Lesson X Practical Exercise

STUDENT NAME

School of Public Affairs: Pennsylvania State University

PADM 504, Section 001: Data Analysis for Policy and Administration

Professor Lauren Azevedo

DATE

**Purpose**

Orient the reader to the purpose of your analysis. Provide the research question and how you are answering it.

**Method**

**Chi-square test**

This test is used to determine if two categorical variables are independent or if they are in fact related to one another. If two categorical variables are independent, then the value of one variable does not change the probability distribution of the other.

**Multinomial Logistic Regression**

Multinomial logistic regression is a [classification](https://en.wikipedia.org/wiki/Statistical_classification) method that generalizes [logistic regression](https://en.wikipedia.org/wiki/Logistic_regression) to [multiclass problems](https://en.wikipedia.org/wiki/Multiclass_classification), i.e. with more than two possible discrete outcomes. That is, it is a model that is used to predict the probabilities of the different possible outcomes of a [categorically distributed](https://en.wikipedia.org/wiki/Categorical_distribution) [dependent variable](https://en.wikipedia.org/wiki/Dependent_variable), given a set of [independent variables](https://en.wikipedia.org/wiki/Independent_variable).

**Missing value**

In the presence of missing value, multinomial logistic regression can provide misleading result. The missing values are excluded using SPSS (version 25) for the purpose of correct result.

**Results**

Table-01: Descriptive Statistics of the demographic variables of the respondents

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Descriptive statistics** | **Measure** | **Variables** | | | | |
| Age | Sex | Race | Degree | Job satisfaction |
| Central tendency | Mean | 46.43 |  |  |  |  |
| Median | 44 |  |  | High school | Moderately satisfied |
| Mode | 30 | Female | White | Less than High School | Very satisfied |
| Dispersion | Minimum | 18 | 1 | 1 | 0 | 1 |
| Maximum | 89 | 2 | 3 | 4 | 4 |
| Range | 71 | 1 | 2 | 4 | 3 |
| Variance | 308.78 |  |  |  |  |
| Standard deviation | 17.572 |  |  |  |  |
| Shape characteristics | Skewness | 0.391 |  |  |  |  |
| Kurtosis | -0.805 |  |  |  |  |

As age is a continuous variable, the mean age of the respondents is found 46.43 years with standard deviation 17.572 years. In addition to, sex, race, degree & job satisfaction are categorical variables, we cannot calculate mean for these variable. Only one central tendency is appropriate for these categorical variable is mode. Mode is the value of a variable for which the frequency is maximum.

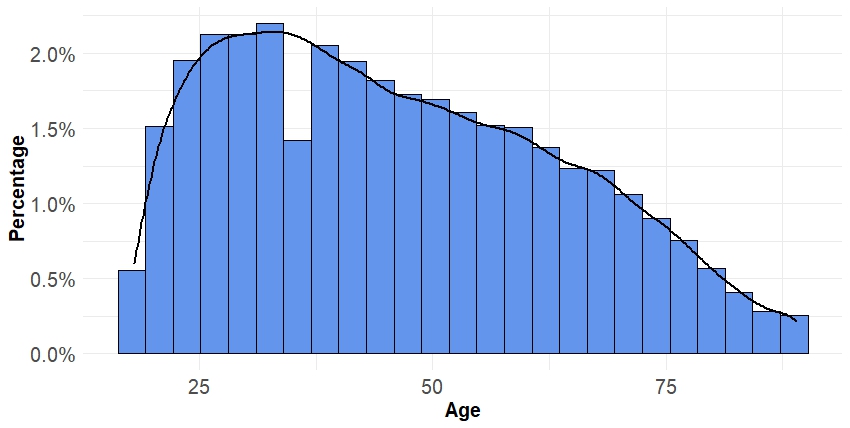


Figure-01: Histogram of age of the respondents

From the above histogram, we can notice that the most of the respondents have age between 25 to 50. The mean age of the respondents is 46.43 years.

