550: Final Report

Tom, XinyaoFan, Nikloas

April 12, 2018

**Summary**

This project aims to explore how demographic, organizational and learning factors are associated with South Korean senior worker’s skill usage including numeracy usage and literacy usage. Due to the likert scale design of our response variable, we recommend the use of an ordinal regression model. We then discuss the results of data analysis, and do a diagnostic for our model.

# 1 Introduction

Rapidly aging population is a phenomenon happening in South Korean nowadays and recent literature shows senior workers continue to participate in the labor market. In order to further develop their workplace skills, seniors often engage in educational programs. Our research objective is to figure out how demographic, organizational and learning factors are associated with South Korean senior worker’s skill usage including numeracy usage and literacy usage. It is an important issue to identify these factors, since skill usage of workers is a key area to improve both individual and organizational performance.

# Our report is organized as follows: Section 3 is the exploratory analysis of data; Section 4 gives some interpretation of the ordinal regression model we use and Section 5 describe the process of model selection; Section 6 is the diagnostics of our models. Finally we draw some meaningful conclusions in section 7.

# 2 Exploratory Analysis

We obtained the primary dataset from the open source of the Programme for the International Assessment of Adult Competencies (PIAAC). From August 2011 to March 2012, the organization collected this survey data set from 24 countries. The individuals’ age participated in this survey range from 16 to 56. We focus on employers in South Korea whose age is from 50 to 65 based on the research purpose. Also, we averaged the variables measured by Likert scale for assessing each of the literacy usage and the numeracy usage. The new averaged Likert scale for the numeracy usage and literacy usage have values ranging from 1 to 5, including decimal numbers. We then rounded off the decimal points and aggregated the values to four integers:1,2,3 and 4. The averaged Likert scale value 5 was aggregated into 4 for the simplicity and holding model assumptions. Table 1 is the data description which has been modified for the analysis.

Table 1: Data description

|  |  |  |  |
| --- | --- | --- | --- |
| Type | Variable | Scale | Explanation |
| Demographic  Factor | AGE\_R | Continuous | Age of individual. 50 to 65. |
| Years\_wk | Number of years at work |
| GENDER\_R | Categorical | Gender. 1-male, 2-female |
| ED\_Level | Education Level. 1-middle, 2-high, 3-college, 4-graduate |
| Full\_part | Employment type. 1-fulltime, 2-part-time |
| Organizational Contexts | Work\_flexM | Continuous | Work flexibility. |
| Work\_lrnM | Learning opportunity. |
| pub\_priv | Categorical | Public vs. private sector. 1-private, 2-public |
| Learning/  Education | NFE12 | Categorical | Participation in non-formal education |
| act\_lrn | Continuous | Active learning strategies.  Averaging 6 sub-items of likert scale(1,2,3,4,5) |
| Response  variable | lit\_use\_new | Categorical  (Ordinal) | Literacy skills usage (1,2,3,4) |
| num\_use\_new | Numeracy skills usage (1,2,3,4) |

Visualization of the data is presented in Figure 1 below. Plot A is the histogram of response variables. We can see the values are not evenly distributed. There are many values of 1 and the number of each value is decreasing.

