DSC 423: DATA ANALYSIS AND REGRESSION

FINAL PROJECT

BOSTON HOUSING DATASET

Swathi BABU ID: 2061690

TABLE OF CONTENTS

I. INTRODUCTION

The objective of this report was to predict the median value of houses occupied by the owners using SAS. The dataset used for this analysis is the Boston Housing Data collected by Harrison D. and Rubinfield D.L. who used it in the paper "Hedonic prices and the demand for clean air", Journal of Environmental Economics and Management, Vol. 5, Issue 1, 1978. [1]

The dataset that has been used here is a subset of the original set and it has 375 observations and 14 variables. There is one dependent variable, medv and 13 independent variables. One objective is to find the variables that would be a part of the model that is predicted.

To understand the dataset and how the variables affect each other, looking at articles regarding median value of houses helped at setting things in perspective. It helped in understanding what the important variables are and how we would incorporate it in the model. For example, one of the independent variables, dis, which is the weighted distance from five employment centers in Boston, is very important when looking for places. Proximity to work opportunities drastically improves the condition of living in a place. So, it is kept in mind that this is an important variable and must be included in the model. [2]

II. METHODOLOGY

To predict the dependent variable, medv, median value of owner-occupied homes, we use linear regression and build a model with significant independent variables. As there are multiple independent variables to predict one dependent variable, we use multiple linear regression. [3]

The dataset is imported under the data file "boston" using the infile function and using procedure print the first 10 observations of the imported dataset is checked (Refer Appendix A, fig. 1). Before proceeding with the process of model prediction, the dataset is analyzed first. Relationship between variables between dependent variable and independent variables and between independent variables themselves are looked through and further analysis is done.

The regression model is constructed after that. This report uses two methods to predict the model and concludes with one model based on various parameters. The two methods exercised are fitting the model with log transformation and through quadratic polynomial regression. The final model is obtained by comparing all these models. The results are further explained in detail under each section in this report.

III. EXPLORATORY DATA ANALYSIS

There are 13 independent variables – crime, zn, indus, chas, nox, rm, age, dis, rad, tax, ptratio, minor and lstat. Variables crime, zn, indus, nox, rm, age, dis, rad, tax, ptratio, minor and lstat are numeric variables. Variable chas is categorical binary variable that has values 0 and 1. Although there are no dummy variables added in the code, there is a binary variable. There aren't any variables that go well together enough to create interaction variables.

The important descriptive of all the variables are checked using the procedure means. The mean, minimum, maximum, standard deviation. standard error, first quartile, second quartile or median, third quartile, confidence intervals and mode (Refer Appendix A, fig. 2). The crime variable has a minimum of 0.00632, mean of 0.7182351, maximum of 18.4982 and Q3 is 0.537. This shows that 75% of the observations have crime rate below 0.537 and the maximum crime rate is much further away than the Q3. There could be some issues such as outliers on the right side there. Similar effect can be seen in zn, indus, rad and minor. For the zn variable, the minimum is 0, maximum is 100, mean is 15.334, Q2 is 0 and Q3 is 22. So, in this case most observations might have the value 0 for zn variable. For indus variable, minimum is 0.46, maximum is 25.65 and Q3 is 10.59. There are chances of outliers on the right side. The rad variable has minimum 1 and maximum 24. 75% of the observations have rad variable 5 or less than 5. The maximum is larger here as well. Chance of outliers on the right side can be identified from that. The variable minor has minimum 70.8, maximum 369.9 and Q3 is 396.06. This shows that there might be outliers on the left side of the graph. Also, from the standard deviations, the variables zn, age, tax and minor look like they would add a lot to the model since they have high standard deviations compared to the other variables. But that could also be because these variables have a higher range than the other variables i.e., they are not normalized. [3] Moreover, there are no missing values in the dataset that was used.
To check the distribution of data with respect to all the variables, histograms are plotted using procedure univariate. From the crime histogram, it can be seen that the graph is skewed to the right with possible outliers on the right side and it is unimodal (Refer Appendix A, fig. 3). The zn is also skewed to the right and is unimodal (Refer Appendix A, fig. 4). The indus variable is skewed to the right and is also unimodal (Refer Appendix A, fig. 5). Chas is a categorical binary variable and there are more observations with without track bounded rivers than with them (Refer Appendix A, fig. 6). The nox variable is slightly skewed to the right due to outliers and it is unimodal (Refer Appendix A, fig. 7). The rm variable is slightly skewed to the left due to some outliers and it is also unimodal (Refer Appendix A, fig. 8). The age variable is skewed to the left and it is unimodal (Refer Appendix A, fig. 9). The variable dis is skewed to the right, has some outliers on the right and is unimodal (Refer Appendix A, fig. 10). The variable rad is also skewed to the right due to some observations that have extreme values (outliers) and it is unimodal (Refer Appendix A, fig. 11). The tax variable is unimodal, skewed to the right and has some outliers on the right side of the graph (Refer Appendix A, fig. 12). The ptratio variable is skewed to the left, has outliers on the left side of the graph and it is also unimodal (Refer Appendix A, fig. 13). The minor is extremely skewed to the left, has outliers on the left side and is unimodal as well (Refer Appendix A, fig. 14). The lstat is skewed to the right, has outliers on the right and

is unimodal (Refer Appendix A, fig. 15). The medv is skewed to the right (Refer Appendix A, fig. 16).

Moving on to boxplots, boxplots and histograms convey similar information but boxplots are clearer on the outliers whereas histograms are not, but histograms are clearer on modal distribution of the graphs. There are many outliers on the right side with respect to crime variable. More than 15 observations higher crime rate than the estimated maximum (Refer Appendix A, fig. 17). There are 9 outliers on the right side for the variable zn. As already seen, most of the observations, around 240 of observations have the value 0 for zn (Refer Appendix A, fig. 18). The indus variable has 2 outliers on the right side (Refer Appendix A, fig. 19). Chas variable is a categorical variable so no point in checking univariate boxplot (Refer Appendix A, fig. 20). The nox variable has 3 outliers on the right side (Refer Appendix A, fig. 21). For rm variable, there are more than 5 outliers on the right side and there are exactly 3 on the left side (Refer Appendix A, fig. 22). The age variable has no outliers (Refer Appendix A, fig. 23). The variable dis has one outlier on the right side (Refer Appendix A, fig. 24). The rad variable has 3 outliers on the right side and 2 on the left (Refer Appendix A, fig. 25). The tax variable has 1 outlier on the right (Refer Appendix A, fig. 26). There are no outliers for ptratio (Refer Appendix A, fig. 27). There are more than 15 outliers on the left side for minor variable (Refer Appendix A, fig. 28). There are around 8 outliers on the right for lstat variable (Refer Appendix A, fig. 29). The boxplot for medv has also been plotted (Refer Appendix A, fig. 30). Boxplot is an important analysis step that tells the user a lot about the data that are crucial for further steps. [3]

Scatterplots are an important way to look at linearity in the relationships between dependent variable medv and the independent variables. The variable crime seems to slightly linear with respect to the dependent variable medv with possibly very low negative correlation (Refer to Appendix A, fig. 31). The zn variable seems to have a linear relationship with medv variable. They seem to have a low positive correlation (Refer to Appendix A, fig. 32). The indus variable also has a slight linear relation with respect to medv and also has low negative correlation (Refer to Appendix A, fig. 33). Chas variable is a categorical binary variable so no point of checking for linearity (Refer to Appendix A, fig. 34). The nox variable also only a slight linear relationship with medv and has low negative correlation (Refer to Appendix A, fig. 35). The rm clearly has a linear relationship with medv and they have high positive correlation (Refer to Appendix A, fig. 36). The age has a very subtle linear relationship with variable medv and has low negative correlation (Refer to Appendix A, fig. 37). The dis variable also has very slight linear relationship with medv and has very low negative correlation (Refer to Appendix A, fig. 38). The rad variable which is actually an index of values 1 to 24 is discrete and not continuous. Stating that, it has low linear relationship and has low positive correlation (Refer to Appendix A, fig. 39). The tax variable also has slight linear relationship with medv and has low negative correlation (Refer to Appendix A, fig. 40). The ptratio has a linear relationship with medv and has moderate positive correlation (Refer to Appendix A, fig. 41). The minor variable also has low linearity with respect to medv and has low positive correlation (Refer to Appendix A, fig. 42). The lstat variable has a good linear relationship with medv and has moderate negative correlation (Refer to Appendix A, fig. 43).

Another sure way of analyzing the linear relationship between the variables is Pearson's correlation coefficients. High correlation values between independent variables can also cause issues like multicollinearity etc.… Looking at the correlation table, the dependent variable medv has low positive correlation with zn, chas, rad, minor, has high positive correlation with rm, has low negative correlation with crime, indus, nox, age, dis, tax, has moderate negative correlation with ptratio, lstat. The variables crime and rad, tax have high positive correlation. Dis and zn seem to have moderate positive correlation. Dis and indus seem to have moderate negative correlation. Nox and indus also seem to have high positive correlation. Dis and nox have high negative correlation. Age and nox have moderate positive correlation. Lstat and rm have moderate negative correlation. Age and dis have high negative correlation. Rad and tax have high positive correlation. (Refer to Appendix A, fig. 44) Another important point to note is that there aren't any independent variables that have correlation as high as +/- 0.9 which means that there probably aren't any variables that are collinear. [3]

IV. FULL MODEL

The full model with all 13 independent variables was modeled using the procedure regression in SAS to predict the medv variable. All the parameters, assumptions and model diagnostics are checked for the full model.

The model obtained has an R^2 value of 0.7801 and adjusted R^2 of 0.7722 which means that 78.01% of the variance of medv variable can be explained by the 13 independent variables and also 77.22%of the variance of medv variable is explained by the 13 independent variables crime, zn, indus, chas, nox, rm, age, dis, rad, tax, ptratio, minor and lstat. The error RMSE value is 4.18606 which is pretty low.

The hypothesis to check if the model is good or not is framed and a goodness of fit test is run. The null hypothesis H_0 is that no independent variable has an effect on the dependent variable medv i.e., all $\beta_i = 0$. The alternate hypothesis H_1 is that there is at least one independent variable that has an effect on the dependent medv variable i.e., at least one $\beta_i \neq 0$. On running the goodness of fit test, we get F-value as 98.53 which is not that high but p-value is less 0.05 which means that the null hypothesis is rejected and there is at least one independent variable that has an effect on medv.

Using the signs of the parameter estimates or beta coefficients, it can be said that the independent variables crime, zn, indus, chas, rm, rad and minor have a positive effect on the dependent variable medv and the variables nox, dis, age, tax, ptratio, lstat have a negative effect on medv. The effect can be expanded into if the per capita crime rate of the town increase by 1%, then the median value of the house will increase by (1.28883*1000) = 1288.83$. This can be done for all the variables in the model. Logically speaking, increase in crime rate might not lead to increase in home value so there could be some inconsistency here.

Based on standardized beta estimates, rm is the independent variable with the highest influence on medv followed by lstat, dis, crime, nox, rad, tax, ptratio, zn, minor, indus, age

and finally chas. This is estimated using the absolute value of the standardized beta estimates. Variable rm had the highest value followed by the other variables in the order mentioned above. [3]

The model equation is *medv = 12.63324 + 1.28883crime + 0.03579zn + 0.07362indus + 0.12124chas – 17.35683nox + 6.41932rm – 0.00650age – 1.18892dis + 0.39406rad – 0.01565tax – 0.71652ptratio + 0.01167minor – 0.042714lstat, where chas = 1 if tract bounds river, 0 otherwise.*

Significance test is also run on each independent variable. From the p-values for these variables, it can be seen that the variables chas, age and indus are insignificant as their p-values are greater than 0.05 (Refer to Appendix B, table 1).

The four assumptions of the regression model are checked using student residual plots and normality plot. The linearity assumption can be checked using scatterplots. It was addressed in the scatterplot section of data exploration. The assumption is satisfied. The constant variance and independency assumption go hand in hand. The residual plot for variable crime has a cone kind of shape and hence does not satisfy the constant variance and independency assumptions (Refer to Appendix B, fig. 2). The residual plot for zn variable does not have randomly scattered points. It does not satisfy the constant variation and independency assumptions (Refer to Appendix B, fig. 3). The indus variable seems to have a tunnel shape and hence does not satisfy constant variance and independency (Refer to appendix B, fig. 4). The assumptions need not be checked for chas variable as it is categorical (Refer to Appendix B, fig. 5). The nox variable also has a tunnel shape. Therefore, the assumptions constant variance and independency are not satisfied (Refer to Appendix B, fig. 6). Rm has a U-shape (Refer to Appendix B, fig. 7). Age has a tunnel shape (Refer to Appendix B, fig. 8). Variable dis also has a tunnel shape (Refer to Appendix B, fig. 9). They do not satisfy the constant variance and independence assumption. Variable rad has index values so they are discrete ad do not satisfy the assumptions (Refer to Appendix B, fig. 10). The variables pat, ptratio and minor have a tunnel shape. So, they do not satisfy the assumptions either (Refer to Appendix B, fig. 11,12,13). The variable lstat has a U -shape (Refer to Appendix B, fig. 14). The predicted student residual plot also has a U-shape (Refer Appendix B, fig. 15). The normal probability plot seems to have a slight S shape but it is almost a straight line (Refer to Appendix B, fig. 16). [3]

Using the VIF, r, influence add-ons in the model, the model diagnostics can be assessed. All the variables have variation inflation factor (VIF) lesser than 10. So, there is no issue of collinearity which was observed through correlation values as well (Refer to Appendix B, table 1). There are outliers and influential points for this model. Observations that are both outliers and influential points are observation number 215, 357, 365, 366, 369, 372, 373 and 375 (Refer to Appendix B, fig. 1). There are many more influential points but those are left in the model as removing them will lead to too much loss of data. Moving forward, we will be removing the above-mentioned observations from the boston dataset and create a new one without these- “boston_new” which is what will be used for the rest of the steps. [3]

This model is a moderate model with 78% accuracy but there are assumptions that are not met and many outliers. Proceeding to running the model again on "boston_new" dataset after removing the outliers.

The model built after removing outliers have R^2 value of 0.8599 and adjusted R^2 of 0.8547 and that means that 85.99% of the variance of medv is explained by the independent variables and 85.47% of the variance of medv is explained by them. This is better than the previous model with almost 5% increase in accuracy. The error RMSE is 3.25627 which has also decreased from the previous model.

The model null hypothesis and alternate hypothesis remains the same and it can be seen here that F-value is 166.65, it has increased as well and the p-value is less than 0.05. Hence, the null hypothesis is rejected. So, this model is good and has at least one independent variable that influences the dependent variable medv.

Using the signs of the parameter estimates or beta coefficients, it can be said that the independent variables crime, zn, indus, chas, rm, rad and minor have a positive effect on the dependent variable medv and the variables nox, dis, age, tax, ptratio, lstat have a negative effect on medv. The effect can be expanded into if the per capita crime rate of the town increase by 1%, then the median value of the house will increase by (1.725577*1000) = 1725.77$. This can be done for all the variables in the model.

Based on standardized beta estimates, rm is the independent variable with the highest influence on medv followed by crime, dis, lstat, nox, tax, ptratio, rad, age, zn, minor, indus, and chas. This is estimated using the absolute value of the standardized beta estimates (Refer to Appendix B, table 2).

The model equation is *medv = -8.01046+ 1.72577 crime + 0.02307zn + 0.05365indus+ 1.08971chas– 12.53887nox +8.51415rm– 0.02453 age –0.90588dis + 0.21037 rad – 0.01464tax – 0.57630ptratio + 0.01385minor – 0.24693 lstat, where chas = 1 if tract bounds river, 0 otherwise.*

Significance test is also run on each independent variable. From the p-values for these variables, it can be seen that the variables chas and indus are insignificant as their p-values are greater than 0.05 (Refer to Appendix B, table 2).

The four assumptions of the regression model are checked using student residual plots and normality plot. The linearity assumption can be checked using scatterplots. It was addressed in the scatterplot section of data exploration. The assumption is satisfied. The constant variance and independency assumption go hand in hand. The residual plots haven't changed from the previous model much. The assumptions constant variance and independency are not satisfied. The normal probability plot looks like a straight line. So, the assumption of normality is satisfied (Refer to Appendix B, fig. 18 – 32).

Using the VIF, r, influence add-ons in the model, the model diagnostics is assessed. All the variables have variation inflation factor (VIF) lesser than 10. So, there is no issue of

collinearity in this model as well. There are outliers and influential points for this model. Observations that are both outliers and influential points are observation numbers 162, 182, 365 and 367 (Refer to Appendix B, fig. 17). Removing these points does not add much to the R^2. There are many more influential points but those are left in the model as removing them will lead to too much loss of data here as well.

This model with 86% accuracy is better than the previous one but still not the best one as the assumptions constant variance and independency are violated. To resolve this, two methods are undertaken here as mentioned in the methodology section.

V. LOG TRANSFORMATION

A new dataset "boston_log" from the "boston_new" dataset was created with the log variable i.e., the variable logmedv which is log of medv is added as the independent variable in the new dataset. The full model with all 13 variables is run to predict logmedv rather than medv.

The model obtained has an R^2 value of 0.8777 and adjusted R^2 of 0.8732 which means that 87.77% of the variance of logmedv variable can be explained by the 13 independent variables and also 87.32%of the variance of logmedv variable is explained by them. This is not a huge increase in the values but this model is better. The error RMSE value is 0.11013 which is a decrease from the previous model RMSE.

The hypothesis to check if the model is good or not is framed and a goodness of fit test is run. The null hypothesis H_O is that no independent variable has an effect on the dependent variable logmedv i.e., all $\beta_i = 0$. The alternate hypothesis H_1 is that there is at least one independent variable that has an effect on the dependent logmedv variable i.e., at least one $\beta_i \neq 0$. On running the goodness of fit test, we get F-value as 194.96 which has also increased from the previous model and p-value is less 0.05 which means that the null hypothesis is rejected and there is at least one independent variable that has an effect on logmedv. [4]

Using the signs of the parameter estimates or beta coefficients, it can be said that the independent variables crime, zn, indus, chas, rm, rad and minor have a positive effect on the dependent variable logmedv and the variables nox, dis, age, tax, ptratio, lstat have a negative effect on logmedv. The effect can be expanded into if the per capita crime rate of the town increase by 1%, then the median value of the house will increase by $(e^{0.04567}-1)$ *100% i.e., 4.67%. This can be done for all the variables in the model. [4]

Based on standardized beta estimates, rm is the independent variable with the highest influence on logmedv followed by lstat, dis, nox, crime, ptratio, tax, rad, age, minor, zn, indus and chas. This is estimated using the absolute value of the standardized beta estimates. (Refer to Appendix C, table 1)

The model equation is *log(medv) = 2.51620 + 0.04567crime + 0.00074642zn +0.00227indus+ 0.03477chas – 0.57152nox + 0.25083rm – 0.00092026age – 0.03512dis + 0.01240rad –*

0.00057326tax – 0.02534ptratio + 0.00060553minor – 0.01516lstat , where chas = 1 if tract bounds river, 0 otherwise.

Significance test is also run on each independent variable. From the p-values for these variables, it can be seen that the variables chas and indus are insignificant as their p-values are greater than 0.05 in this model also(Refer to Appendix C, table 1).

The four assumptions of the regression model are checked using student residual plots and normality plot. The linearity assumption can be checked using scatterplots. The assumption is satisfied. The variables student residual plots are more random around the origin now than the previous model especially the student residuals vs predicted values plot has significant improvement with the transformation (Refer to Appendix C, fig. 2-16). [4]

Using the VIF, r, influence add-ons in the model, the model diagnostics can be assessed. All the variables have variation inflation factor (VIF) lesser than 10. So, there is no issue of collinearity (Refer to Appendix C, table 1). There are outliers and influential points for this model. Observations that are both outliers and influential points are observation numbers 8, 182 and 367 (Refer to Appendix C, fig. 1). The outliers are not removed as they don't add much R^2 to the model. Removing influential points could lead to too much loss in data. This model with 87% accuracy is a pretty good model and it is better than the previous one because it satisfies all the assumptions and model diagnostics.

After looking at the model, the dataset is split into training and testing through the holdout method with 80% training and 20% testing with seed value 317380 using procedure surveyselect and a new dataset is created for the training called "boston_all_log" where a new variable "new_medv" is created to assign the logmedv value to it if it part of the training. There are 294 observations for training and 73 for testing.

Further, two model selection methods Stepwise and backward selection is run on the new dataset "boston_all_log". On running both these methods, they arrive at the same models with 12 variables- crime, zn, chas, nox, rm, dis, rad, tax, ptratio, minor and lstat.

The model obtained has an R^2 value of 0.8854 and adjusted R^2 of 0.8805 which means that 87.77% of the variance of new_medv variable can be explained by the 12 independent variables and also 88.54% of the variance of new_medv variable is explained by them. The error RMSE value is 0.10332 which is a decrease from the previous model RMSE, although not much.

The hypothesis to check if the model is good or not is framed and a goodness of fit test is run. The null hypothesis H_0 is that no independent variable has an effect on the dependent variable new_medv i.e., all $\beta_i = 0$. The alternate hypothesis H_1 is that there is at least one independent variable that has an effect on the dependent new_medv variable i.e., at least one $\beta_i \neq 0$. On running the goodness of fit test, we get F-value as 180.98 which has also increased from the previous model and p-value is less 0.05 which means that the null hypothesis is rejected and there is at least one independent variable that has an effect on new_medv.

Using the signs of the parameter estimates or beta coefficients, it can be said that the independent variables crime, zn, chas, rm, rad and minor have a positive effect on the dependent variable new_medv and the variables nox, dis, age, tax, ptratio, lstat have a negative effect on new_medv. The effect can be expanded into if the per capita crime rate of the town increase by 1%, then the median value of the house will increase by ($e^{0.05996}$-1)*100% i.e., 6.18%. This can be done for all the variables in the model.

Based on standardized beta estimates, rm is the independent variable with the highest influence on logmedv followed by dis, crime, nox, lstat, ptratio, tax, rad, age, minor, zn and chas. This is estimated using the absolute value of the standardized beta estimates. (Refer to Appendix C, table 2)

The model equation is *new_medv = log(medv) = 2.47906 + 0.05996crime + 0.00076665zn + 0.04569chas – 0.66316nox + 0.25313rm – 0.00123age – 0.03781dis + 0.01079rad – 0.00048448tax - 0.02516ptratio + 0.00074009minor – 0.01197lstat , where chas = 1 if tract bounds river, 0 otherwise.*

Significance test is also run on each independent variable. From the p-values for these variables, it can be seen that all variables in the model are significant as their p-values are less than 0.05 in this model (Refer to Appendix C, table 2).

The four assumptions of the regression model are checked using student residual plots and normality plot. The linearity assumption is satisfied as confirmed by the scatterplots. The variables student residual plots are random around the origin. The student residuals vs predicted values plot has significant improvement with the transformation as already seen (Refer to Appendix C, fig. 18-30). The normal probability plot is a straight line. [4]

Using the VIF, r, influence add-ons in the model, the model diagnostics can be assessed. All the variables have variation inflation factor (VIF) lesser than 10. So, there is no issue of collinearity (Refer to Appendix C, table 2). There are outliers and influential points for this model. Observations that are both outliers and influential points are observation numbers 8, and 185 (Refer to Appendix C, fig. 17). The outliers are removed, and the model is run again.

On doing that, we get a model with R^2 value of 0.8938 and adjusted R^2 0.8893. The values have slightly increased from the previous model. The RMSE is 0.09977 which has decreased from before. For the same null and alternate hypothesis, this has a F-value of 195.72 and p-value is less than 0.05. So, the null hypothesis is rejected, and this model passes the goodness of fit test.

The signs of the effect of the variables using parameter estimates remains the same from the previous model. By increasing crime rate by 1%, the medv increases by ($e^{0.0621}$ – 1)*100 = 6.41%. This can be calculated for all variables in the model.

Based on standardized beta estimates, rm is the independent variable with the highest influence on logmedv followed by dis, crime, lstat, nox, ptratio, age, tax, rad, minor, zn and chas. This is estimated using the absolute value of the standardized beta estimates. (Refer to Appendix C, table 3)

The model equation is *new_medv = log(medv) = 2.42099 + 0.0621crime + 0.0008005zn + 0.04938chas – 0.63361nox + 0,25741rm – 0.00143age – 0.03883dis + 0.00988rad – 0.00045523tax _ 0.02333ptratio + 0.0007221minor – 0.01217lstat, where chas = 1 if tract bounds rive, otherwise 0.*

Significance test is also run on each independent variable. From the p-values for these variables, it can be seen that all variables in the model are significant as their p-values are less than 0.05 in this model (Refer to Appendix C, table 3).

All four assumptions – Linearity, Constant Variance, Independency and Normality are satisfied which can be seen from the residual plots (Refer to Appendix C, fig. 32-45)

The model diagnostics have been checked. All VIF values are less than 10 so there is no collinearity issue. There is only one outlier, observation number 148, which can be left in the model.

This model is tested on test set and then the test model parameters are computed. The R^2 value for the test set is $(0.92171)^2 = 0.8495$. The error terms MAE has a value of 22.6199 and RMSE of 9.57828. The cross-validated R^2 is calculated by taking the absolute value of the difference between the squares of training R^2 and testing R^2. The cross-validated R^2 value for this model is 0.0772 which is lesser than 0.3. Hence, it is a good model based on this parameter. (Refer to Appendix C, table 4)

Comparing the training and testing performance, the training set has higher R^2 and adjusted R^2 value but there is not a huge difference. Training does better with respect to RMSE as well. Overall, although training seems to be doing better than test, testing still performed well and hence this is a good model.

VI. QUADRATIC POLYNOMIAL REGRESSION

Based on the student residuals vs predicted plot with the U-shape, quadratic polynomial regression could be an option. This method is being exercised to see if this gets better results than the models with no transformation or with log transformation.

On running the full model using procedure glmselect, the resulting model had 91 independent variables and 1 dependent variable medv.

The R^2 value of the model is 0.9522 and Adj- R^2 is 0.9364 that means that 95.22% of the variance of the medv variable is explained by the 91 independent variables. The error terms

RMSE has value 2.15487, AIC has 1010.60656, AICC has 1074.65051 and SBC has a value of 1000.89985.

The hypothesis to check if the model is good or not is framed and a goodness of fit test is run. The null hypothesis H_0 is that no independent variable has an effect on the dependent variable medv i.e., all $\beta_i = 0$. Alternate hypothesis is that there is at least 1 variable that has an effect on the independent variable. The F-value for this model is 60.20 and the corresponding p-value is less than 0.05. So, the null hypothesis is rejected and hence this model satisfies the goodness of fit test (Refer to Appendix D, table 1).

The data is then split in to training and testing sets through holdout method with 75% training and 25% testing using procedure surveyselect with seed 456729. A new variable is created for training new_medv for just the training set in the new dataset "boston_all_pol".

Following this, two model selection process was run on the training dataset – stepwise and forward model selection. Two methods arrive at two different models.

Looking at the model obtained by stepwise method, model 1, it has the independent variables rm, dis*tax (dis_tax), rm*ptratio (rm_ptratio), lstat, rm*lstat (rm_lstat). Creating a new dataset with these variables along with the existing ones, the model is run again using the procedure regression to predict medv (Refer to Appendix D, table 2).

The model obtained has an R^2 value of 0.8635 and adjusted R^2 of 0.8610 which means that 86.35% of the variance of new_medv variable can be explained by the 5 independent variables and also 86.10% of the variance of new_medv variable is explained by them. The error RMSE value is 3.11001 which is an increase from the full model although not by much.

The hypothesis to check if the model is good or not is framed and a goodness of fit test is run. The null hypothesis H_0 and the alternate hypothesis H_1 is the same. On running the goodness of fit test, we get F-value as 341.57 which is a great increase from the full model and p-value is less 0.05 which means that the null hypothesis is rejected and there is at least one independent variable that has an effect on new_medv.

Using the signs of the parameter estimates or beta coefficients, it can be said that the independent variables rm and lstat have a positive effect on the dependent variable new_medv and the variables dis_tax, rm_ptratio, rm_lstat have a negative effect on new_medv. The effect can be expanded into if the average number of rooms increases by 1, then the median home value will increase by 12566.05$. This can be done for all the variables in the model.

Based on standardized beta estimates, rm_lstat is the independent variable with the highest influence on medv followed by lstat, rm, dis_tax, and rm_ptraio. This is estimated using the absolute value of the standardized beta estimates. (Refer to Appendix D, table 2)

The model equation is *new_medv = medv = -40.38857 + 12.56605rm – 0.00151dis*tax – 0.06595rm*ptratio + 2.33246lstat – 0.46491rm*lstat.*

Significance test is also run on each independent variable. From the p-values for these variables, it can be seen that all variables in the model are significant as their p-values are less than 0.05 in this model (Refer to Appendix D, table 2).

The four assumptions of the regression model are checked using student residual plots and normality plot. The linearity assumption is satisfied as confirmed by the scatterplots. The variables student residual plots are random around the origin. The student residuals vs predicted values plot has significant improvement with the transformation as already seen (Refer to Appendix D, fig. 2-7). The normal probability plot seems to have a very slight S shape but that could be due to outliers.

Using the VIF, r, influence add-ons in the model, the model diagnostics can be assessed. The variables rm, dis_tax, rm_ptratio have VIF less than 10 so they are not collinear but the variables lstat and rm_lstat have VIF of 81.51468 and 69.53019 which are much greater than 10. So, these variables are collinear, but we can safely ignore this as one variable is actually the product of the other variable. So, this is to be expected and hence can be ignored [5]. There are outliers and influential points for this model. Observations that are both outliers and influential points are observation numbers 8,89,182,365 (Refer to Appendix D, fig. 1). The outliers are removed, and the model is run again.

The model obtained has an R^2 value of 0.8965 and adjusted R^2 of 0.8946 which means that 89.65% of the variance of new_medv variable can be explained by the 5 independent variables and also 89.46% of the variance of new_medv variable is explained by them. The error RMSE value has dropped to 2.67161.

The hypothesis to check if the model is good or not is framed and a goodness of fit test is run. The null hypothesis H_0 and the alternate hypothesis H_1 is the same. On running the goodness of fit test, we get F-value as 461.05 which is an increase from the previous model and p-value is less 0.05 which means that the null hypothesis is rejected and there is at least one independent variable that has an effect on new_medv.

