Analysis of Unemployment Rate:

Data by County

in Washington State and Oregon State from March 2021

TMATH 390 (Spring 2021)

Armstrong Kannika and Jessie De Jesus

June 4, 2021

Introduction:

The COVID-19 Pandemic has dominated the media with updates on the drastic impact it has had on the economy, society, and healthcare within the United States. Despite the all-encompassing effects of the virus and the pandemic, the true impact of rural communities seems to be underrepresented and uniquely impacted [1]. The vulnerability of the rural communities causes a disparity in unemployment rates relative to urban communities [1]. This is relevant as rural areas have higher rates of poverty, fewer job opportunities, and a lack of adequate health care relative to urban areas [1,2]. We will explore the unemployment rates between Washington and Oregon State to analyze the impact of unemployment between the rural and urban communities.

Method:

Data

The reported unemployment rate from the US Bureau of Labor Statistics (BLS), focusing on data in seventy-five counties in Washington and Oregon State, from March 2021.

The reported rural versus urban counties are determined by Census.gov. This measurement is defined as counties with less than fifty percent of the population living in rural areas are classified as *urban* counties. Therefore, those counties with less than fifty percent of the population living in urban areas are classified as *rural*.

Methodology

Given the small pool of data, this study utilizes a t-distribution of 2-sample comparing the two populations: the Population 1: All counties in Urban Area in Washington and Oregon, and the Population 2: All counties in Rural Area in Washington and Oregon

Analyzing the samples from two different groups can determine the difference between their population mean values. This requires calculating a point estimate for the difference between the mean values, and an estimate of the standard error of the difference between the mean values.

The estimated difference between mean is $\mu_1 - \mu_2$.

The Hypothesis Testing:

$$\mathbf{H_0: \mu_1 - \mu_2 = 0}$$

$$\mathbf{H_1: \mu_1 - \mu_2 \neq 0}$$

Where: μ_1 : the population mean for the unemployment rate in the urban area

μ_2 : the population mean for the unemployment rate in the rural area

The observed value of the t-test statistic.

$$t_{obs} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where: $\overline{x_1}$: sample mean of the unemployment rate in the urban area.

$\overline{x_2}$: sample mean of the unemployment rate in the rural area.

s_1 : standard deviation of the unemployment rate in the urban area.

s_2 : standard deviation of the unemployment rate in the rural area.

n_1 : number of the counties in the urban area.

n_2 : number of the counties in the rural area.

The degree of freedom for this case, which is two independent samples with unequal population variances. We used Welch's formula:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left[\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}\right]}$$

Where: s_1: standard deviation of the unemployment rate in the urban area.

s_2 : standard deviation of the unemployment rate in the rural area.

n_1 : number of the counties in the urban area.

n_2 : number of the counties in the rural area.

The confidence interval for the difference between the means. The confidence interval is 95% confidence interval.

$$(\overline{x_1} - \overline{x_2}) - t_{\frac{\alpha}{2}} * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\overline{x_1} - \overline{x_2}) + t_{\frac{\alpha}{2}} * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Where: $\overline{x_1}$: sample mean of the unemployment rate in the urban area.

$\overline{x_2}$: sample mean of the unemployment rate in the rural area.

$t_{\frac{\alpha}{2}}$: t-critical value for a 2-sided test.

α : the significant level, which is 0.05 for 95% confidence interval.

s_1 : standard deviation of the unemployment rate in the urban area.

s_2 : standard deviation of the unemployment rate in the rural area.

n_1 : number of the counties in the urban area.

n_2 : number of the counties in the rural area.

P-value for 2-tailed testing:

$$P - value = P(t \geq |t_{obs}|)$$

Where: t_{obs} is observed value of the t-test statistic

Use R to analyze the data and create a graphical model.

Result:

The histogram (Figure 1) shows that the center of the unemployment rate's distribution in the urban area is less than that of the rural area, with a notably different spread. The unemployment rate in the urban area falls between 4.5 - 9 percent while the unemployment rate in the rural area range is between 4 – 11 percent. Displayed in the histogram and the plot it is evident that the distribution of unemployment rate in both categories follow the normal distribution.

The QQ plot (Figure 3) shows that the distribution of the unemployment rate in the urban area and the rural area in Washington and Oregon in March 2021 follow the normal distribution. But the slope of the QQ plot in the rural area is slightly different from the urban area, creating a similar trend in standard deviation.

The spread of the boxplot (Figure 2), evident by the range of the data shows that there is more variability in the unemployment rate's distribution in the rural area (range = 5.9 percent) in comparison to the unemployment rate's distribution in the urban area (range = 4.3 percent). On the other hand, if we look at the IQR, which measures the variability only among the middle 50% of the distribution, we see more spread in the unemployment rate in the rural area (IQR = 1.8) than the unemployment rate in the urban area (IQR = 1.2).

Furthermore, the center of the boxplot plot shows that the distribution of unemployment rate in the rural area is higher than in the urban area. This is supported by the numeric measure. The median unemployment rate's distribution in the urban area is 6.75 percent, which is lower than in the rural area which is 7.4 percent. In addition, the third quartile of the unemployment rate's distribution in the urban area listed at 7.1 percent is lower than the median of the unemployment rate's distribution in the rural area. Therefore, we

conclude that the general unemployment rate in the urban area is lower than that of the rural area.

