**OMIS 645 - 1 Applied Business Analytics SAS**

Project Report

**Ajay Rama Raju Datla (Z1979893)**

**Vaishnavi Peddiraj (Z1979763)**

**Gayathri Parupalli (Z2006525)**

**Professor Kishen Iyenger**

**12/04/2023**

**SOCIOECONOMIC ANALYSIS OF COUNTIES IN UNITED STATES**

**Introduction**

The purpose of this project is to analyze various socioeconomic factors in counties across United States and understand their relationships. The dataset includes information on population, poverty rate, home ownership, multi-unit housing, unemployment rate, education levels, per capita income, median household income, the number of people working, and smoking ban status.

In this project, a diverse set of variables was employed, categorized into different types based on their nature and measurement scales. Categorical variables, such as "State," "Metro," and "Smoking Ban," provided descriptive information about the counties' geographical location, metropolitan status, and smoking policies. Continuous variables, including "Population," "Per Capita Income," "Median Household Income," and "No of People Working," offered numerical data with a true zero point, allowing for precise quantitative comparisons. Discrete variables, such as "Poverty Rate," "Homeownership," "Multi-Unit," and "Unemployment Rate," represented percentages and proportions related to socioeconomic factors.

**Data Cleaning**

After searching over all the government datasets, we have finally found a dataset that includes the information of all the counties located in United States. Our dataset mainly contains Economic data like median household income, per-capita income, Unemployment rate, home ownership, poverty, no of people working, etc., But we have also included some other interesting factors that could influence the above economic factors like Health information like smoking ban, urbanization information like multi unit and metro. Education information like median education.

**Sample of our dataset:**

A screenshot of a computer

Description automatically generated

1. **ANOVA for Median Educational Level and Per Capita Income:**

To assess the impact of median education level (Categorical, independent) on per capita income(numerical, dependent), we have conducted analysis of variance (ANOVA). As the p-value is less than alpha (0.05), we reject the null and conclude that not all means are equal. The results revealed a statistically significant model, indicating that the means of different educational attainment groups are not equal. From the Anova table, we can say that group mean of per capita income of counties having bachelor’s degree, some college degree, high school diploma (hs\_diploma) and below high school (below\_hs) are different from each other.

A screenshot of a computer

Description automatically generated

1. **Correlation Analysis of per capita income with poverty, Homeownership, unemployment rate and median household income:**

We investigated the correlations between several key variables. Notably, we explored whether poverty, home ownership, unemployment rate, and median household income are correlated with per capita income. The findings indicate that, except for home ownership, all variables exhibit a correlation with per capita income. Specifically, poverty and unemployment\_rate demonstrate a negative relationship with per capita income, suggesting that higher rates of poverty and unemployment are associated with lower per capita income. Home ownership has very weak correlation with per capita income. Median household income has strong positive correlation with per capita income and poverty has strong negative correlation.

