A regression analysis was conducted by a junior data scientist to determine the relationship between the amounts a hospital charges for a medical service (AVE\_ave\_provider\_charge), the amount a hospital is reimbursed by Medicare (AVE\_ave\_medicare\_payment), and the number of services a hospital provides (AVE\_num\_service).

The model formula was specified as:

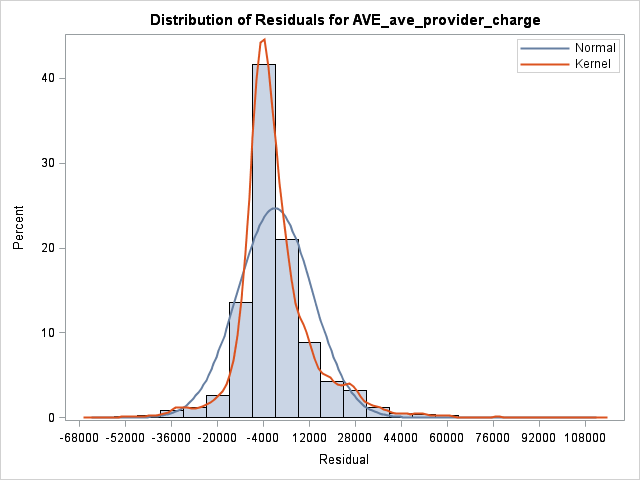
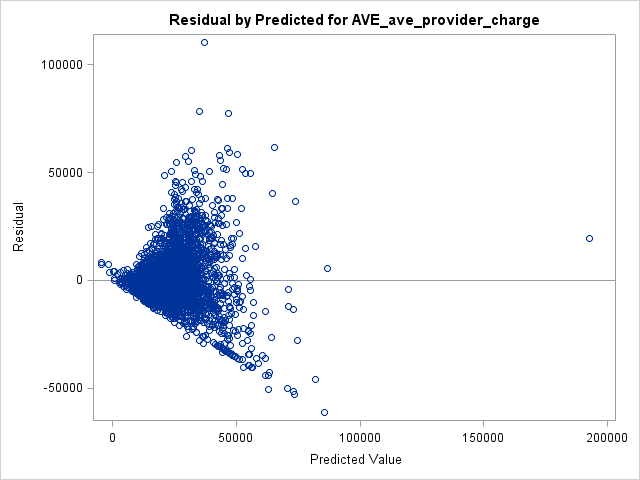
AVE\_ave\_provider\_charge ~ AVE\_ave\_medicare\_payment + AVE\_num\_service

Among many other tables and plots, the following information was provided by the statistical software package after training the traditional regression model:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Analysis of Variance |  |  |  |  |  |
| Source | DF | Sum of  Squares | Mean  Square | F Value | Pr > F |
| Model | 2 | 3.85E+11 | 1.92E+11 | 1148.9 | <.0001 |
| Error | 3334 | 5.58E+11 | 167376011 |  |  |
| Corrected Total | 3336 | 9.43E+11 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Root MSE | 12937 | R-Square | 0.408 |
| Dependent Mean | 24721 | Adj R-Sq | 0.4076 |
| Coeff Var | 52.33355 |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Parameter Estimates |  |  |  |  |  |  |  |
| Variable | Label | DF | Parameter  Estimate | Standard  Error | t Value | Pr > |t| | Variance  Inflation |
| Intercept | Intercept | 1 | -1219.43 | 598.38 | -2.04 | 0.0416 | 0 |
| AVE\_ave\_medicare\_payment | Average Medicare Payment | 1 | 3.83 | 0.08 | 47.88 | <.0001 | 1.02 |
| AVE\_num\_service | Number of Services | 1 | -5.84 | 1.17 | -4.96 | <.0001 | 1.02 |

1.) **(2 pts.)** State the exact interpretation of the presented standard R-Square statistic.

2.) **(2 pts.)** State the exact interpretation of the presented parameter estimate for AVE\_ave\_medicare\_payment.

3.) **(3 pts.)** As you may have noticed, there is a serious problem with this regression analysis. Given the information provided what is the technical term that describes this problem?

4.) **(3 pts.)** The presented output states that the parameter describing the linear relationship between the target variable and both input variables is statistically different from zero at the default alpha=0.05 level for the parameter ***t***-tests. Given the problem identified in 3 above, will the ***t***-tests remain unbiased?