Eelisa Jones  
UW-Milwaukee  
SOCIOL461  
Fall ‘22

**3.7.1 Exercise 1: Sentence Length for Convicted Robbery Offenders**

Contents

[Q1.) From the provided data, use multiple regression to test the research hypothesis that age, prior criminal record, and ascribed socioeconomic class influence severity of punishment. 1](#_Toc120532370)

[Descriptive Statistics 1](#_Toc120532371)

[Selecting Reference Variables 2](#_Toc120532372)

[Checking Independent Variables for Correlations 2](#_Toc120532373)

[Regression with Dummy Variables 3](#_Toc120532374)

[R2 Value 4](#_Toc120532375)

[Statistical Significance (Numeric Variable: Age) 4](#_Toc120532376)

[B-value Interpretation & Real-World Significance (Numeric Variable: Age) 4](#_Toc120532377)

[Statistical Significance (Categorical Variables: Record, Socioeconomic Class) 4](#_Toc120532378)

[B-value Interpretation & Real-World Significance (Categorical Variables: Record & Socioeconomic Status) 6](#_Toc120532379)

[Q2.) From the multiple regression in Question 1, predict the sentence length for a 65-year-old working class offender with a prior record. Should you trust this predicted value? Explain. 6](#_Toc120532380)

## Introduction

Information on severity of punishment (measured as length of sentence in years), age at the time of arrest, presence of a prior criminal record (yes or no), and ascribed socioeconomic class (middle, working, and low) was collected for 37 convicted robbery offenders. This information has been entered into SPSS for analysis.

# Q1.) From the provided data, use multiple regression to test the research hypothesis that age, prior criminal record, and ascribed socioeconomic class influence severity of punishment.

## Descriptive Statistics

|  |  |  |  |
| --- | --- | --- | --- |
| Statistics for “length” and “age” | | | |
|  | | length | age |
| N | Valid | 37 | 37 |
| Missing | 0 | 0 |
| Mean | | 13.46 | 28.89 |
| Median | | 14.00 | 24.00 |
| Std. Deviation | | 6.862 | 8.627 |
| Range | | 24 | 28 |
| Minimum | | 1 | 21 |
| Maximum | | 25 | 49 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Statistics for “record” | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | No | 17 | 45.9 | 45.9 | 45.9 |
| Yes | 20 | 54.1 | 54.1 | 100.0 |
| Total | 37 | 100.0 | 100.0 |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Statistics for “SES” (Socioeconomic Status) | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Low | 11 | 29.7 | 29.7 | 29.7 |
| Middle | 10 | 27.0 | 27.0 | 56.8 |
| Working | 16 | 43.2 | 43.2 | 100.0 |
| Total | 37 | 100.0 | 100.0 |  |

## Selecting Reference Variables

Our experiment contains two categorical values, “record” and “SES” (socioeconomic class). For our “record” variable, we can choose “No” as our reference variable value. For our “SES” variable, we can choose “Working class” as our reference variable value.

In our final dataset for analysis, the “yes\_record\_dummy” variable will represent the categorical variable “record”. The “low\_ses\_dummy” and “middle\_ses\_dummy” variables will represent the categorical variable “SES”.

## Checking Independent Variables for Correlations

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Independent Variable Correlations** | | | | | |
|  | | age | yes\_record\_dummy | low\_ses\_dummy | middle\_ses\_dummy |
| age | Pearson Correlation | 1 | .326\* | .126 | -.100 |
| Sig. (2-tailed) |  | .049 | .456 | .558 |
| N | 37 | 37 | 37 | 37 |
| yes\_record\_dummy | Pearson Correlation | .326\* | 1 | -.112 | -.172 |
| Sig. (2-tailed) | .049 |  | .508 | .310 |
| N | 37 | 37 | 37 | 37 |
| low\_ses\_dummy | Pearson Correlation | .126 | -.112 | 1 | -.396\* |
| Sig. (2-tailed) | .456 | .508 |  | .015 |
| N | 37 | 37 | 37 | 37 |
| middle\_ses\_dummy | Pearson Correlation | -.100 | -.172 | -.396\* | 1 |
| Sig. (2-tailed) | .558 | .310 | .015 |  |
| N | 37 | 37 | 37 | 37 |
| \*. Correlation is significant at the 0.05 level (2-tailed). | | | | | |

From our SPSS bivariate correlation results, we can see that no two independent variables have a correlation approaching an absolute value of .7. Our highest correlation is -.396 between our socio-economic dummy variables which, in any case, is not very informative. None of the correlations among our independent variables appear remarkable.

## Regression with Dummy Variables

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Summary** | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | .715a | .511 | .449 | 5.092 |
| a. Predictors: (Constant), middle\_ses\_dummy, age, yes\_record\_dummy, low\_ses\_dummy | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 865.561 | 4 | 216.390 | 8.346 | <.001b |
| Residual | 829.628 | 32 | 25.926 |  |  |
| Total | 1695.189 | 36 |  |  |  |
| a. Dependent Variable: length | | | | | | |
| b. Predictors: (Constant), middle\_ses\_dummy, age, yes\_record\_dummy, low\_ses\_dummy | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | 20.046 | 3.124 |  | 6.417 | <.001 |
| age | -.251 | .106 | -.316 | -2.375 | .024 |
| yes\_record\_dummy | 6.841 | 1.852 | .504 | 3.694 | <.001 |
| low\_ses\_dummy | -4.941 | 2.065 | -.334 | -2.393 | .023 |
| middle\_ses\_dummy | -5.778 | 2.114 | -.379 | -2.734 | .010 |
| a. Dependent Variable: length | | | | | | |

### R2 Value

Our R2 value (.511) suggests that our model is only a moderate fit to our observation data, meaning that (1) there are additional independent variables (Xs) that could help explain variability in our dependent variable and (2) our model predictions may be relatively inaccurate given actual data. As such, our model can explain about 51.1% of variability in sentence length with our given Xs.

