **Grouping Information Using the Tukey Method and 95% Confidence**

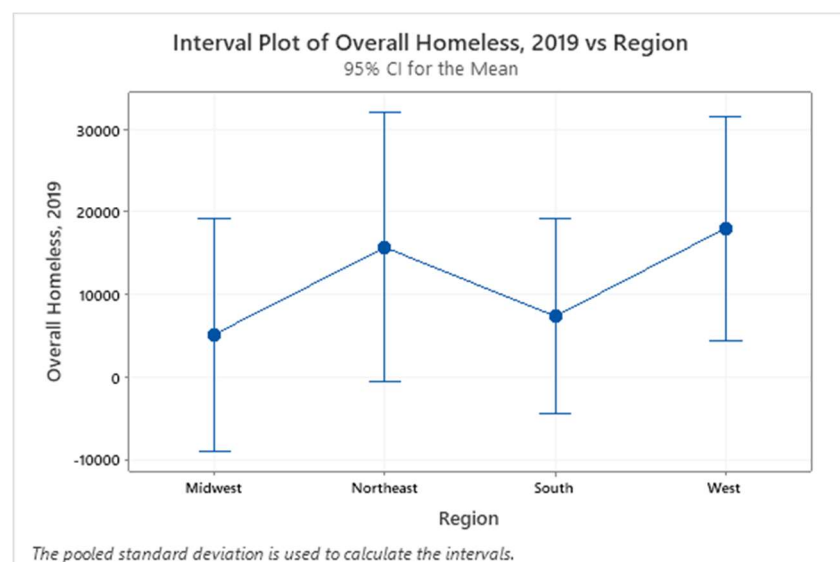
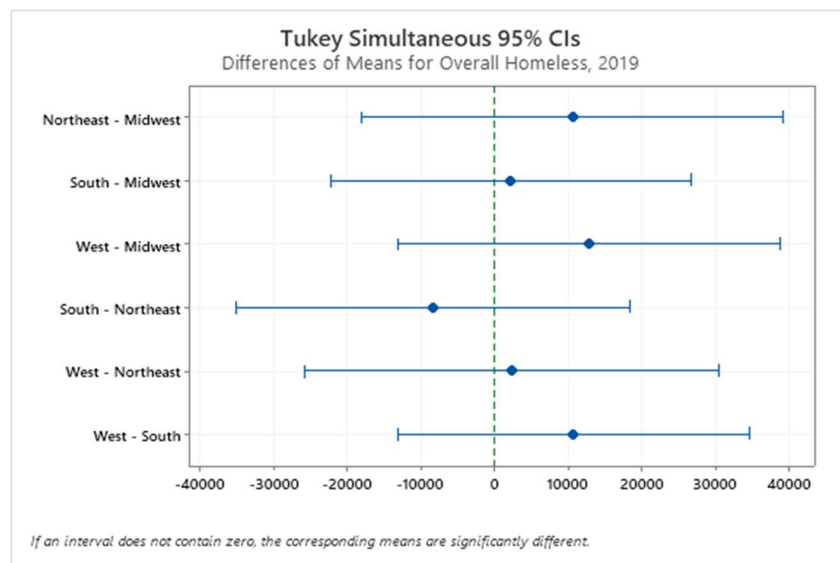
<u>Region</u>	<u>N</u>	<u>Mean</u>	<u>Grouping</u>
West	13	18008	A
Northeast	9	15700	A
South	17	7346	A
Midwest	12	5158	A

*Means that do not share a letter are significantly different.*

This is supported by the p-value of 0.482, which is greater than 0.05, indicating that loosely speaking, there is a 48.2% chance that the null hypothesis is true, well above the 5% threshold we are looking for. In addition, as apparent by the interval plot for the Tukey comparison, the intervals for each geographic region overlap, showing the graphical representation of the grouping provided above.

### Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Region	3	1473780509	491260170	0.83	0.482
Error	47	27715462695	589690696		
Total	50	29189243205			



### Multiple Regression Analysis: Overall Homeless, 2019 vs. MHI 2019, URATE 2019

Whilst testing for the association between the overall homeless count in 2019, the median household income in 2019, and the unemployment rate in 2019, it was discovered that there is statistically significant data to suggest that there is no association between the overall homeless count and the MHI and URATE in 2019. Both MHI and URATE for 2019 have p-values of 0.249 and 0.255 respectively, greater than 0.05, indicating that we fail to reject the null hypothesis that there is no association between the variables.

#### Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-30987	28667	-1.08	0.285	
MHI 2019	0.348	0.298	1.17	0.249	1.05
URATE 2019 (% of Pop)	4954	4304	1.15	0.255	1.05

In addition, only 4.39% of the variability in the overall homeless count in 2019 is explained by the MHI and URATE. This result is puzzling as it would be expected that the median household income and unemployment rate would have some sort of influence on the overall homeless count in 2019, however, the data suggests that this is not the case.

#### Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
24112.4	4.39%	0.41%	0.00%

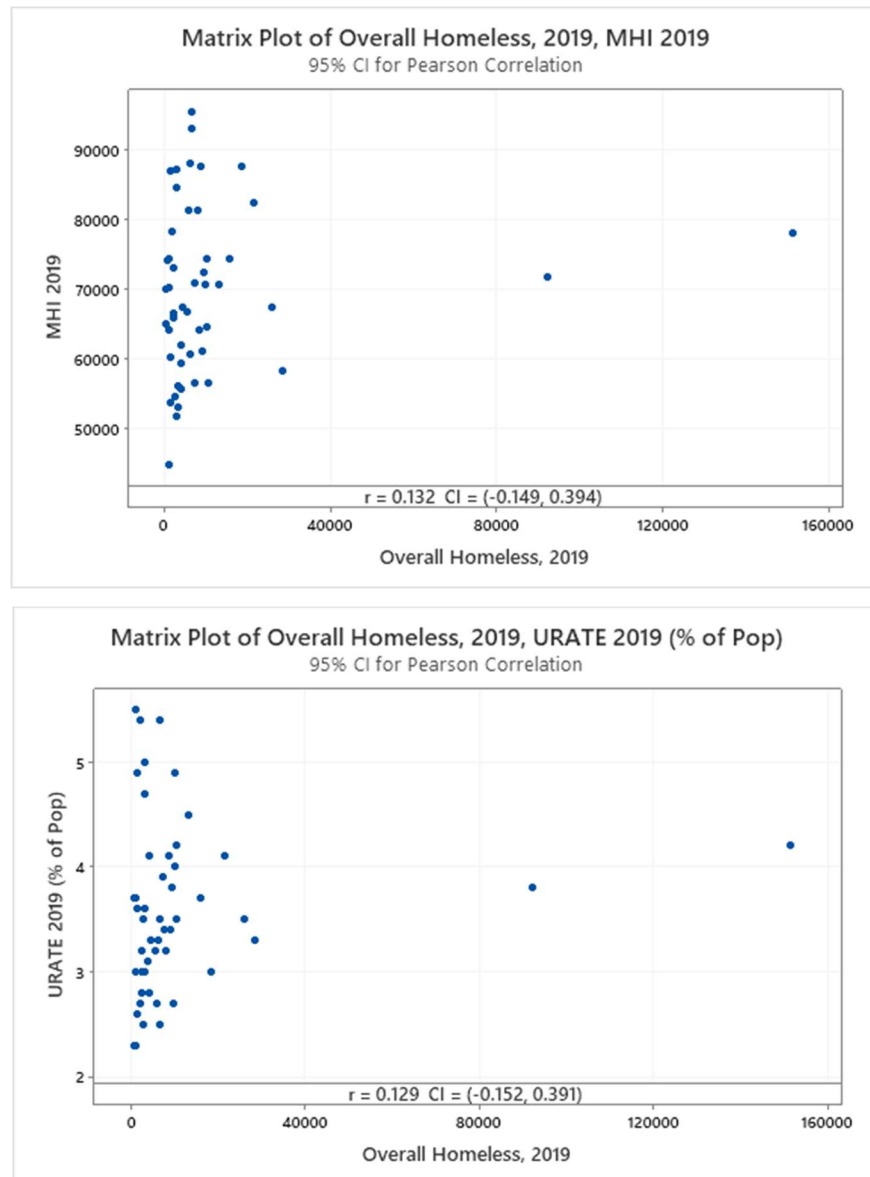
Finally, the regression equation calculated is shown below:

