**Data Science – I Final Project Regression Output**

**Overall Model (Best Regression Model):**

|  |
| --- |
| **The SAS System** |

|  |  |
| --- | --- |
| **The REG Procedure** | |
| **Model: MODEL1** | |
| **Dependent Variable: New\_Total\_Enroll** | |
| **Number of Observations Read** | 33022 |
| **Number of Observations Used** | 32959 |
| **Number of Observations with Missing Values** | 63 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of** **Squares** | **Mean** **Square** | **F Value** | **Pr > F** |
| **Model** | 4 | 125245962 | 31311491 | 1056.85 | <.0001 |
| **Error** | 32954 | 976332531 | 29627 |  |  |
| **Corrected Total** | 32958 | 1101578493 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 172.12534 | **R-Square** | 0.1137 |
| **Dependent Mean** | 369.90425 | **Adj R-Sq** | 0.1136 |
| **Coeff Var** | 46.53240 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter** **Estimate** | **Standard** **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | 1 | 385.93810 | 3.78651 | 101.92 | <.0001 |
| **employ\_county** | 1 | 0.00024828 | 0.00000521 | 47.68 | <.0001 |
| **percap\_income\_county** | 1 | -0.00179 | 0.00010008 | -17.90 | <.0001 |
| **total\_officers** | 1 | -0.01017 | 0.00152 | -6.71 | <.0001 |
| **violent\_crime\_rate** | 1 | 0.14189 | 0.00407 | 34.86 | <.0001 |

**Interpretation:**

The overall model is statistically significant as indicated by the F-statistic (1056.85) and associated p-value (<.0001) which is less than the conventional significance level of 5%.

The R-squared value is 0.1137 which means that about 11.37% variability in the outcome variable (dependent variable), i.e., total school enrollment (New\_Total\_Enroll) can be explained by the regressors (independent variables), i.e., economic factors like employment (employ\_county) and per capita income (percap\_income\_county) of the county; law enforcement factor like total number of officers (total\_officers); and crime statistics like violent crime rates (violent\_crime\_rate). The R-squared value determines the explanatory power of the model. In this case, it is very less. Hence, it can be concluded that the model may not be a very good model.

The beta coefficient of employ\_county is 0.00024828 with the standard error of 0.00000521. It is statistically significant as indicated by the t-statistic (47.68) and associated p-value (<.0001) which is less than the conventional significance level of 0.05. This parameter estimate is positive which suggests that there is a positive relationship between the employment in the county (employ\_county) and the overall enrollment in the schools (New\_Total\_Enroll). Specifically, one unit increase in the employment (employ\_county) increases the total enrollment by 0.00024828 units, holding other variables constant.

The beta coefficient of percap\_income\_county is -0.00179 with the standard error of 0.00010008. It is statistically significant as indicated by the t-statistic (-17.90) and associated p-value (<.0001) which is less than the conventional significance level of 0.05. This parameter estimate is negative which suggests that there is a negative relationship between the per capita income of the residents in the county (percap\_income\_county) and the overall enrollment in the schools (New\_Total\_Enroll). Specifically, one unit increase in the per capita income (percap\_income\_county) decreases the total enrollment (New\_Total\_Enroll) by 0.00179 units, holding other variables constant.

The beta coefficient of total\_officers is -0.01017 with the standard error of 0.00152. It is statistically significant as indicated by the t-statistic (-6.71) and associated p-value (<.0001) which is less than the conventional significance level of 0.05. This parameter estimate is negative which suggests that there is a negative relationship between the total number of officers in the county (total\_officers) and the overall enrollment in the schools (New\_Total\_Enroll). Specifically, one unit increase in the total number of officers in the county (total\_officers) decreases the total enrollment (New\_Total\_Enroll) by 0.01017 units, holding other variables constant.

The beta coefficient of violent\_crime\_rate is 0.14189 with the standard error of 0.00407. It is statistically significant as indicated by the t-statistic (34.86) and associated p-value (<.0001) which is less than the conventional significance level of 0.05. This parameter estimate is positive which suggests that there is a positive relationship between the violent crime rate in the county (violent\_crime\_rate) and the overall enrollment in the schools (New\_Total\_Enroll). Specifically, one unit increase in the violent crime rate (violent\_crime\_rate) increases the total enrollment by 0.14189 units, holding other variables constant.