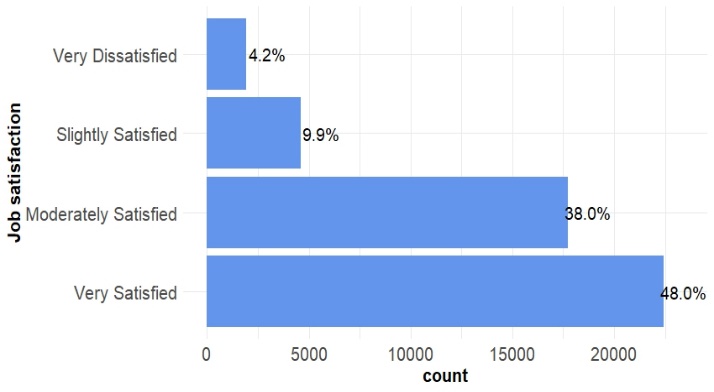
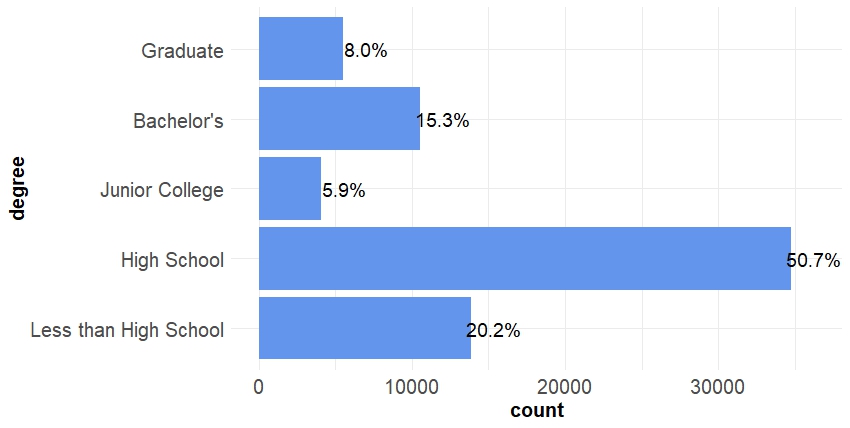
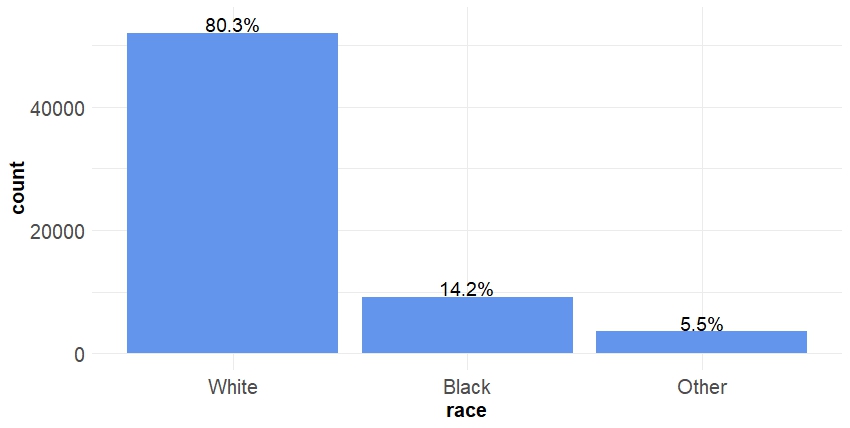
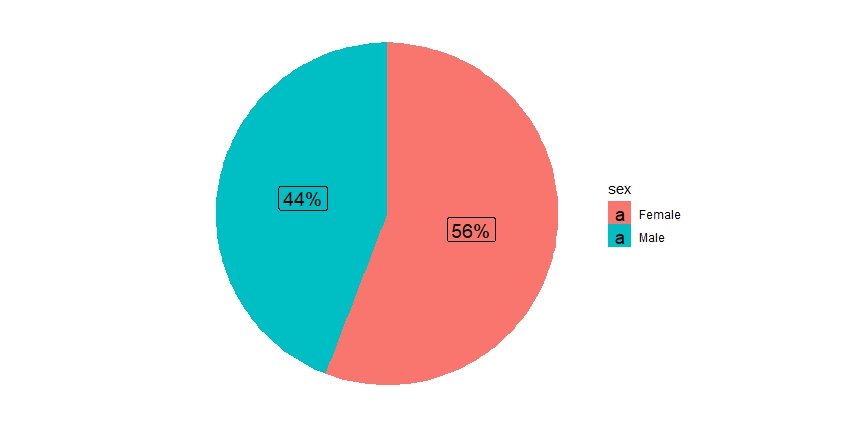


Figure-02: Demographic characteristics of the respondents

To begin with above bar graph, we notice that 56% of the respondents are male and rest 44% of the respondents are female. Additionally, it is also seen that the most of respondents are white (80.3%). Furthermore, the most of the respondents have high school level degree (50.7%) and finally, the most of the respondents are very satisfied (48%) & moderately satisfied (38%) with their job.

