Stats 506 Final Project Report

Alyssa Yang

Introduction

Mental health disorders such as anxiety, bipolar disorder, and depression have grown increasingly prominent in recent years, and with them have arrived public health challenges. These can have severe implications on an individual's quality of life, both impeding their productivity and increasing healthcare costs. Thus, understanding the factors that influence the prominence of these mental health disorders is crucial in providing preventative measures and more support to those impacted by them.

One of these potential factors is unemployment due to the stress and uncertainty of financial strain. Thus, in this project, I aim to explore the relationship between unemployment rates and mental health disorder prevalence. More specifically, I want to determine whether there is a relationship between unemployment rates and the percentage of Medicare beneficiaries that have mental health disorders by state from 2018-2022, focusing on anxiety, bipolar disorder, other depressive mood disorders, and major depressive affective disorder. By examining the relationships between these factors, this project seeks to contribute to a deeper understanding of how economic conditions, such as unemployment, may influence mental health outcomes.

Data and Methodology

In order to explore this relationship, I will be using the "Medicare Provider Utilization and Payment Data" and "Unemployment in America per US State" datasets.

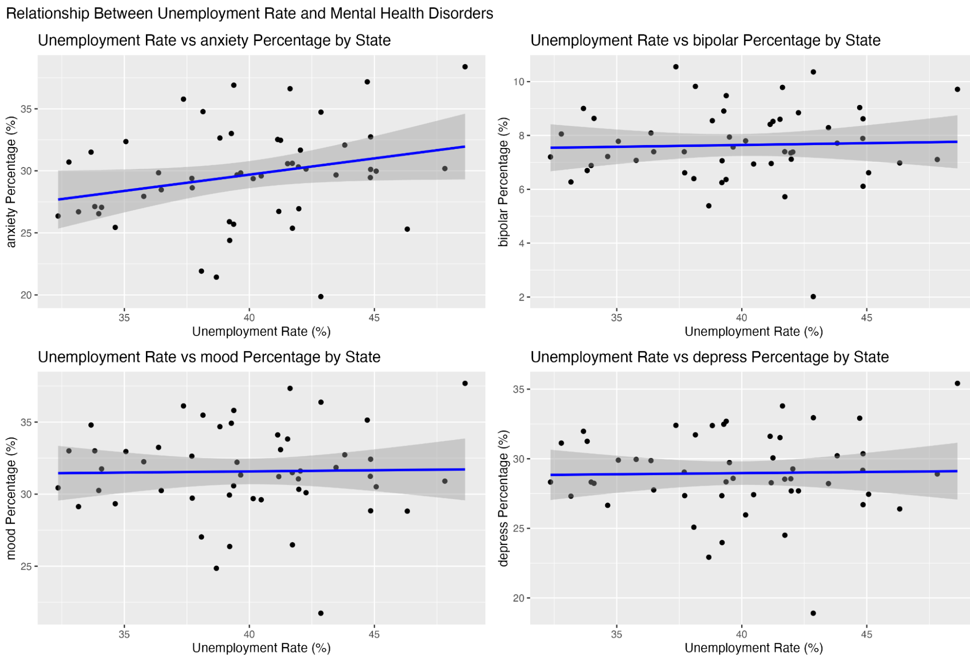
I utilized the Medicare data from 2018-2022 and included variables pertaining to the FIPS code, total number of beneficiaries, and the percentage of beneficiaries with each mental health disorder. To calculate the total percentage of beneficiaries with each disorder by state, I converted the percentages to raw counts, replaced missing values with the median, aggregated the data by summing the totals, and then recalculated the percentages.

For the unemployment data, I did not include institutionalized civilians as they may have limited access to the same resources or job opportunities as those who are non-institutionalized, and thus, including them could potentially skew the results. I included variables pertaining to the FIPS code, year, total number of civilians, percentage of civilians eligible for employment, and percentage of employed civilians. I then converted the last variable to those who are unemployed, and to match the scope of the Medicare dataset, I restricted it to only the years 2018-2022. Finally, I calculated the total percentage of people per state who are employed and who are eligible for employment using the same process as the Medicare data above.

To analyze the data, I performed a left join of the unemployment dataset onto the Medicare dataset, using the FIPS codes as the index and keeping only US states (FIPS codes 01-56). I then created scatterplots to visualize the relationships between unemployment rates and the prevalence of each mental health disorder by state. After this, I created linear regression models to predict the percentage of Medicare beneficiaries with each mental health disorder, using state unemployment rates as the primary predictor. To account for potential variations in unemployment rates due to eligibility factors, I also included the percentage of civilians eligible for employment as an additional predictor.

Results

Below are the visualizations for the relationships between unemployment rate and prevalence of each mental health disorder by state.



We can see that for anxiety, there is a slight positive relationship, suggesting that higher unemployment rates may be associated with a greater percentage of individuals experiencing anxiety within a state. However, for all other mental health disorders, unemployment rates appear to have no relationship with the percentage of individuals affected. The regression lines for them are nearly flat, indicating that changes in unemployment rates do not significantly impact the percentage of people with that disorder.

