DISCUSSION GROUP PROBLEMS

# CHAPTER 12: CONTROLLING FOR A THIRD VARIABLE: MULTIPLE OLS REGRESSION

1. You have data on four variables for a sample of 24 convicted burglars:

*Y*: Sentence length for convicted burglars (in months)

*X*1: Defendant’s race (non-White = 1; White = 0)

*X*2: Number of prior convictions

*X*3: Dollar amount of money/goods stolen (in thousand dollar units)

Your dependent variable is the length of sentence received. Here are the data:

|  |  |  |  |
| --- | --- | --- | --- |
| Sentence Length (*Y*) | Defendant’s Race (*X*1) | # Prior Convictions (*X*2) | $ Amount Stolen (in Thousands) (*X*3) |
| 12.00 | 1.00 | 0.00 | 5.00 |
| 32.00 | 0.00 | 2.00 | 5.00 |
| 14.00 | 1.00 | 4.00 | 1.00 |
| 18.00 | 1.00 | 2.00 | 2.00 |
| 22.00 | 0.00 | 4.00 | 2.00 |
| 17.00 | 1.00 | 0.00 | 8.00 |
| 43.00 | 0.00 | 4.00 | 7.00 |
| 23.00 | 1.00 | 5.00 | 7.00 |
| 33.00 | 1.00 | 6.00 | 4.00 |
| 37.00 | 0.00 | 3.00 | 5.00 |
| 27.00 | 1.00 | 7.00 | 3.00 |
| 29.00 | 0.00 | 6.00 | 8.00 |
| 29.00 | 1.00 | 3.00 | 3.00 |
| 19.00 | 1.00 | 0.00 | 3.00 |
| 33.00 | 1.00 | 5.00 | 5.00 |
| 37.00 | 1.00 | 3.00 | 4.00 |
| 48.00 | 0.00 | 6.00 | 8.00 |
| 49.00 | 0.00 | 5.00 | 7.00 |
| 50.00 | 0.00 | 5.00 | 5.00 |
| 67.00 | 0.00 | 9.00 | 12.00 |
| 76.00 | 0.00 | 7.00 | 10.00 |
| 56.00 | 0.00 | 4.00 | 8.00 |
| 60.00 | 0.00 | 8.00 | 9.00 |
| 71.00 | 0.00 | 8.00 | 6.00 |

1. Do a scatterplot for the relationship between sentence length and the number of prior convictions and the dollar amount stolen in the crime, and also between your two *x* variables and determine if these are good variables to use in a multiple regression. [The correlations are *ryx*2 = .72, *ryx*3 = .69 and *rx*2x3 = .48]
2. What is the mean sentence length given to White defendants and non-White defendants and determine if race of defendant seems like a good variable to include in a multiple regression.
3. The following shows the relationship between sentence length and each independent variable in a bivariate regression. Interpret the slope coefficient, determine if you would reject the null hypothesis that each β = 0 with an alpha of .05, and interpret the value of R2.

|  |
| --- |
| **Coefficientsa** |
| Model | Unstandardized Coefficients | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
|  | (Constant) | 49.231 | 3.752 |  | 13.122 | .000 |
| Race of Defendant | -25.413 | 5.542 | -.699 | -4.586 | .000 |
| a. Dependent variable: sentence length |

R2 = .49

|  |
| --- |
| **Coefficientsa** |
| Model | Unstandardized Coefficients | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | 14.198 | 5.494 |  | 2.584 | .017 |
| Number of Prior Convictions | 5.295 | 1.086 | .721 | 4.875 | .000 |
| a. Dependent variable: sentence length |

*R*2 = .52

|  |
| --- |
| **Coefficientsa** |
| Model | Unstandardized Coefficients | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | 11.316 | 6.572 |  | 1.722 | .099 |
| Dollar Amount Stolen | 4.602 | 1.041 | .686 | 4.421 | .000 |
| a. Dependent variable: sentence length |

*R*2 = .47

1. Here is the result of the multivariate regression that has all three independent variables included. Interpret each slope coefficient, determine if you would reject the null hypothesis that each β = 0 with an alpha of .05, and interpret the value of *R*2. Why is the *R*2 value in this model not simply the sum of the three bivariate model *R*2s? Which variable has the greatest effect on sentence length?

|  |
| --- |
| **Coefficientsa** |
| Model | Unstandardized Coefficients | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | 18.156 | 7.733 |  | 2.348 | .029 |
| Race of Defendant | -12.378 | 5.096 | -.340 | -2.429 | .025 |
| Number of Prior Convictions | 3.103 | .977 | .422 | 3.176 | .005 |
| Dollar Amount Stolen | 1.997 | .949 | .298 | 2.105 | .048 |
| a. Dependent variable: sentence length |

*R*2 = .75