Table 1. Summary Statistics for Measured Data.

The County Category	n	Mean	Median	Standard Deviation	Q_1	Q_2	Q_3	IQR
Urban	52	6.61	6.75	1.00	5.90	6.75	7.10	1.20
Rural	23	7.33	7.40	1.41	6.30	7.40	8.10	1.80

Table 2. Confidence Interval Estimation and Hypothesis Test for the Difference Between the Mean of Urban Counties and Rural Counties; Welch's degree of freedom is 32.

Mean difference	Standard error for the difference	Confidence interval for the difference (95%CI)	t_{obs}	p-value	Decision
-0.72	0.325	-1.3832 and - 0.554	-2.207	0.0346	Reject H_0

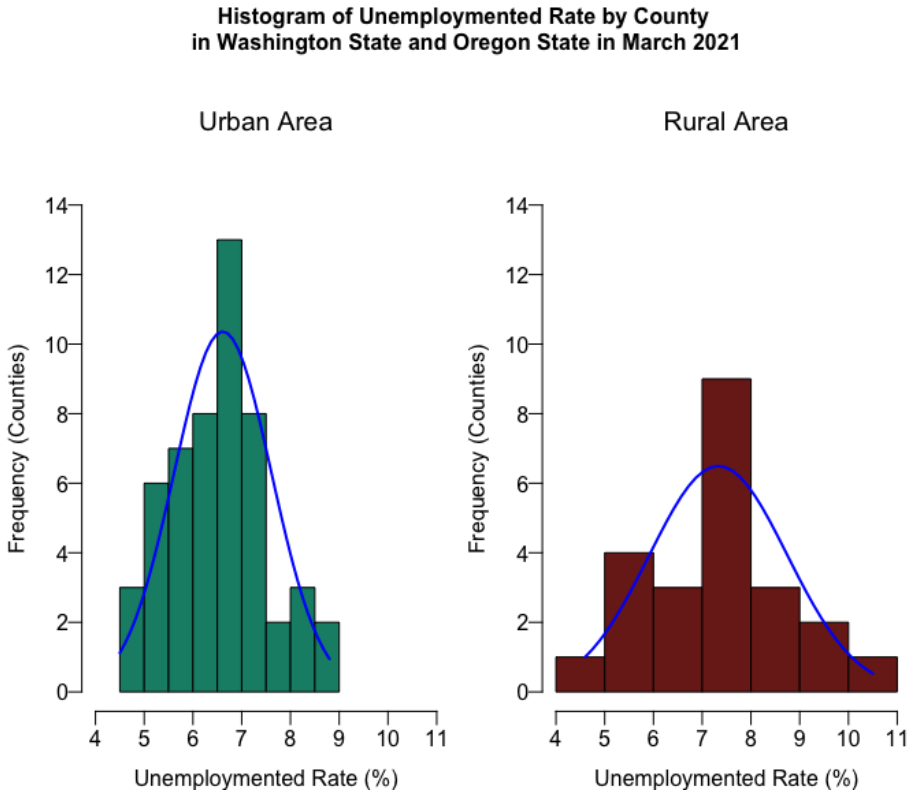

Figure 1. Histogram of Unemployment Rate by County in Washington State and Oregon State in March 2021

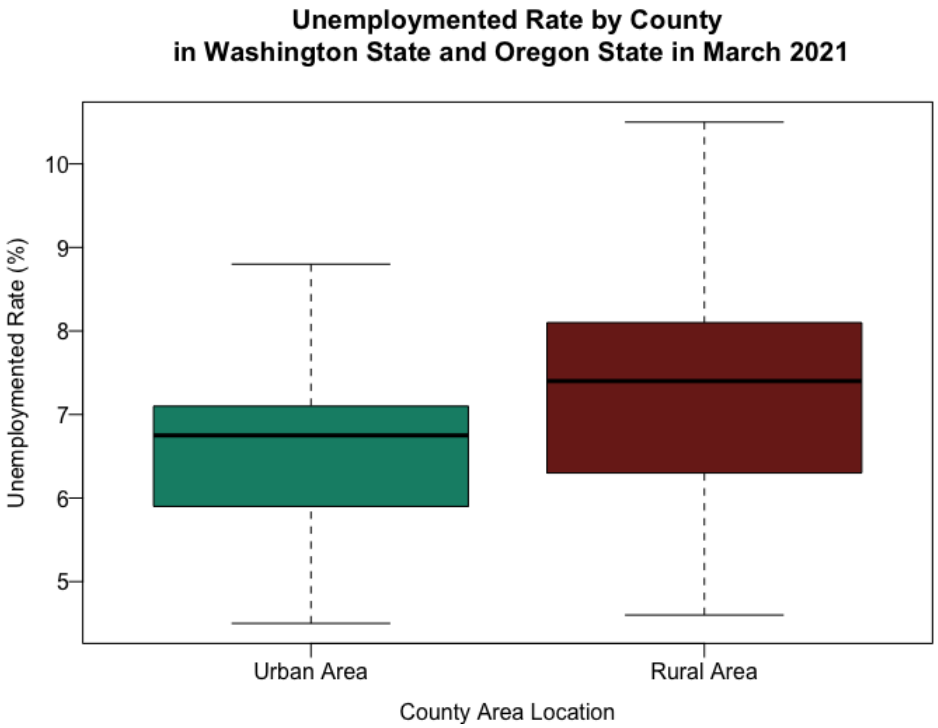

Figure 2. Boxplot of Unemployment Rate by County in Washington State and Oregon State in March 2021