Table-02: Chi test of the variables

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Category** | **Job satisfaction** | | | | **Chi square Statistic** | **P-value** |
| Very satisfied | Moderately satisfied | A little satisfied | Very dissatisfied |
| Sex | Male | 9683 | 7916 | 1999 | 816 | 11.309 | 0.010 |
| Female | 12734 | 9807 | 2612 | 1123 |
| Race | White | 18692 | 13941 | 3410 | 1428 | 378.423 | 0.000 |
| Black | 2579 | 2666 | 906 | 405 |
| Other | 1146 | 1116 | 295 | 106 |
| Degree | Less than High School | 3873 | 3163 | 932 | 434 | 322.798 | 0.000 |
| High School | 11282 | 9508 | 2559 | 1088 |
| Junior College | 1430 | 1044 | 249 | 79 |
| Bachelor’s | 3666 | 2820 | 636 | 246 |
| Graduate | 2120 | 1141 | 227 | 90 |

Now, we need to test the following hypothesis,

: There is no relation between two variables vs : There is a relation between two variables

Since we can notice from above table that all the p-value of the Chi-square test is less than 0.05, we can reject null hypothesis at 5% level of significance. That means, there exists a relation between Job satisfaction and Sex, Race & Degree, respectively.

Now, considering job satisfaction as a response variable and age, sex, race & degree as predictor variable a multinomial logistic regression model has been fitted in the following table.

**Table-03:** Estimate of multinomial logistic regression model

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Job statisfaction |  | B | Std error | Wald Statistic | df | p-value | Exp(B) |
| Very Statisfied | Intercept | 1.766 | .165 | 115.179 | 1 | .000 \*\*\* |  |
| age | .031 | .002 | 293.149 | 1 | .000 \*\*\* | 1.031 |
| Male | .056 | .049 | 1.342 | 1 | .247 | 1.058 |
| White | .051 | .107 | .228 | 1 | .633 | 1.052 |
| Black | -.556 | .116 | 22.885 | 1 | .000 \*\*\* | .574 |
| Less than High School | -.941 | .120 | 61.748 | 1 | .000 \*\*\* | .390 |
| High school | -.627 | .113 | 30.713 | 1 | .000 \*\*\* | .534 |
| Associated Junior College | -.038 | .159 | .058 | 1 | .810 | .962 |
| Bachelor’s | -.308 | .127 | 5.907 | 1 | .015 | .735 |
|  |  |  |  |  |  |  |  |
| Moderately satisfied | Intercept | 2.116 | .166 | 162.894 | 1 | .000 \*\*\* |  |
| age | .012 | .002 | 46.843 | 1 | .000 \*\*\* | 1.013 |
| Male | .097 | .049 | 3.915 | 1 | .048 \* | 1.101 |
| White | -.144 | .106 | 1.826 | 1 | .177 | .866 |
| Black | -.472 | .116 | 16.611 | 1 | .000 \*\*\* | .624 |
| Less than High School | -.504 | .121 | 17.266 | 1 | .000 \*\*\* | .604 |
| High school | -.263 | .115 | 5.255 | 1 | .022 \* | .769 |
| Associated Junior College | .164 | .161 | 1.033 | 1 | .310 | 1.178 |
| Bachelor’s | -.028 | .129 | .046 | 1 | .830 | .973 |
|  |  |  |  |  |  |  |  |
| A little dissatisfied | Intercept | .976 | .187 | 27.251 | 1 | .000 \*\*\* |  |
| age | .002 | .002 | 1.014 | 1 | .314 | 1.002 |
| Male | .050 | .055 | .813 | 1 | .367 | 1.051 |
| White | -.177 | .119 | 2.227 | 1 | .136 | .838 |
| Black | -.225 | .129 | 3.037 | 1 | .081 | .799 |
| Less than High School | -.146 | .138 | 1.117 | 1 | .291 | .864 |
| High school | -.038 | .131 | .087 | 1 | .768 | .962 |
| Associated Junior College | .253 | .181 | 1.968 | 1 | .161 | 1.288 |
| Bachelor’s | .051 | .146 | .123 | 1 | .725 | 1.053 |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The above regression model can be interpreted in the following way:

**Very satisfied relative to very dissatisfied**

This is the multinomial logit estimate for very satisfiedrelative to very dissatisfied when the predictor variables in the model are evaluated at zero. For females, other race, graduate with zero, the logit for preferring very satisfied to very dissatisfied is 1.766.

