Discussion-WriteUp

Sophia Li

2024-12-16

Discussion

The results section does not include a discussion regarding the odds of educational levels on the combined variable. Based on the model, the interpretation would be as follows: Individuals with primary education have 2.28 times higher odds of high fatality compared to those with no formal education (exp(0.8227)). For secondary education, the odds ratio is 11.94 (exp(2.4811)), indicating a substantial increase in the likelihood of high fatality compared to the baseline group. Conversely, tertiary education does not show a significant impact on fatality outcomes ($p > 0.05$). However, this interpretation may be incorrect due to a possible interaction between education level and the hospitalized variable.

The proportional bar plot indicates that individuals with tertiary education are more likely to receive treatment compared to all other groups, and those with secondary education are also more likely to be hospitalized. In contrast, patients with no formal education or primary education are the least likely to receive hospital treatment after a suicide attempt. Consequently, when the hospitalized variable is included as a response variable in the binary logistic regression model, it causes distorted coefficients due to multicollinearity. While the graph highlights the importance of education, the table results from the final model may be somewhat misleading. Therefore, I believe it is inappropriate to interpret the impact of education levels as described earlier, and I have chosen not to discuss their effects on the outcome in detail.

For the missing data, I directly removed or dropped records with "unknown" or "unspecified" values among our predictors. However, the missing data for the occupation variable could be categorized as missing at random (MAR). There are approximately 160 cases with unknown occupations, which I believe may be influenced by geographic factors. Shandong Province is predominantly agricultural, with limited opportunities for tourism, service, or sales-related jobs. Additionally, access to education in the region is highly competitive and limited, resulting in many individuals having only primary or even no formal education. As a result, the unknown occupations are likely due to MAR, as missingness can be associated with geographic and socioeconomic factors.On the other hand, the missing data for the suicide method variable, labeled as "unknown" or "unspecified," could be missing completely at random (MCAR). This is likely because certain suicide methods may not have been accurately identified or recorded at the time of the incident.While there are different types of missing data, I chose to remove the "unknown" values directly to maintain the clarity of graphs and table results. However, alternative methods for handling missing data, such as imputation, could potentially yield better outcomes.

On the other hand, focusing solely on the farming and household groups is more convenient and allows for clearer conclusions. However, there may be differences in the odds between these two groups, which raises the question of whether it is appropriate to combine them simply because they make up the majority of the data and share some similarities compared to other very different occupations, such as students or business/service providers. It would be beneficial to obtain a larger sample with a broader representation of occupations and social groups. This would allow me to extend the model to a more diverse and comprehensive population, leading to more generalizable conclusions instead of exclusively focusing on farmers.

Another concern I have is about the data itself. While it contains around 2.6k observations, which is generally considered a large dataset under usual circumstances, representing China's population—especially Shandong

province—could be particularly challenging due to its complexity and long history. Shandong province has 137 counties, yet the dataset is derived from only three counties. It is unclear whether these counties are sufficiently representative or diverse to reflect the entire Shandong population, let alone China as a whole. Additionally, the dataset is heavily composed of the farming population, which raises questions about the data collection process. I suspect there may be convenience bias, or the dataset appears as it does simply because farming populations were easier to access or record data from.

From the results, it is evident that different suicide methods have significantly varying odds of high fatality. For instance, poisoning is about four times more likely to result in high fatality compared to cutting. There may be potential collinearity between occupation and suicide methods or relationships with other demographic factors. For example, farmers are more likely to use pesticides or poisoning as suicide methods because these substances are readily accessible and familiar to them. The limited number of cases involving jumping could be attributed to the lower heights of buildings in rural areas, where high-rise structures are uncommon. If individuals jump from mountains, it is plausible that they might not be found or could be classified as missing. While hanging may initially seem less probable, traditional Chinese architecture often includes *liang* (girders), which could make this method more feasible in rural settings.