#### Regression Equation

$$\text{Overall Homeless, 2019} = -30987 + 0.348 \text{ MHI 2019} + 4954 \text{ URATE 2019 (\% of Pop)}$$

From the regression equation, it is apparent that there is a positive correlation between the MHI and overall homeless count, indicating that as the MHI increases in 2019, so does the overall homeless count, and a positive correlation between the URATE and overall homeless count, once again indicating that as the URATE increases, so does the overall homeless count.

Upon examination of the correlations between these variables, the results showed that both URATE and MHI held only weak positive relationships with the overall homeless count in 2019, further corroborating the results found in the multiple regression test.



### One-Way ANOVA: Change in OH, 2007-2019 vs. Region

The next test performed explored the hypothesis that the change in the overall homeless count from 2007 to 2019 across the geographic regions was equal. The results showed that there was statistically significant data to suggest that the homeless counts had not changed equally over the time period of 2007 to 2019, with a confidence level of 95%. This was supported by a p-value for the ANOVA Tukey test of 0.031, a value falling below our accepted threshold of 0.05, meaning we reject the null hypothesis.

### Analysis of Variance

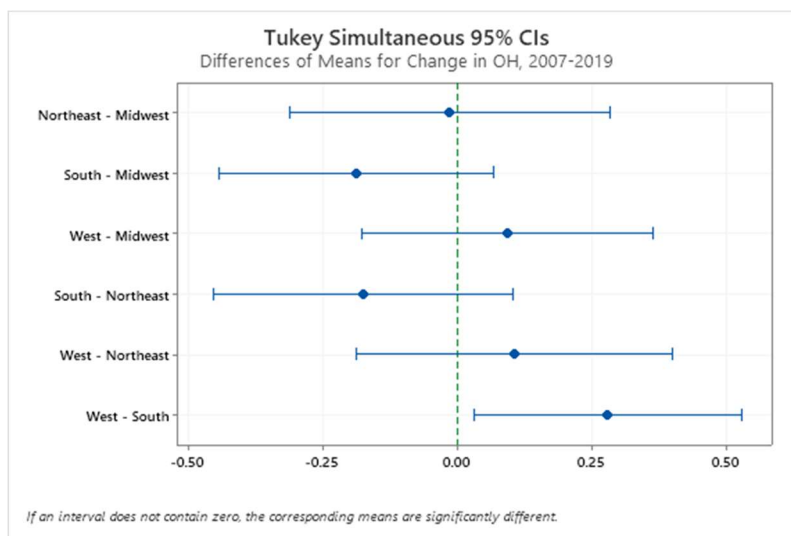
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Region	3	0.6194	0.20647	3.23	0.031
Error	47	3.0050	0.06394		
Total	50	3.6244			