Using the signs of the parameter estimates or beta coefficients, it can be said that the independent variables rm and lstat have a positive effect on the dependent variable new_medv and the variables dis_tax, rm_ptratio, rm_lstat have a negative effect on new_medv. The effect can be expanded into if the average number of rooms increases by 1, then the median home value will increase by (12.74890*1000)$ = 12748.90$. This can be done for all the variables in the model.

Based on standardized beta estimates, rm_lstat is the independent variable with the highest influence on medv followed by lstat, rm, rm_ptratio and dis_tax. This is estimated using the absolute value of the standardized beta estimates. (Refer to Appendix D, table 3)

The model equation is *new_medv = medv = -41.58632 + 12.74890rm – 0.00137dis*tax – 0.065953rm*ptratio + 2.37504lstat – 0.46988rm*lstat.*

Significance test is also run on each independent variable. From the p-values for these variables, it can be seen that all variables in the model are significant as their p-values are less than 0.05 in this model (Refer to Appendix D, table 3).

The four assumptions of the regression model are checked using student residual plots and normality plot. The linearity assumption is satisfied as confirmed by the scatterplots. The variables student residual plots are random around the origin. The student residuals vs predicted values plot has significant improvement with the transformation as already seen (Refer to Appendix D, fig. 9-15). The normal probability plot is a straight line now.

Using the VIF, r, influence add-ons in the model, the model diagnostics can be assessed. The variables lstat and rm_lstat are still collinear and it can be ignored here as well [5]. There are outliers and influential points for this model. Observations that are both outliers and influential points are observation numbers 183, 184 and 339(Refer to Appendix D, fig. 8). The outliers are not removed.

This model is tested on test set and then the test model parameters are computed. The R^2 value for the test set is $(0.92611)^2 = 0.8577$. The error terms MAE has a value of 2.36930 and RMSE of 3.46342. The cross-validated R^2 value for this model is 0.068 which is lesser than 0.3. Hence, it is a good model based on this parameter. (Refer to Appendix D, table 4)

Comparing the training and testing performance, the training set has higher R^2 and adjusted R^2 value but there is not a huge difference. Testing has lower RMSE than training. Overall, looking at all parameters, it is good model.

Moving on to the second model that was obtained by forward selection method, it has 7 independent variables – rm, age*dis (age_dis), dis*tax (dis_tax), rm*ptratio (rm_ptratio), lstat, nox*lstat (nox_lstat) and rm*lstat (rm_lstat). Creating a new dataset like the previous model, regression is carried out again through procedure regression.

The model obtained has an R^2 value of 0.8723 and adjusted R^2 of 0.8690 which means that 87.23% of the variance of new_medv variable can be explained by the 7 independent variables and also 86.90% of the variance of new_medv variable is explained by them. The error RMSE value is 3.01872 which is an increase from the full model although not by much.

The hypothesis to check if the model is good or not is framed and a goodness of fit test is run. The null hypothesis H_0 and the alternate hypothesis H_1 is the same. On running the goodness of fit test, we get F-value as 261.62 and p-value is less 0.05 which means that the null hypothesis is rejected and there is at least one independent variable that has an effect on new_medv.

Using the signs of the parameter estimates or beta coefficients, it can be said that the independent variables rm and lstat have a positive effect on the dependent variable new_medv and the variables age_dis, dis_tax, rm_ptratio, nox_lstat, rm_lstat have a negative effect on new_medv. The effect can be expanded into if the average number of rooms

increases by 1, then the median home value will increase by (13.17032 *1000) = 13170.32$. This can be done for all the variables in the model.

Based on standardized beta estimates, lstat is the independent variable with the highest influence on medv followed by rm_lstat, rm, rm_ptratio, dis_tax, and age_dis. This is estimated using the absolute value of the standardized beta estimates. (Refer to Appendix D, table 4)

The model equation is *new_medv = medv = - 42.33980 + 13.17032rm – 0.00482age*dis – 0.00165dis*tax – 0.8043rm*ptratio + 2.99927lstat – 0.59786nox*lstat – 0.50394rm*lstat*

Significance test is also run on each independent variable. From the p-values for these variables, it can be seen that all variables in the model are significant as their p-values are less than 0.05 in this model (Refer to Appendix D, table 4).

The four assumptions of the regression model are checked using student residual plots and normality plot. The linearity assumption is satisfied as confirmed by the scatterplots. The variables student residual plots are random around the origin. The student residuals vs predicted values plot has improvement as already seen (Refer to Appendix D, fig. 17-25). The normal probability plot seems to have a very slight S shape but that could be due to outliers.

Using the VIF, r, influence add-ons in the model, the model diagnostics can be assessed. The variables rm, age_dis, dis_tax, rm_ptratio have VIF less than 10 so they are not collinear but the variables lstat, nox_lstat and rm_lstat have VIF greater than 10. So, these variables are collinear, but we can safely ignore it for the same reason as before. [5] There are outliers and influential points for this model. Observations that are both outliers and influential points are observation numbers 8, 182 and 365 (Refer to Appendix D, fig. 16). The outliers are removed, and the model is run again.

The model obtained has an R^2 value of 0.9021 and adjusted R^2 of 0.8995 which means that 90.21% of the variance of new_medv variable can be explained by the 7 independent variables and also 89.95% of the variance of new_medv variable is explained by them. The error RMSE value has dropped to 2.60421.

The hypothesis to check if the model is good or not is framed and a goodness of fit test is run. The null hypothesis H_0 and the alternate hypothesis H_1 is the same. On running the goodness of fit test, we get F-value as 348.75 which is an increase from the previous model and p-value is less 0.05 which means that the null hypothesis is rejected and there is at least one independent variable that has an effect on new_medv.

Using the signs of the parameter estimates or beta coefficients, it can be said that the independent variables rm and lstat have a positive effect on the dependent variable new_medv and the variables dis_tax, rm_ptratio, rm_lstat have a negative effect on new_medv. The effect can be expanded into if the average number of rooms increases by 1,

then the median home value will increase by (13.17032*1000)$ = 13170.32$. This can be done for all the variables in the model.

Based on standardized beta estimates, the order of influence of the independent variables on the dependent variable is same as the previous model. (Refer to Appendix D, table 5)

The model equation is *new_medv = medv = - 43.37581 + 13.33592rm – 0.00549age*dis – 0.00144dis*tax – 0.08550rm*ptratio + 3.03364lstat – 0.61369nox*lstat – 0.50391rm*lstat.*

Significance test is also run on each independent variable. From the p-values for these variables, it can be seen that all variables in the model are significant as their p-values are less than 0.05 in this model (Refer to Appendix D, table 5).

The four assumptions of the regression model are checked using student residual plots and normality plot. The linearity assumption is satisfied as confirmed by the scatterplots. The variables student residual plots are random around the origin. The student residuals vs predicted values plot has significant improvement with the transformation as already seen (Refer to Appendix D, fig.27-35). The normal probability plot is a straight line now.

Using the VIF, r, influence add-ons in the model, the model diagnostics can be assessed. The variables lstat, nox_lstat and rm_lstat are still collinear and it can be ignored here as well. [5]There are outliers and influential points for this model. Observations that are both outliers and influential points are observation numbers 88 and 185(Refer to Appendix D, fig. 26). The outliers are not removed.

This model is tested on test set and then the test model parameters are computed. The R^2 value for the test set is $(0.93033)^2 = 0.8655$. The error terms MAE has a value of 2.22168 and RMSE of 3.36385. The cross-validated R^2 value for this model is 0.064 which is lesser than 0.3. Hence, it is a good model based on this parameter. (Refer to Appendix D, table 6)

Comparing the training and testing performance, the training set has higher R^2 and adjusted R^2 value but there is not a huge difference. Training has lower RMSE than training. Overall, looking at all parameters, training seems to be doing better than testing but it still is a good model as there isn't any huge disparities.

Comparing the stepwise model 1 and forward selection model 2, looking at the training test parameters for both models, model 2 has higher R^2 and adj R^2 than 1, RMSE is also lower for model 2. F-value is higher for model 1. Multicollinearity is ignored in both. Considering all this, training for model 2 is better than 1. Comparing the test model parameters of both model, R^2 is higher for model 2. MAE, RMSE are lower for model 2. Both satisfy the cross-validated R^2 test. Looking at all this, model 2 test is better than 1. So, overall, in quadratic polynomial regression, model 2 is better.

In the next section, the log model and the polynomial model is compared and a conclusion is decided.

VII. COMPARING THE TWO FITTED MODELS

Comparing the training models, the polynomial model has the higher R^2 and Adj R^2 than the log transformation model although just 0.01 difference. The log transformation model has lower RMSE than the polynomial one. Polynomial model has the higher F-value and both of them satisfy the goodness of fit test. The polynomial model has multicollinearity issues but the other one doesn't.

Comparing the testing model parameters, polynomial model has R^2 0.02 higher than the log transformation model. MAE and RMSE models have lower values for polynomial model. Both the models satisfy the cross-validated R^2 condition.

In spite of polynomial model working better than log model, log model also has good parameters and is a good model. Log transformation model is the simpler one out of the two. If both models work well, then choosing the simpler one is the good option. Hence, the final fitted model is the log transformed model(Refer to Appendix C, table 1).

VIII. THE FINAL MODEL

The final model is *log(medv) = 2.50502 + 0.04650crime + 0.00070974zn + 0.03864chas – 0.52470nox + 0.24922rm – 0.00094060age – 0.03668dis + 0.01153rad – 0.00052892tax - 0.02459ptratio + 0.00060319minor – 0.01494lstat, where chas = 1 if tract bounds rive, otherwise 0.* (Run on the full dataset boston_log) (Refer to Appendix E, table 1)

All assumptions are satisfied. The model diagnostics was checked and there are some outliers which can be ignored. The effect of the variables remains the same. But the variable chas is insignificant as the p-value is higher than 0.05, it is 0.0723. Removing that variable and running the model again, we get the model equation
Log(medv) = 2.50664 + 0.04657crime + 0.00071733zn – 0.52204nox + 0.25034rm – 0.00092748age – 0.03722dis + 0.01231rad – 0.00054065tax - 0.02506ptratio + 0.00060768minor – 0.01482lstat (Refer to Appendix E, table 2)

All assumptions – linearity, constant variance, independency and normality are satisfied. The model diagnostics also look good. There are some outliers and influential points which we can ignore. There is no issue of multicollinearity. The other parameters such as R^2 etc. also look good.

Interpreting the variables, increase in crime rate by 1% will lead to $(e^{0.04657} – 1) *100$ = 4.77% increase in median home value. Increase in zn variable by 1 will lead to $(e^{0.00070974} – 1)*100$ = 0.07% increase in medv. When nitric oxide concentration increases by 1, then the medv will decrease by $(e^{0.52204} -1)*100$ = 68.55%. When the average number of rooms increases by 1, the median home value will increase by $(e^{0.25034} -1)*100$ = 28.45%. When the age proportion increases by 1, the medv variable decreases by $(e^{0.00092748} -1)*100$ = 0.093%. When the weighted distance to the employment center increases by 1, then the medv will decrease by $(e^{0.03722} – 1)*100$ = 3.79%. When the index of accessibility to highways increases by 1, then the medv will increase by $(e^{0.01231}-1) *100$ = 1.24%. When the tax

variable increases by 1 i.e., 10000$, the medv variable decreases by ($e^{0.00054065} - 1$) *100 = 0.0054%. When the pupil-teacher ratio increases by 1, then the medv variable will decrease by ($e^{0.02506} - 1$) *100 = 2.54%. When the minor variable increases by 1, the medv variable increases by ($e^{0.00060768}$ -1)*100 = 0.06%. When the lower status population increases by 1%, the medv variable decreases by ($e^{0.014182} - 1$)*100 = 1.43%.

Two predictions were made using this model. When the crime rate is 0.5%, proportion for zoned residential land is 15, nitric oxides concentration(parts per 10 million) is 0.5, average number of rooms per dwelling is 3, proportion of homes built prior to 1940 is 50, weighted distance to closest employment centers is 4, with index for accessibility to highways as 4, paying 400(per 10000$) tax, with pupil-teacher ratio as 15, with minor as 300 and percentage of lower status homes as 25% , the median value for home is ($e^{2.1043}$) = 8.20136 i.e., 8201.36$ is the median value of the home with 95% confidence interval equal to ($e^{2.0220}$, $e^{2.1865}$) i.e., (7.5534, 8.9039) and prediction interval equal to to ($e^{1.8716}$, $e^{2.3370}$) i.e., (6.4987,10.3501) which means that the predicted value for medv can increase from 6498.7$ to 10350.1$. (Refer to Appendix E, fig. 1)

Another prediction is when the crime rate is 18.5%, proportion for zoned residential land is 25, nitric oxides concentration(parts per 10 million) is 0.4, average number of rooms per dwelling is 4, proportion of homes built prior to 1940 is 70, weighted distance to closest employment centers is 5, with index for accessibility to highways as 10, paying 350(per 10000$) tax, with pupil-teacher ratio as 20, with minor as 250 and percentage of lower status homes as 25% , the median value for home is ($e^{3.1416}$) = 23.14086 i.e., 23140.86$ is the median value of the home95% confidence interval equal to ($e^{2.8739}$, $e^{3.4094}$) i.e., (17.7059, 30.24709) and prediction interval equal to to ($e^{2.7966}$, $e^{3.4867}$) i.e., (16.38883, 32.67793) which means that the predicted value for medv can increase from 16388.83$ to 32677.93$. (Refer to Appendix E, fig. 1)

IX. CONCLUSION

The initial model did not satisfy certain assumptions. So, log transformation and polynomial regression was adopted for this dataset. After various steps, the log transformed model was chosen as the final model.

Based on the model, there seems to be some inconsistency as increase in crime would not lead to increase in the value of a home. Collecting more data could solve this issue. Also, the data is outdated. There are variables like nox, crime, dis, tax, rad etc. that changes over time and as cities and states develop, a lot of these parameters change as well. For example, there are probably more employment centers in Boston now than in 1978. Tax policies could have changed, more highways would've been constructed. Crime rates have also changed since 1978. All these factors make the model not useful for the present.

The important avenue with respect to this dataset and model for the future is to keep collecting and updating data as this is more subjective to time and the model would have to

be adopted to the changes, if not, it becomes useless. This shows that the current result isn't enough to make decisions or give recommendations for the present.

All these topics and suggestion were easier to identify and make because there was lots of background research involved. As already mentioned, the articles helped put things into perspective on what's important and what's not. The scientific articles gave some detailed insights on the statistical side of this process.

REFERENCES

[1] D. L. R. David Harrison Jr., "Hedonic housing prices and the demand for clean air," *Journal of Environmental Economics and Management ,* vol. 5, no. 1, 1978.

[2] J. Gomaz, "8 critical factors that influence a home's value," 2019.

[3] M. M. J. E. M. a. P. M. Tranmer, "Multiple Linear Regression (2nd Edition)," *Cathie Marsh Institute Working Paper 2020-01,* 2020.

[4] H. W. N. L. T. C. H. H. Y. L. X. M. T. Changyong Feng, "Log-transformation and its implications for data analysis," *Shanghai Arch Psychiatry,* vol. 26, no. 2, pp. 105 - 109, 2014.

[5] P. Allison, "When Can You Safely Ignore Multicollinearity?," 2012.

APPENDIX

APPENDIX A – Data Exploratory Stage

Fig 1.

Fitted model assumptions and Diagnostics check

Obs	crime	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	minor	lstat	medv
1	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.98	24.0
2	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	21.6
3	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
4	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
5	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.33	36.2
6	0.02985	0.0	2.18	0	0.458	6.430	58.7	6.0622	3	222	18.7	394.12	5.21	28.7
7	0.08829	12.5	7.87	0	0.524	6.012	66.6	5.5605	5	311	15.2	395.60	12.43	22.9
8	0.14455	12.5	7.87	0	0.524	6.172	96.1	5.9505	5	311	15.2	396.90	19.15	27.1
9	0.21124	12.5	7.87	0	0.524	5.631	100.0	6.0821	5	311	15.2	386.63	29.93	16.5
10	0.17004	12.5	7.87	0	0.524	6.004	85.9	6.5921	5	311	15.2	386.71	17.10	18.9

Fig. 2

Descriptives

The MEANS Procedure

Variable	Mean	Minimum	Maximum	Std Dev	Std Error	25th Pctl	50th Pctl	75th Pctl	Lower 95% CL for Mean	Upper 95% CL for Mean	Mode
crime	0.7182351	0.0063200	18.4982000	1.7627831	0.0910297	0.0612900	0.1405200	0.5370000	0.5392408	0.8972293	0.0150100
zn	15.3333333	0	100.0000000	25.9505048	1.3400783	0	0	22.0000000	12.6983009	17.9683657	0
indus	8.8374400	0.4600000	25.6500000	6.3134986	0.3260277	4.0500000	6.9100000	10.5900000	8.1963629	9.4785171	19.5800000
chas	0.0933333	0	1.0000000	0.2912876	0.0150420	0	0	0	0.0637558	0.1229109	0
nox	0.5194659	0.3850000	0.8710000	0.1096108	0.0056603	0.4379000	0.4930000	0.5470000	0.5083359	0.5305958	0.5380000
rm	6.3685067	3.5610000	8.7800000	0.7220313	0.0372855	5.9270000	6.2450000	6.7260000	6.2951911	6.4418222	6.1270000
age	61.9800000	2.9000000	100.0000000	28.7008645	1.4821063	36.6000000	65.1000000	89.8000000	59.0656941	64.8943059	100.0000000
dis	4.3744125	1.1296000	12.1265000	2.1365919	0.1103331	2.5961000	4.0123000	5.8700000	4.1574615	4.5913636	3.4952000
rad	5.4586667	1.0000000	24.0000000	4.5653629	0.2357543	4.0000000	5.0000000	5.0000000	4.9950965	5.9222368	5.0000000
tax	328.6986667	187.0000000	666.0000000	102.2730296	5.2813565	270.0000000	307.0000000	398.0000000	318.3137917	339.0835416	307.0000000
ptratio	17.8581333	12.6000000	22.0000000	2.2173916	0.1145056	16.4000000	18.0000000	19.6000000	17.6329778	18.0832888	14.7000000
minor	379.4128533	70.8000000	396.9000000	41.5068023	2.1434021	380.3400000	392.2000000	396.0600000	375.1982236	383.6274831	396.9000000
lstat	10.4724533	1.7300000	37.9700000	6.0966282	0.3148285	6.0500000	9.3800000	13.2800000	9.8533974	11.0915092	6.3600000
medv	25.1944000	11.8000000	50.0000000	8.7709472	0.4529298	19.5000000	22.8000000	28.7000000	24.3037919	26.0850081	50.0000000

Fig. 3

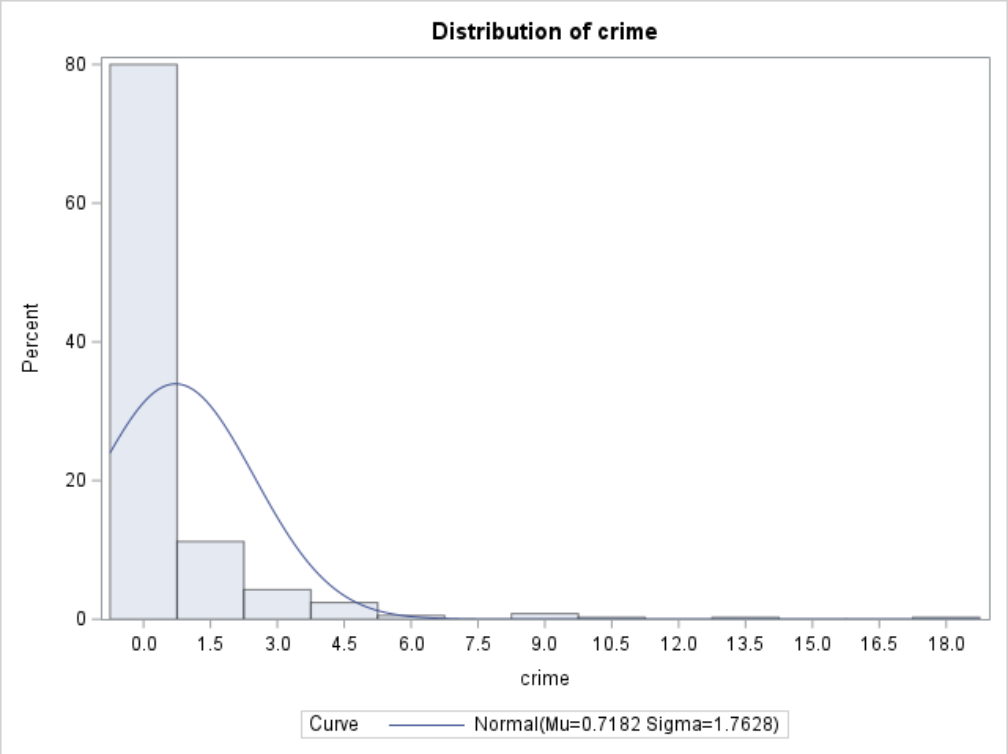

Distribution of crime
Percent
80
60
40
20
0
0.0
1.5
3.0
4.5
6.0
7.5
9.0
10.5
12.0
13.5
15.0
16.5
18.0
crime
Curve
Normal(Mu=0.7182 Sigma=1.7628)

Fig. 4

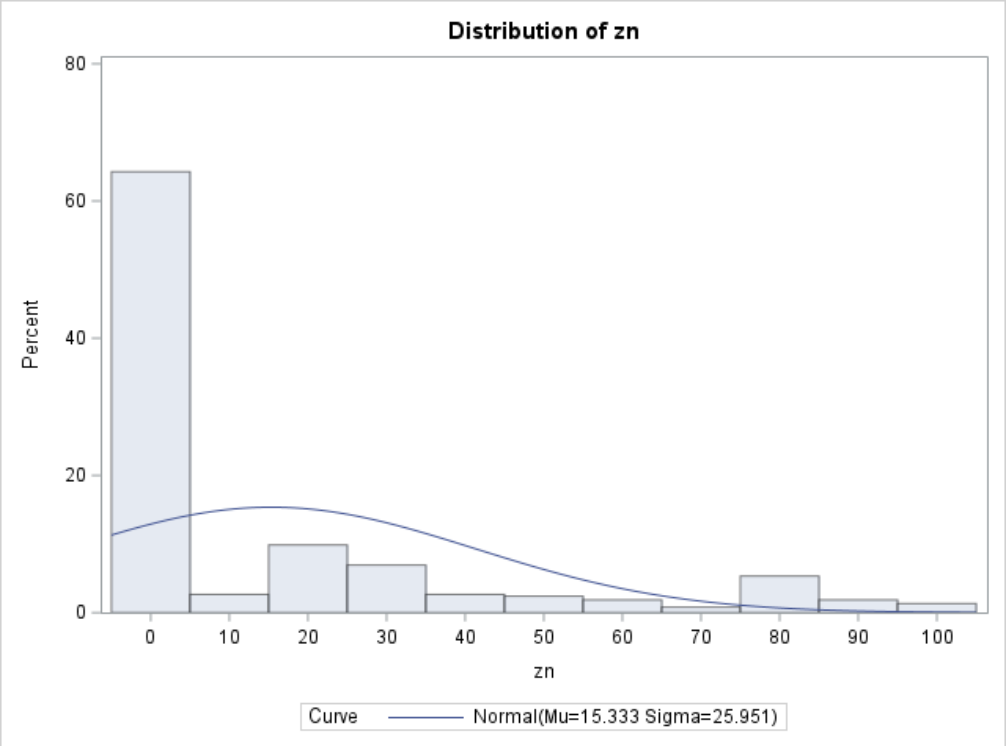

Distribution of zn
Percent
80
60
40
20
0
0
10
20
30
40
50
60
70
80
90
100
zn
Curve
Normal(Mu=15.333 Sigma=25.951)

Fig. 5

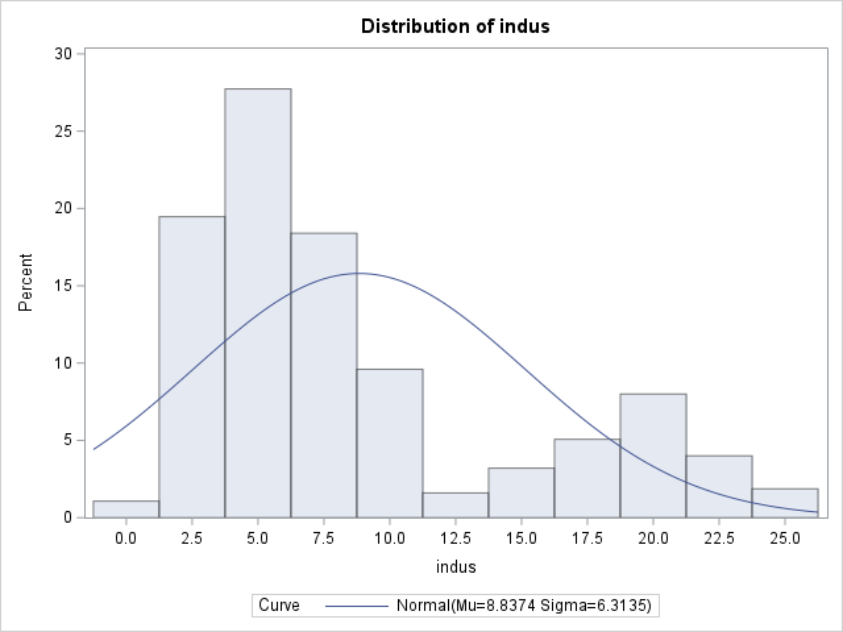

Distribution of indus
Percent
30
25
20
15
10
5
0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
indus
Curve
Normal(Mu=8.8374 Sigma=6.3135)

Fig. 6

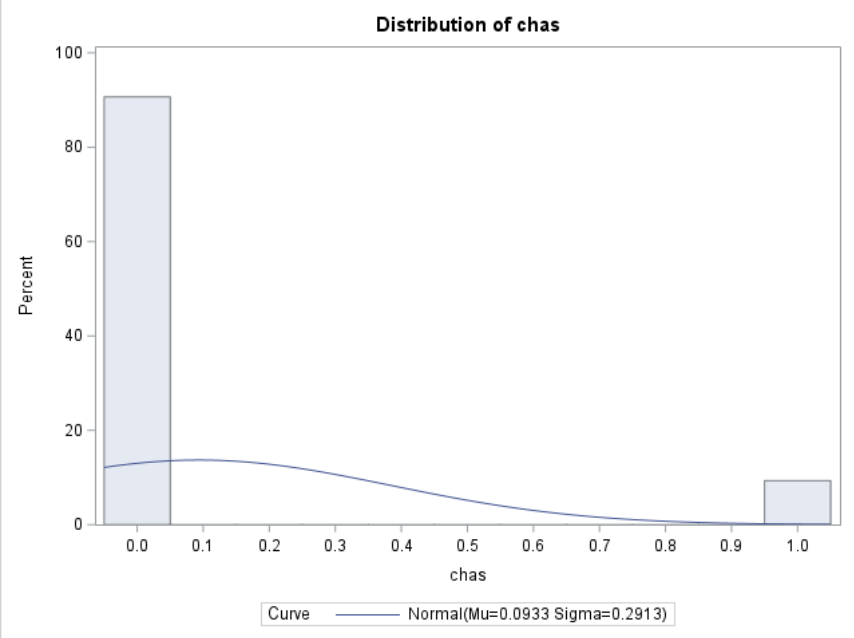

Distribution of chas
Percent
100
80
60
40
20
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
chas
Curve
Normal(Mu=0.0933 Sigma=0.2913)

Fig. 7

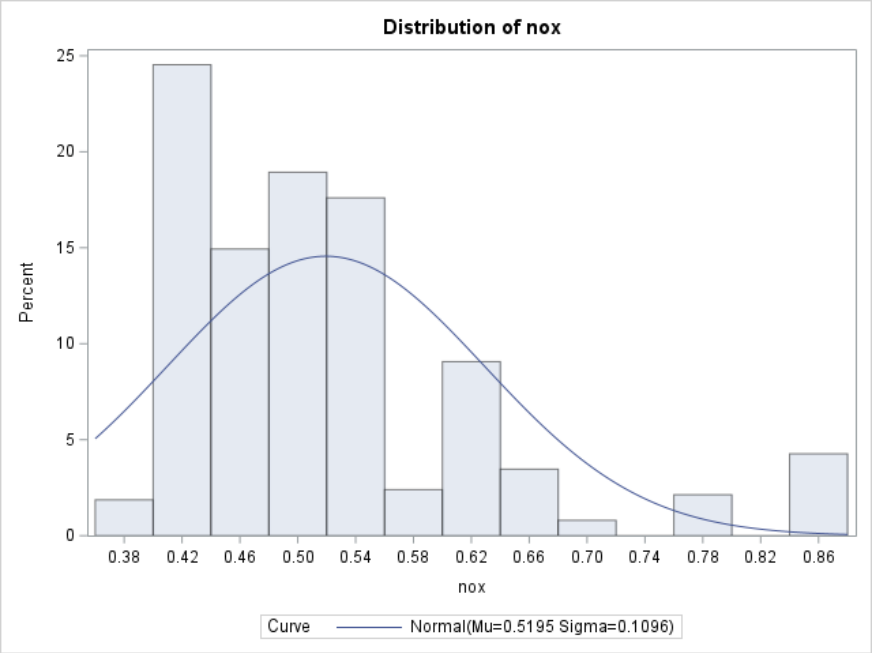

Distribution of nox
Percent
25
20
15
10
5
0
0.38
0.42
0.46
0.50
0.54
0.58
0.62
0.66
0.70
0.74
0.78
0.82
0.86
nox
Curve
Normal(Mu=0.5195 Sigma=0.1096)

