**PSM 2 Assessment 2**

**Introduction**  
  
The Egypt dataset contains 1164 observations and 22 categorical variables (e.g. demographic, societal, economic and educational). One strength of this dataset is that it has a relatively large sample size which covers both the demographic and societal factors (e.g. employment, perceptions of corruption).

However, the dataset also presents certain limitations. For instance, the dataset has a lack of diversity, as the majority of participants are Egyptian (1163) and has missing data in variables such as income (80) and ethnic group (66). Additionally, only 17 respondents were placed in the high income level.

Noteworthy statistics in the dataset include a nearly balanced gender distribution (610 males, 554 females), with the majority of respondents having medium income levels (909) and most individuals having upper secondary education (424).

**Methods**

Firstly, an ordinal logistic regression model was used for the first model question based on the outcome variable (perception\_corruption) being measured on a 1 to 10 ordinal scale. Since this model only involved two variables, the (dont\_carry\_money) variable was the only explanatory variable. The model was chosen to handle the ordered nature of the outcome variable and to estimate the odds ratios of corruption based on security concerns.

Secondly, both linear regression and count regression models were used for the second model question based on the assumption of the outcome variable (perception\_corruption) being measured on count and numeric data. After running a performance test between the linear and the poisson model, results indicated that the linear model was a better fit model, but another point was noted on top of this. The demographic variables (age, sex, social class) were fit to perceptions influence on security and corruption.

Thirdly, a binary logistic model was used for the third model question due to the binary nature of the outcome variable (dont\_carry\_money) producing yes or no responses. The model was chosen to estimate the probability between avoiding carrying money due to security concerns based on demographics and perceptions of corruption.

**Results**

Both the tidyverse and easystats library packages were utilized across all 4 models (Ordinal, Linear, Count, Binary) for data wrangling, data visualization and data transformation. However, the binary logistic model did not make use of any additional libraries.

For the first question, the MASS library was used to create the ordinal logistic model with the polr function and the brant library was used to verify the proportional odds assumption of the ordinal model.

In the second question, the psych library was used to run assumption checks on the model (e.g. normality assessment) and the sjsmc library was used to perform data wrangling in preparation for the regression analysis.

**Ordinal Logistic Model (Beta)**

Table 1: Odds ratios for model in perception of security and subjective corruption

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| | Parameter | | --- | | Odds Ratio | SE | t(1151) | p-value |
| dont carry money [Yes] | 1.31 | 0.16 | 2.19 | 0.029 |

The Brant test results of the model indicated that the proportional odds assumption does hold for this model, following these statistical results (X^2 = 5.07, p = 0.75, df = 8) indicating no rejection of the null hypothesis. This suggests that the relationship between the explanatory variable and the dependent variable is consistent across all levels of the dependent variable.

Following the results of the table, all else being equal, individuals who did not carry money due to security were predicted to have a 31% odds increase of perceiving corruption in their country compared to those who did carry money, with this difference being statistically significant (SE = 0.16, t(1151)=2.19, p=0.029).

**Linear Regression Model**

For the second model question, I fit a linear model and ran the following assumption checks required for a linear model:

* A Breusch–Pagan test for non-constant variance in the model residuals, which showed significant heteroscedasticity (p<0.001).
* A Shapiro–Wilk test for normality of the residuals, which showed the residuals were significantly non-normal (p<0.001).
* A multicollinearity check, which showed that all variance inflation factors (VIFs) were less than 5 (the largest VIF was 1.06).
* A check for influential outliers which did not work due to an error indicating no numeric variables found and that there was no data to check for outliers.
* A Durbin–Watson test for autocorrelation of residuals, which showed significant autocorrelation (p<0.001).

Since the linear model did not meet any of the assumptions required for the results to be considered reliable (except for the lack of multicollinearity), the results produced by this model cannot be trusted.

Adding on to this model, the question also required fitting count models (Poisson distribution and Negative Binomial distribution). After running the compare performance test between the two distributions, the results showed that the Poisson model had better performance statistics (AIC = 5125.1, BIC = 5165.6).

**Poisson Model**

Table 2: Incidence rate ratios for model between demographic characteristics, the perception of security and people’s subjective corruption

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameter | IRR | SE | z | p |
| (Intercept) | 8.48 | 0.26 | 69.83 | < .001 |
| don’t carry money [Yes] | 1.03 | 0.02 | 1.27 | 0.203 |
| social class [Working class] | 1.04 | 0.03 | 1.49 | 0.135 |
| social class [Lower middle class] | 1.02 | 0.03 | 0.81 | 0.418 |
| social class [Upper middle class] | 1.04 | 0.04 | 0.87 | 0.383 |
| sex [Female] | 0.99 | 0.02 | -0.65 | 0.514 |
| age 3 cat [30-49 years] | 0.99 | 0.02 | -0.25 | 0.799 |
| age 3 cat [50 and more years] | 0.93 | 0.03 | -2.58 | 0.010 |

Starting with the age category (50 and above), the incidence rate ratio for this variable is 0.93, which implies that individuals who fall within this category are 0.93 times likely of perceiving corruption in comparison to individuals who belong to the age category (30 and below) on a statistically significant scale (p = 0.010), all else being equal.