A screenshot of a computer

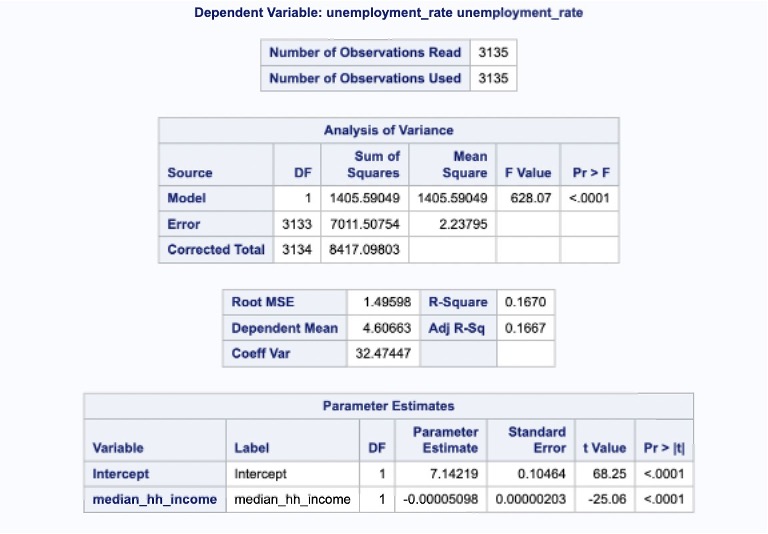
Description automatically generated

1. **Linear Regression:**

**Independent variable:** unemployment\_rate (Numerical)

**Dependent Variable:** median\_hh\_income (Numerical)

The analysis focused on examining the relationship between the independent variable, unemployment rate, and the dependent variable, median household income. With an F value that is less than the predetermined alpha level, we infer the significance of the model. The R-square value of 0.167 indicates that approximately 16.7% of the variation in median household income is explained by changes in the unemployment rate within the county. The observed negative coefficient of -0.00005 suggests that for each additional unit increase in the unemployment rate, the median household income of the county decreases by 0.00005. This finding underscores the influence of unemployment on median household income, providing a quantitative insight into the extent of its impact within the context of the analyzed data.



1. **Correlation of poverty, home ownership, unemployment rate, per capita income Influencing Median Household Income:**

We extended our analysis to examine factors influencing median household income. By correlating poverty, home ownership, multi-unit housing, unemployment rate, and per capita income with median household income, we found that all variables, except for home ownership, show a correlation with median household income. Notably, poverty and unemployment rate exhibit a negative relationship with median household income with poverty being strong and unemployment rate being weak indicating that higher levels of poverty and unemployment are associated with lower median household income. Per capita income and multiunit has positive relationship with median household income, while per capita income has strong correlation and multi-unit has weak.

A screenshot of a computer

Description automatically generated

1. **Linear regression to find relationship between homeownership and poverty:**

**Dependent variable:** Poverty

**Independent variable:** Homeownership

We delved into the intricate relationship between homeownership and poverty levels, uncovering a significant negative relationship. The linear regression results indicated that as homeownership rates increased, there was a corresponding 27% decrease in poverty. This regression analysis not only reaffirms the statistical significance of the negative relation but also offers a quantitative measure of the impact of homeownership on poverty. These findings underscore the socioeconomic importance of promoting homeownership and present valuable insights for policymakers aiming to design targeted interventions to alleviate poverty and foster economic stability within communities. The R-Square value indicates that the model predicts that 10.85 change in poverty is explained by homeownership.

A screenshot of a computer

Description automatically generated

1. **Moderating Effect of Smoking Ban on Poverty and Median Household Income:**

**Moderator:** Smoking\_ban

**Independent variable:** poverty

**Dependent variable:** median\_hh\_income

Our investigation revealed a dynamic in the socioeconomic landscape of counties, uncovering that smoking\_ban status serves as a crucial moderator in the relationship between poverty and median\_hh\_income. Analyzing the interplay, we found that counties experiencing poverty without smoking regulations exhibited significantly lower median household incomes compared to those with both poverty and a partial smoking\_ban. This underscores the moderating influence of smoking\_ban policies, suggesting that their absence exacerbates the economic challenges associated with poverty. These findings emphasize the potential positive impact of implementing or strengthening smoking regulations in regions facing economic adversity, providing policymakers with valuable insights for targeted interventions aimed at improving both health and economic outcomes.

A screenshot of a computer

Description automatically generated

1. **Chi-Square test of independence:**

Our analysis delved into the potential dependence between the smoking\_ban status and median\_education level in the counties under study. The chi-square test is a powerful statistical tool used to assess the independence or dependence of categorical variables. In our case, the calculated chi-square (0.00005) < alpha (0.05), a significant model. The significance level of the test allows us to reject the null hypothesis, suggesting that smoking\_ban status and median\_education level in these counties are dependent.

A screenshot of a graph

Description automatically generatedA screenshot of a calculator

Description automatically generatedA screenshot of a computer

Description automatically generated

**Conclusion:**

In conclusion, this project offers a comprehensive examination of socioeconomic factors in United states counties, providing valuable insights into the complex relationships between various variables. From uncovering correlations to exploring predictive models and moderating effects, the analysis contributes to a deeper understanding of the socioeconomic landscape. These findings can inform policymakers, researchers, and community stakeholders in developing targeted interventions and strategies to address the multifaceted challenges faced by different counties in United states.