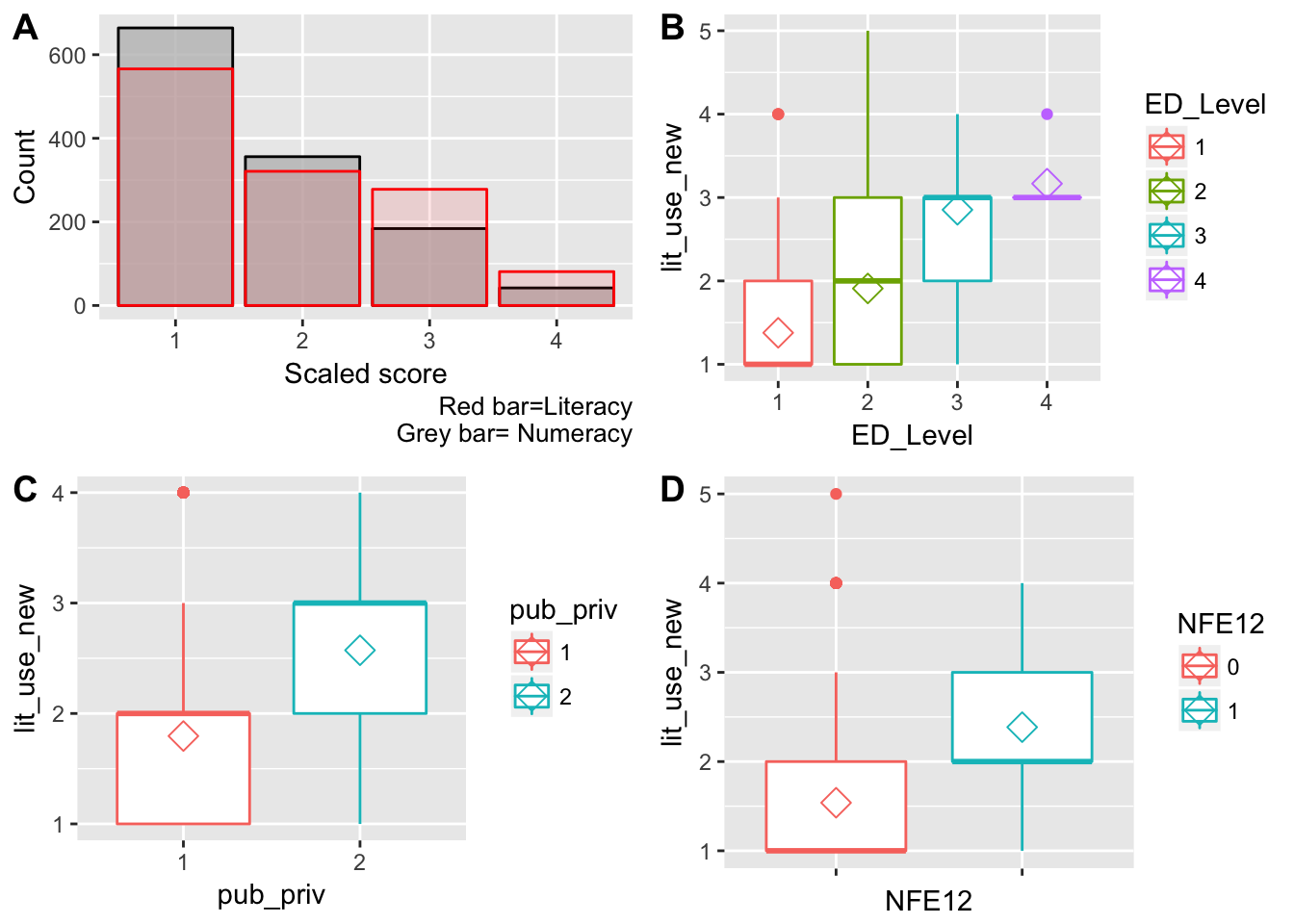


Figure 1: Visualiation of data

V

All of the boxplots in B, C, and D show that the three categorical variables(ED\_level, NFE12, and pub\_priv) are significant of explaining literacy skills usage variances. The other variables are not significant of literacy skills. For the numeracy skills usage, none of the variables shows significance from their boxplots.

# 3 Ordinal Regression Interpretation

# 4 Model Selection

# 5 Model Diagnostic

## 5.1 Proportion Odds Assumption

In ordinal logistic regression model there is an important assumption which belongs to ordinal odds. According to this assumption, the coefficients that describe the relationship between the lowest versus all higher categories of the response variable are the same as those that describe the relationship between the next lowest category and all higher categories, etc[1]. We first test the literacy model. Based on the results, the p\_values of all variables are greater than 0.05 except the variables act\_lrn and work\_lrnM. The p\_values of these two variables are 0.02690 and 0.03914, which slightly violates the proportional odds assumption. Since ordinal regression model is robust to the slightly violation of assumption, our model is sensible.

We then test the numeracy model, the p\_values of variable act\_lrn is 0.004999, which is much smaller than 0.05.It suggests that the proportional odds assumption does not hold for the predictor act\_lrn. That would indicate that the effect of the number of hours attending active activity is different for the transition from level “1” to “2” and “2” to “3" of numeracy usage.

**5.1 Proportion Odds Assumption**

Both models convergent successfully, and the result shows the numerical error in the parameter estimates is really small. It indicates the parameter estimates of our models are accurate. The AIC, BIC for literacy model are 2185.488, 2261.926 while the AIC, BIC for numeracy model are 2190.701, 2262.043. Since we do a model selection based on AIC, our models have the smallest AIC values. That indicates our model has a good fit to the truth but with few parameters.

# 6 Conclusions