**Individual Models:**

1. **Economic Factors: Employment & Per Capita Income**

|  |
| --- |
| **The SAS System** |

|  |  |
| --- | --- |
| **The REG Procedure** | |
| **Model: MODEL1** | |
| **Dependent Variable: New\_Total\_Enroll** | |
| **Number of Observations Read** | 33022 |
| **Number of Observations Used** | 32959 |
| **Number of Observations with Missing Values** | 63 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of** **Squares** | **Mean** **Square** | **F Value** | **Pr > F** |
| **Model** | 2 | 88423322 | 44211661 | 1438.12 | <.0001 |
| **Error** | 32956 | 1013155171 | 30743 |  |  |
| **Corrected Total** | 32958 | 1101578493 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 175.33586 | **R-Square** | 0.0803 |
| **Dependent Mean** | 369.90425 | **Adj R-Sq** | 0.0802 |
| **Coeff Var** | 47.40034 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter** **Estimate** | **Standard** **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | 1 | 433.84774 | 3.59481 | 120.69 | <.0001 |
| **employ\_county** | 1 | 0.00026635 | 0.00000503 | 52.97 | <.0001 |
| **percap\_income\_county** | 1 | -0.00226 | 0.00010061 | -22.50 | <.0001 |

**Interpretation:**

The overall model is statistically significant as indicated by the F-statistic (1438.12) and associated p-value (<.0001) which is less than the conventional significance level of 5%.

The R-squared value is 0.0803 which means that about 8% variability in the outcome variable (dependent variable), i.e., total school enrollment (New\_Total\_Enroll) can be explained by the regressors (independent variables), i.e., economic factors like employment (employ\_county). The R-squared value determines the explanatory power of the model. In this case, it is very less. Hence, it can be concluded that the model may not be a very good model.

The beta coefficient of employ\_county is 0.00026635 with the standard error of 0.00000503. It is statistically significant as indicated by the t-statistic (52.97) and associated p-value (<.0001) which is less than the conventional significance level of 0.05. This parameter estimate is positive which suggests that there is a positive relationship between the employment in the county (employ\_county) and the overall enrollment in the schools (New\_Total\_Enroll). Specifically, one unit increase in the employment (employ\_county) increases the total enrollment by 0.00026635 units, holding other variables constant.

The beta coefficient of percap\_income\_county is -0.00226 with the standard error of 0. 00010061. It is statistically significant as indicated by the t-statistic (-22.50) and associated p-value (<.0001) which is less than the conventional significance level of 0.05. This parameter estimate is negative which suggests that there is a negative relationship between the per capita income of the residents in the county (percap\_income\_county) and the overall enrollment in the schools (New\_Total\_Enroll). Specifically, one unit increase in the per capita income (percap\_income\_county) decreases the total enrollment (New\_Total\_Enroll) by 0.00226 units, holding other variables constant.

1. **Law Enforcement: Total Number of Officers**

|  |
| --- |
| **The SAS System** |

|  |  |
| --- | --- |
| **The REG Procedure** | |
| **Model: MODEL1** | |
| **Dependent Variable: New\_Total\_Enroll** | |
| **Number of Observations Read** | 33022 |
| **Number of Observations Used** | 32959 |
| **Number of Observations with Missing Values** | 63 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of** **Squares** | **Mean** **Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 2245429 | 2245429 | 67.32 | <.0001 |
| **Error** | 32957 | 1099333064 | 33357 |  |  |
| **Corrected Total** | 32958 | 1101578493 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 182.63786 | **R-Square** | 0.0020 |
| **Dependent Mean** | 369.90425 | **Adj R-Sq** | 0.0020 |
| **Coeff Var** | 49.37436 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter** **Estimate** | **Standard** **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | 1 | 369.25678 | 1.00910 | 365.93 | <.0001 |
| **total\_officers** | 1 | 0.01240 | 0.00151 | 8.20 | <.0001 |

**Interpretation:**

The overall model is statistically significant as indicated by the F-statistic (67.32) and associated p-value (<.0001) which is less than the conventional significance level of 5%.

The R-squared value is 0.0020 which means that about 0.2% variability in the outcome variable (dependent variable), i.e., total school enrollment (New\_Total\_Enroll) can be explained by the regressor (independent variable), i.e., law enforcement factor like total number of officers (total\_officers). The R-squared value determines the explanatory power of the model. In this case, it is very less. Hence, it can be concluded that the model may not be a very good model.

The beta coefficient of total\_officers is 0.01240 with the standard error of 0.00151. It is statistically significant as indicated by the t-statistic (8.20) and associated p-value (<.0001) which is less than the conventional significance level of 0.05. This parameter estimate is positive which suggests that there is a positive relationship between the total number of officers in the county (total\_officers) and the overall enrollment in the schools (New\_Total\_Enroll). Specifically, one unit increase in the total number of officers in the county (total\_officers) increases the total enrollment (New\_Total\_Enroll) by 0.01240 units, holding other variables constant.