In contrast, looking at the other demographic variables such as social class, none of the social hierarchies within this variable fall on the statistically significant scale. For instance, looking at the upper middle class category, it gives an incidence rate ratio of 1.04, suggesting that individuals who fall within this social class category are 1.04 times likely of perceiving corruption in comparison to individuals who in the lower class category on a statistically significant scale (p = 0.383), all else being equal.

In the end, I ran a test performance analysis between the linear model and the Poisson model in which the results showed that the linear model performed better (AIC = 4648.7, BIC = 4648.7). Despite the linear model being a better model, it is not the best model to use as its assumptions have been violated making it unreliable.

**Binary Logistic Model**

Table 3: Odds ratios for model between demographic characteristics, subjective corruption and people's perception of security

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Odds Ratio | SE | p |
| (Intercept) | 0.06 | 0.07 | 0.014 |
| perception corruption [2] | 1.80 | 2.94 | 0.718 |
| perception corruption [3] | 0.96 | 1.31 | 0.976 |
| perception corruption [4] | 0.36 | 0.55 | 0.505 |
| perception corruption [5] | 1.08 | 1.26 | 0.946 |
| perception corruption [6] | 0.93 | 1.06 | 0.946 |
| perception corruption [7] | 1.53 | 1.70 | 0.701 |
| perception corruption [8] | 1.41 | 1.57 | 0.755 |
| perception corruption [9] | 1.58 | 1.76 | 0.683 |
| perception corruption [10 There is abundant corruption in my country] | 1.66 | 1.83 | 0.644 |
| social class [Working class] | 2.31 | 0.55 | < .001 |
| social class [Lower middle class] | 3.63 | 0.82 | < .001 |
| social class [Upper middle class] | 4.25 | 1.22 | < .001 |
| sex [Female] | 1.87 | 0.26 | < .001 |
| age 3 cat [30-49 years] | 1.00 | 0.17 | 0.992 |
| age 3 cat [50 and more years] | 0.94 | 0.19 | 0.772 |

For the final model question, I fit a binary logistic model and ran the following assumption checks required for a binary logistic model:

* The model predicted intervals included observed data points with most of the points falling within the intervals .
* Most of the average residuals fell within the binned residuals bound, indicating minimal error.
* A multicollinearity check showed that all variance inflation factors (VIFs) were less than 5 (the largest VIF was 1.07).
* All the data points fell inside the contour lines of the standard residuals and leverage points, hinting that there were no influential outliers.
* The residuals of the model aligned perfectly with the standard uniform distribution line, which suggested a goodness of fit.

Since the linear model met nearly all of the assumptions required for the results to be considered reliable (except for the binned residuals), the results produced by this model can be trusted.

Looking at the table results produced by the binary logistic model, all the social class levels are statistically significant. In comparison to lower class individuals, working class individuals are 2.31 more times likely to have a higher perception of security (p < .001), lower middle class individuals are 3.63 more times likely to have a higher perception of security (p < .001) and upper middle class individuals are 4.25 more times likely to have a higher perception of security (p < .001) all else being equal.

On a non statistically significant scale, all the perception corruption levels had high p values along each level. This is similar in the case for the age categories as well in comparison to individuals aged below 30, all else being equal.

A binary logistic model was chosen for this model question given the outcome nature of the outcome variable producing a yes or no response.

**Discussion**

The results reveal relevant analytical insights to policymakers. Higher social class levels lead to increased security due to financial stability and safer environments (World Bank, 2019). Policymakers can address this by focusing improving infrastructures and introducing social welfare programs. Secondly, the perception of corruption exhibited no significant correlation with security, suggesting that anti-corruption policies should be prioritized to restore public trust in governance (Transparency International, 2022). Thirdly, the high perception of security among women indicates a need for gender-specific research to better understand cultural and societal dynamics (UNDP, 2020).

**Real world insights**

Addressing socioeconomic inequalities through strategically implemented policies may significantly improve community welfare (World Bank, 2014). The normalization of corruption perceptions underscores the need for transparent anti-corruption actions to restore public trust (Transparency International, 2022). Gender-specific perceptions of security indicate that societal standards and vulnerability to threats differ, necessitating additional qualitative investigations (UNICEF, 2020).

**References**

1. World Bank. (2019). *Understanding poverty and inequality in Egypt*. Retrieved from <https://documents.worldbank.org/curated/en/351121575640799887/pdf/Understanding-Poverty-and-Inequality-in-Egypt.pdf>

2. World Bank. (2014). *Inside inequality in the Arab Republic of Egypt: Facts and perceptions across people, time, and space*. Retrieved from <https://www.worldbank.org/content/dam/Worldbank/egypt-inequality-book.pdf>

3. Transparency International. (2022). *Corruption perceptions index 2022*. Retrieved from <https://www.transparency.org/en/publications/corruption-perceptions-index-2022>

4. United Nations Development Programme. (2020). *Gender, climate & security: Sustaining inclusive peace on the frontlines of climate change*. Retrieved from <https://climate-diplomacy.org/sites/default/files/2020-10/UNEP%2C%20UN%20Women%2C%20UNDP%2C%20UNDPPA-PBSO%202020%20-%20Gender%20Climate%20and%20Security%20report.pdf>

5. United Nations Children’s Fund. (2020). *Situational analysis of women and girls in the Middle East and North Africa*. Retrieved from <https://www.unicef.org/mena/reports/situational-analysis-women-and-girls-middle-east-and-north-africa>

\*Note: Matt said that tables were not included in the word count, only table headings