Income-Depression Relationship Among Japanese During COVID-19 Pandemic

Sooyeon Oh, Sicheng Qian, Hongjian Wang, Jingchao Yang

GitHub Link

Abstract

This study examines the relationship between household income and depression severity among Japanese adults during the COVID-19 pandemic using data from 2,708 participants. Depression severity was assessed with the Patient Health Questionnaire-9 (PHQ-9), and household income was analyzed as a continuous variable. Proportional odds logistic regression showed a significant negative association between household income and depression severity (OR = 0.93, p = 0.046). Machine learning models including Extreme Gradient Boosting (XGBoost), highlighted additional factors such as age, marital status, and economic impact. These findings emphasize the role of socioeconomic factors in mental health, highlighting the need for targeted interventions during crises.

Introduction

The COVID-19 pandemic has significantly impacted individuals' mental health globally, driven by uncertainties and quarantine periods [1]. Due to an unexpected outbreak of an infectious disease, there were insufficient metrics or methods to understand and prevent its spread. Studies have shown that sociodemographic factors, such as income, educational level, and female sex were strongly associated with higher anxiety levels among Germans [2]. Investigating the impact of income on depressive symptoms among Japanese individuals will enhance global understanding of sociodemographic influences on mental health during the pandemic.

Methods

Study Population

The data for this study were collected from a longitudinal web-based survey conducted by Macromill, Inc., Japan, spanning from July 2020 to January 2021. The predefined inclusion criteria were participants who were 20 to 69 years old and lived in prefectures under special COVID-19 precautionary measures. The survey targeted Japanese adults to investigate sociodemographic factors and their potential impact on mental health during the COVID-19 pandemic. A total of 2,708 participants were included in the analysis.

Variables

The primary predictor in this study is household income, a key indicator of socioeconomic status that influences one's mental well-being. Depression severity, the outcome of interest, was assessed using the Patient Health Questionnaire-9 (PHQ-9) scores, ranging from 0 to 27 based on DSM symptom criteria, and

categorized into five groups: 'Minimal or None' (0–4), 'Mild' (5–9), 'Moderate' (10–14), 'Moderately Severe' (15–19), and 'Severe' (20–27) [3],[4]. To control for confounding factors, several adjustment covariates were included. These are selected based on the prior literature [5].

Data Imputation

The missingness mechanism for the income variable was evaluated using the mcar_test() function, which rejected the null hypothesis of Missing Completely at Random (MCAR). Further analysis was conducted using logistic regression, where the missingness indicator (1 = missing, 0 = observed) was modeled as the dependent variable. The results indicated that the probability of missingness was significantly associated with these observed variables, supporting the assumption of Missing at Random (MAR).

To address this, multiple imputation using the Predictive Mean Matching (PMM) method was applied, generating five imputed datasets. These datasets were pooled using the pool() function to account for the uncertainty introduced by imputation, providing robust estimates for subsequent modeling.

Models and Statistical Analysis

Ordinal Logistic Regression

We used an ordinal logistic regression model to examine the relationship between household income, the primary predictor, and depression severity. Depression severity, categorized into five ordered levels based on PHQ-9 scores, was treated as an ordinal outcome. This approach was chosen for its suitability in analyzing ordered categories while adjusting for covariates such as age, employment status, and others to account for their potential influence on the outcome.

$$\begin{aligned}\log\left(\frac{P(Y_i \leq j)}{P(Y_i > j)}\right) =& \beta_{0j} + \beta_1 \times \text{Household Income} + \beta_2 \times \text{Age} + \beta_3 \times \text{Employment Status} + \\ & \beta_4 \times \text{Sex} + \beta_5 \times \text{Residential Area} + \beta_6 \times \text{Underlying Disease} + \\ & \beta_7 \times \text{Marital Status (Married)} + \beta_8 \times \text{Child Status (Has Children)} + \\ & \beta_9 \times \text{Economic Impact}\end{aligned} \tag{1}$$

Figure 1: Logistic Regression

Machine Learning: Extreme Gradient Boosting

We implemented an Extreme Gradient Boosting (XGBoost) model to predict depression severity, using household income as the primary predictor and other covariates. The XGBoost model was selected for its ability to handle complex relationships and interactions. Data preprocessing involved imputing missing values in household income with the median and encoding categorical variables into numeric formats. Interaction terms, including household income with sex, age, and employment status, were created to capture potential moderating effects. To optimize model performance, we conducted a grid search over a range of hyperparameters, including the number of boosting rounds (nrounds), maximum tree depth (max_depth), learning rate (eta), and others. Five-fold cross-validation was applied to evaluate each hyperparameter combination, selecting the best configuration based on cross-validated accuracy. Probabilities for each class were calculated with the test data, and the model's discriminative ability was evaluated using Receiver Operating Characteristic (ROC) curves. Variable importance was visualized to identify the most influential predictors, providing interpretability for the model's results.