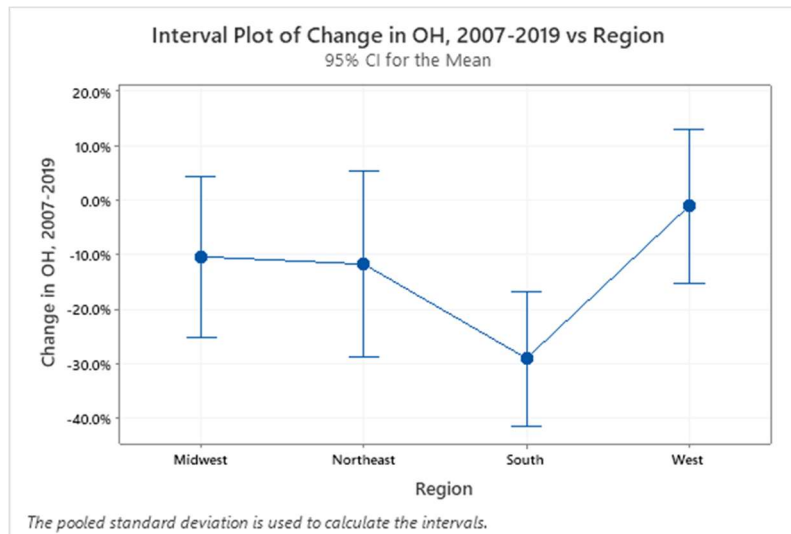
In addition to the p-value for the Tukey ANOVA comparison, the tests grouping clusters showed that there was a statistically significant difference between the West and South regions. Finally, the graphical representation shown by the interval plot shows that two clusters are created, the first containing the regions Midwest, Northeast and South, and the second containing the regions West, Northeast and Midwest. The South and West regions are distinctly different from each other, however, just barely.

### Grouping Information Using the Tukey Method and 95% Confidence

Region	N	Mean	Grouping
West	13	-0.0116	A
Midwest	12	-0.1047	A B
Northeast	9	-0.118	A B
South	17	-0.2910	B

*Means that do not share a letter are significantly different.*





### One-Way ANOVA: OH – Under 18, 18 – 24, Over 24

The final test I ran explored the difference in the homeless counts in 2019 of the three age groups. From this test, there was not statistically significant data to suggest that the overall homeless counts across the three age groups was equal in 2019. This was supported by the p-value of 0.003, less than the 0.05 allowed by our 95% confidence interval, indicating we must reject the null hypothesis.

#### Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	2	1505387984	752693992	6.10	0.003
Error	150	18509285099	123395234		
Total	152	20014673083			

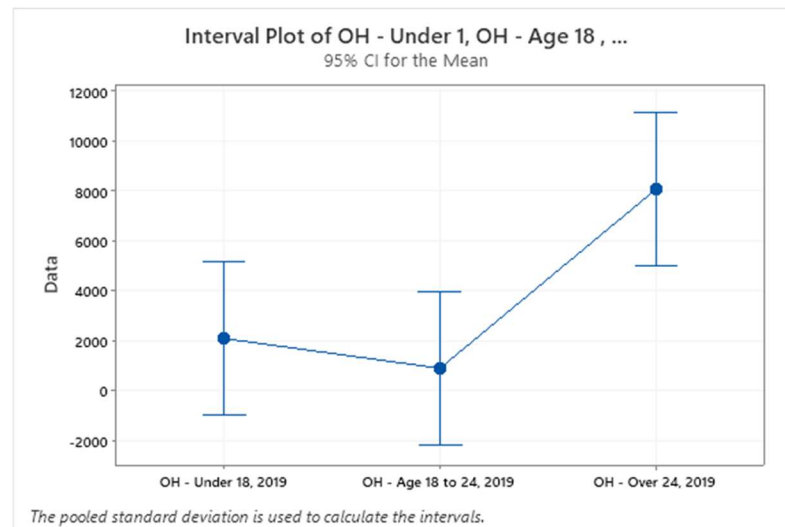
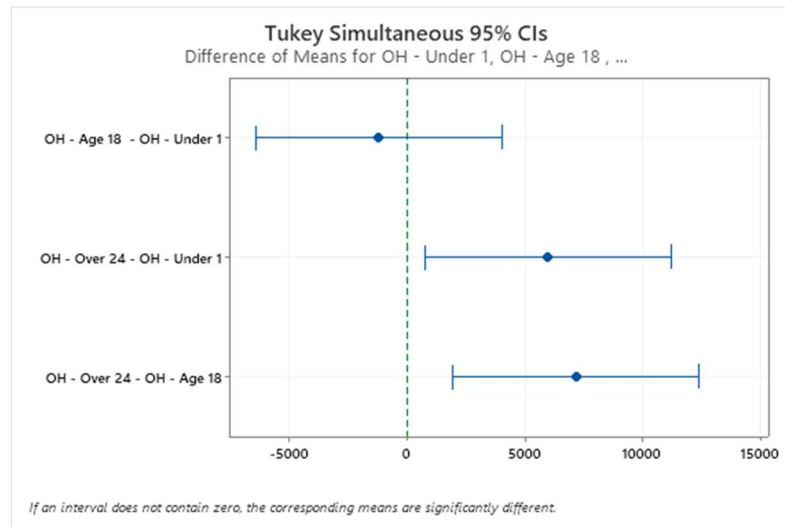
Further supporting this analysis was the Tukey comparison cluster displayed, showing that the overall homeless count in the age group over 24 was significantly different from those in the age groups under 18 and 18 – 24. Finally, the graphical representation displayed by the test showed that the two clusters, with the first containing the over 24 population, and the second containing those under 18 and in the 18 – 24 populations, were statistically different as explained above.