### Statistical Significance (Numeric Variable: Age)

Our model has revealed that the one numeric independent variable, age, appears to have statistically significant effect on our dependent variable while holding other Xs constant. Given a confidence level of 95%, this variable’s p-value of .024, which is less than our .05 cutoff, suggests that age has a significant impact on sentence length.

### B-value Interpretation & Real-World Significance (Numeric Variable: Age)

Our b-value for age indicates that an increase of one unit of age (i.e. 1 year) will translate into a decrease of sentence length by .251 years. An increase of one unit of age appears real-world significant given this independent variable’s descriptive statistics (mean = 28.89, min = 21, max = 49). A decrease of .251 years in sentence length (Y), however, does not appear real-world significant given this dependent variable’s descriptive statistics (mean = 13.36, min = 1, max = 25).

### Statistical Significance (Categorical Variables: Record, Socioeconomic Class)

All the p-values for our categorical dummy variables are less than 0.05, suggesting statistical significance within the categories themselves given a confidence level of 95% while holding other independent variables constant (p-value yes\_record\_dummy = <.001; p-value low\_ses\_dummy = .023; p-value middle\_ses\_dummy = 0.010).

Since our categorical variable “record” has only two potential category values, we can conclude statistical significance overall for this variable based on its p-value as derived from the output’s t-statistic.

However, because our socioeconomic status variable - “SES” - consists of more than two possible categorical values, we will need to derive an F-statistic to consider overall statistical significance for the variable as a whole. To do so, we will need to run a regression without the dummy “SES” variables so that we can collect its sum of squared residuals (SSR) – this will be referred to as SSR0:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Summary** | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | .606a | .368 | .331 | 5.614 |
| a. Predictors: (Constant), yes\_record\_dummy, age | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 623.486 | 2 | 311.743 | 9.890 | <.001b |
| Residual | 1071.703 | 34 | 31.521 |  |  |
| Total | 1695.189 | 36 |  |  |  |
| a. Dependent Variable: length | | | | | | |
| b. Predictors: (Constant), yes\_record\_dummy, age | | | | | | |

From this SPSS output, we observe an SSR0 value of 1071.703 and from our full model output, we’ve observed a SSR value of 829.628 which we will label SSR1. Finally, we will need to calculate (1) the degrees of freedom for our full model which we will call df1 and (2) the number of dummy variables for the “SES” categorical variable which we call s.

We can calculate df1 as:

N – (#Xs in full model) – 1 = 37 – 4 – 1 = 32

We also know that we’re representing the “SES” categorical variable with two dummy variables, so our s will equal 2.

With the above information, we can now calculate our f-statistic as:

F = = =

From the text’s F-table with (2, 32) where 2 is our degrees of freedom in the numerator and 32 is our degrees of freedom in the denominator, we can estimate that a F-statistic of 4.669 roughly translates to a p-value of somewhere between .025 and .010, indicating that our “SES” variable is statistically significant overall while holding other independent variables constant.

### B-value Interpretation & Real-World Significance (Categorical Variables: Record & Socioeconomic Status)

For our “record” variable, our regression output reveals that individuals with a prior criminal record are likely to receive sentence lengths that are 6.841 years longer than individuals without a prior record. This difference is statistically significant given the dummy variable’s individual p-value (<.001) and appears real-world significant given the descriptive statistics for the sentence length dependent variable (mean = 13.36, min = 1, max = 25)

For our “SES” (socioeconomic status) variable, we can visualize the differences between variables with the below table (used instead of number line for technical ease):

|  |  |
| --- | --- |
| **Socioeconomic Status** | **Spacing from Reference (Working)** |
| Low | -4.941 |
| Middle | -5.778 |
| Working | 0 |

All calculations were completed while holding the other model independent variables constant. From our table, we can see that individuals belonging to a low socioeconomic status receive 4.941 fewer years in sentence length compared to working status individuals. Individuals of middle status receive 5.778 fewer years in sentence length compared to working status individuals. Finally, individuals of middle status receive .837 fewer years in sentence length compared to individuals of low status. In sum, sentence length is lowest for middle status individuals, next lowest for low status individuals, and highest for working status individuals. The differences in sentence length between (1) middle and working and (2) low and working status mentioned above appear real-world significant given the descriptive statistics of sentence length (mean = 13.36, min = 1, max = 25); the difference between low and middle class does not appear real-world significant given the same descriptive statistics, although subjectively speaking may be of some significance to an individual who would receive that additional sentence time.

# Q2.) From the multiple regression in Question 1, predict the sentence length for a 65-year-old working class offender with a prior record. Should you trust this predicted value? Explain.

Our model has produced the below regression formula:

Ŷ = a - .251\*(age) + 6.841\*(yes\_record\_dummy) – 4.941\*(low\_ses\_dummy) – 5.778\*(middle\_ses\_dummy)

The provided observation data will thus yield:

Ŷ = 20.046 - .251\*65 + 6.841\*1 - 4.941\*0 - 5.778\*0 = 10.572

As displayed above, our model predicts that a 65-year-old working class offender with a prior record would receive a sentence length of 10.572 years. Because our model has an R2 value of .511 and the age of the observation is outside the historical maximum of the variable (max = 49), we should place very limited trust in this prediction.