This is the multinomial logit estimate for a one unit increase in age score for very satisfied relative to very dissatisfied given the other variables in the model are held constant. If a subject were to increase his age by one year, the multinomial log-odds of preferring very satisfied relative to very dissatisfied would be expected to increase by 0.031 unit while holding all other variables in the model constant.

This is the multinomial logit estimate comparing males to females for very satisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for males relative to females is 0.056 unit higher for preferring very satisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing white to other race for very satisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for white relative to other race is 0.051 unit higher for preferring very satisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing black to other race for very satisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for black relative to other race is 0.556 unit lower for preferring very satisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing less than high school to graduate for very satisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for less than high school relative to graduate is 0.941 unit lower for preferring very satisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing high school to graduate for very satisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for high school relative to graduate is 0.627 unit lower for preferring very satisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing junior college to graduate for very satisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for junior college relative to graduate is 0.308 unit lower for preferring very satisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing bachelor’s to graduate for very satisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for bachelor’s relative to graduate is 0.308 unit lower for preferring very satisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

**Moderately satisfied relative to very dissatisfied**

This is the multinomial logit estimate for moderately satisfiedrelative to very dissatisfied when the predictor variables in the model are evaluated at zero. For females, other race, graduate with zero, the logit for preferring very satisfied to very dissatisfied is 2.116.

This is the multinomial logit estimate for a one unit increase in age score for moderately satisfied relative to very dissatisfied given the other variables in the model are held constant. If a subject were to increase his age by one year, the multinomial log-odds of preferring moderately satisfied relative to very dissatisfied would be expected to increase by 0.012 unit while holding all other variables in the model constant.

This is the multinomial logit estimate comparing males to females for moderately satisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for males relative to females is 0.097 unit higher for preferring moderately satisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing white to other race for moderately satisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for white relative to other race is 0.144 unit lower for preferring moderately satisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing black to other race for moderately satisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for black relative to other race is 0.472 unit lower for preferring moderately satisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing less than high school to graduate for moderately satisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for less than high school relative to graduate is 0.504 unit lower for preferring moderately satisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing high school to graduate for moderately satisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for high school relative to graduate is 0.263 unit lower for preferring moderately satisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing junior college to graduate for moderately satisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for junior college relative to graduate is 0.164 unit increase for preferring moderately satisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing bachelor’s to graduate for moderately satisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for bachelor’s relative to graduate is 0.028 unit lower for preferring moderately satisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

**A little dissatisfied relative to very dissatisfied**

This is the multinomial logit estimate for a little dissatisfiedrelative to very dissatisfied when the predictor variables in the model are evaluated at zero. For females, other race, graduate with zero, the logit for preferring a little dissatisfied to very dissatisfied is 0.976.

This is the multinomial logit estimate for a one unit increase in age score for a little dissatisfied relative to very dissatisfied given the other variables in the model are held constant. If a subject were to increase his age by one year, the multinomial log-odds of preferring a little dissatisfied relative to very dissatisfied would be expected to increase by 0.002 unit while holding all other variables in the model constant.

This is the multinomial logit estimate comparing males to females for a little dissatisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for males relative to females is 0.05 unit higher for preferring a little dissatisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing white to other race for a little dissatisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for white relative to other race is 0.177 unit lower for preferring a little dissatisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing black to other race for a little dissatisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for black relative to other race is 0.225 unit lower for preferring a little satisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing less than high school to graduate for a little dissatisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for less than high school relative to graduate is 0.146 unit lower for preferring a little dissatisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing high school to graduate for a little dissatisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for high school relative to graduate is 0.038 unit lower for preferring a little dissatisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing junior college to graduate for a little dissatisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for junior college relative to graduate is 0.253 unit increase for preferring a little dissatisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

This is the multinomial logit estimate comparing bachelor’s to graduate for a little dissatisfied relative to very dissatisfied given the other variables in the model are held constant. The multinomial logit for bachelor’s relative to graduate is 0.051 unit higher for preferring a little dissatisfiedrelative to very dissatisfied given all other predictor variables in the model are held constant.

**Assumptions**

Since sex, race & degree are categorical variable, we can not check the multicolinearity, heteroscedasticity & normality. To check multicolinearity, heteroscedasticity & normality, the variables are must be in continuous variable. It is also added that the multinomial logistic regression is a non-linear regression.

**Discussion and Policy Implications/Recommendations**

Most of the practical exercises ask you to offer policy implications or recommendations based on your findings. First discuss the overall meaning of your analysis and then present your policy implications and recommendations to the reader.

**References** (if used)