Shiny App

We have developed a Shiny app to display the results of our predictive modeling, and is accessible at this link: https://depreesionrlincome.shinyapps.io/shiny/. Users can upload their clinical dataset, choose the type of model, and adjust its parameters. Once the data is uploaded, the app generates predictions and provides outputs that include model performance metrics with the variable importance plots.

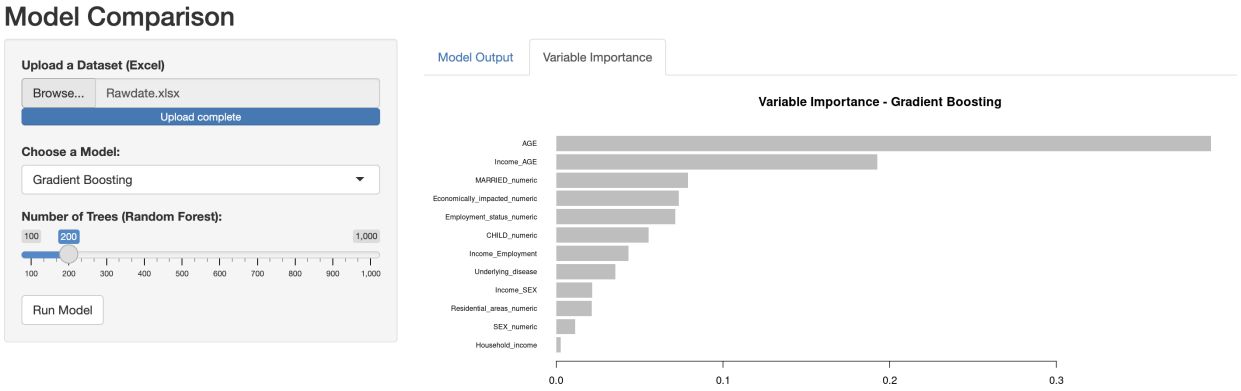

Figure 2: Shiny App

Results

The relationship between household income and depression severity during the COVID-19 pandemic, along with other relevant predictors, has been examined by ordinal logistic regression. Household income was negatively associated with depression severity (OR = 0.93, 95% CI: 0.86-1.01, p = 0.046), with statistical significance. This suggests that higher income may reduce the likelihood of severe depression. The model also adjusted for covariates such as age, employment status, and other relevant factors to control for their potential influence.

Table 1: Odds Ratios of Predictors

Predictors	Odds.Ratios	X95..CI	p.value
Household income	0.93	0.86 – 1.00	0.046
Minimal or None/Mild	0.16	0.10 – 0.25	<0.001
Mild/Moderate	0.67	0.42 – 1.07	0.095
Moderate/Moderately Severe	1.85	1.15 – 2.98	0.011
Moderately Severe/Severe	6.29	3.73 – 10.60	<0.001
Age	0.97	0.96 – 0.97	<0.001
Employment status	1.12	1.05 – 1.19	0.001
Sex	1.11	0.93 – 1.33	0.261
Residential areas	0.94	0.80 – 1.09	0.41
Underlying disease	2.03	1.60 – 2.58	<0.001
Married	0.51	0.40 – 0.64	<0.001
Child	0.85	0.66 – 1.08	0.171
Economically impacted	1.83	1.59 – 2.10	<0.001

The tuned XGBoost model demonstrated the best predictive accuracy for depression severity among the

models evaluated. The optimized hyperparameters included 100 boosting rounds, a maximum tree depth of 5, a learning rate (eta) of 0.3, a gamma value of 5, a column sampling ratio (colsample_bytree) of 0.5, a minimum child weight of 5, and a subsample ratio of 1. These settings minimized overfitting while maintaining high model performance. Model evaluation on the test dataset showed that XGBoost achieved the lowest Mean Squared Error (MSE =1.304267) and Root Mean Squared Error (RMSE =1.142045), outperforming Random Forest (MSE = 1.346939, RMSE = 1.160577) and Support Vector Machine (MSE = 1.330241, RMSE = 1.153361). These results highlight XGBoost's superior ability to model complex relationships and interactions within the data. The variable importance analysis revealed that age, the interaction between income and age (Income_AGE), and economic impact were the most influential predictors of depression severity. Interestingly, household income itself, while included as a key predictor, ranked below these interaction terms and other covariates, suggesting the importance of contextual factors in determining mental health outcomes.