Fig. 8

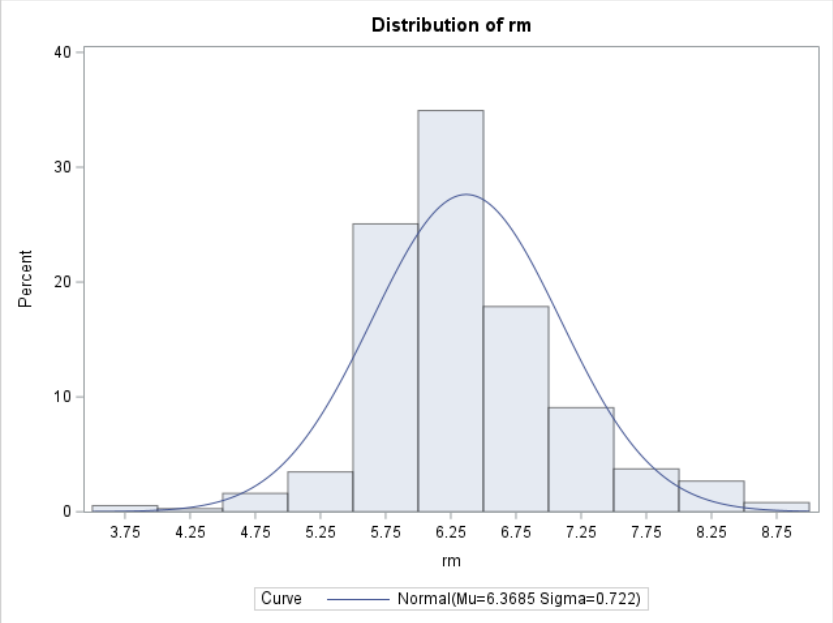

Distribution of rm
Percent
40
30
20
10
0
3.75
4.25
4.75
5.25
5.75
6.25
6.75
7.25
7.75
8.25
8.75
rm
Curve
Normal(Mu=6.3685 Sigma=0.722)

Fig. 9

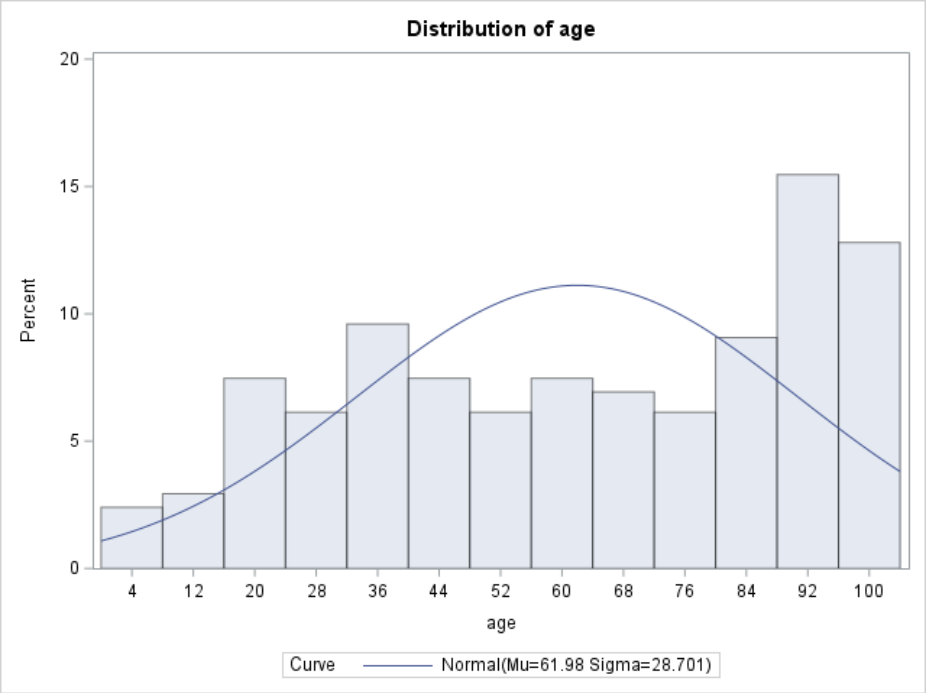

Distribution of age
Percent
20
15
10
5
0
4
12
20
28
36
44
52
60
68
76
84
92
100
age
Curve
Normal(Mu=61.98 Sigma=28.701)

Fig. 10

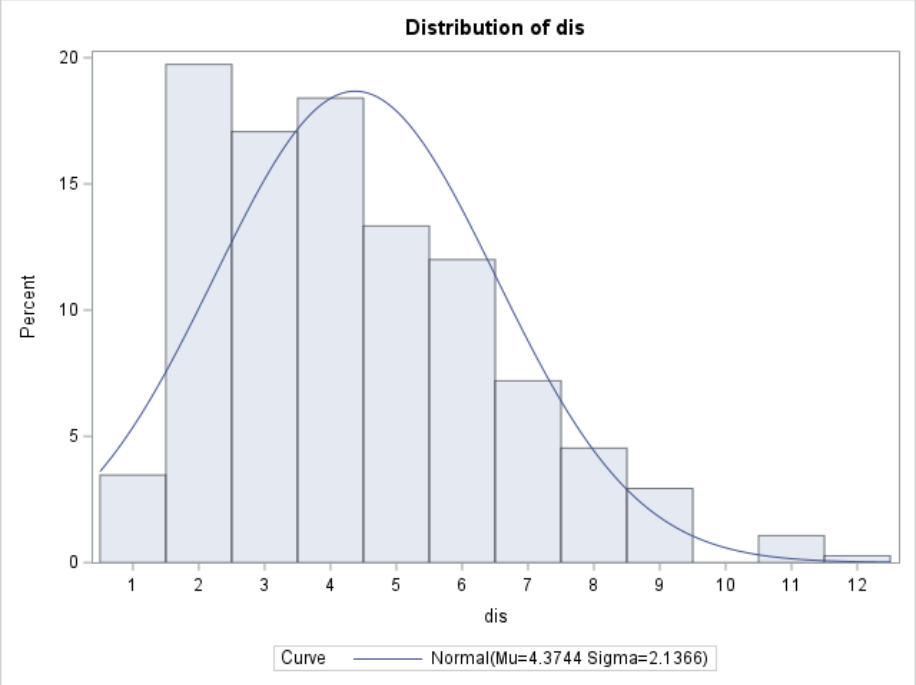

Distribution of dis
Percent
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
11
12
dis
Curve
Normal(Mu=4.3744 Sigma=2.1366)

Fig. 11

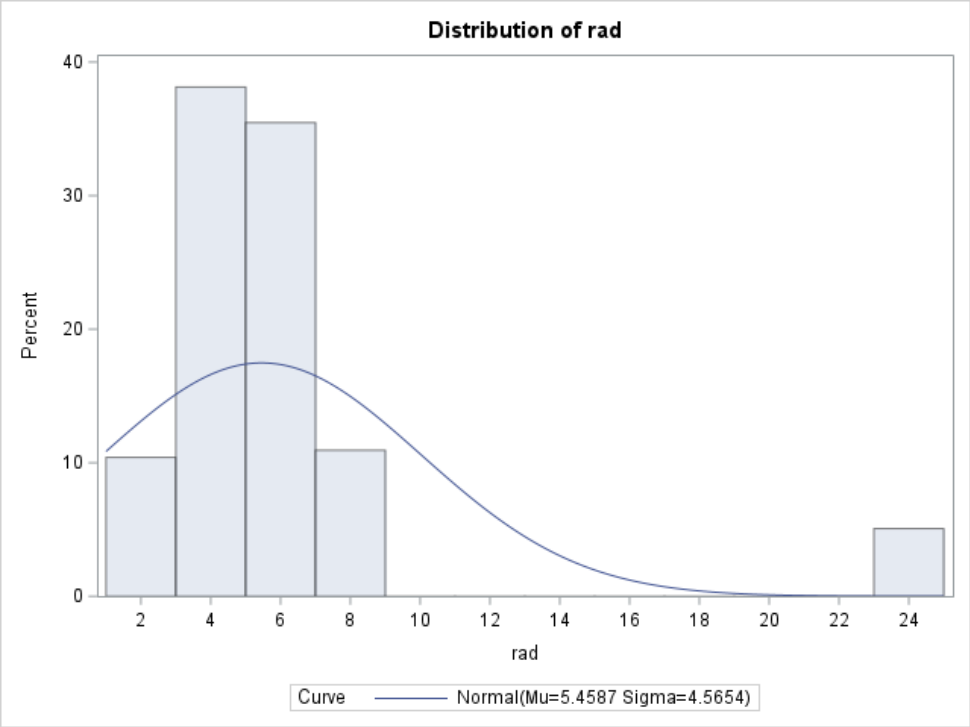

Distribution of rad
Percent
40
30
20
10
0
2
4
6
8
10
12
14
16
18
20
22
24
rad
Curve
Normal(Mu=5.4587 Sigma=4.5654)

Fig. 12

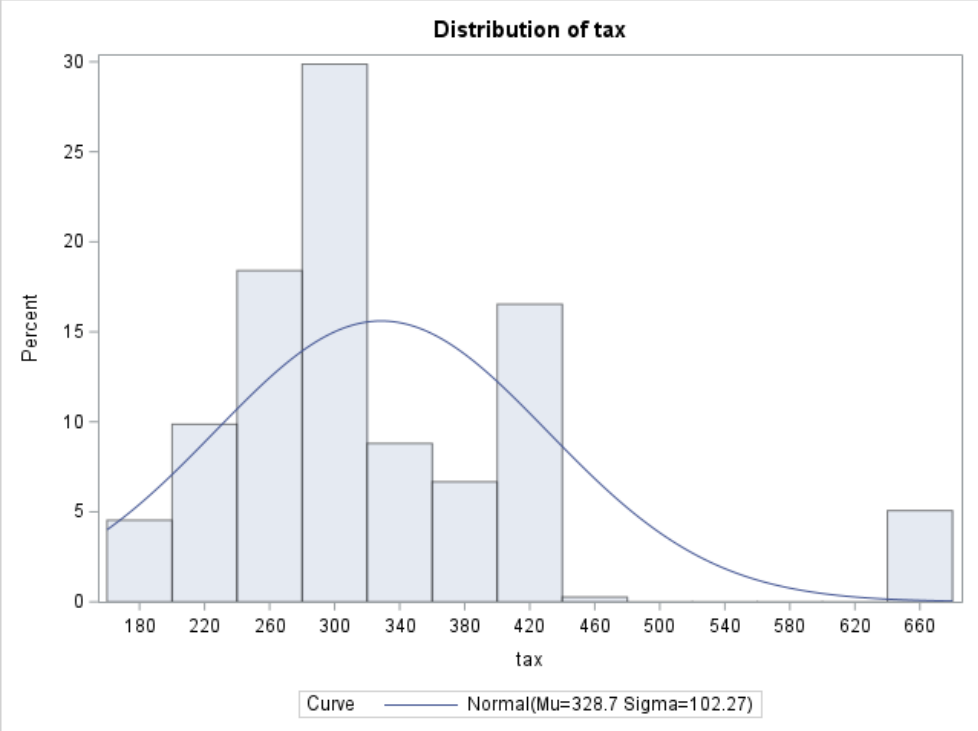

Distribution of tax
Percent
30
25
20
15
10
5
0
180
220
260
300
340
380
420
460
500
540
580
620
660
tax
Curve
Normal(Mu=328.7 Sigma=102.27)

Fig. 13

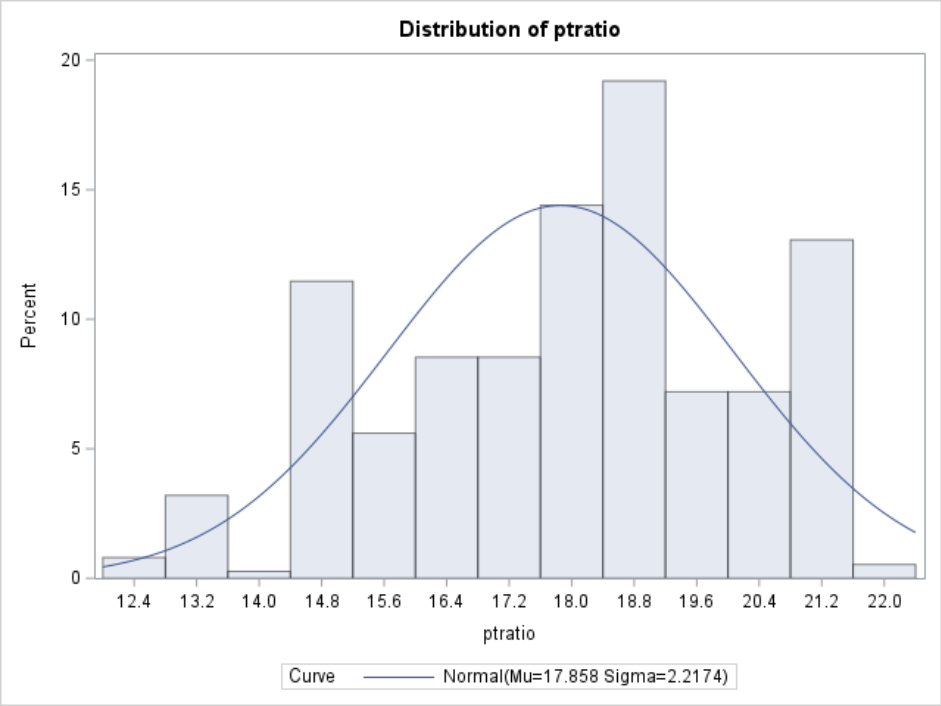

Distribution of ptratio
Percent
20
15
10
5
0
12.4
13.2
14.0
14.8
15.6
16.4
17.2
18.0
18.8
19.6
20.4
21.2
22.0
ptratio
Curve
Normal(Mu=17.858 Sigma=2.2174)

Fig. 14

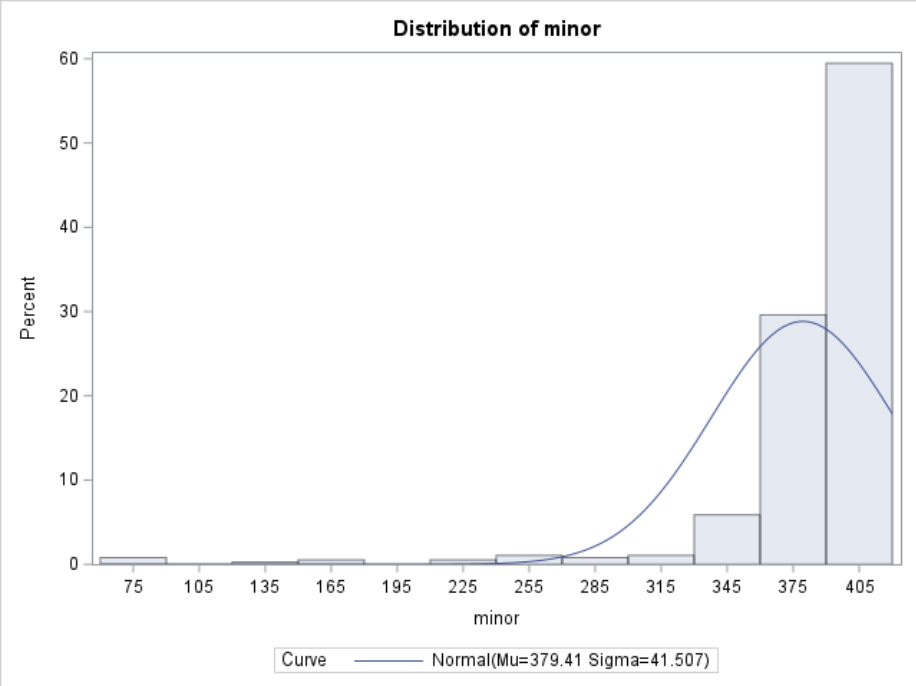

Distribution of minor
Percent
60
50
40
30
20
10
0
75
105
135
165
195
225
255
285
315
345
375
405
minor
Curve
Normal(Mu=379.41 Sigma=41.507)

Fig. 15

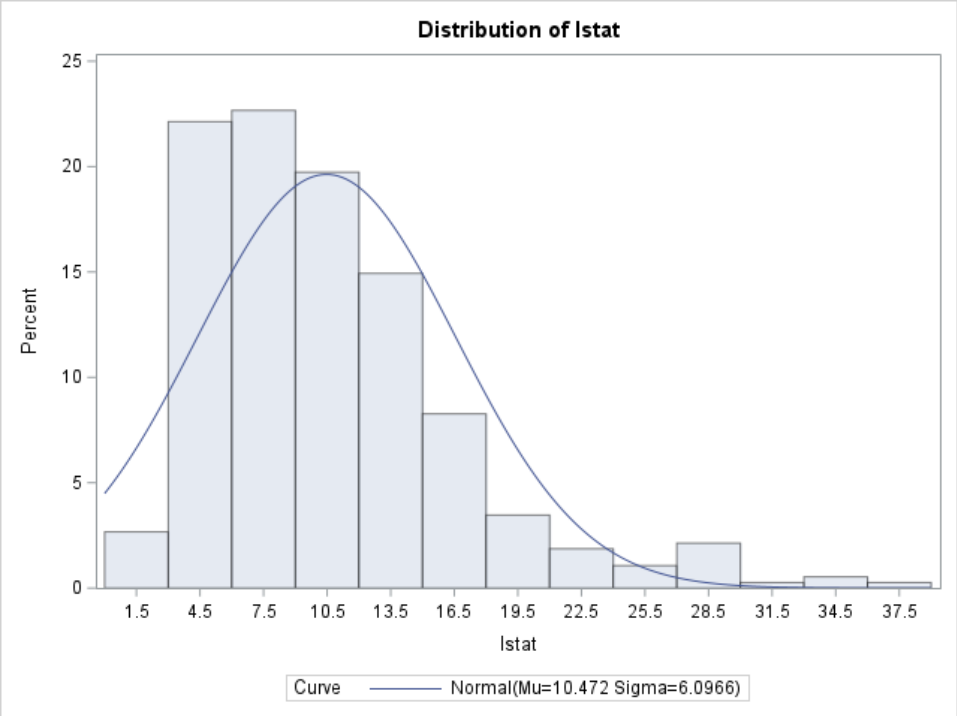

Distribution of lstat
Percent
25
20
15
10
5
0
1.5
4.5
7.5
10.5
13.5
16.5
19.5
22.5
25.5
28.5
31.5
34.5
37.5
lstat
Curve
Normal(Mu=10.472 Sigma=6.0966)

Fig. 16

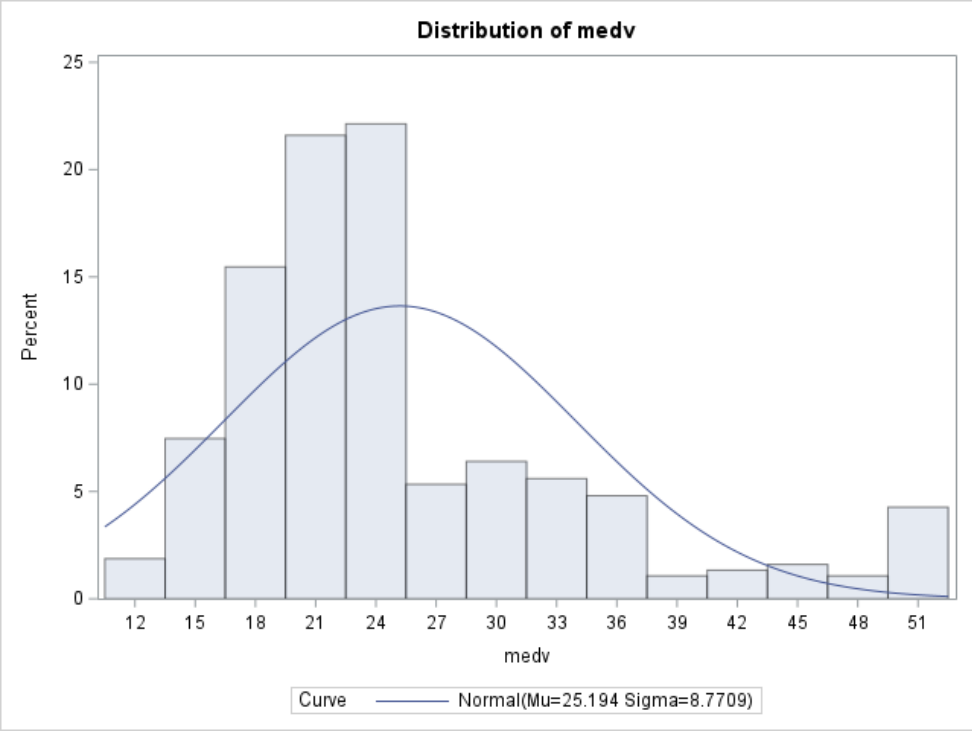

Distribution of medv
Percent
25
20
15
10
5
0
12
15
18
21
24
27
30
33
36
39
42
45
48
51
medv
Curve
Normal(Mu=25.194 Sigma=8.7709)

Fig. 17

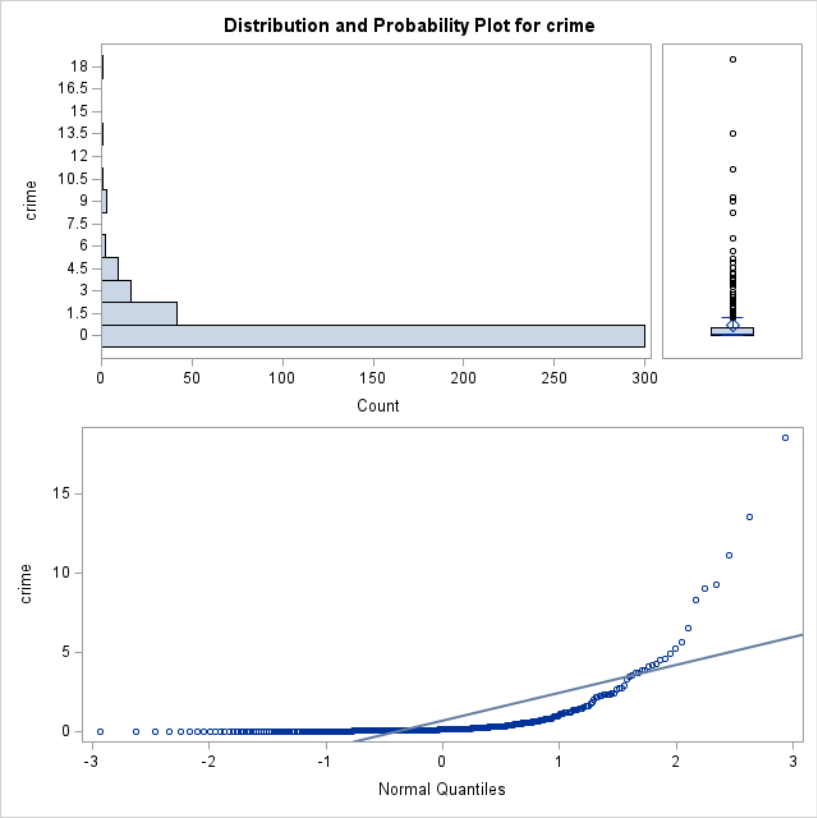

Distribution and Probability Plot for crime
crime
Count
Normal Quantiles

Fig. 18

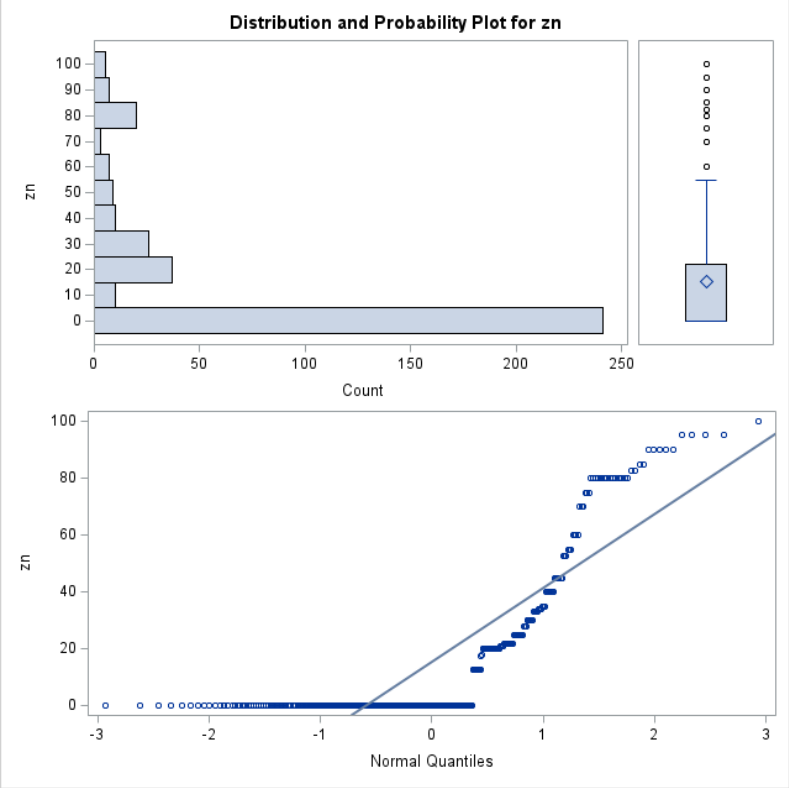

Distribution and Probability Plot for zn
zn
Count
Normal Quantiles

Fig. 19

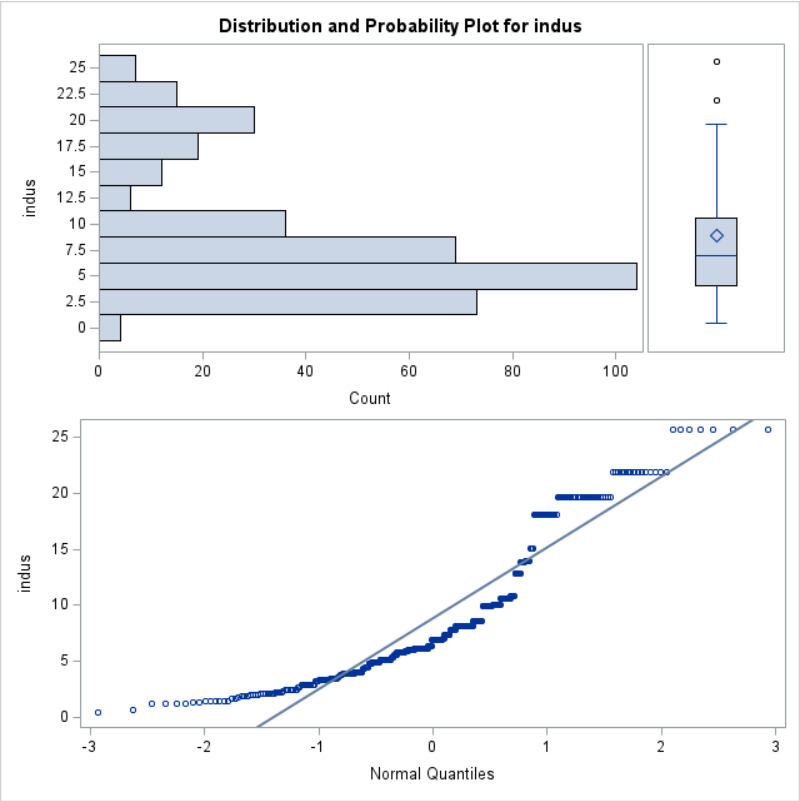

Distribution and Probability Plot for indus
indus
25
22.5
20
17.5
15
12.5
10
7.5
5
2.5
0
0
20
40
60
80
100
Count
indus
25
20
15
10
5
0
-3
-2
-1
0
1
2
3
Normal Quantiles

Fig. 20

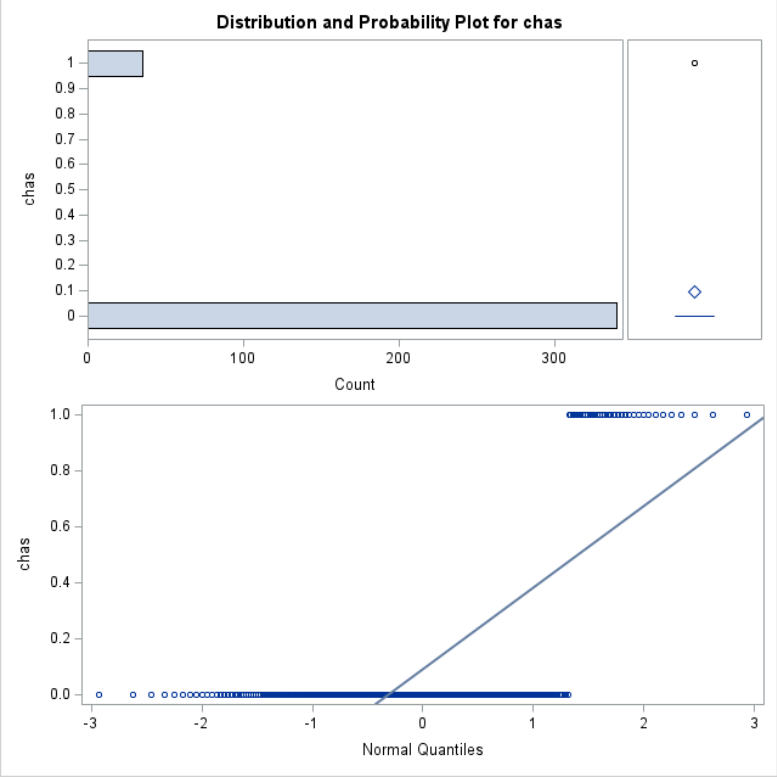

Distribution and Probability Plot for chas
chas
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
100
200
300
Count
chas
1.0
0.8
0.6
0.4
0.2
0.0
-3
-2
-1
0
1
2
3
Normal Quantiles