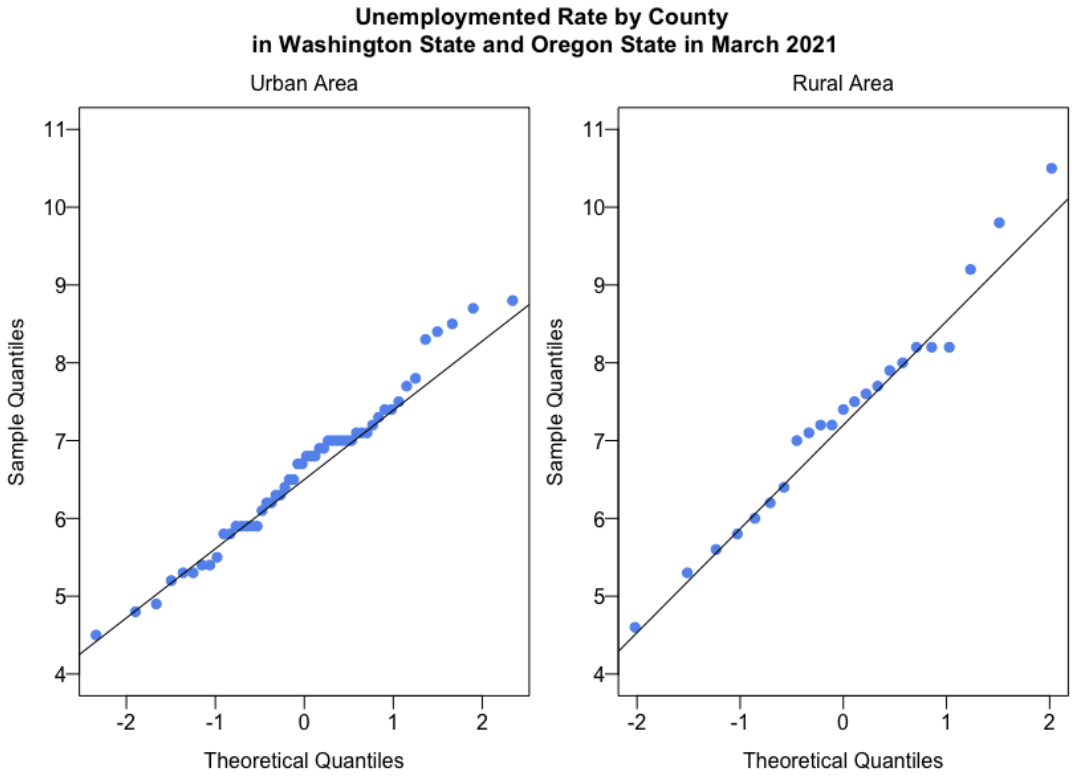

Figure 3. QQ-Plot of Unemployment Rate by County in Washington State and Oregon State in March 2021

Discussion:

We are 95% confident that the true difference in mean of the unemployment rate in the urban counties and the rural counties in Washington State and Oregon State in March 2021 is somewhere between -1.3832 and - 0.554. As the interval does not include 0, it can be concluded that the true mean values do differ.

The p-value is 0.0346, less than the significance level of 0.05, creating sufficient evidence to reject the null hypothesis. It can therefore conclude that the true difference in mean of the unemployment rate in the urban counties and the rural counties in Washington State and Oregon State, in March 2021 are different.

The result of this study shows that the unemployment rate in March 2021, in the urban counties is lower than that of the rural county counterparts in Washington State and Oregon State. Despite the outcome of our study, there are variables such as sample size and length of

data collection that can impact the results. Given the opportunity to expand the data throughout the pacific northwest region of the United States to pool a larger data set, our data may be sounder. In spite of the interest of the impact of the COVID 19 Pandemic on urban versus rural communities, the data collected is insufficient to be able to determine this impact. Nonetheless, the study can highlight the general disparity of unemployment rates in the rural and urban counties amongst Washington and Oregon State.

References:

1. Lakhani HV, Pillai SS, Zehra M, Sharma I, Sodhi K. Systematic Review of Clinical Insights into Novel Coronavirus (CoVID-19) Pandemic: Persisting Challenges in U.S. Rural Population. International Journal of Environmental Research and Public Health. 2020; 17(12):4279. https://doi.org/10.3390/ijerph17124279

2. C. Henning-Smith, The unique impact of COVID-19 on older adults in rural areas. J. Aging Soc. Policy **32**, 396–402 (2020) https://doi.org/10.1080/08959420.2020.1770036