### Grouping Information Using the Tukey Method and 95% Confidence

Factor	N	Mean	Grouping
OH - Over 24, 2019	51	8057	A
OH - Under 18, 2019	51	2079	B
OH - Age 18 to 24, 2019	51	887	B

Means that do not share a letter are significantly different.



## Section 5: Discussion and Conclusion

The homeless crisis has persisted for decades despite efforts to solve the issue. To explore the issue, I gathered data on the homeless counts, unemployment rates, and median household income for all 50 states and Washington D.C., for the years 2007 to 2019. To test the relationship and changes in these figures over the time period, I

conducted three ANOVA Tukey comparison tests and a multiple regression analysis for the four hypotheses I developed for my analysis. After conducting all four tests, there was statistically significant data to suggest that the homeless counts in 2019 in the four regions, defined by the Census Bureau, were equal, that there was no association between homeless counts and the unemployment rate and median household income, and finally, that the change in homeless counts throughout the four regions was not equal from 2007 to 2019. While these results reflected an analysis of the data from all 50 states and the capital regarding homeless counts, URATE and MHI, the association between these three variables was not supported statistically according to the multiple regression test. This result was unexpected, as I had initially thought that there would be a strong association between the unemployment rate and homeless count specifically, however, the data suggested no significant association. As a result, the URATE and MHI provided no predictive power with regard to the homeless counts in each state. Therefore, to further the analysis of this data, further research and data collection into the factors that affect homelessness will be needed. Some points of future research may include the length of unemployment, the average cost of living in each state, and the availability of affordable housing in each state. In addition, it may be interesting to look at information regarding each state/region's economy, looking at potentially how the GDP of each state affects these other variables. Overall, it is clear that the homeless crisis is a real issue in the United States, and many of our attempts to solve the problem have not had the lasting effects we desire. Through the analysis of data on this topic, we can further explain the trends and look at how different variables may affect the homeless counts, focusing our legislative efforts on specific areas.

Works Cited

*Homelessness Statistics by State*, United States Interagency Council on Homelessness (USICH), [Homelessness Statistics by State | United States Interagency Council on Homelessness \(USICH\)](#).

“Local Area Unemployment Statistics Home Page.” *U.S. Bureau of Labor Statistics*, U.S. Bureau of Labor Statistics, [www.bls.gov/lau/](http://www.bls.gov/lau/).

“2019 Median Household Income in the United States.” *The United States Census Bureau*, The United States Census Bureau, 17 Sept. 2020, [www.census.gov/library/visualizations/interactive/2019-median-household-income.html](http://www.census.gov/library/visualizations/interactive/2019-median-household-income.html).