Fig. 21

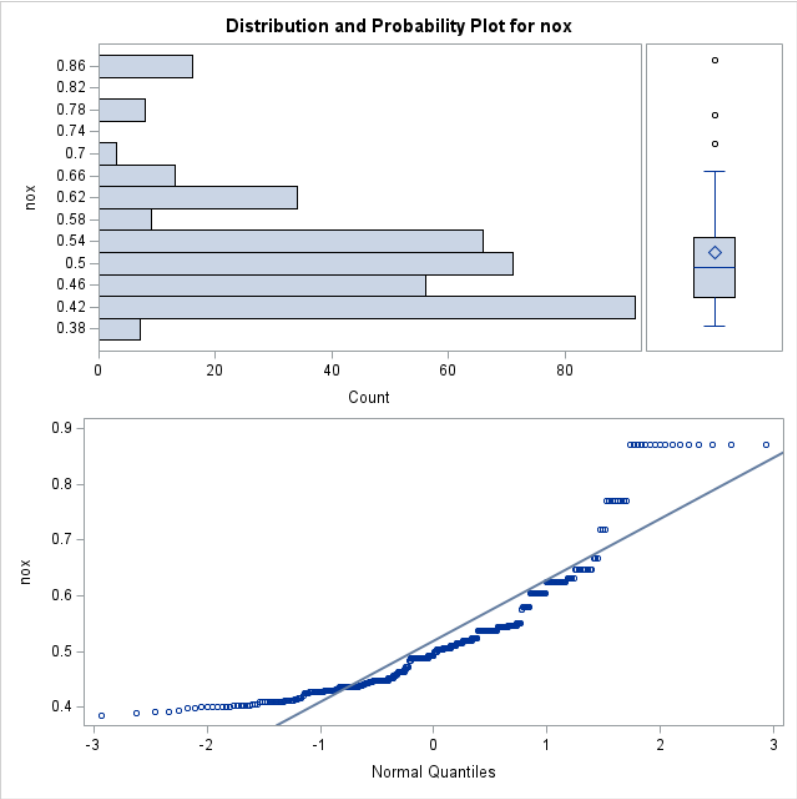

Distribution and Probability Plot for nox
nox
0.86
0.82
0.78
0.74
0.7
0.66
0.62
0.58
0.54
0.5
0.46
0.42
0.38
0
20
40
60
80
Count
nox
0.9
0.8
0.7
0.6
0.5
0.4
-3
-2
-1
0
1
2
3
Normal Quantiles

Fig. 22

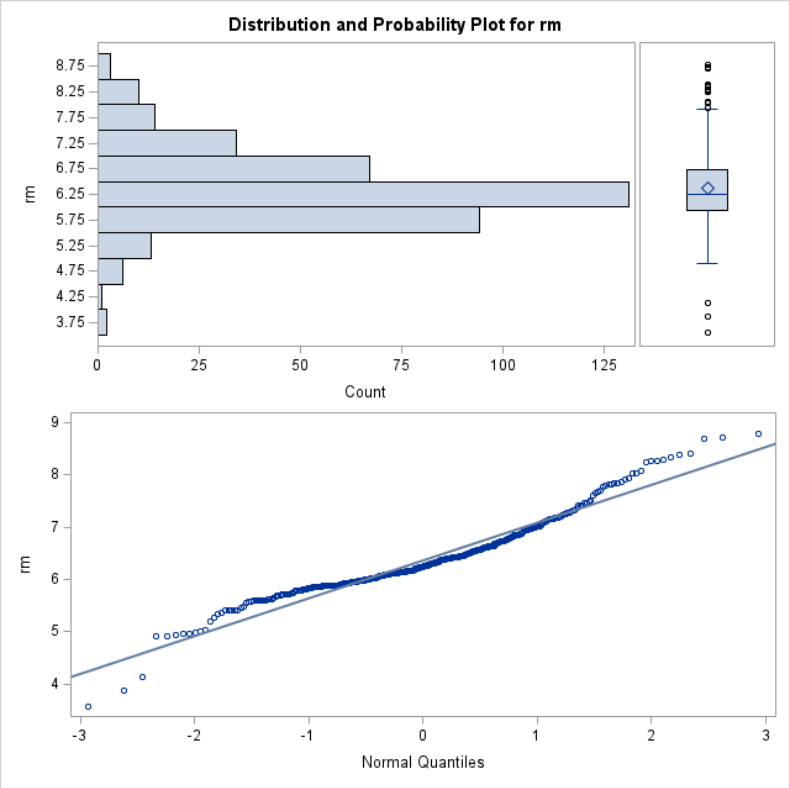

Distribution and Probability Plot for rm
rm
8.75
8.25
7.75
7.25
6.75
6.25
5.75
5.25
4.75
4.25
3.75
0
25
50
75
100
125
Count
rm
9
8
7
6
5
4
-3
-2
-1
0
1
2
3
Normal Quantiles

Fig. 23

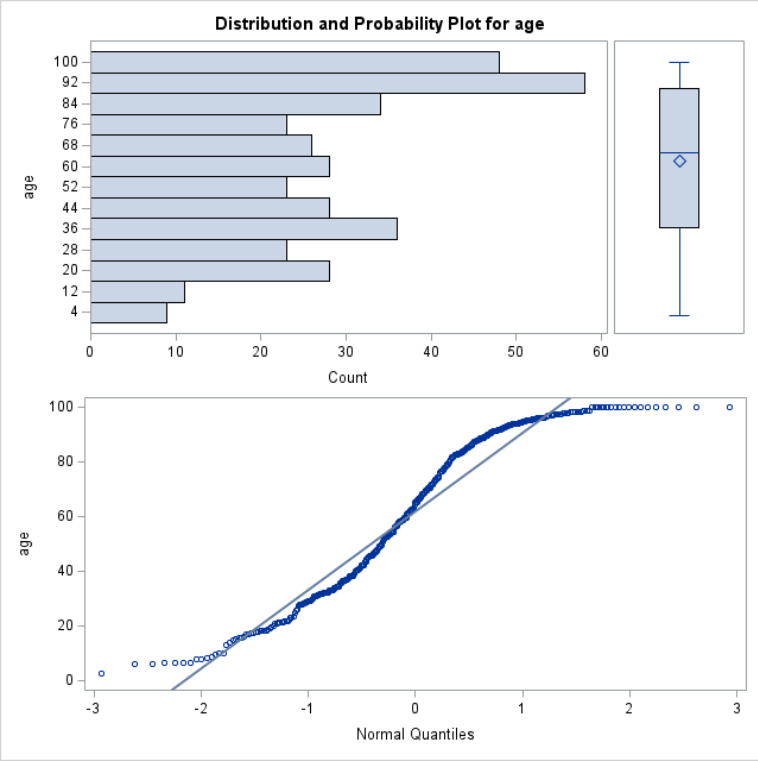

Distribution and Probability Plot for age
age
100
92
84
76
68
60
52
44
36
28
20
12
4
0
10
20
30
40
50
60
Count
age
100
80
60
40
20
0
-3
-2
-1
0
1
2
3
Normal Quantiles

Fig. 24

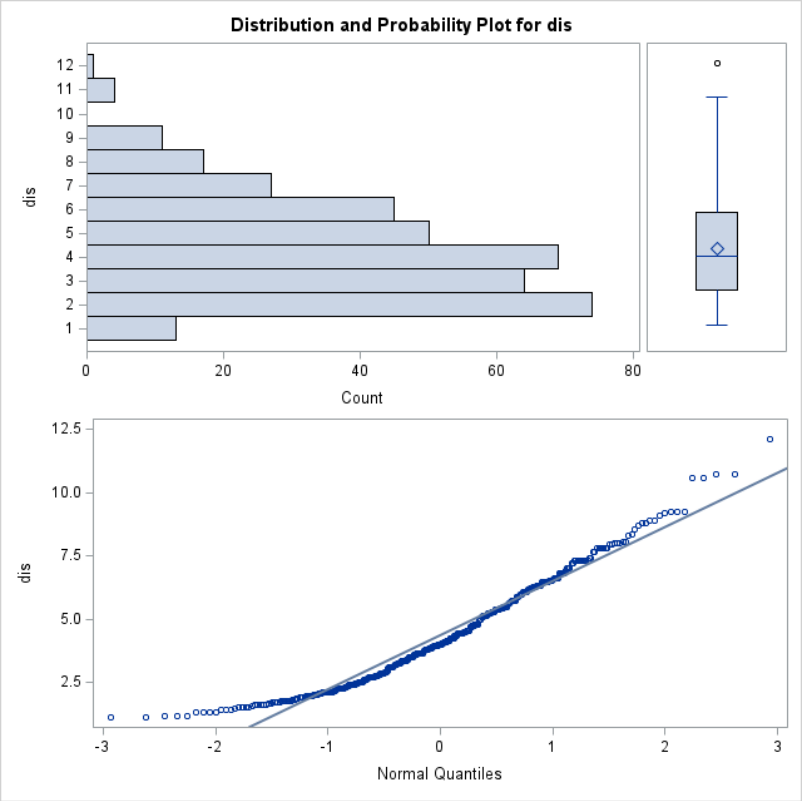

Distribution and Probability Plot for dis
dis
Count
Normal Quantiles

Fig. 25

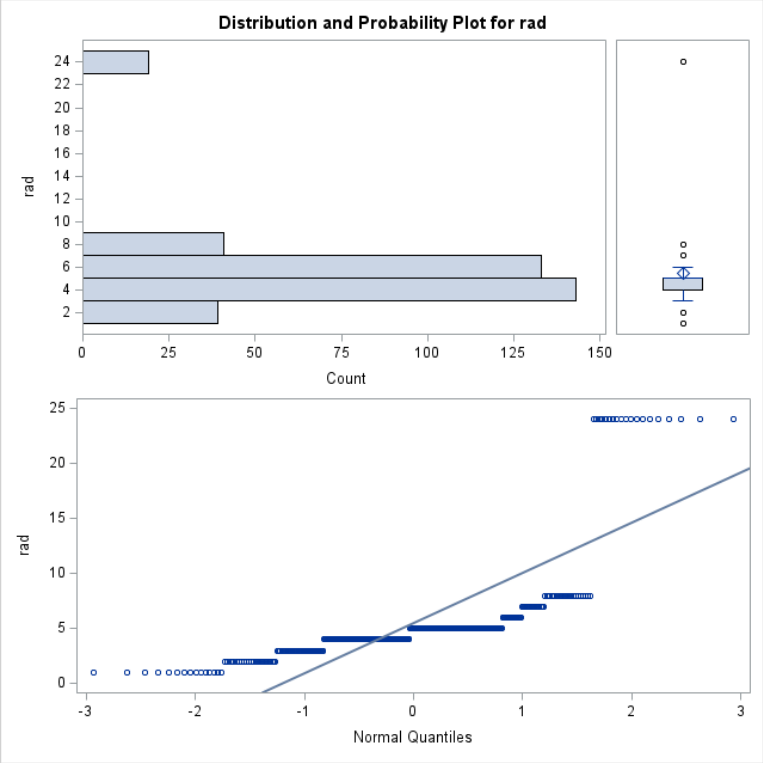

Distribution and Probability Plot for rad
rad
Count
Normal Quantiles

Fig. 26

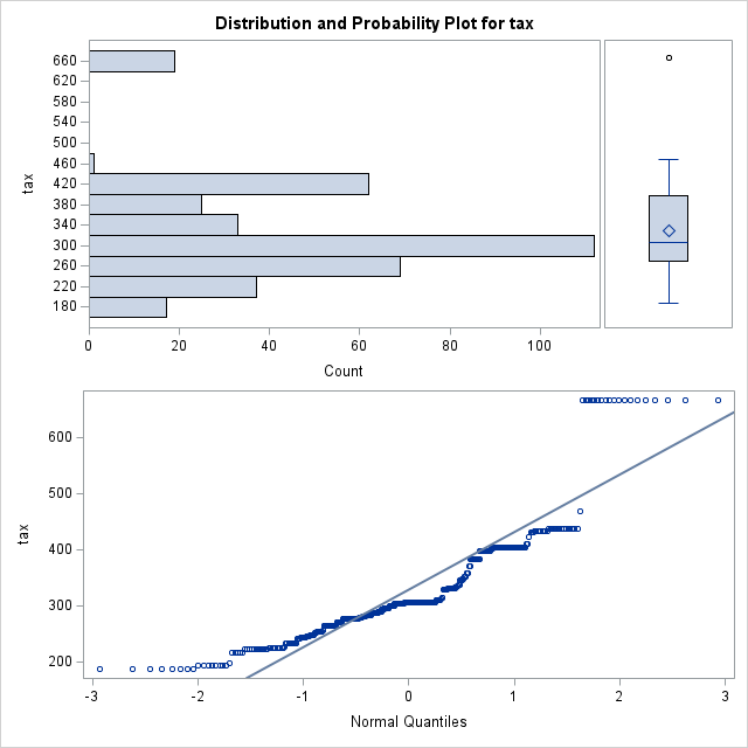

Distribution and Probability Plot for tax
tax
660
620
580
540
500
460
420
380
340
300
260
220
180
0
20
40
60
80
100
Count
tax
600
500
400
300
200
-3
-2
-1
0
1
2
3
Normal Quantiles

Fig. 27

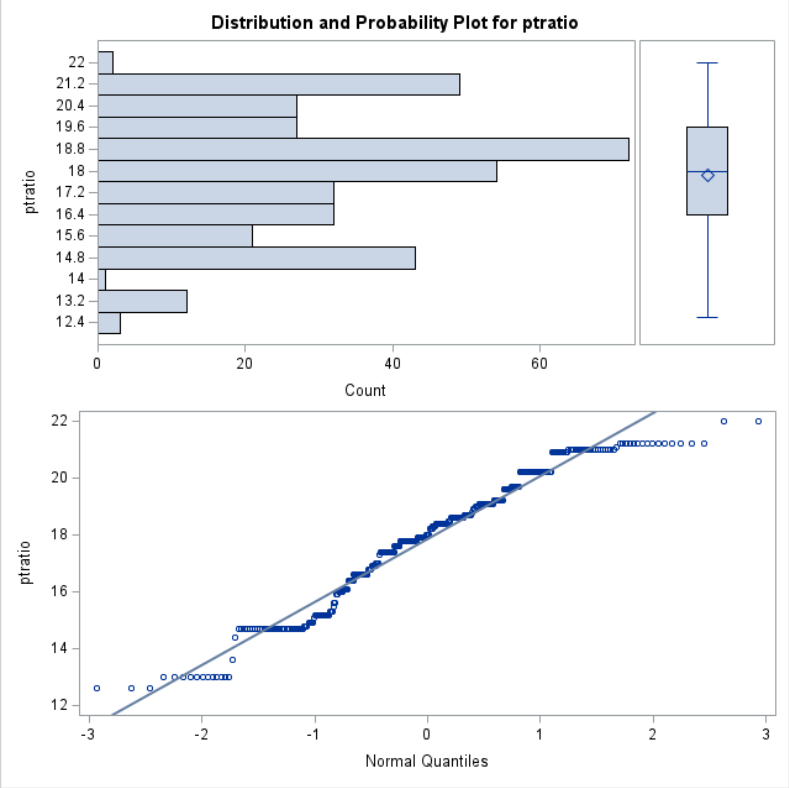

Distribution and Probability Plot for ptratio
ptratio
22
21.2
20.4
19.6
18.8
18
17.2
16.4
15.6
14.8
14
13.2
12.4
0
20
40
60
Count
ptratio
22
20
18
16
14
12
-3
-2
-1
0
1
2
3
Normal Quantiles

Fig. 28

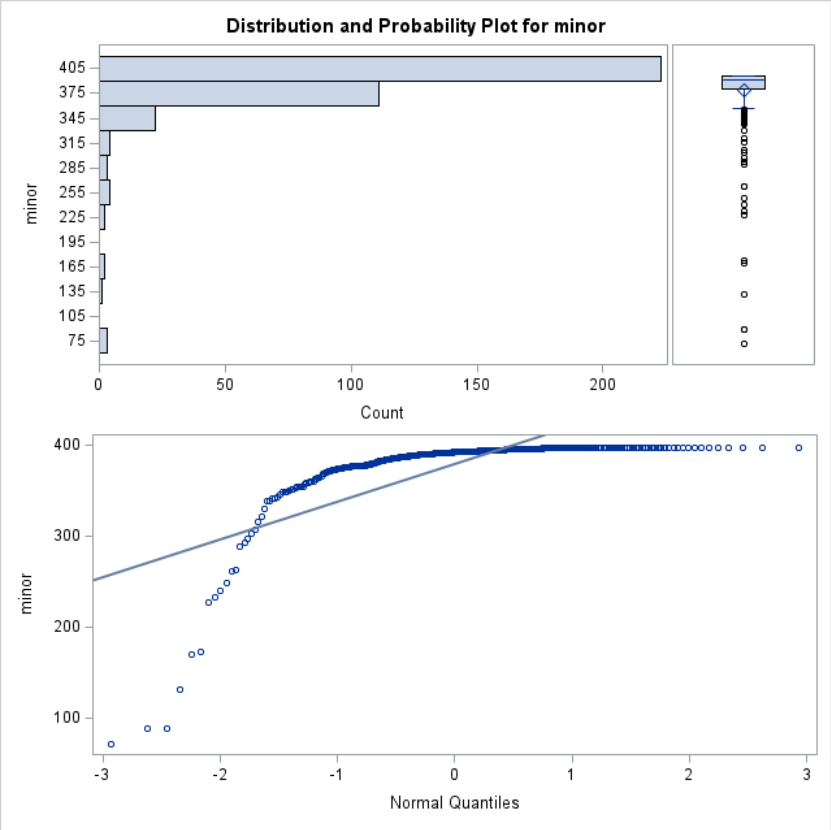

Distribution and Probability Plot for minor
minor
405
375
345
315
285
255
225
195
165
135
105
75
0
50
100
150
200
Count
400
300
200
100
-3
-2
-1
0
1
2
3
Normal Quantiles

Fig. 29

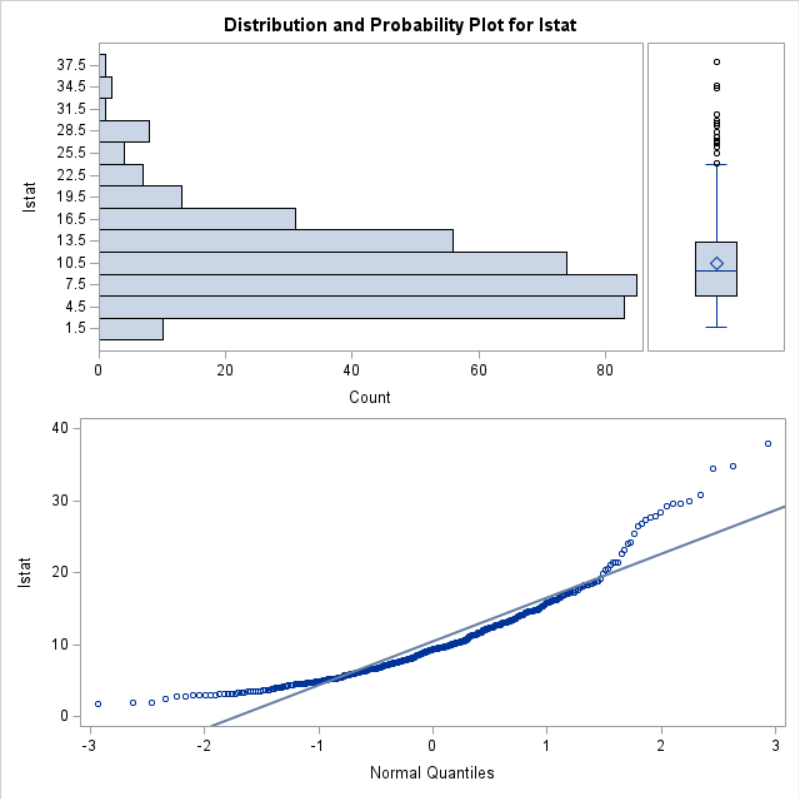

Distribution and Probability Plot for lstat
lstat
37.5
34.5
31.5
28.5
25.5
22.5
19.5
16.5
13.5
10.5
7.5
4.5
1.5
0
20
40
60
80
Count
40
30
20
10
0
-3
-2
-1
0
1
2
3
Normal Quantiles

Fig. 30

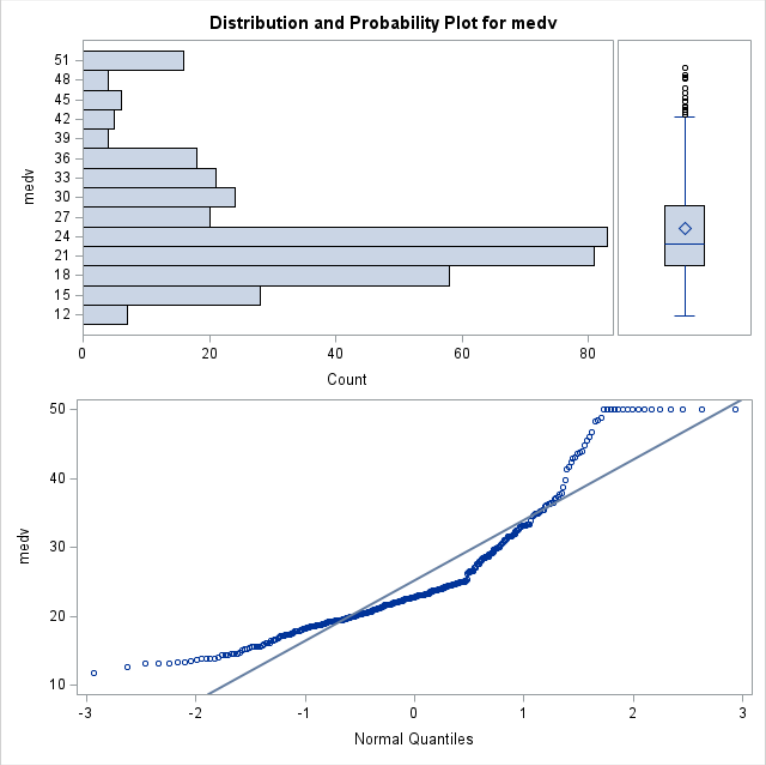


Fig. 31

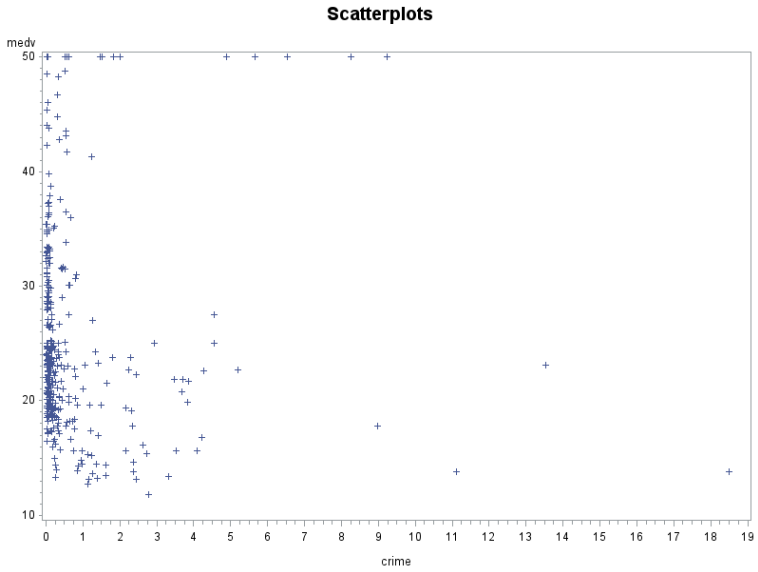


Fig. 32

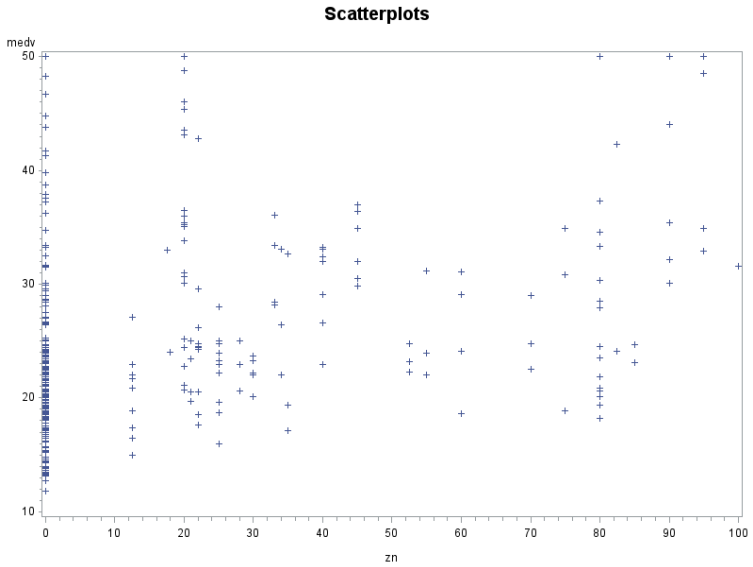

Scatterplots
medv
50
40
30
20
10
0
10
20
30
40
50
60
70
80
90
100
zn

Fig. 33

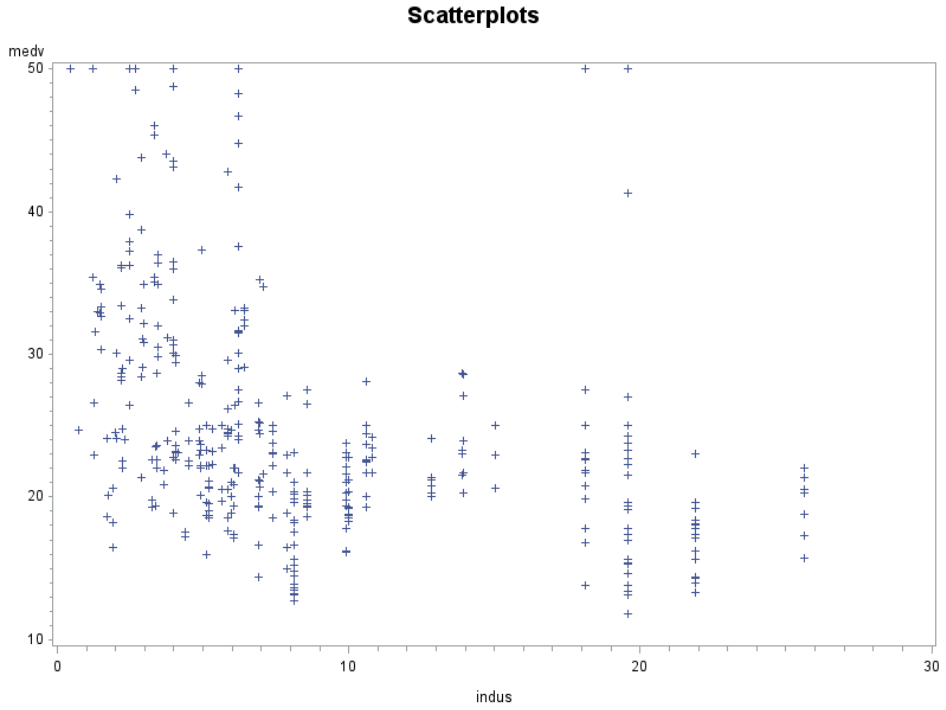

Scatterplots
medv
50
40
30
20
10
0
10
20
30
indus

Fig. 34

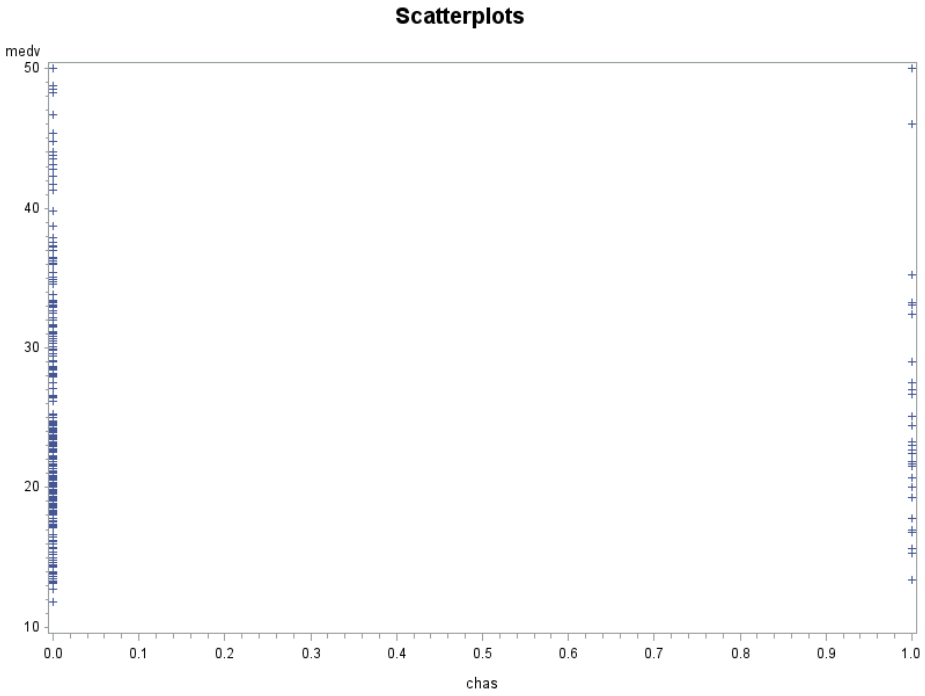

Scatterplots
medv
50
40
30
20
10
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
chas

Fig. 35

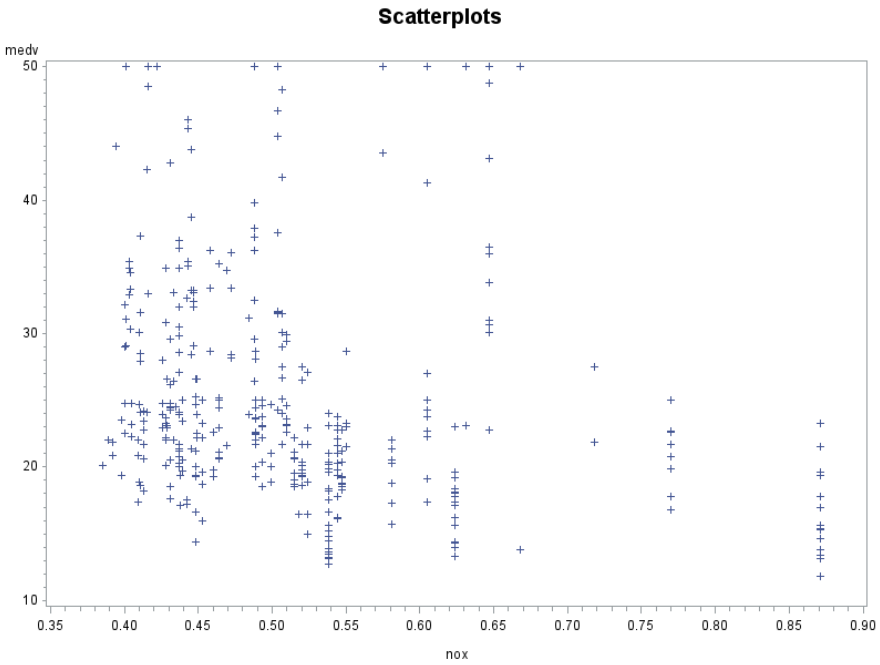

Scatterplots
medv
50
40
30
20
10
0.35
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
nox