1. **Crime Factors: Violent Crime Rate**

|  |
| --- |
| **The SAS System** |

|  |  |
| --- | --- |
| **The REG Procedure** | |
| **Model: MODEL1** | |
| **Dependent Variable: New\_Total\_Enroll** | |
| **Number of Observations Read** | 33022 |
| **Number of Observations Used** | 32959 |
| **Number of Observations with Missing Values** | 63 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of** **Squares** | **Mean** **Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 55644334 | 55644334 | 1753.33 | <.0001 |
| **Error** | 32957 | 1045934159 | 31736 |  |  |
| **Corrected Total** | 32958 | 1101578493 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 178.14693 | **R-Square** | 0.0505 |
| **Dependent Mean** | 369.90425 | **Adj R-Sq** | 0.0505 |
| **Coeff Var** | 48.16028 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter** **Estimate** | **Standard** **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | 1 | 329.62905 | 1.37406 | 239.89 | <.0001 |
| **violent\_crime\_rate** | 1 | 0.17348 | 0.00414 | 41.87 | <.0001 |

**Interpretation:**

The overall model is statistically significant as indicated by the F-statistic (1753.33) and associated p-value (<.0001) which is less than the conventional significance level of 5%.

The R-squared value is 0.0505 which means that about 5.1% variability in the outcome variable (dependent variable), i.e., total school enrollment (New\_Total\_Enroll) can be explained by the regressor (independent variable), i.e., crime statistics like violent crime rates (violent\_crime\_rate). The R-squared value determines the explanatory power of the model. In this case, it is very less. Hence, it can be concluded that the model may not be a very good model.

The beta coefficient of violent\_crime\_rate is 0.17348 with the standard error of 0.00414. It is statistically significant as indicated by the t-statistic (41.87) and associated p-value (<.0001) which is less than the conventional significance level of 0.05. This parameter estimate is positive which suggests that there is a positive relationship between the violent crime rate in the county (violent\_crime\_rate) and the overall enrollment in the schools (New\_Total\_Enroll). Specifically, one unit increase in the violent crime rate (violent\_crime\_rate) increases the total enrollment by 0.17348 units, holding other variables constant.

1. **Law Enforcement and Crime factors: Total number of officers and violent crime rate**

|  |
| --- |
| **The SAS System** |

|  |  |
| --- | --- |
| **The REG Procedure** | |
| **Model: MODEL1** | |
| **Dependent Variable: New\_Total\_Enroll** | |
| **Number of Observations Read** | 33022 |
| **Number of Observations Used** | 32959 |
| **Number of Observations with Missing Values** | 63 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of** **Squares** | **Mean** **Square** | **F Value** | **Pr > F** |
| **Model** | 2 | 56593817 | 28296908 | 892.41 | <.0001 |
| **Error** | 32956 | 1044984677 | 31708 |  |  |
| **Corrected Total** | 32958 | 1101578493 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 178.06875 | **R-Square** | 0.0514 |
| **Dependent Mean** | 369.90425 | **Adj R-Sq** | 0.0513 |
| **Coeff Var** | 48.13915 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter** **Estimate** | **Standard** **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | 1 | 329.57922 | 1.37349 | 239.96 | <.0001 |
| **total\_officers** | 1 | 0.00808 | 0.00148 | 5.47 | <.0001 |
| **violent\_crime\_rate** | 1 | 0.17188 | 0.00415 | 41.40 | <.0001 |

**Interpretation:**

The overall model is statistically significant as indicated by the F-statistic (892.41) and associated p-value (<.0001) which is less than the conventional significance level of 5%.

The R-squared value is 0.0514 which means that about 5.14% variability in the outcome variable (dependent variable), i.e., total school enrollment (New\_Total\_Enroll) can be explained by the regressors (independent variables), i.e., law enforcement factor like total number of officers (total\_officers); and crime statistics like violent crime rates (violent\_crime\_rate). The R-squared value determines the explanatory power of the model. In this case, it is very less. Hence, it can be concluded that the model may not be a very good model.

The beta coefficient of total\_officers is 0.00808 with the standard error of 0.00148. It is statistically significant as indicated by the t-statistic (5.47) and associated p-value (<.0001) which is less than the conventional significance level of 0.05. This parameter estimate is positive which suggests that there is a positive relationship between the total number of officers in the county (total\_officers) and the overall enrollment in the schools (New\_Total\_Enroll). Specifically, one unit increase in the total number of officers in the county (total\_officers) increases the total enrollment (New\_Total\_Enroll) by 0.00808 units, holding other variables constant.

The beta coefficient of violent\_crime\_rate is 0.17188 with the standard error of 0.00415. It is statistically significant as indicated by the t-statistic (41.40) and associated p-value (<.0001) which is less than the conventional significance level of 0.05. This parameter estimate is positive which suggests that there is a positive relationship between the violent crime rate in the county (violent\_crime\_rate) and the overall enrollment in the schools (New\_Total\_Enroll). Specifically, one unit increase in the violent crime rate (violent\_crime\_rate) increases the total enrollment by 0.17188 units, holding other variables constant.