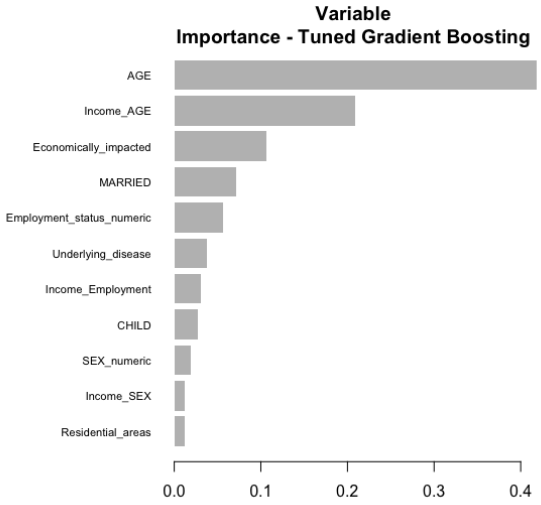

Figure 3: Variable Importance

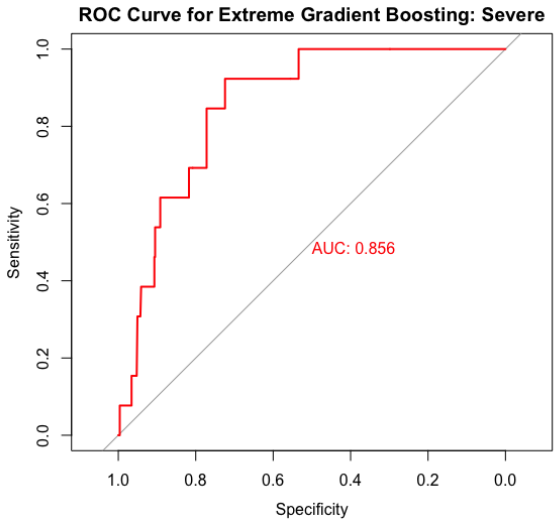

Figure 4: ROC Curve

Receiver Operating Characteristic (ROC) curves were used to assess the model's discriminative ability across the five depression severity classes. For the "Severe" depression severity class, the Area Under the Curve (AUC) was 0.856, indicating strong model performance in distinguishing this class from others. This result underscores the robustness of the XGBoost model for capturing patterns of severe depression.

Conclusion

This study investigated the relationship between household income and depression severity among Japanese adults during the COVID-19 pandemic using a combination of statistical and machine learning models. The findings suggest a potential trend that higher income levels may reduce depression severity, as indicated by a proportional odds logistic regression model (OR = 0.93, p = 0.046). However, the association, while statistically significant, was weaker compared to other predictors. Factors such as age, marital status demonstrated stronger and more consistent associations with depression severity. The XGBoost model highlighted the importance of interaction effects between income and other variables in predicting mental health.Tailored approaches that address specific socioeconomic circumstances are likely to be more effective.

Limitations

The study relied on a web-based survey, potentially excluding individuals without internet access or technological literacy. This could lead to a sample bias limiting the generalizability of the findings to the broader Japanese population. Additionally, the depression severity and household income were self-reported data. This could potentially introduce recall bias.

References

[1] Lakhan, R., Agrawal, A., & Sharma, M. (2020). Prevalence of depression, anxiety, and stress during COVID-19 pandemic. *Journal of neurosciences in rural practice*, 11(4), 519.
[2] Benke, C., Autenrieth, L. K., Asselmann, E., & Pané-Farré, C. A. (2020). Lockdown, quarantine measures, and social distancing: Associations with depression, anxiety and distress at the beginning of the COVID-19 pandemic among adults from Germany. *Psychiatry research*, 293, 113462.
[3] Kroenke, K., Spitzer, R.L., and Williams, J.B.W. (2001). The PHQ-9. *Journal of General Internal Medicine*, 16: 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x
[4] Núñez, C., Delgadillo, J., Barkham, M., & Behn, A. (2024). Understanding symptom profiles of depression with the PHQ-9 in a community sample using network analysis. *European Psychiatry*, 67(1), e50. doi: 10.1192/j.eurpsy.2024.1756
[5] Geprägs, A., Bürgin, D., Fegert, J. M., Brähler, E., & Clemens, V. (2022). The impact of mental health and sociodemographic characteristics on quality of life and life satisfaction during the second year of the COVID-19 pandemic—Results of a population-based survey in Germany. *International Journal of Environmental Research and Public Health*, 19(14), 8734. https://doi.org/10.3390/ijerph19148734

Contribution

Sooyeon Oh: literature review, data analysis, writing report; Hongjian Wang: dataset collection, writing report, literature review; Jingchao Yang: R-coding, data imputation, report writing; Sicheng Qian: R-shiny app, data analysis, writing report;