Fig. 36

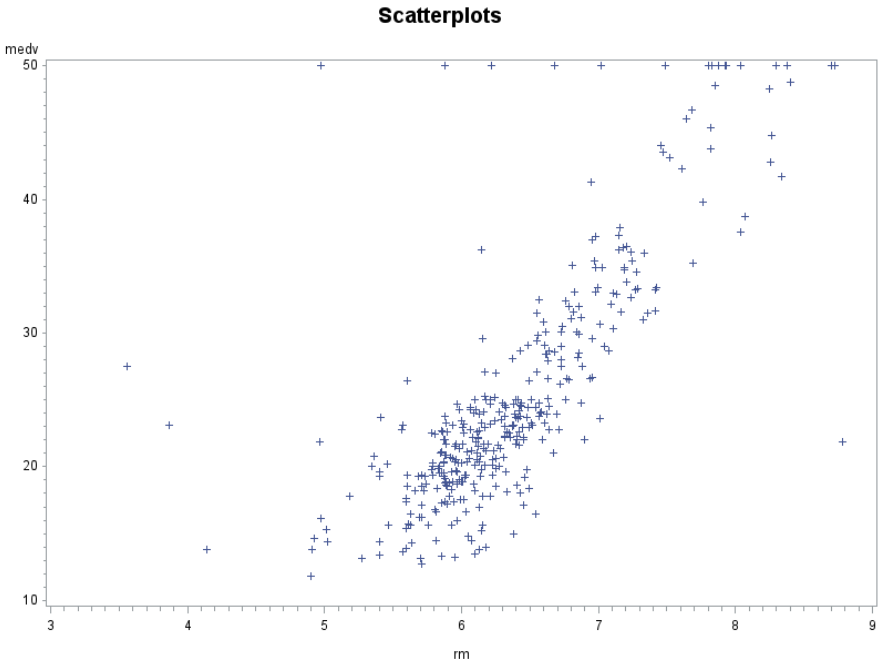

Scatterplots
medv
50
40
30
20
10
3
4
5
6
7
8
9
rm

Fig. 37

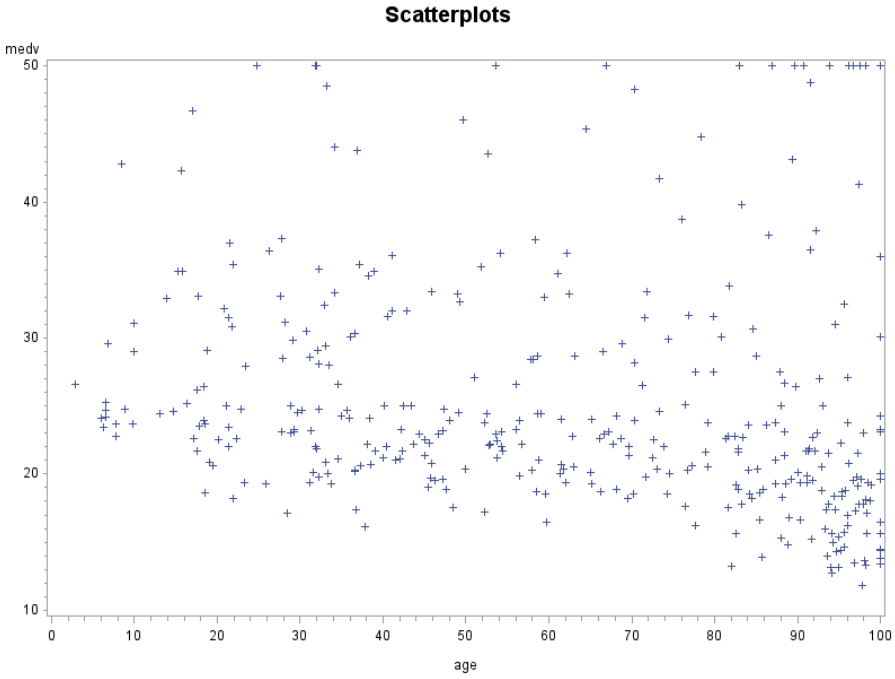

Scatterplots
medv
50
40
30
20
10
0
10
20
30
40
50
60
70
80
90
100
age

Fig. 38

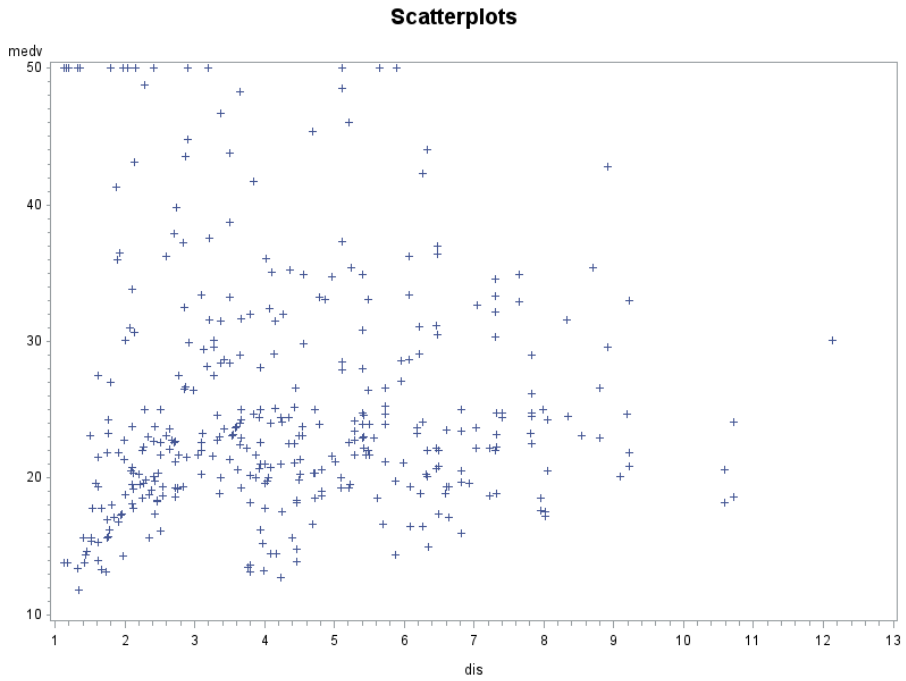

Scatterplots
medv
50
40
30
20
10
1
2
3
4
5
6
7
8
9
10
11
12
13
dis

Fig. 39

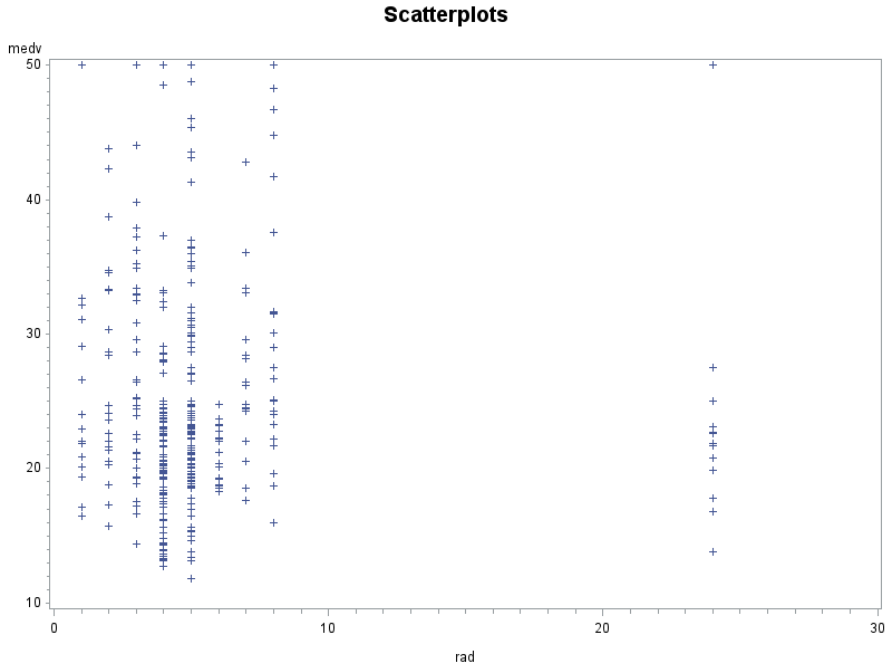

Scatterplots
medv
50
40
30
20
10
0
10
20
30
rad

Fig. 40

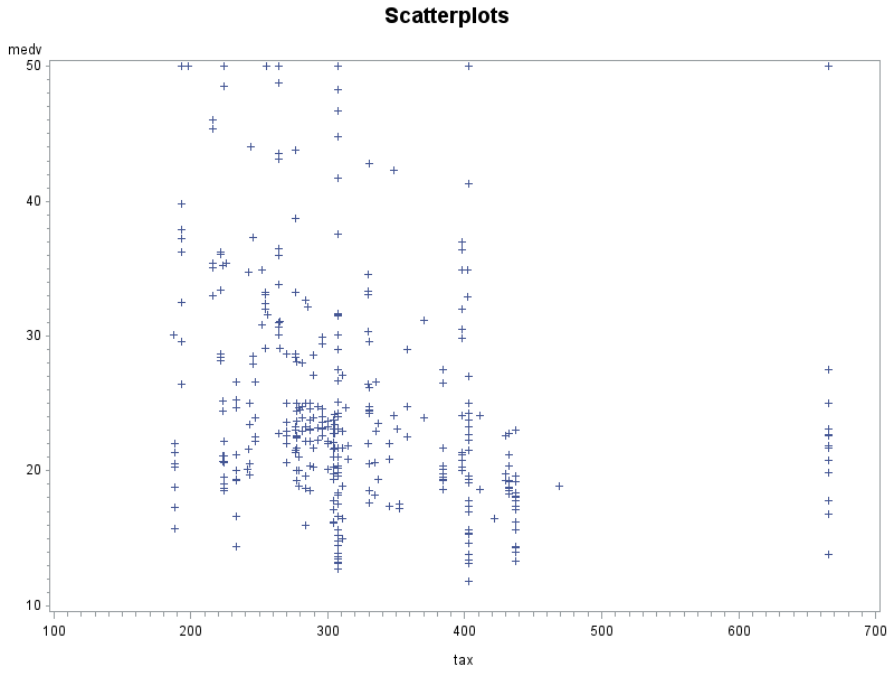

Scatterplots
medv
50
40
30
20
10
100
200
300
400
500
600
700
tax

Fig. 41

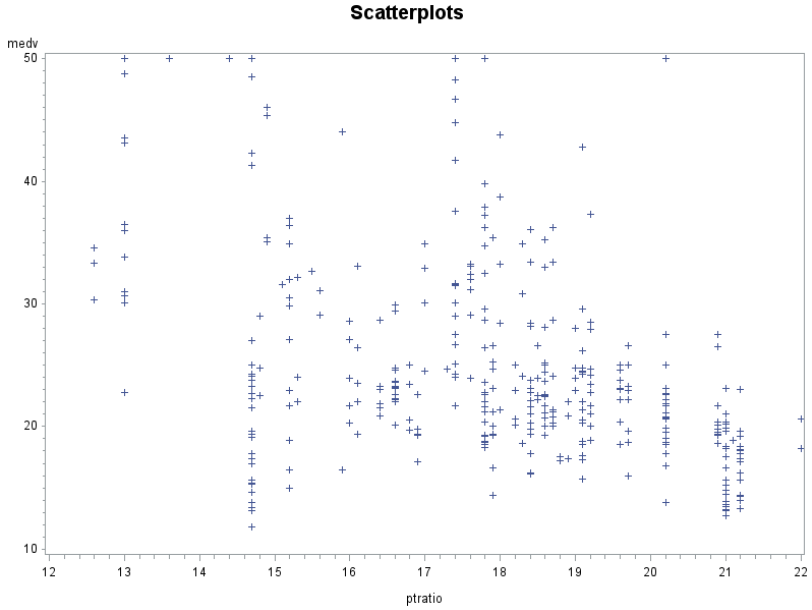

Scatterplots
medv
50
40
30
20
10
12
13
14
15
16
17
18
19
20
21
22
ptratio

Fig. 42

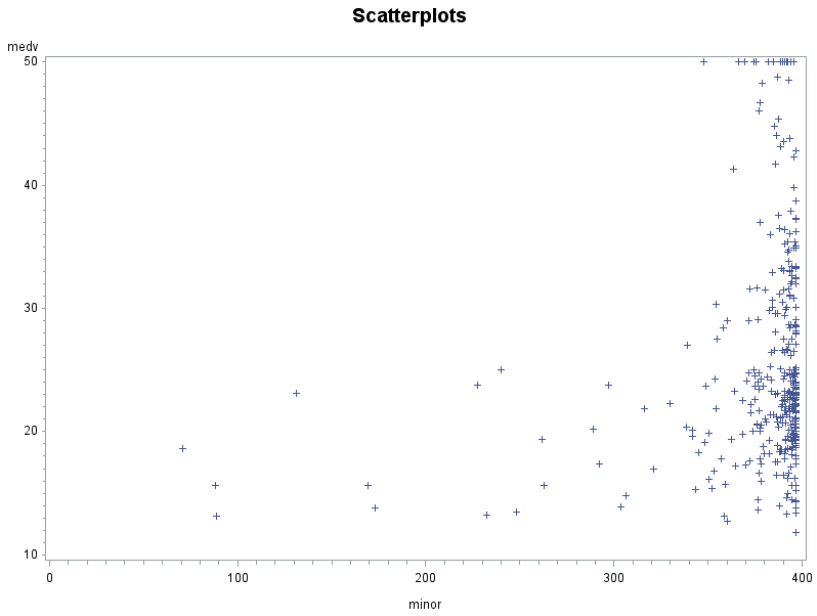

Scatterplots
medv
50
40
30
20
10
0
100
200
300
400
minor

Fig. 43

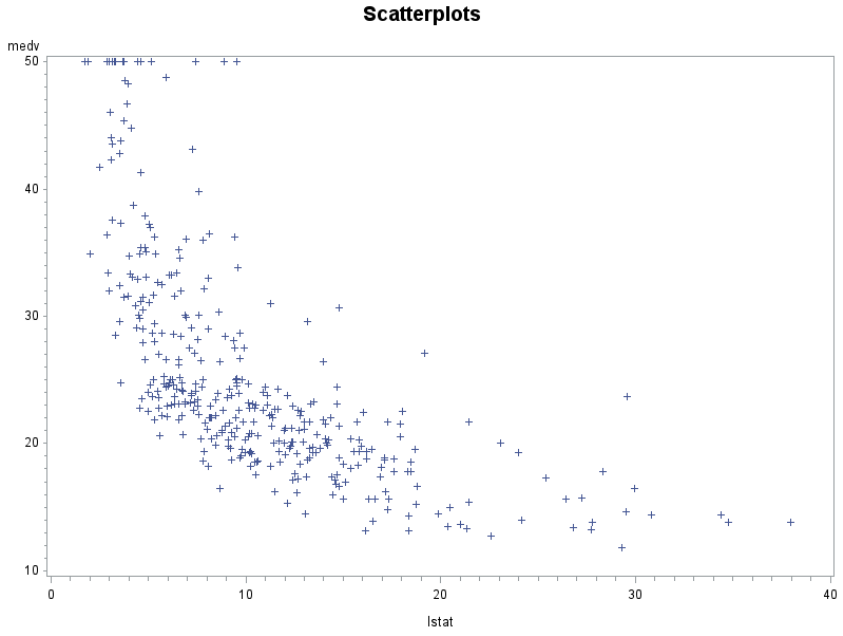

Scatterplots
medv
50
40
30
20
10
0
10
20
30
40
lstat

Fig. 44

Pearson Correlation Coefficients, N = 375 Prob > |r| under H0: Rho=0														
	crime	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	minor	lstat	medv
crime	1.00000	-0.21399 <.0001	0.45386 <.0001	0.20657 <.0001	0.53360 <.0001	-0.31044 <.0001	0.37829 <.0001	-0.39877 <.0001	0.75936 <.0001	0.69373 <.0001	0.11588 0.0248	-0.30261 <.0001	0.32048 <.0001	-0.01481 0.7749
zn	-0.21399 <.0001	1.00000	-0.48816 <.0001	-0.09433 0.0681	-0.47234 <.0001	0.29961 <.0001	-0.54172 <.0001	0.63397 <.0001	-0.19114 0.0002	-0.16831 0.0011	-0.30445 <.0001	0.14393 0.0052	-0.37513 <.0001	0.27888 <.0001
indus	0.45386 <.0001	-0.48816 <.0001	1.00000	0.19753 0.0001	0.71166 <.0001	-0.38523 <.0001	0.56979 <.0001	-0.63166 <.0001	0.32548 <.0001	0.51616 <.0001	0.16211 0.0016	-0.31283 <.0001	0.48576 <.0001	-0.28576 <.0001
chas	0.20657 <.0001	-0.09433 0.0681	0.19753 0.0001	1.00000	0.21678 <.0001	0.06723 0.1939	0.17373 0.0007	-0.20220 <.0001	0.27133 <.0001	0.18081 0.0004	-0.05313 0.3048	-0.04966 0.3376	0.04054 0.4338	0.11888 0.0213
nox	0.53360 <.0001	-0.47234 <.0001	0.71166 <.0001	0.21678 <.0001	1.00000	-0.27901 <.0001	0.68604 <.0001	-0.72091 <.0001	0.41944 <.0001	0.53015 <.0001	-0.07488 0.1479	-0.39915 <.0001	0.45703 <.0001	-0.19826 0.0001
rm	-0.31044 <.0001	0.29961 <.0001	-0.38523 <.0001	0.06723 0.1939	-0.27901 <.0001	1.00000	-0.20323 <.0001	0.12517 0.0153	-0.15020 0.0036	-0.26861 <.0001	-0.34207 <.0001	0.20477 <.0001	-0.64347 <.0001	0.77073 <.0001
age	0.37829 <.0001	-0.54172 <.0001	0.56979 <.0001	0.17373 0.0007	0.68604 <.0001	-0.20323 <.0001	1.00000	-0.70018 <.0001	0.27608 <.0001	0.34772 <.0001	0.09062 0.0797	-0.23857 <.0001	0.53111 <.0001	-0.19727 0.0001
dis	-0.39877 <.0001	0.63397 <.0001	-0.63166 <.0001	-0.20220 <.0001	-0.72091 <.0001	0.12517 0.0153	-0.70018 <.0001	1.00000	-0.29710 <.0001	-0.33734 <.0001	-0.01774 0.7321	0.23309 <.0001	-0.34094 <.0001	-0.00473 0.9273
rad	0.75936 <.0001	-0.19114 0.0002	0.32548 <.0001	0.27133 <.0001	0.41944 <.0001	-0.15020 0.0036	0.27608 <.0001	-0.29710 <.0001	1.00000	0.76603 <.0001	0.22047 <.0001	-0.12609 0.0145	0.07382 0.1536	0.10109 0.0504
tax	0.69373 <.0001	-0.16831 0.0011	0.51616 <.0001	0.18081 0.0004	0.53015 <.0001	-0.26861 <.0001	0.34772 <.0001	-0.33734 <.0001	0.76603 <.0001	1.00000	0.18418 0.0003	-0.24978 <.0001	0.20910 <.0001	-0.11132 0.0311
ptratio	0.11588 0.0248	-0.30445 <.0001	0.16211 0.0016	-0.05313 0.3048	-0.07488 0.1479	-0.34207 <.0001	0.09062 0.0797	-0.01774 0.7321	0.22047 <.0001	0.18418 0.0003	1.00000	0.07069 0.1719	0.20737 <.0001	-0.38236 <.0001
minor	-0.30261 <.0001	0.14393 0.0052	-0.31283 <.0001	-0.04966 0.3376	-0.39915 <.0001	0.20477 <.0001	-0.23857 <.0001	0.23309 <.0001	-0.12609 0.0145	-0.24978 <.0001	0.07069 0.1719	1.00000	-0.20106 <.0001	0.17426 0.0007
lstat	0.32048 <.0001	-0.37513 <.0001	0.48576 <.0001	0.04054 0.4338	0.45703 <.0001	-0.64347 <.0001	0.53111 <.0001	-0.34094 <.0001	0.07382 0.1536	0.20910 <.0001	0.20737 <.0001	-0.20106 <.0001	1.00000	-0.65111 <.0001
medv	-0.01481 0.7749	0.27888 <.0001	-0.28576 <.0001	0.11888 0.0213	-0.19826 0.0001	0.77073 <.0001	-0.19727 0.0001	-0.00473 0.9273	0.10109 0.0504	-0.11132 0.0311	-0.38236 <.0001	0.17426 0.0007	-0.65111 <.0001	1.00000

APPENDIX B – Full Model

Table 1

Full model - Model assumptions and diagnostics

The REG Procedure
Model: MODEL1
Dependent Variable: medv

Number of Observations Read	375
Number of Observations Used	375

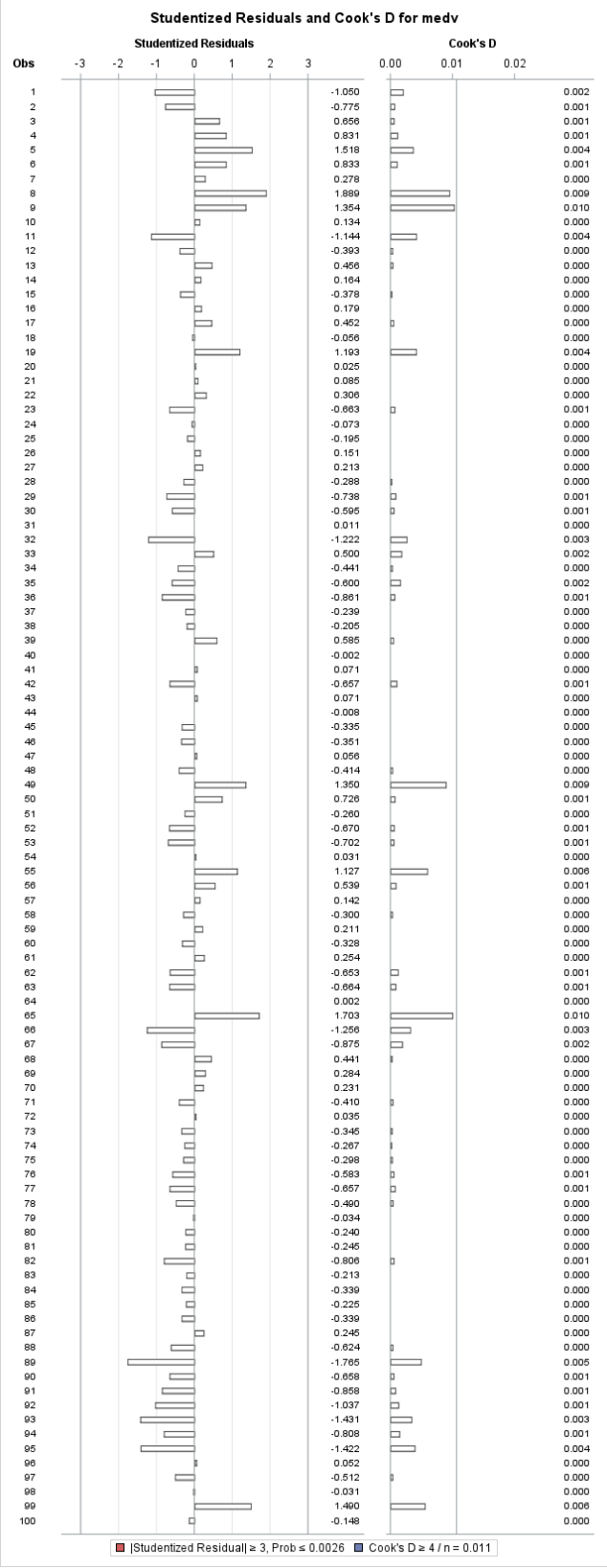
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	13	22446	1726.60036	98.53	<.0001
Error	361	6325.83351	17.52308		

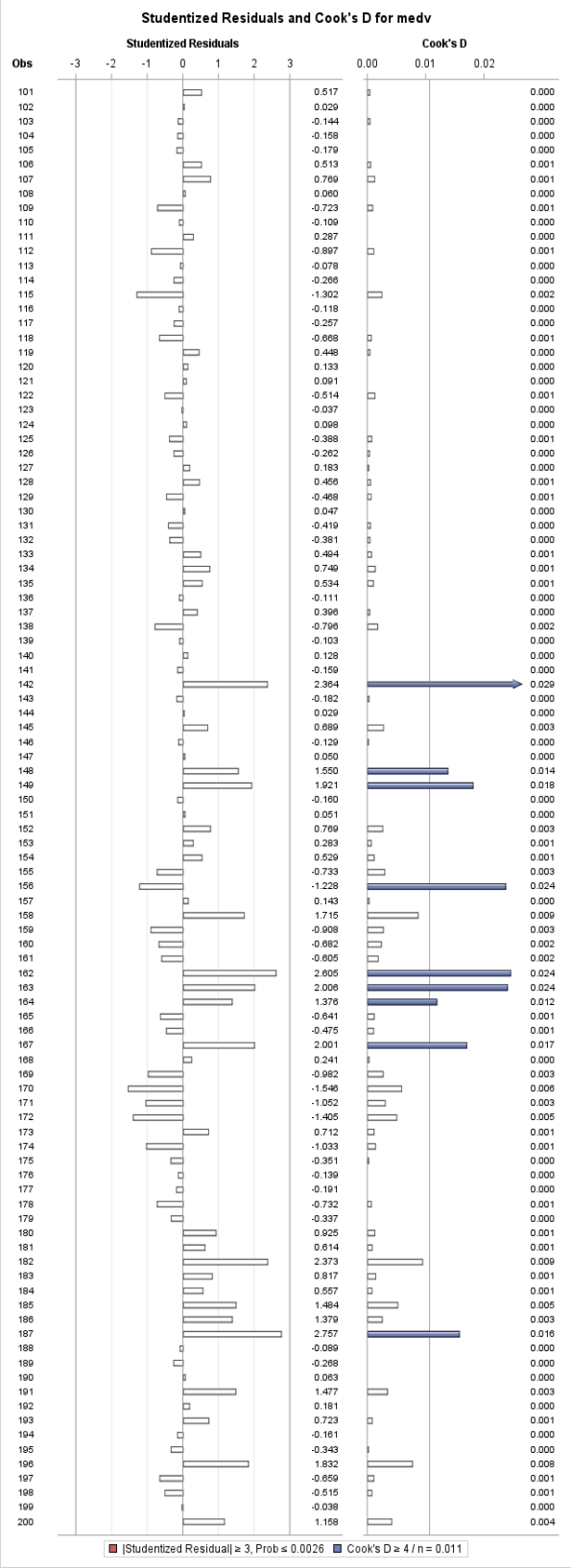
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Corrected Total	374	28772			

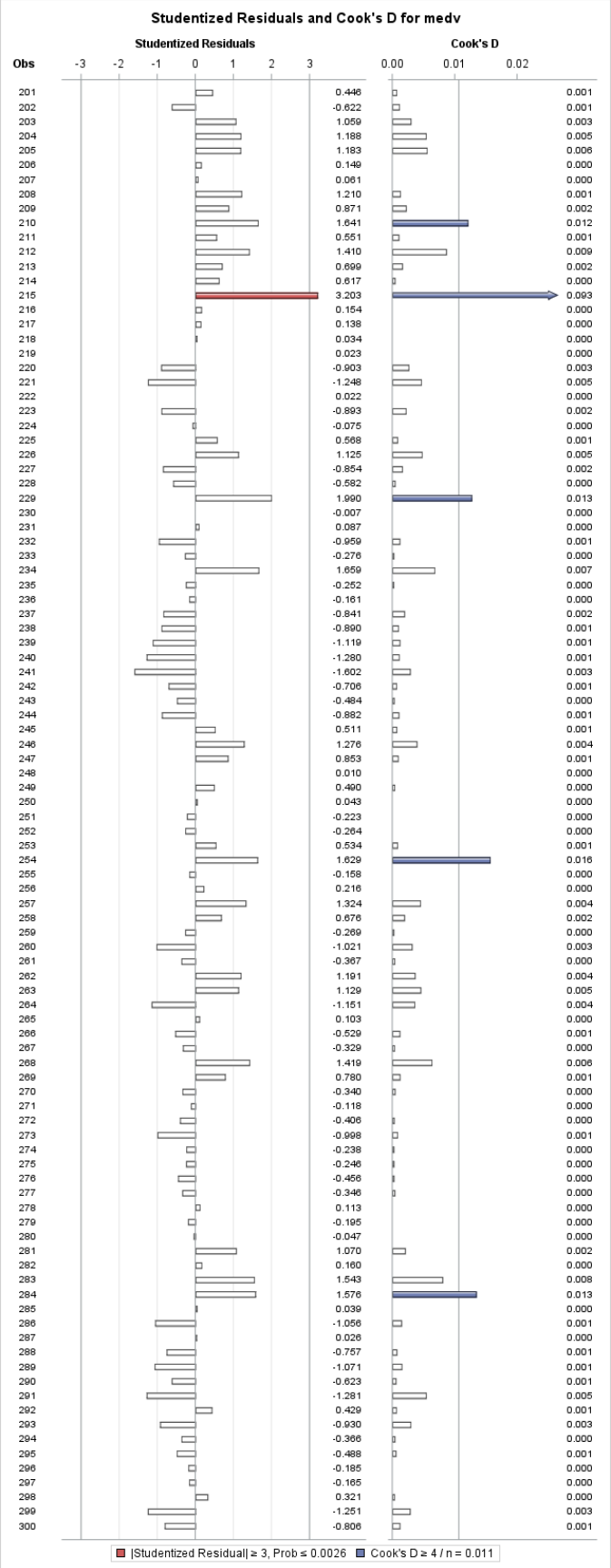
Root MSE	4.18606	**R-Square**	0.7801
Dependent Mean	25.19440	**Adj R-Sq**	0.7722
Coeff Var	16.61503		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standardized Estimate**	**Variance Inflation**
Intercept	**1**	12.63324	5.55329	2.27	0.0235	0	0
crime	**1**	1.28883	0.22380	5.76	<.0001	0.25903	3.32187
zn	**1**	0.03579	0.01241	2.88	0.0042	0.10588	2.21198
indus	**1**	0.07362	0.05727	1.29	0.1994	0.05300	2.79005
chas	**1**	0.12124	0.79815	0.15	0.8794	0.00403	1.15365
nox	**1**	-17.35683	3.98974	-4.35	<.0001	-0.21691	4.08185
rm	**1**	6.41932	0.44839	14.32	<.0001	0.52844	2.23708
age	**1**	-0.00650	0.01269	-0.51	0.6088	-0.02127	2.83139
dis	**1**	-1.18892	0.18716	-6.35	<.0001	-0.28962	3.41311
rad	**1**	0.39406	0.09801	4.02	<.0001	0.20511	4.27283
tax	**1**	-0.01565	0.00389	-4.02	<.0001	-0.18253	3.37728
ptratio	**1**	-0.71652	0.12293	-5.83	<.0001	-0.18114	1.58574
minor	**1**	0.01167	0.00591	1.97	0.0491	0.05521	1.28351
lstat	**1**	-0.42714	0.05824	-7.33	<.0001	-0.29690	2.69046