The results of the OLS models showed that none of the predictors were statistically significant at the **α** = 0.05 level. The p-values for the coefficients representing unemployment rates and employment eligibility percentages were all above 0.05, indicating that these factors do not have a statistically significant effect on the prevalence of mental health disorders across states.

To further explore potential relationships, I labeled each state by region (South, West, Midwest, and Northeast) and constructed a mixed-effects model using region as a grouping factor. Despite accounting for additional regional variability, the t-values for all coefficients in this model remained greater than 0.05, leading to the same conclusion as above: unemployment rates and employment eligibility percentages do not significantly affect the prevalence of mental health disorders. Detailed outputs for all models are included in the Appendix below.

Conclusion

Based on the analysis of the Medicare and unemployment datasets, we conclude that unemployment rates, even when accounting for work eligibility, do not significantly impact the prevalence of anxiety, bipolar disorder, depressive mood disorders, or major depressive affective disorder.

However, this result may be influenced by the absence of other factors that may also contribute to the prevalence of mental health disorders such as access to healthcare and other socioeconomic variables like income or state poverty rates. As unemployment rates alone do not account for much of the variability (evidenced by the results above), incorporating these additional factors into the analysis may yield different results. Thus, to further explore these relationships, future research could incorporate these additional factors, and addressing them may aid in more targeted interventions to support mental health outcomes in impacted populations.

References

[1] Medicare Provider Utilization and Payment Data:
https://data.cms.gov/provider-summary-by-type-of-service/medicare-physician-other-practitioners/medicare-physician-other-practitioners-by-provider
[2] Unemployment in America per US State:
https://www.kaggle.com/datasets/justin2028/unemployment-in-america-per-us-state

Appendix

GitHub repository link: https://github.com/alyssawyang/stats506-final-project
OLS models:

```
Call:
lm(formula = anxiety_pct ~ unemployed_pct + eligible_pct, data = combined_data)

Residuals:
     Min       1Q   Median       3Q      Max
-10.9653  -1.3319  -0.1909   2.3607   7.1120

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    23.99531  114.16918   0.210    0.835
unemployed_pct  0.23920    1.08317   0.221    0.826
eligible_pct   -0.05674    1.13072  -0.050    0.960

Residual standard error: 3.927 on 41 degrees of freedom
Multiple R-squared:  0.08814,	Adjusted R-squared:  0.04365
F-statistic: 1.981 on 2 and 41 DF,  p-value: 0.1509
```

```
Call:
lm(formula = bipolar_pct ~ unemployed_pct + eligible_pct, data = combined_data)

Residuals:
    Min      1Q  Median      3Q     Max
-5.8384 -0.7234 -0.0230  0.7723  2.5568

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    -59.1779    42.1637  -1.404    0.168
unemployed_pct   0.6480     0.4000   1.620    0.113
eligible_pct     0.6514     0.4176   1.560    0.126

Residual standard error: 1.45 on 41 degrees of freedom
Multiple R-squared:  0.06247,	Adjusted R-squared:  0.01674
F-statistic: 1.366 on 2 and 41 DF,  p-value: 0.2665
```

```
Call:
lm(formula = mood_pct ~ unemployed_pct + eligible_pct, data = combined_data)

Residuals:
    Min      1Q  Median      3Q     Max
-10.206  -1.555   0.002   1.631   5.539

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    30.222366  91.395237   0.331    0.743
unemployed_pct  0.037574   0.867105   0.043    0.966
eligible_pct    0.001823   0.905171   0.002    0.998

Residual standard error: 3.144 on 41 degrees of freedom
Multiple R-squared:  0.00225,	Adjusted R-squared:  -0.04642
F-statistic: 0.04623 on 2 and 41 DF,  p-value: 0.9549
```

```
Call:
lm(formula = depress_pct ~ unemployed_pct + eligible_pct, data = combined_data)

Residuals:
     Min       1Q   Median       3Q      Max
-10.3697  -1.1796  -0.2174   1.9625   5.7456

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     78.8360    87.3500   0.903    0.372
unemployed_pct  -0.4476     0.8287  -0.540    0.592
eligible_pct    -0.5041     0.8651  -0.583    0.563

Residual standard error: 3.005 on 41 degrees of freedom
Multiple R-squared:  0.01001,	Adjusted R-squared:  -0.03829
F-statistic: 0.2072 on 2 and 41 DF,  p-value: 0.8137
```

Mixed-effects model:

```
Linear mixed model fit by REML ['lmerMod']
Formula: anxiety_pct ~ unemployed_pct + eligible_pct + (1 | region)
   Data: combined_data

REML criterion at convergence: 259.8

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.29884 -0.65129  0.02465  0.60837  1.96107

Random effects:
 Groups   Name        Variance Std.Dev.
 region   (Intercept) 9.871    3.142
 Residual             8.477    2.912
Number of obs: 51, groups:  region, 4

Fixed effects:
               Estimate Std. Error t value
(Intercept)    12.85679   81.46009   0.158
unemployed_pct  0.35747    0.77123   0.463
eligible_pct    0.04187    0.80822   0.052

Correlation of Fixed Effects:
            (Intr) unmpl_
unmplyd_pct -0.994
eligibl_pct -0.998  0.986
```