Fig. 1


Studentized Residuals and Cook's D for medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.011


Studentized Residuals and Cook's D for medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.011


Studentized Residuals and Cook's D for medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02
201 0.446 0.001
202 -0.622 0.001
203 1.059 0.003
204 1.188 0.005
205 1.183 0.006
206 0.149 0.000
207 0.061 0.000
208 1.210 0.001
209 0.871 0.002
210 1.641 0.012
211 0.551 0.001
212 1.410 0.009
213 0.699 0.002
214 0.617 0.000
215 3.203 0.093
216 0.154 0.000
217 0.138 0.000
218 0.034 0.000
219 0.023 0.000
220 -0.903 0.003
221 -1.248 0.005
222 0.022 0.000
223 -0.893 0.002
224 -0.075 0.000
225 0.568 0.001
226 1.125 0.005
227 -0.854 0.002
228 -0.582 0.000
229 1.990 0.013
230 -0.007 0.000
231 0.087 0.000
232 -0.959 0.001
233 -0.276 0.000
234 1.659 0.007
235 -0.252 0.000
236 -0.161 0.000
237 -0.841 0.002
238 -0.890 0.001
239 -1.119 0.001
240 -1.280 0.001
241 -1.602 0.003
242 -0.706 0.001
243 -0.484 0.000
244 -0.882 0.001
245 0.511 0.001
246 1.276 0.004
247 0.853 0.001
248 0.010 0.000
249 0.490 0.000
250 0.043 0.000
251 -0.223 0.000
252 -0.264 0.000
253 0.534 0.001
254 1.629 0.016
255 -0.158 0.000
256 0.216 0.000
257 1.324 0.004
258 0.676 0.002
259 -0.269 0.000
260 -1.021 0.003
261 -0.367 0.000
262 1.191 0.004
263 1.129 0.005
264 -1.151 0.004
265 0.103 0.000
266 -0.529 0.001
267 -0.329 0.000
268 1.419 0.006
269 0.780 0.001
270 -0.340 0.000
271 -0.118 0.000
272 -0.406 0.000
273 -0.998 0.001
274 -0.238 0.000
275 -0.246 0.000
276 -0.456 0.000
277 -0.346 0.000
278 0.113 0.000
279 -0.195 0.000
280 -0.047 0.000
281 1.070 0.002
282 0.160 0.000
283 1.543 0.008
284 1.576 0.013
285 0.039 0.000
286 -1.056 0.001
287 0.026 0.000
288 -0.757 0.001
289 -1.071 0.001
290 -0.623 0.001
291 -1.281 0.005
292 0.429 0.001
293 -0.930 0.003
294 -0.366 0.000
295 -0.488 0.001
296 -0.185 0.000
297 -0.165 0.000
298 0.321 0.000
299 -1.251 0.003
300 -0.806 0.001
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.011

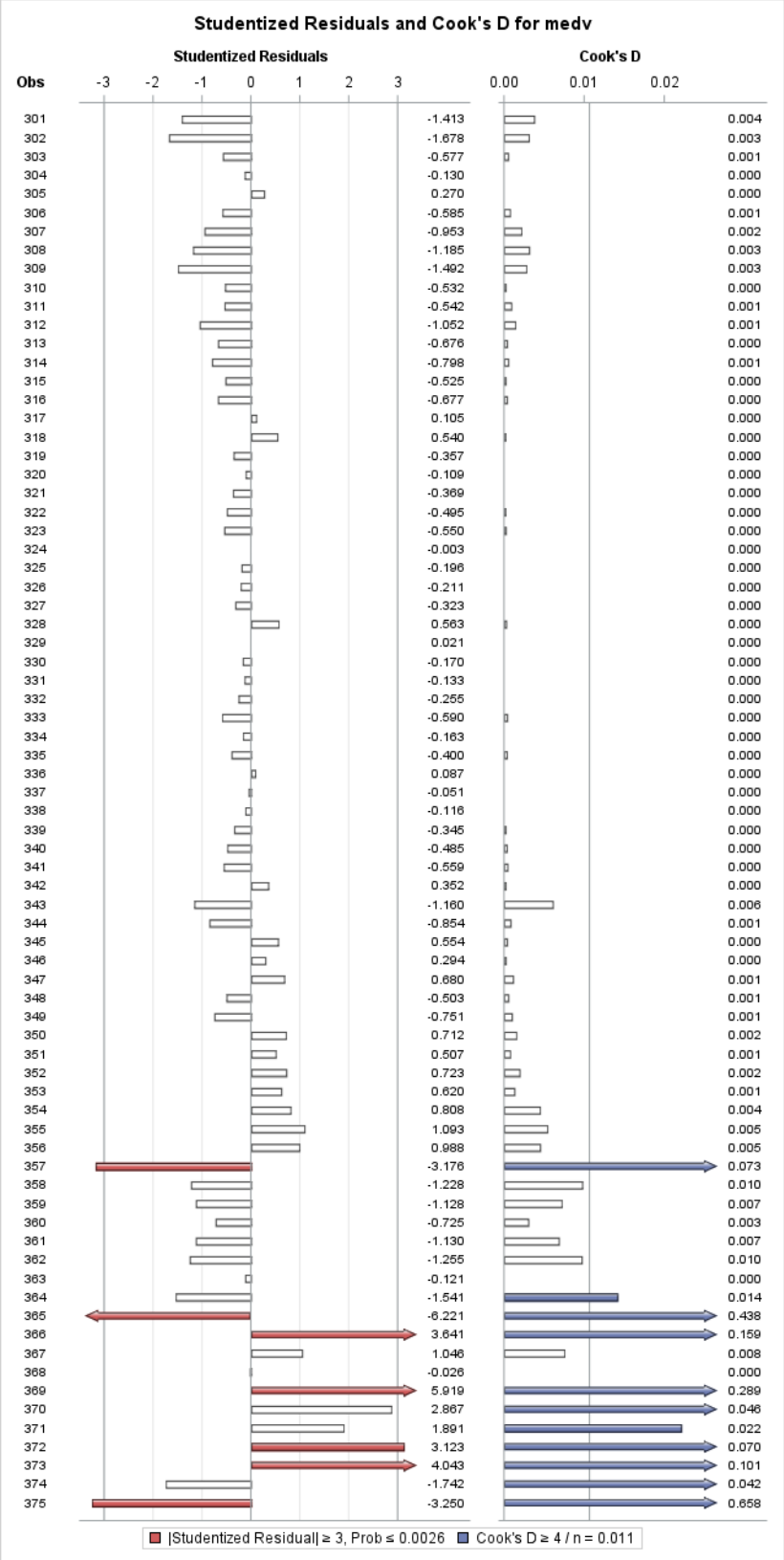

Studentized Residuals and Cook's D for medv
Studentized Residuals
Cook's D
Obs
-3
-2
-1
0
1
2
3
0.00
0.01
0.02
301 -1.413 0.004
302 -1.678 0.003
303 -0.577 0.001
304 -0.130 0.000
305 0.270 0.000
306 -0.585 0.001
307 -0.953 0.002
308 -1.185 0.003
309 -1.492 0.003
310 -0.532 0.000
311 -0.542 0.001
312 -1.052 0.001
313 -0.676 0.000
314 -0.798 0.001
315 -0.525 0.000
316 -0.677 0.000
317 0.105 0.000
318 0.540 0.000
319 -0.357 0.000
320 -0.109 0.000
321 -0.369 0.000
322 -0.495 0.000
323 -0.550 0.000
324 -0.003 0.000
325 -0.196 0.000
326 -0.211 0.000
327 -0.323 0.000
328 0.563 0.000
329 0.021 0.000
330 -0.170 0.000
331 -0.133 0.000
332 -0.255 0.000
333 -0.590 0.000
334 -0.163 0.000
335 -0.400 0.000
336 0.087 0.000
337 -0.051 0.000
338 -0.116 0.000
339 -0.345 0.000
340 -0.485 0.000
341 -0.559 0.000
342 0.352 0.000
343 -1.160 0.006
344 -0.854 0.001
345 0.554 0.000
346 0.294 0.000
347 0.680 0.001
348 -0.503 0.001
349 -0.751 0.001
350 0.712 0.002
351 0.507 0.001
352 0.723 0.002
353 0.620 0.001
354 0.808 0.004
355 1.093 0.005
356 0.988 0.005
357 -3.176 0.073
358 -1.228 0.010
359 -1.128 0.007
360 -0.725 0.003
361 -1.130 0.007
362 -1.255 0.010
363 -0.121 0.000
364 -1.541 0.014
365 -6.221 0.438
366 3.641 0.159
367 1.046 0.008
368 -0.026 0.000
369 5.919 0.289
370 2.867 0.046
371 1.891 0.022
372 3.123 0.070
373 4.043 0.101
374 -1.742 0.042
375 -3.250 0.658
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.011

Fig. 2

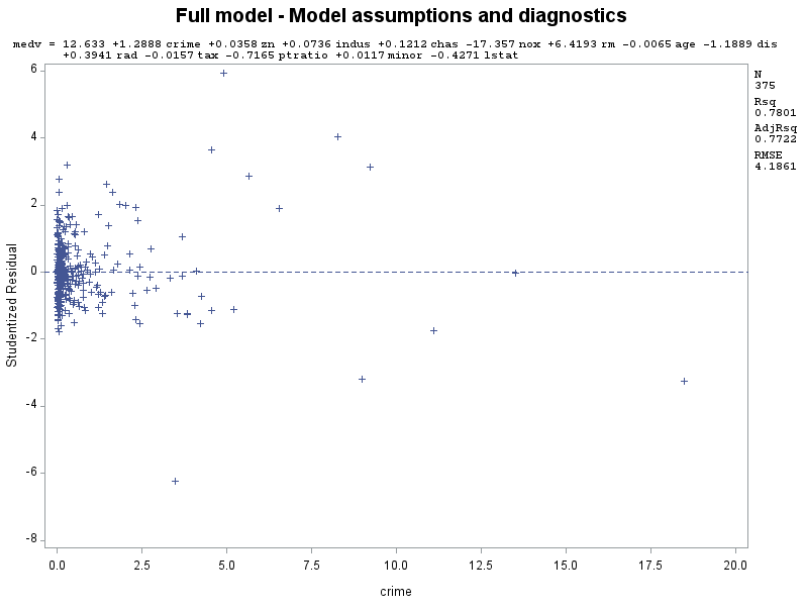

Full model - Model assumptions and diagnostics
medv = 12.633 +1.2888 crime +0.0358 zn +0.0736 indus +0.1212 chas -17.357 nox +6.4193 rm -0.0065 age -1.1889 dis
+0.3941 rad -0.0157 tax -0.7165 ptratio +0.0117 minor -0.4271 lstat
N
375
Rsq
0.7801
AdjRsq
0.7722
RMSE
4.1861
Studentized Residual
crime

Fig. 3

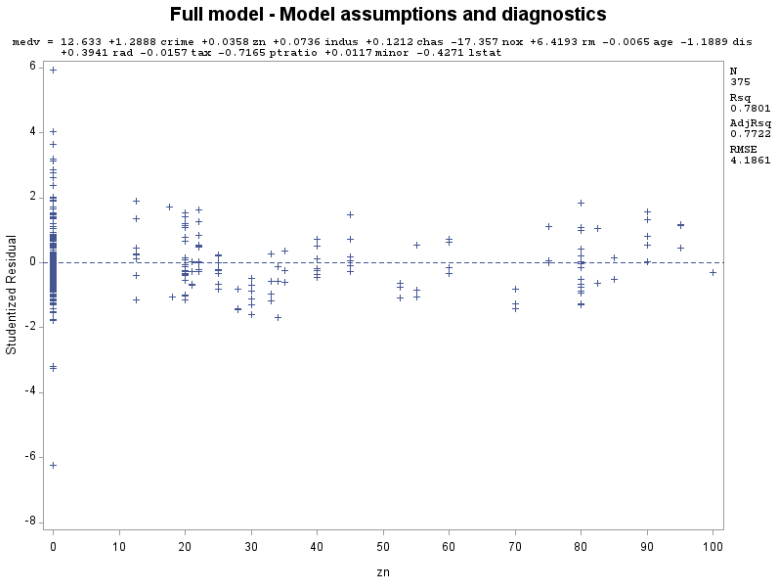

Full model - Model assumptions and diagnostics
medv = 12.633 +1.2888 crime +0.0358 zn +0.0736 indus +0.1212 chas -17.357 nox +6.4193 rm -0.0065 age -1.1889 dis
+0.3941 rad -0.0157 tax -0.7165 ptratio +0.0117 minor -0.4271 lstat
N
375
Rsq
0.7801
AdjRsq
0.7722
RMSE
4.1861
Studentized Residual
zn

Fig. 4

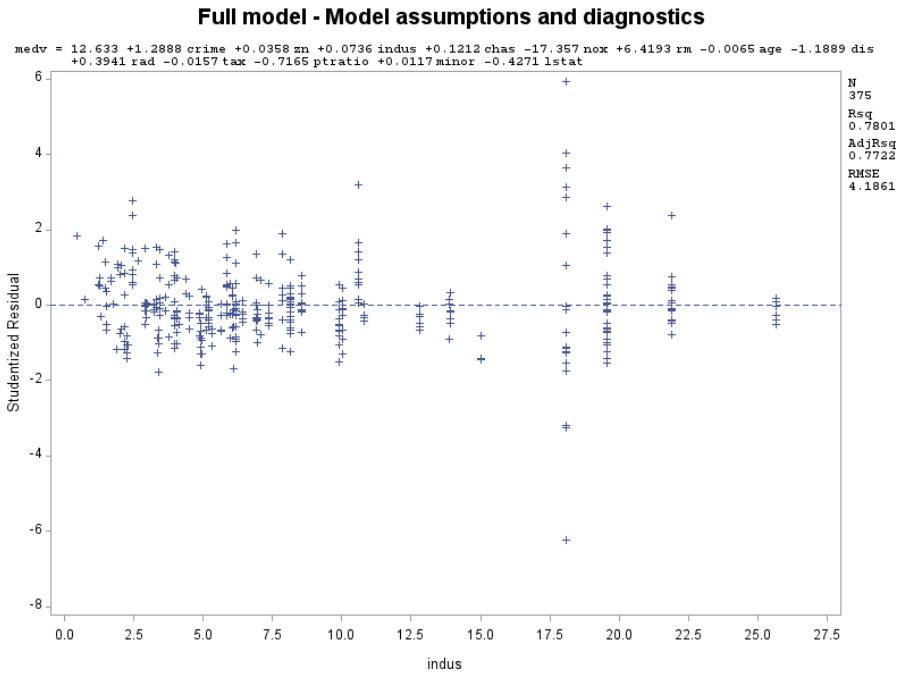

Full model - Model assumptions and diagnostics
medv = 12.633 +1.2888 crime +0.0358 zn +0.0736 indus +0.1212 chas -17.357 nox +6.4193 rm -0.0065 age -1.1889 dis
+0.3941 rad -0.0157 tax -0.7165 ptratio +0.0117 minor -0.4271 lstat
N
375
Rsq
0.7801
AdjRsq
0.7722
RMSE
4.1861
Studentized Residual
indus

Fig. 5

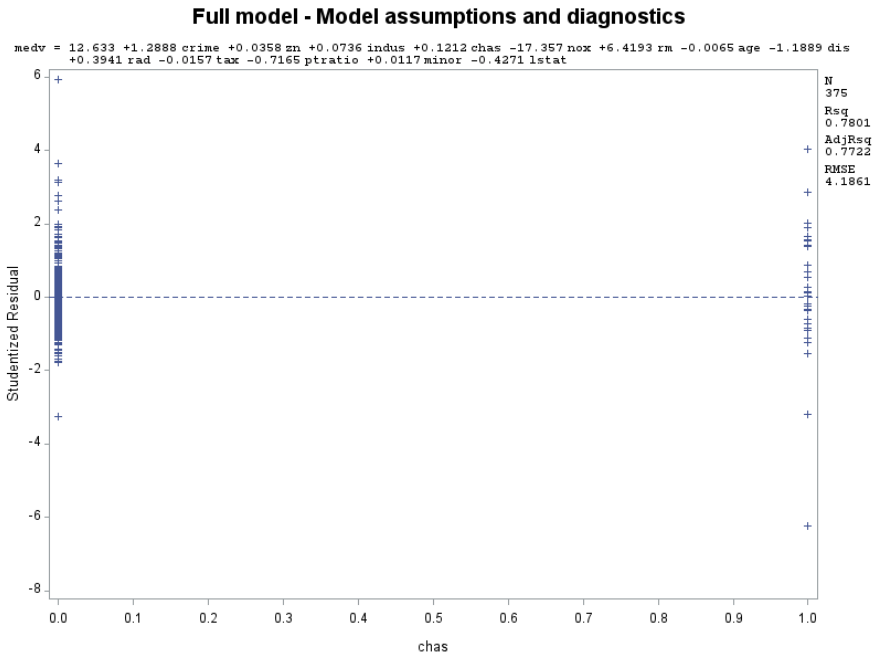

Full model - Model assumptions and diagnostics
medv = 12.633 +1.2888 crime +0.0358 zn +0.0736 indus +0.1212 chas -17.357 nox +6.4193 rm -0.0065 age -1.1889 dis
+0.3941 rad -0.0157 tax -0.7165 ptratio +0.0117 minor -0.4271 lstat
N
375
Rsq
0.7801
AdjRsq
0.7722
RMSE
4.1861
Studentized Residual
chas

Fig. 6

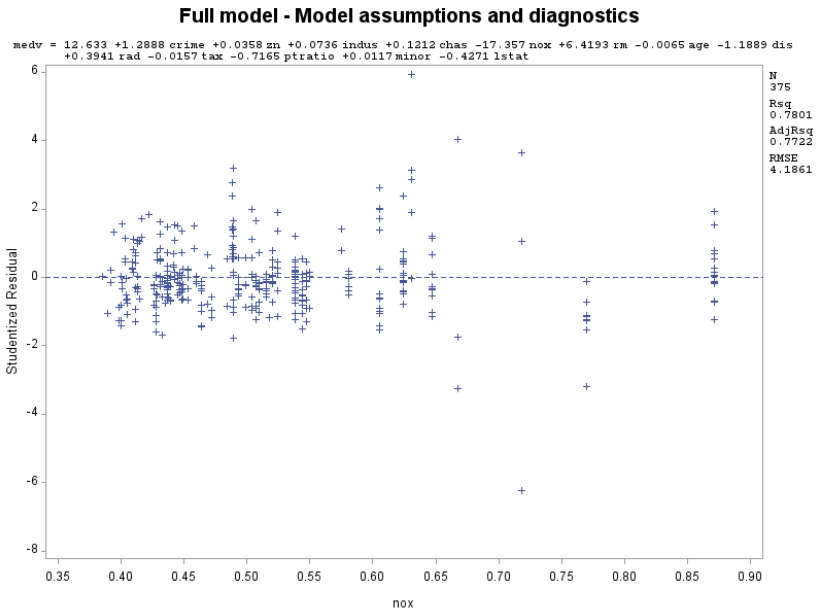

Full model - Model assumptions and diagnostics
medv = 12.633 +1.2888 crime +0.0358 zn +0.0736 indus +0.1212 chas -17.357 nox +6.4193 rm -0.0065 age -1.1889 dis
+0.3941 rad -0.0157 tax -0.7165 ptratio +0.0117 minor -0.4271 lstat
N
375
Rsq
0.7801
AdjRsq
0.7722
RMSE
4.1861
Studentized Residual
nox

Fig. 7

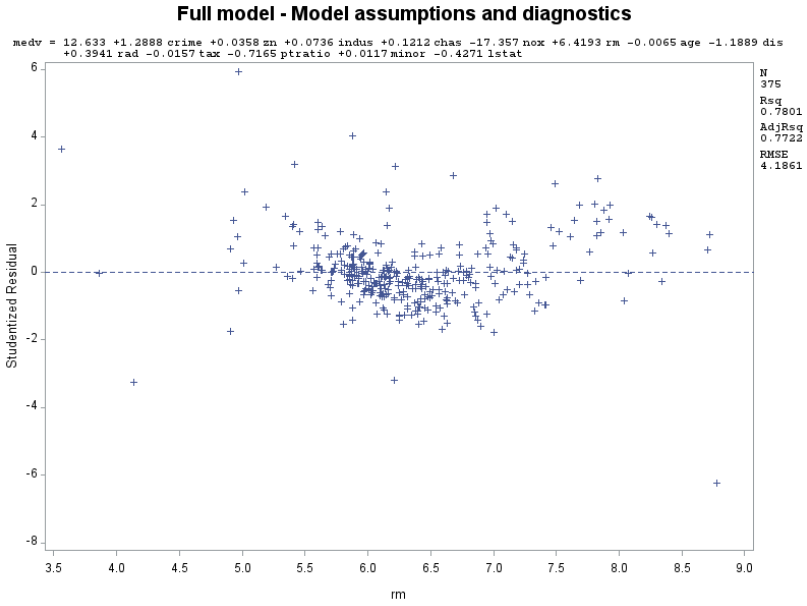

Full model - Model assumptions and diagnostics
medv = 12.633 +1.2888 crime +0.0358 zn +0.0736 indus +0.1212 chas -17.357 nox +6.4193 rm -0.0065 age -1.1889 dis
+0.3941 rad -0.0157 tax -0.7165 ptratio +0.0117 minor -0.4271 lstat
N
375
Rsq
0.7801
AdjRsq
0.7722
RMSE
4.1861
Studentized Residual
rm

Fig. 8

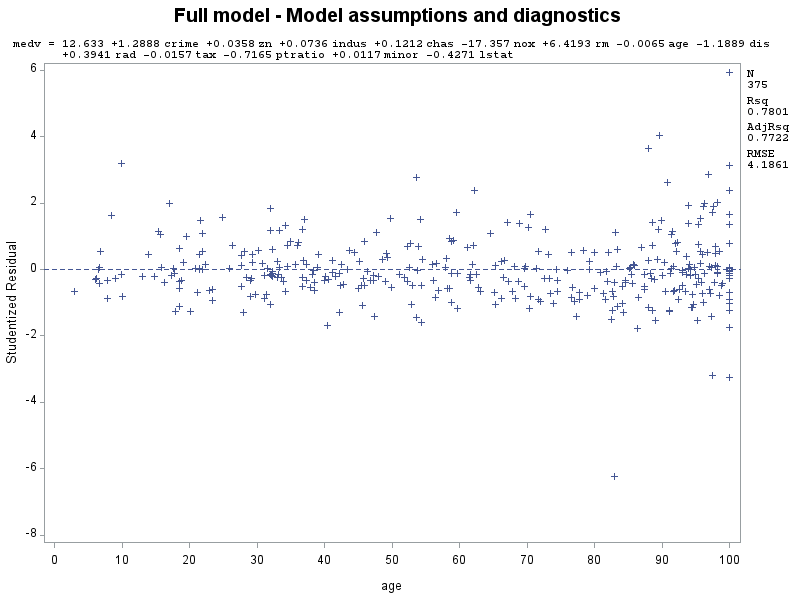

Full model - Model assumptions and diagnostics
medv = 12.633 +1.2888 crime +0.0358 zn +0.0736 indus +0.1212 chas -17.357 nox +6.4193 rm -0.0065 age -1.1889 dis
+0.3941 rad -0.0157 tax -0.7165 ptratio +0.0117 minor -0.4271 lstat
N
375
Rsq
0.7801
AdjRsq
0.7722
RMSE
4.1861
Studentized Residual
age

Fig. 9

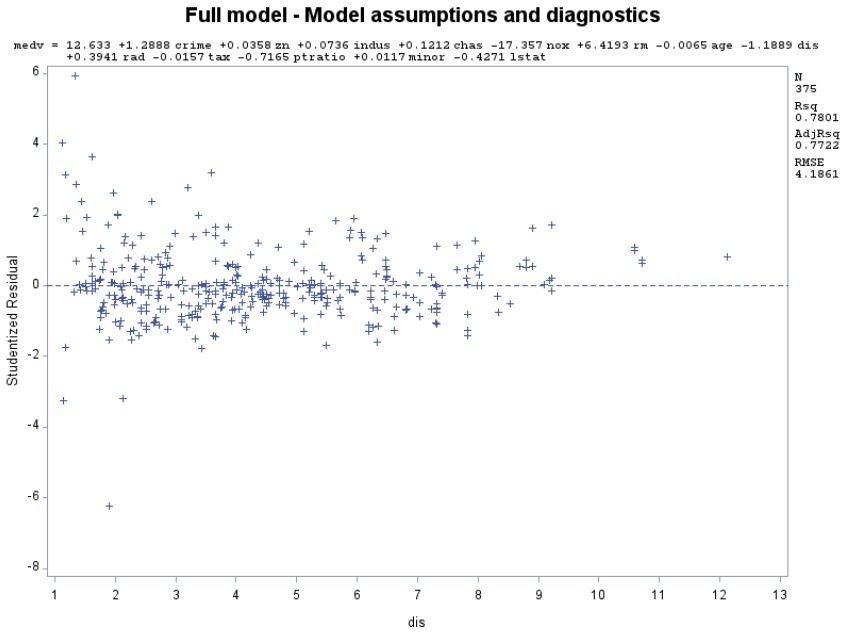

Full model - Model assumptions and diagnostics
medv = 12.633 +1.2888 crime +0.0358 zn +0.0736 indus +0.1212 chas -17.357 nox +6.4193 rm -0.0065 age -1.1889 dis
+0.3941 rad -0.0157 tax -0.7165 ptratio +0.0117 minor -0.4271 lstat
N
375
Rsq
0.7801
AdjRsq
0.7722
RMSE
4.1861
Studentized Residual
dis

Fig. 10

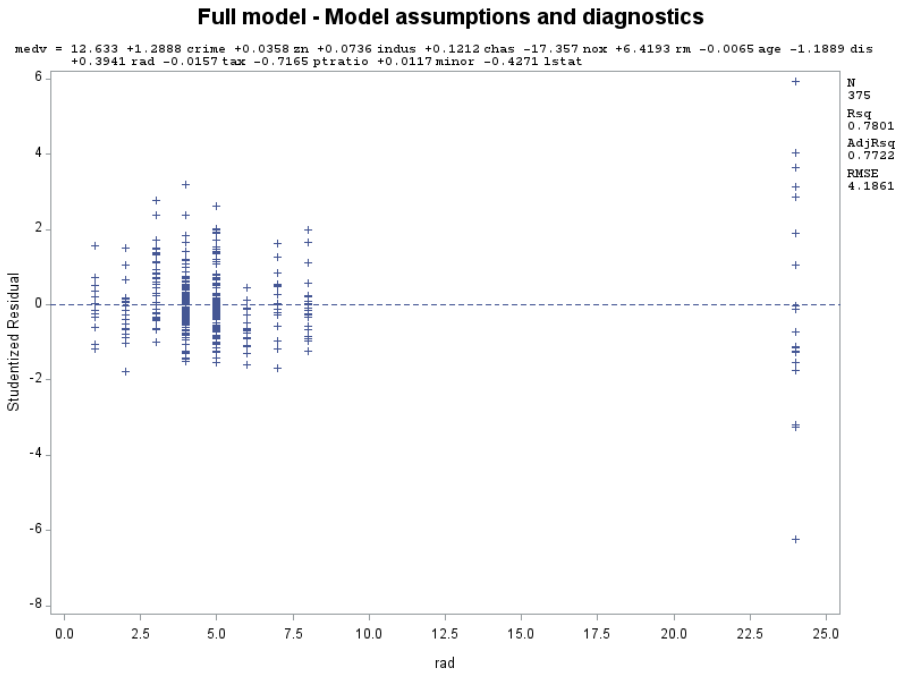

Full model - Model assumptions and diagnostics
medv = 12.633 +1.2888 crime +0.0358 zn +0.0736 indus +0.1212 chas -17.357 nox +6.4193 rm -0.0065 age -1.1889 dis
+0.3941 rad -0.0157 tax -0.7165 ptratio +0.0117 minor -0.4271 lstat
N
375
Rsq
0.7801
AdjRsq
0.7722
RMSE
4.1861
Studentized Residual
rad

Fig. 11

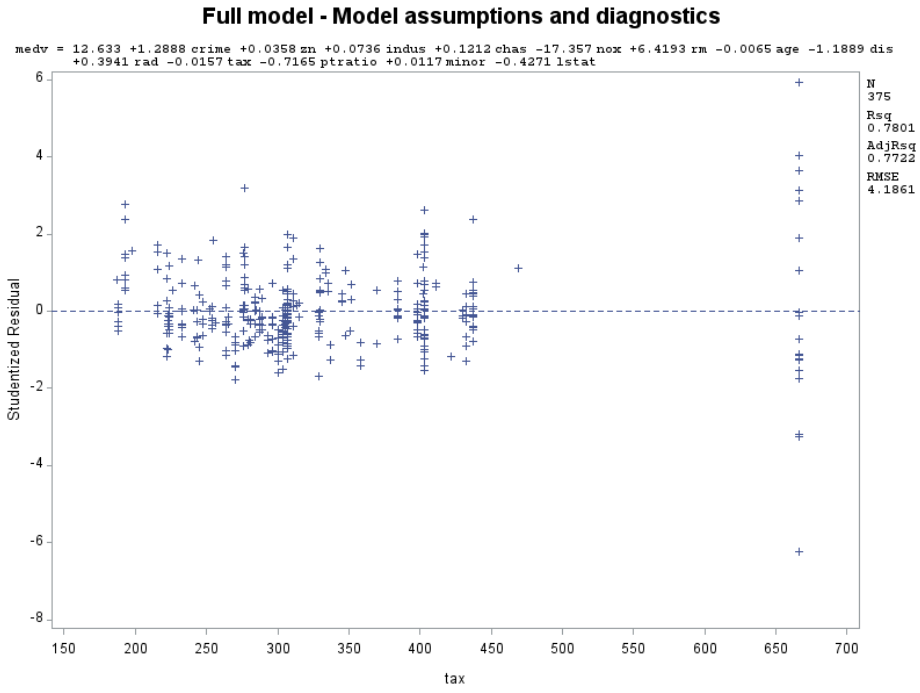

Full model - Model assumptions and diagnostics
medv = 12.633 +1.2888 crime +0.0358 zn +0.0736 indus +0.1212 chas -17.357 nox +6.4193 rm -0.0065 age -1.1889 dis
+0.3941 rad -0.0157 tax -0.7165 ptratio +0.0117 minor -0.4271 lstat
N
375
Rsq
0.7801
AdjRsq
0.7722
RMSE
4.1861
Studentized Residual
tax

Fig. 12

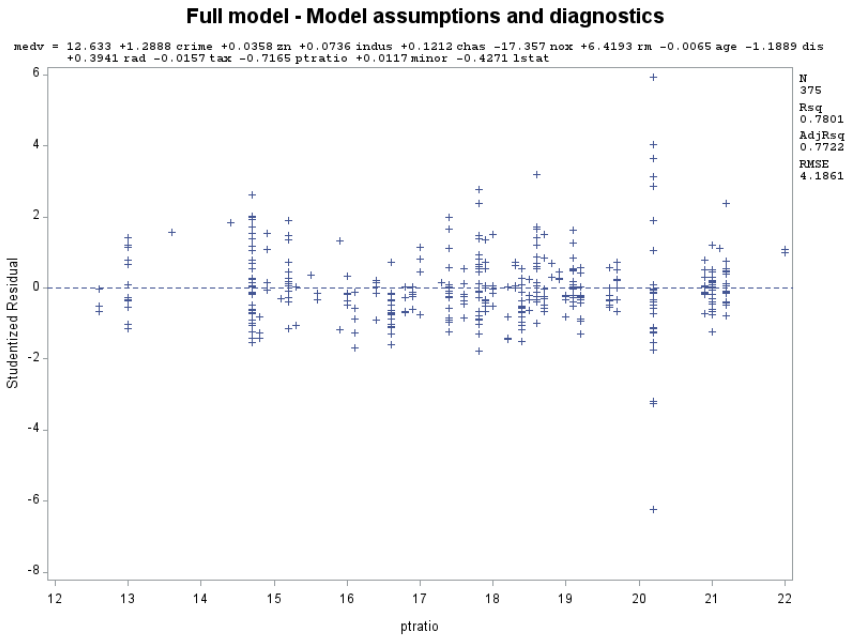

Full model - Model assumptions and diagnostics
medv = 12.633 +1.2888 crime +0.0358 zn +0.0736 indus +0.1212 chas -17.357 nox +6.4193 rm -0.0065 age -1.1889 dis
+0.3941 rad -0.0157 tax -0.7165 ptratio +0.0117 minor -0.4271 lstat
N
375
Rsq
0.7801
AdjRsq
0.7722
RMSE
4.1861
Studentized Residual
ptratio

Fig. 13

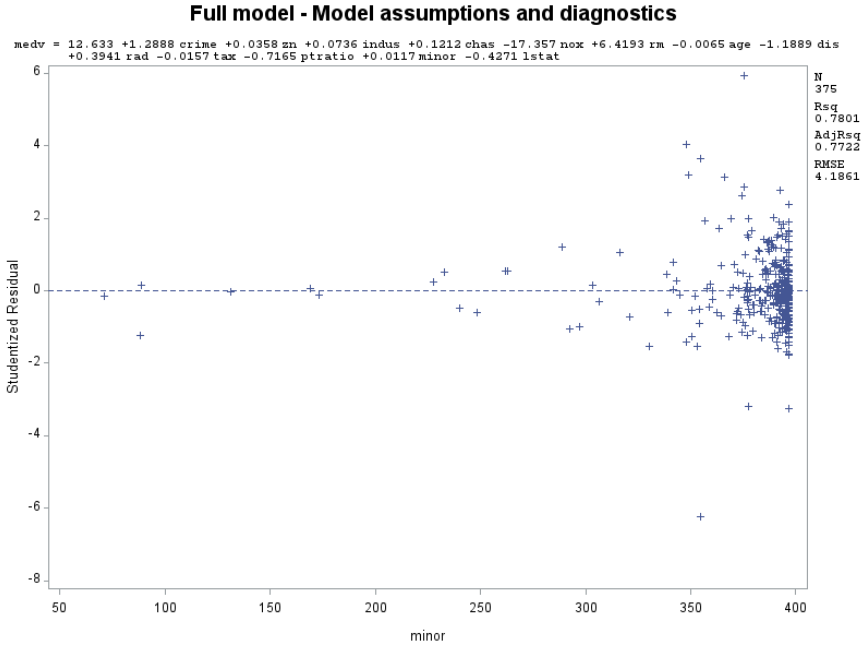

Full model - Model assumptions and diagnostics
medv = 12.633 +1.2888 crime +0.0358 zn +0.0736 indus +0.1212 chas -17.357 nox +6.4193 rm -0.0065 age -1.1889 dis
+0.3941 rad -0.0157 tax -0.7165 ptratio +0.0117 minor -0.4271 lstat
N
375
Rsq
0.7801
AdjRsq
0.7722
RMSE
4.1861
Studentized Residual
minor

Fig. 14

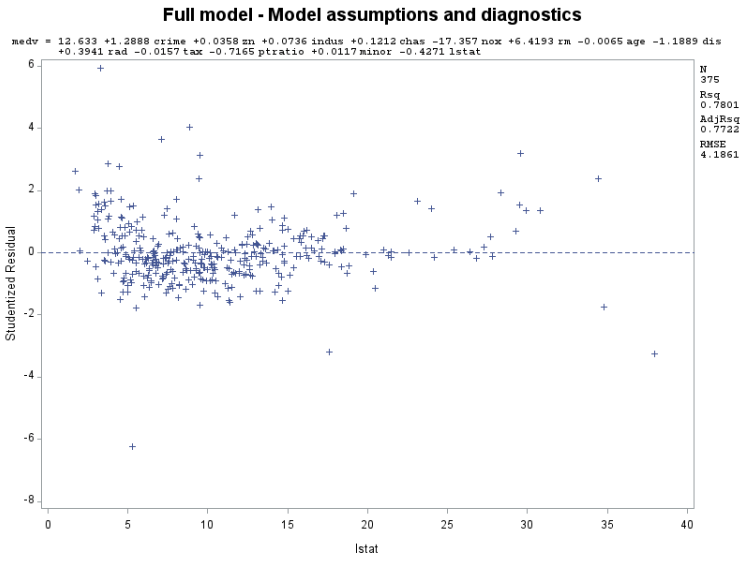

Full model - Model assumptions and diagnostics
medv = 12.633 +1.2888 crime +0.0358 zn +0.0736 indus +0.1212 chas -17.357 nox +6.4193 rm -0.0065 age -1.1889 dis
+0.3941 rad -0.0157 tax -0.7165 ptratio +0.0117 minor -0.4271 lstat
N
375
Rsq
0.7801
AdjRsq
0.7722
RMSE
4.1861
Studentized Residual
lstat

Fig. 15

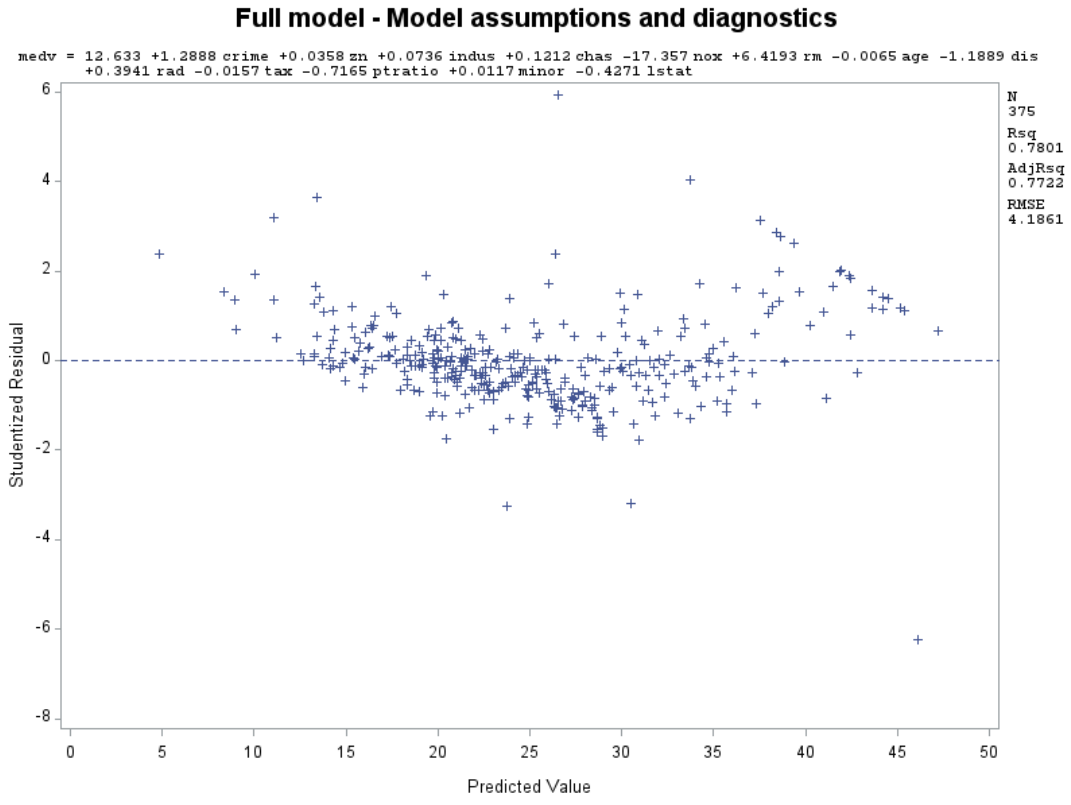

Full model - Model assumptions and diagnostics
medv = 12.633 +1.2888 crime +0.0358 zn +0.0736 indus +0.1212 chas -17.357 nox +6.4193 rm -0.0065 age -1.1889 dis
+0.3941 rad -0.0157 tax -0.7165 ptratio +0.0117 minor -0.4271 lstat
N
375
Rsq
0.7801
AdjRsq
0.7722
RMSE
4.1861
Studentized Residual
Predicted Value

Fig. 16

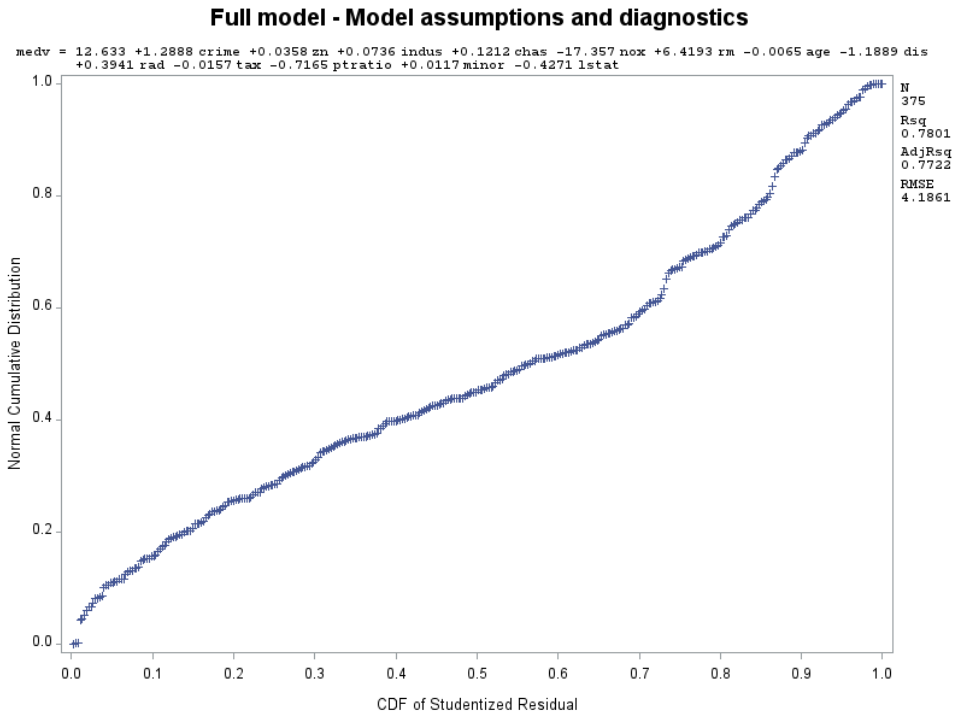


Table 2

Checking Model assumptions and diagnostics after removing outliers

The REG Procedure
Model: MODEL1
Dependent Variable: medv

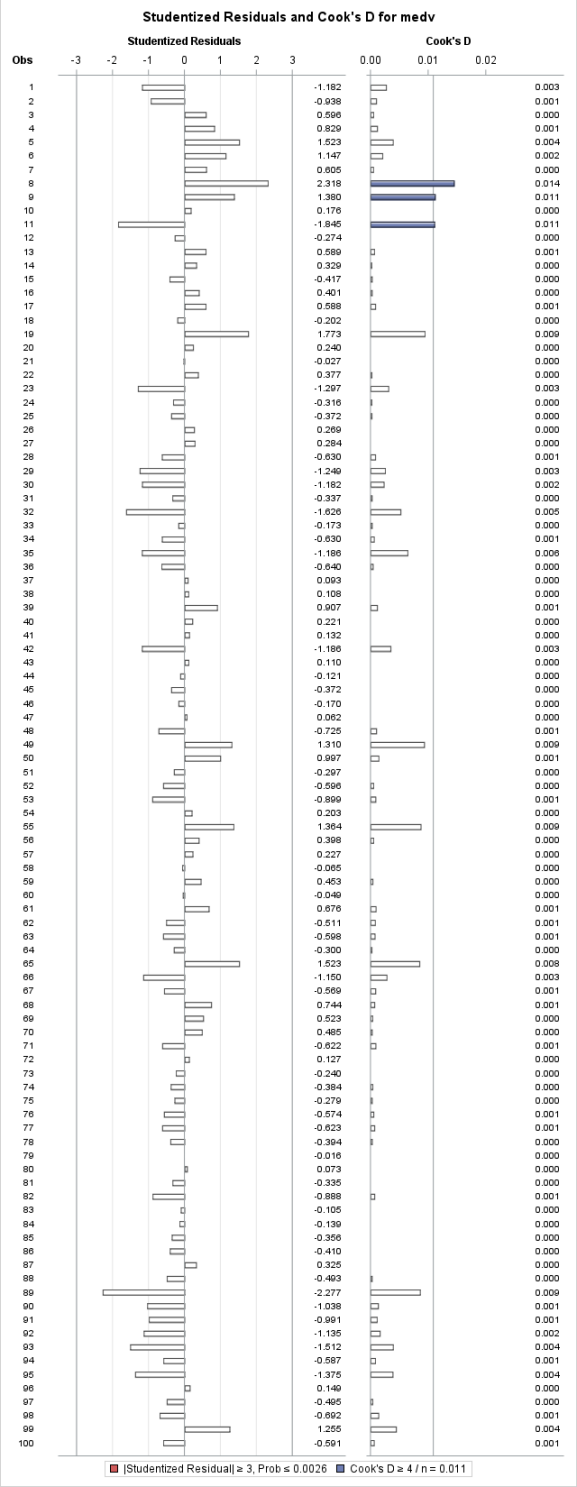
Number of Observations Read	367
Number of Observations Used	367

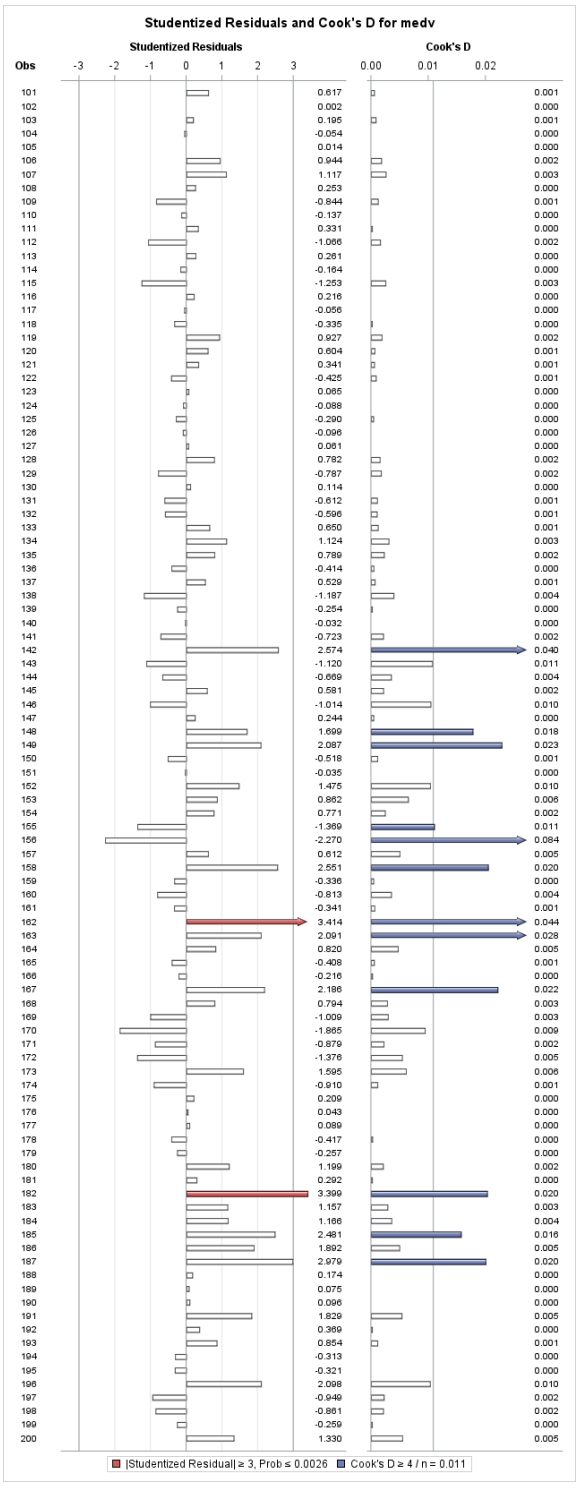
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	13	22972	1767.08614	166.65	<.0001
Error	353	3742.95757	10.60328		
Corrected Total	366	26715			

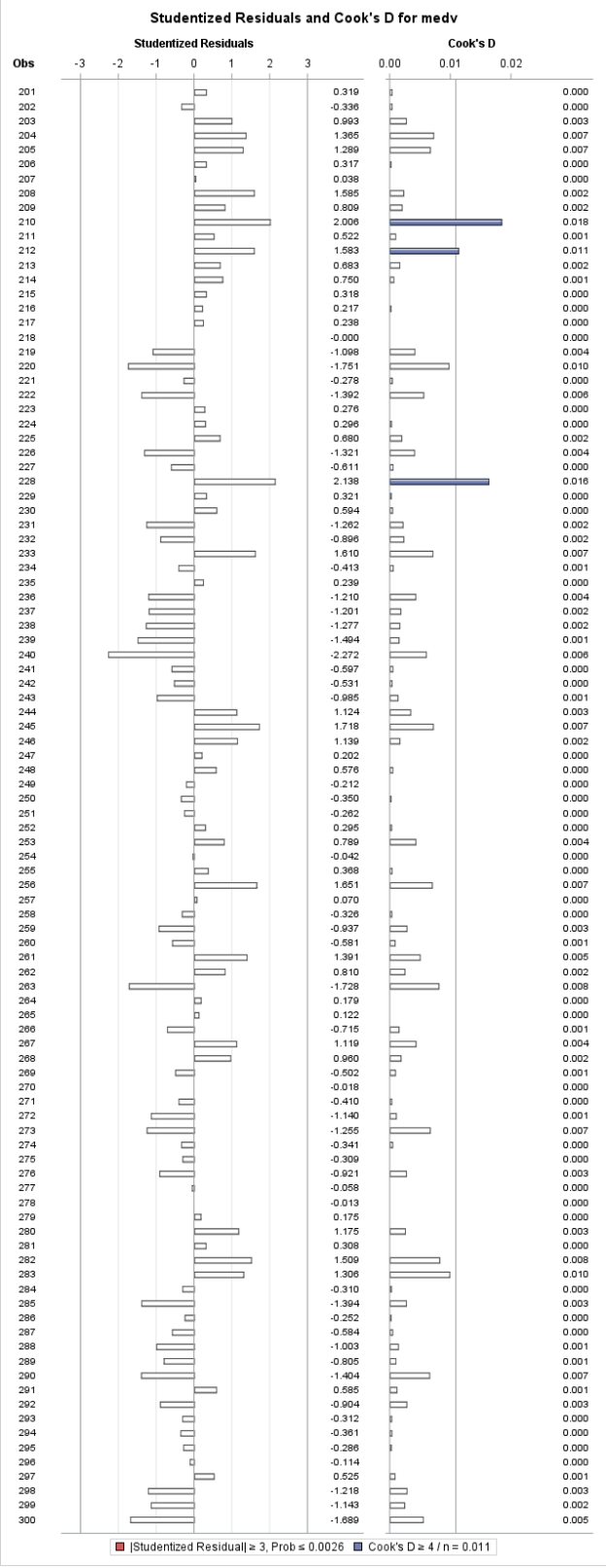
Root MSE	3.25627	**R-Square**	0.8599
Dependent Mean	25.04959	**Adj R-Sq**	0.8547
Coeff Var	12.99928		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standardized Estimate**	**Variance Inflation**
Intercept	**1**	-8.01046	4.57775	-1.75	0.0810	0	0
crime	**1**	1.72577	0.24493	7.05	<.0001	0.25880	3.39914
zn	**1**	0.02307	0.00972	2.37	0.0181	0.07057	2.22582
indus	**1**	0.05365	0.04475	1.20	0.2314	0.03924	2.69961
chas	**1**	1.08971	0.63738	1.71	0.0882	0.03603	1.11915
nox	**1**	-12.53887	3.17136	-3.95	<.0001	-0.15868	4.05816
rm	**1**	8.51415	0.39053	21.80	<.0001	0.68525	2.48915
age	**1**	-0.02453	0.01041	-2.36	0.0190	-0.08189	3.04143
dis	**1**	-0.90588	0.14755	-6.14	<.0001	-0.22487	3.38009
rad	**1**	0.21037	0.08159	2.58	0.0103	0.09407	3.35379
tax	**1**	-0.01464	0.00303	-4.82	<.0001	-0.15767	2.69156
ptratio	**1**	-0.57630	0.09643	-5.98	<.0001	-0.14956	1.57765
minor	**1**	0.01385	0.00477	2.90	0.0039	0.06783	1.37604
lstat	**1**	-0.24693	0.05024	-4.91	<.0001	-0.16980	3.00784

Fig. 17


Studentized Residuals and Cook's D for medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02
1 -1.182 0.003
2 -0.938 0.001
3 0.596 0.000
4 0.829 0.001
5 1.523 0.004
6 1.147 0.002
7 0.605 0.000
8 2.318 0.014
9 1.380 0.011
10 0.176 0.000
11 -1.845 0.011
12 -0.274 0.000
13 0.589 0.001
14 0.329 0.000
15 -0.417 0.000
16 0.401 0.000
17 0.588 0.001
18 -0.202 0.000
19 1.773 0.009
20 0.240 0.000
21 -0.027 0.000
22 0.377 0.000
23 -1.297 0.003
24 -0.316 0.000
25 -0.372 0.000
26 0.269 0.000
27 0.284 0.000
28 -0.630 0.001
29 -1.249 0.003
30 -1.182 0.002
31 -0.337 0.000
32 -1.626 0.005
33 -0.173 0.000
34 -0.630 0.001
35 -1.186 0.006
36 -0.640 0.000
37 0.093 0.000
38 0.108 0.000
39 0.907 0.001
40 0.221 0.000
41 0.132 0.000
42 -1.186 0.003
43 0.110 0.000
44 -0.121 0.000
45 -0.372 0.000
46 -0.170 0.000
47 0.062 0.000
48 -0.725 0.001
49 1.310 0.009
50 0.997 0.001
51 -0.297 0.000
52 -0.596 0.000
53 -0.899 0.001
54 0.203 0.000
55 1.364 0.009
56 0.398 0.000
57 0.227 0.000
58 -0.065 0.000
59 0.453 0.000
60 -0.049 0.000
61 0.676 0.001
62 -0.511 0.001
63 -0.598 0.001
64 -0.300 0.000
65 1.523 0.008
66 -1.150 0.003
67 -0.569 0.001
68 0.744 0.001
69 0.523 0.000
70 0.485 0.000
71 -0.622 0.001
72 0.127 0.000
73 -0.240 0.000
74 -0.384 0.000
75 -0.279 0.000
76 -0.574 0.001
77 -0.623 0.001
78 -0.394 0.000
79 -0.016 0.000
80 0.073 0.000
81 -0.335 0.000
82 -0.888 0.001
83 -0.105 0.000
84 -0.139 0.000
85 -0.356 0.000
86 -0.410 0.000
87 0.325 0.000
88 -0.493 0.000
89 -2.277 0.009
90 -1.038 0.001
91 -0.991 0.001
92 -1.135 0.002
93 -1.512 0.004
94 -0.587 0.001
95 -1.375 0.004
96 0.149 0.000
97 -0.485 0.000
98 -0.892 0.001
99 1.255 0.004
100 -0.591 0.001
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.011


Studentized Residuals and Cook's D for medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.011


Studentized Residuals and Cook's D for medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02
201 0.319 0.000
202 -0.336 0.000
203 0.993 0.003
204 1.365 0.007
205 1.289 0.007
206 0.317 0.000
207 0.038 0.000
208 1.585 0.002
209 0.809 0.002
210 2.006 0.018
211 0.522 0.001
212 1.583 0.011
213 0.683 0.002
214 0.750 0.001
215 0.318 0.000
216 0.217 0.000
217 0.238 0.000
218 -0.000 0.000
219 -1.098 0.004
220 -1.751 0.010
221 -0.278 0.000
222 -1.392 0.006
223 0.276 0.000
224 0.296 0.000
225 0.680 0.002
226 -1.321 0.004
227 -0.611 0.000
228 2.138 0.016
229 0.321 0.000
230 0.594 0.000
231 -1.262 0.002
232 -0.896 0.002
233 1.610 0.007
234 -0.413 0.001
235 0.239 0.000
236 -1.210 0.004
237 -1.201 0.002
238 -1.277 0.002
239 -1.494 0.001
240 -2.272 0.006
241 -0.597 0.000
242 -0.531 0.000
243 -0.985 0.001
244 1.124 0.003
245 1.718 0.007
246 1.139 0.002
247 0.202 0.000
248 0.576 0.000
249 -0.212 0.000
250 -0.350 0.000
251 -0.262 0.000
252 0.295 0.000
253 0.789 0.004
254 -0.042 0.000
255 0.368 0.000
256 1.651 0.007
257 0.070 0.000
258 -0.326 0.000
259 -0.937 0.003
260 -0.581 0.001
261 1.391 0.005
262 0.810 0.002
263 -1.728 0.008
264 0.179 0.000
265 0.122 0.000
266 -0.715 0.001
267 1.119 0.004
268 0.960 0.002
269 -0.502 0.001
270 -0.018 0.000
271 -0.410 0.000
272 -1.140 0.001
273 -1.255 0.007
274 -0.341 0.000
275 -0.309 0.000
276 -0.921 0.003
277 -0.058 0.000
278 -0.013 0.000
279 0.175 0.000
280 1.175 0.003
281 0.308 0.000
282 1.509 0.008
283 1.306 0.010
284 -0.310 0.000
285 -1.394 0.003
286 -0.252 0.000
287 -0.584 0.000
288 -1.003 0.001
289 -0.805 0.001
290 -1.404 0.007
291 0.585 0.001
292 -0.904 0.003
293 -0.312 0.000
294 -0.361 0.000
295 -0.286 0.000
296 -0.114 0.000
297 0.525 0.001
298 -1.218 0.003
299 -1.143 0.002
300 -1.689 0.005
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.011

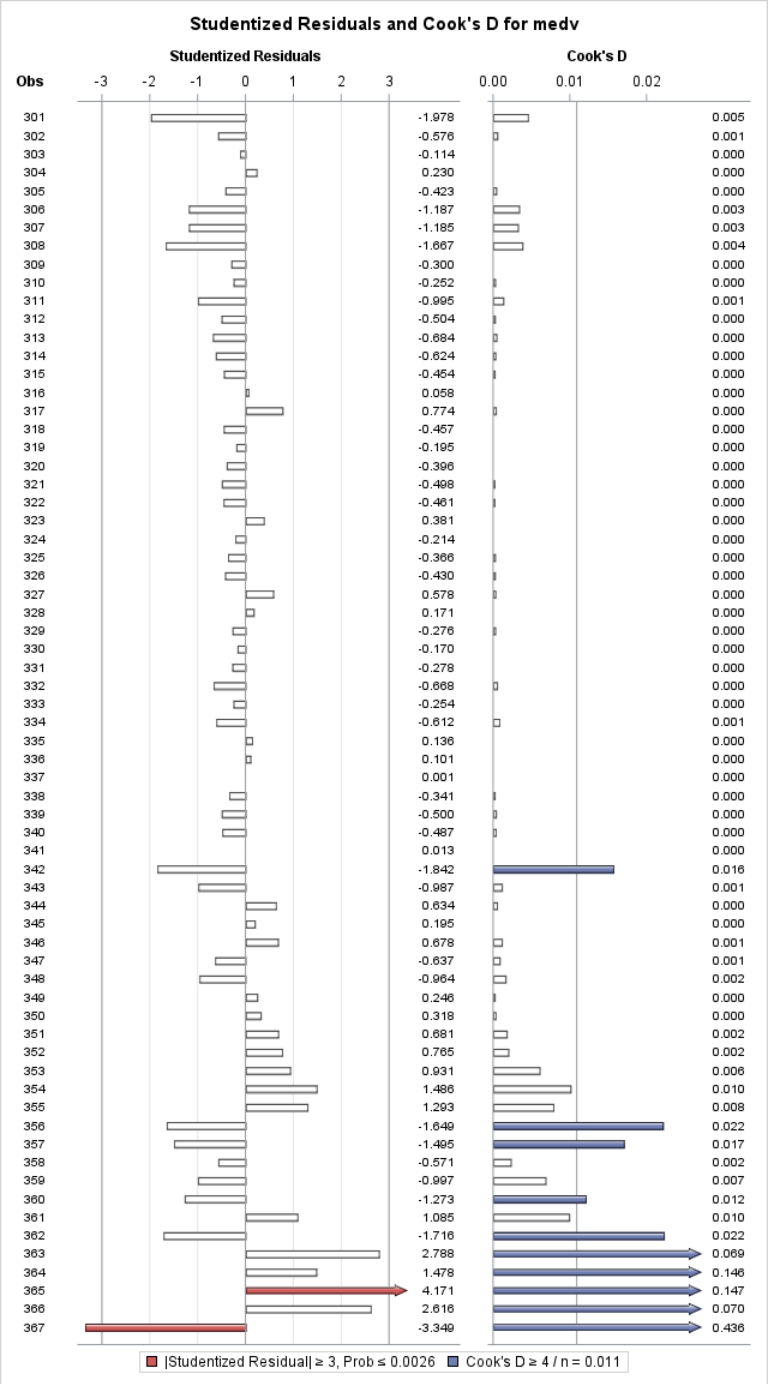

Studentized Residuals and Cook's D for medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02
301 -1.978 0.005
302 -0.576 0.001
303 -0.114 0.000
304 0.230 0.000
305 -0.423 0.000
306 -1.187 0.003
307 -1.185 0.003
308 -1.667 0.004
309 -0.300 0.000
310 -0.252 0.000
311 -0.995 0.001
312 -0.504 0.000
313 -0.684 0.000
314 -0.624 0.000
315 -0.454 0.000
316 0.058 0.000
317 0.774 0.000
318 -0.457 0.000
319 -0.195 0.000
320 -0.396 0.000
321 -0.498 0.000
322 -0.461 0.000
323 0.381 0.000
324 -0.214 0.000
325 -0.366 0.000
326 -0.430 0.000
327 0.578 0.000
328 0.171 0.000
329 -0.276 0.000
330 -0.170 0.000
331 -0.278 0.000
332 -0.668 0.000
333 -0.254 0.000
334 -0.612 0.001
335 0.136 0.000
336 0.101 0.000
337 0.001 0.000
338 -0.341 0.000
339 -0.500 0.000
340 -0.487 0.000
341 0.013 0.000
342 -1.842 0.016
343 -0.987 0.001
344 0.634 0.000
345 0.195 0.000
346 0.678 0.001
347 -0.637 0.001
348 -0.964 0.002
349 0.246 0.000
350 0.318 0.000
351 0.681 0.002
352 0.765 0.002
353 0.931 0.006
354 1.486 0.010
355 1.293 0.008
356 -1.649 0.022
357 -1.495 0.017
358 -0.571 0.002
359 -0.997 0.007
360 -1.273 0.012
361 1.085 0.010
362 -1.716 0.022
363 2.788 0.069
364 1.478 0.146
365 4.171 0.147
366 2.616 0.070
367 -3.349 0.436
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.011

Fig. 18

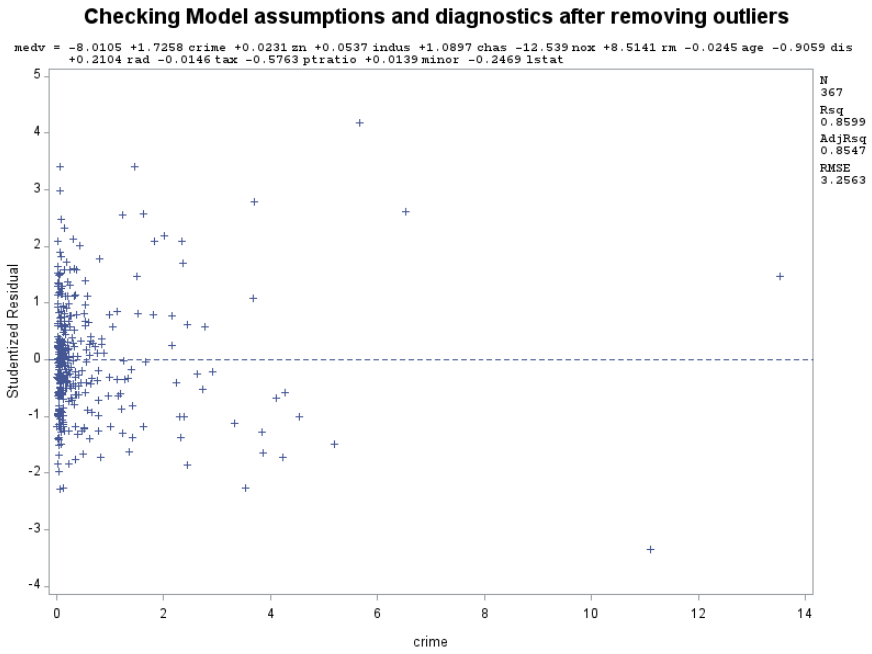

Checking Model assumptions and diagnostics after removing outliers
medv = -8.0105 +1.7258 crime +0.0231 zn +0.0537 indus +1.0897 chas -12.539 nox +8.5141 rm -0.0245 age -0.9059 dis
+0.2104 rad -0.0146 tax -0.5763 ptratio +0.0139 minor -0.2469 lstat
N
367
Rsq
0.8599
AdjRsq
0.8547
RMSE
3.2563
Studentized Residual
crime

Fig. 19

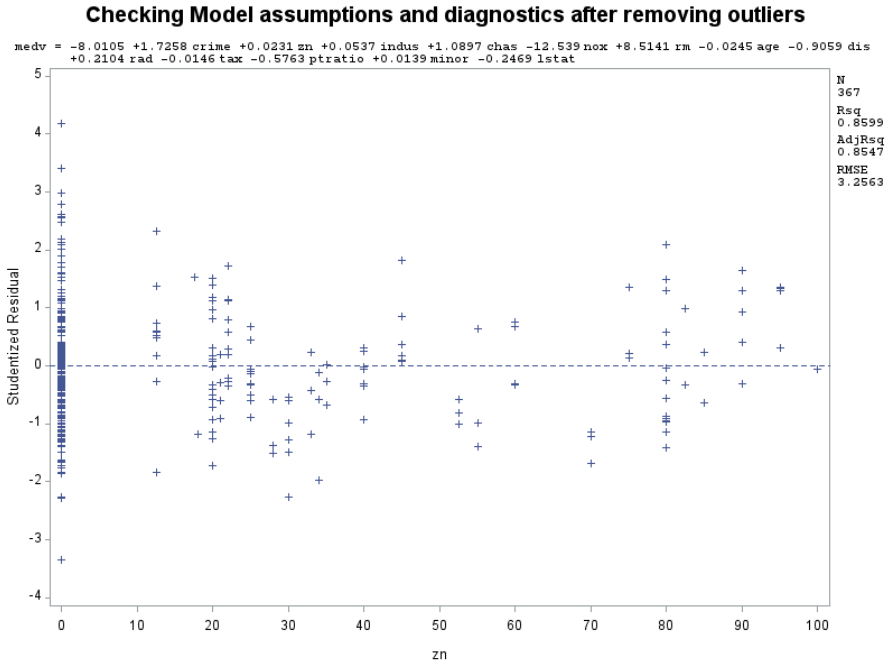

Checking Model assumptions and diagnostics after removing outliers
medv = -8.0105 +1.7258 crime +0.0231 zn +0.0537 indus +1.0897 chas -12.539 nox +8.5141 rm -0.0245 age -0.9059 dis
+0.2104 rad -0.0146 tax -0.5763 ptratio +0.0139 minor -0.2469 lstat
N
367
Rsq
0.8599
AdjRsq
0.8547
RMSE
3.2563
Studentized Residual
zn

Fig. 20

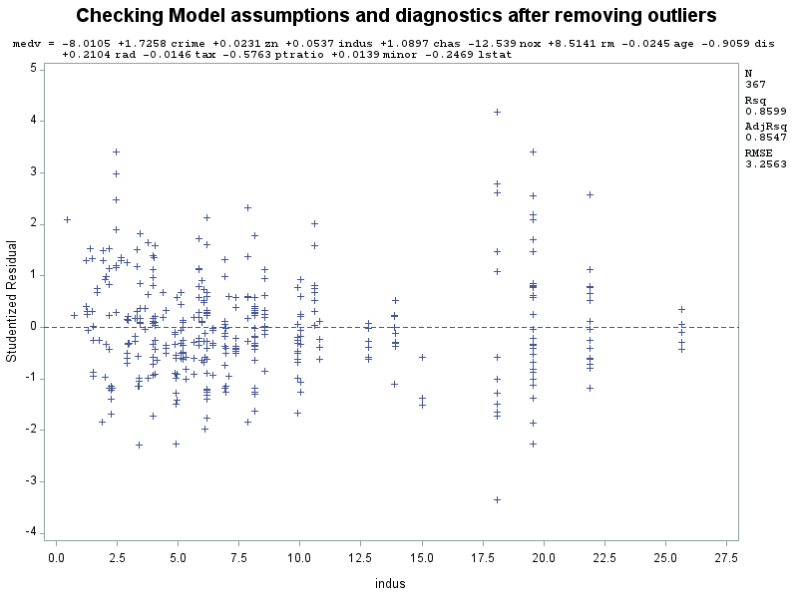

Checking Model assumptions and diagnostics after removing outliers
medv = -8.0105 +1.7258 crime +0.0231 zn +0.0537 indus +1.0897 chas -12.539 nox +8.5141 rm -0.0245 age -0.9059 dis
+0.2104 rad -0.0146 tax -0.5763 ptratio +0.0139 minor -0.2469 lstat
N
367
Rsq
0.8599
AdjRsq
0.8547
RMSE
3.2563
Studentized Residual
indus

Fig. 21

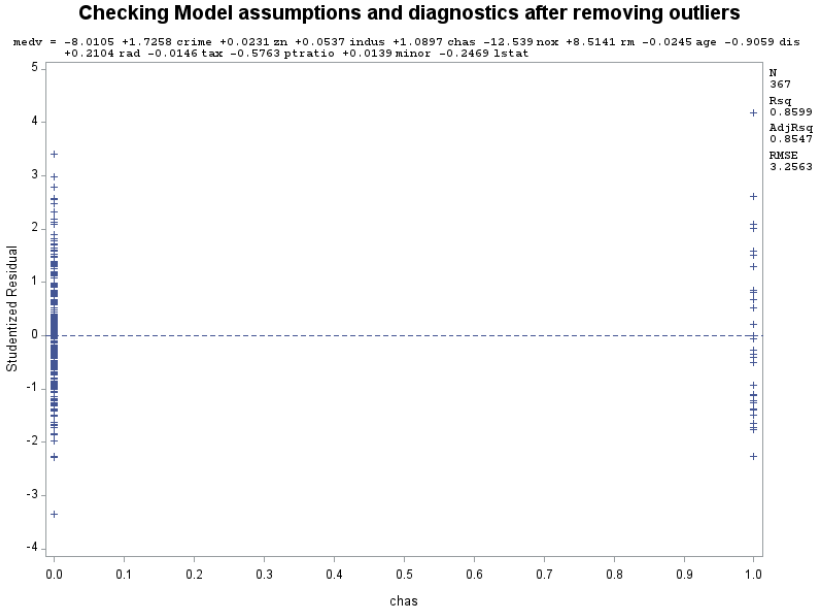

Checking Model assumptions and diagnostics after removing outliers
medv = -8.0105 +1.7258 crime +0.0231 zn +0.0537 indus +1.0897 chas -12.539 nox +8.5141 rm -0.0245 age -0.9059 dis
+0.2104 rad -0.0146 tax -0.5763 ptratio +0.0139 minor -0.2469 lstat
N
367
Rsq
0.8599
AdjRsq
0.8547
RMSE
3.2563
Studentized Residual
chas

Fig. 22

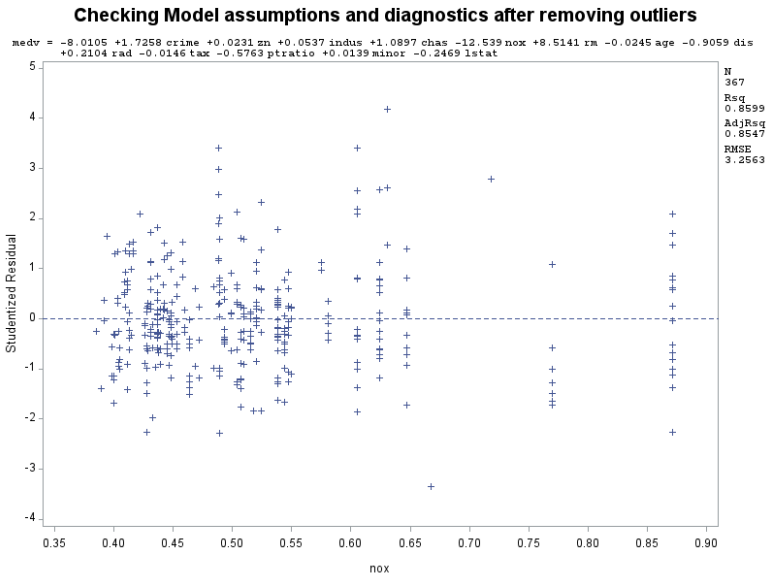

Checking Model assumptions and diagnostics after removing outliers
medv = -8.0105 +1.7258 crime +0.0231 zn +0.0537 indus +1.0897 chas -12.539 nox +8.5141 rm -0.0245 age -0.9059 dis
+0.2104 rad -0.0146 tax -0.5763 ptratio +0.0139 minor -0.2469 lstat
N
367
Rsq
0.8599
AdjRsq
0.8547
RMSE
3.2563
Studentized Residual
nox

Fig. 23

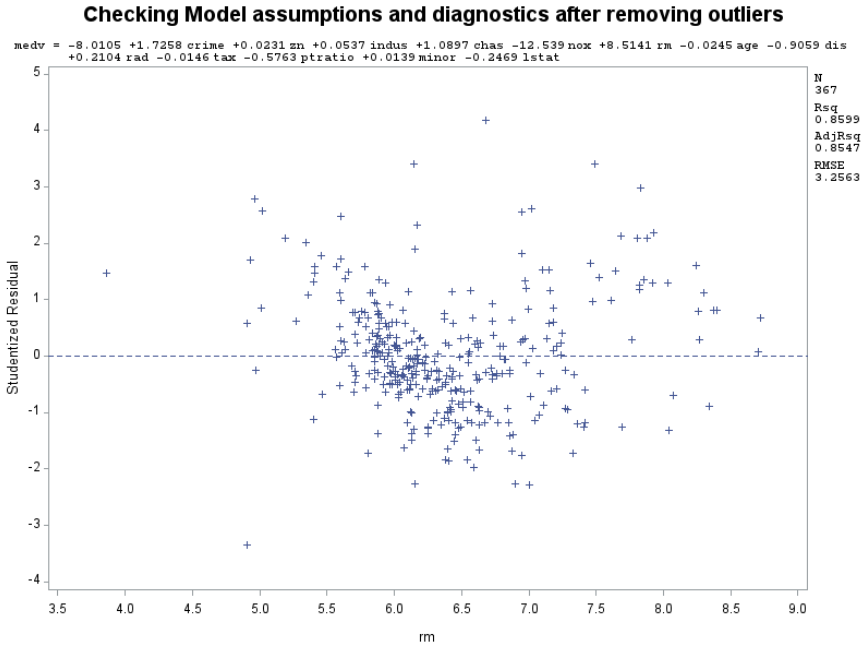

Checking Model assumptions and diagnostics after removing outliers
medv = -8.0105 +1.7258 crime +0.0231 zn +0.0537 indus +1.0897 chas -12.539 nox +8.5141 rm -0.0245 age -0.9059 dis
+0.2104 rad -0.0146 tax -0.5763 ptratio +0.0139 minor -0.2469 lstat
N
367
Rsq
0.8599
AdjRsq
0.8547
RMSE
3.2563
Studentized Residual
rm

Fig. 24

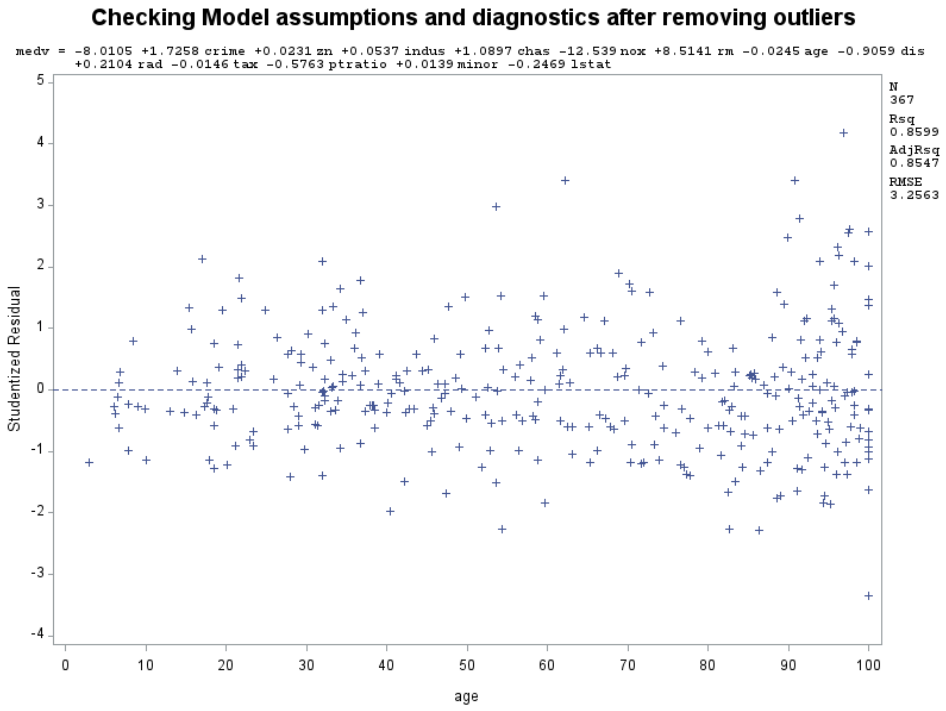

Checking Model assumptions and diagnostics after removing outliers
medv = -8.0105 +1.7258 crime +0.0231 zn +0.0537 indus +1.0897 chas -12.539 nox +8.5141 rm -0.0245 age -0.9059 dis
+0.2104 rad -0.0146 tax -0.5763 ptratio +0.0139 minor -0.2469 lstat
N
367
Rsq
0.8599
AdjRsq
0.8547
RMSE
3.2563
Studentized Residual
age

Fig. 25

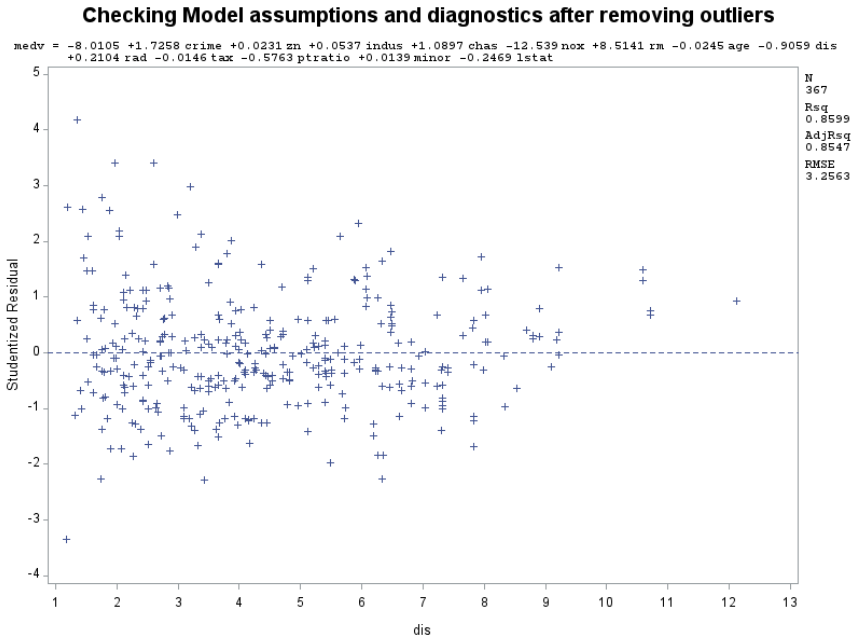

Checking Model assumptions and diagnostics after removing outliers
medv = -8.0105 +1.7258 crime +0.0231 zn +0.0537 indus +1.0897 chas -12.539 nox +8.5141 rm -0.0245 age -0.9059 dis
+0.2104 rad -0.0146 tax -0.5763 ptratio +0.0139 minor -0.2469 lstat
N
367
Rsq
0.8599
AdjRsq
0.8547
RMSE
3.2563
Studentized Residual
dis

Fig. 26

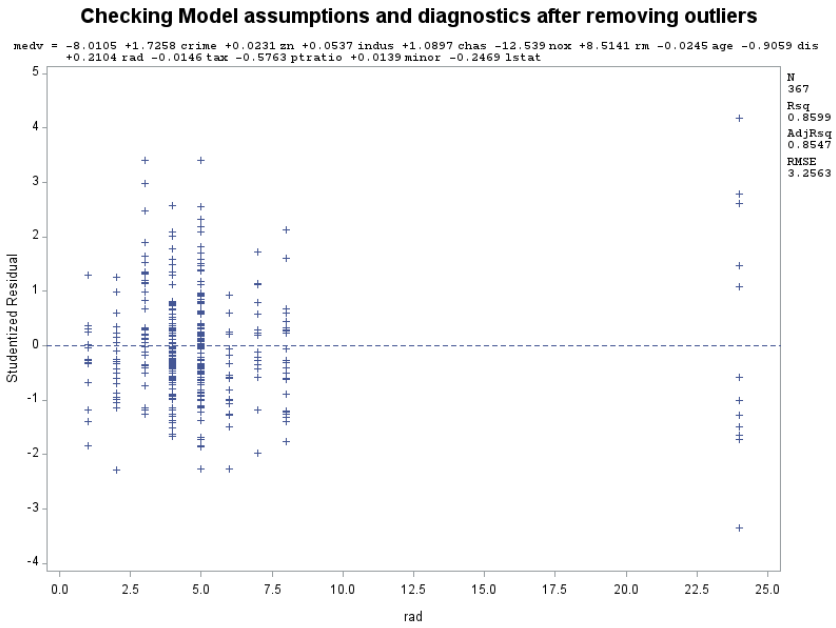

Checking Model assumptions and diagnostics after removing outliers
medv = -8.0105 +1.7258 crime +0.0231 zn +0.0537 indus +1.0897 chas -12.539 nox +8.5141 rm -0.0245 age -0.9059 dis
+0.2104 rad -0.0146 tax -0.5763 ptratio +0.0139 minor -0.2469 lstat
N
367
Rsq
0.8599
AdjRsq
0.8547
RMSE
3.2563
Studentized Residual
rad

Fig. 27

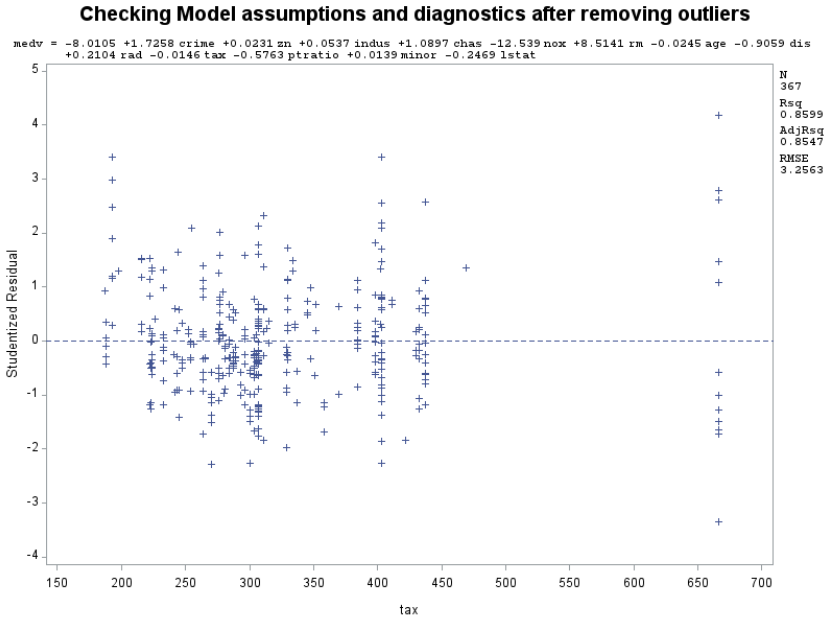

Checking Model assumptions and diagnostics after removing outliers
medv = -8.0105 +1.7258 crime +0.0231 zn +0.0537 indus +1.0897 chas -12.539 nox +8.5141 rm -0.0245 age -0.9059 dis
+0.2104 rad -0.0146 tax -0.5763 ptratio +0.0139 minor -0.2469 lstat
N
367
Rsq
0.8599
AdjRsq
0.8547
RMSE
3.2563
Studentized Residual
tax

Fig. 28

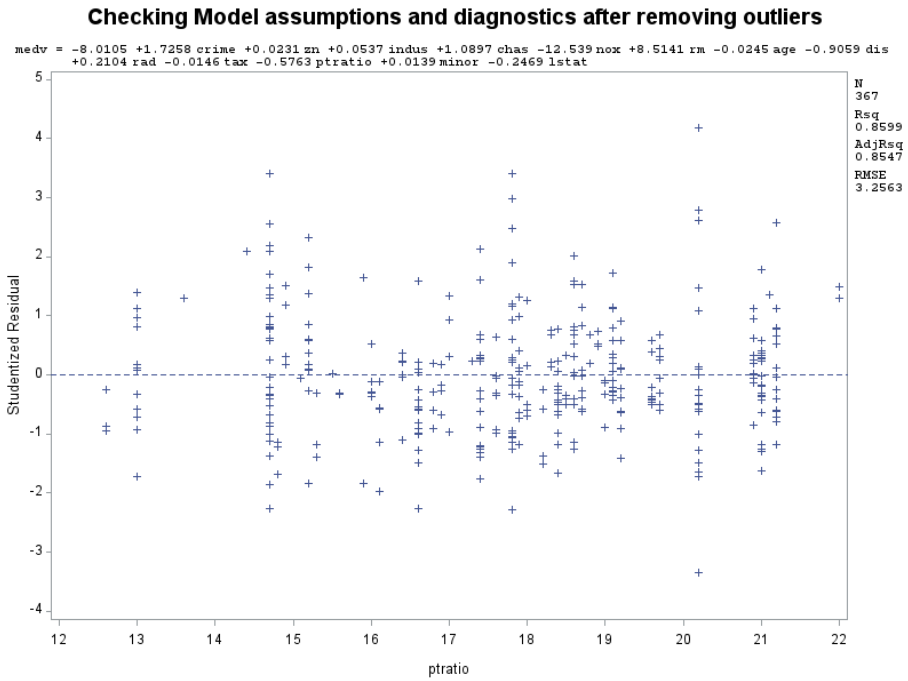

Checking Model assumptions and diagnostics after removing outliers
medv = -8.0105 +1.7258 crime +0.0231 zn +0.0537 indus +1.0897 chas -12.539 nox +8.5141 rm -0.0245 age -0.9059 dis
+0.2104 rad -0.0146 tax -0.5763 ptratio +0.0139 minor -0.2469 lstat
N
367
Rsq
0.8599
AdjRsq
0.8547
RMSE
3.2563
Studentized Residual
ptratio

Fig. 29

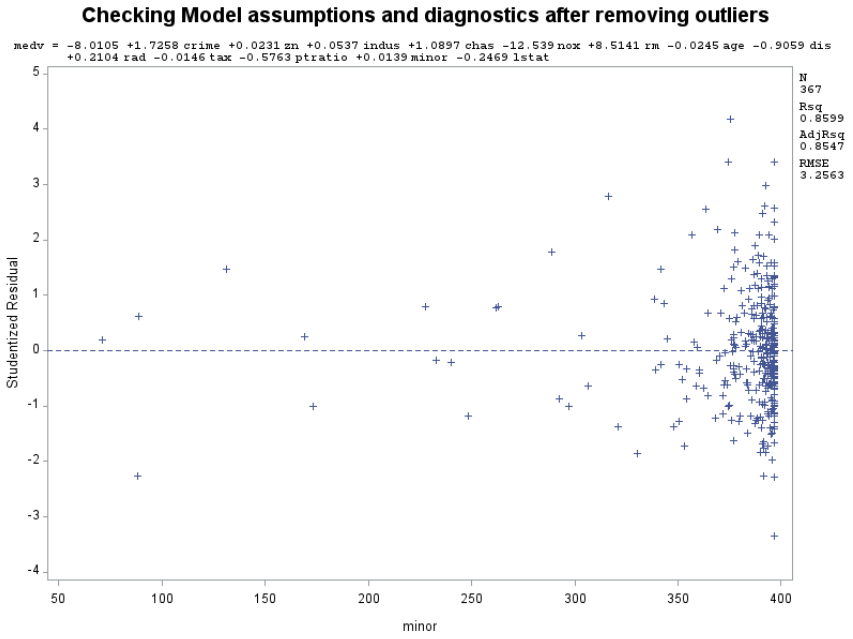

Checking Model assumptions and diagnostics after removing outliers
medv = -8.0105 +1.7258 crime +0.0231 zn +0.0537 indus +1.0897 chas -12.539 nox +8.5141 rm -0.0245 age -0.9059 dis
+0.2104 rad -0.0146 tax -0.5763 ptratio +0.0139 minor -0.2469 lstat
N
367
Rsq
0.8599
AdjRsq
0.8547
RMSE
3.2563
Studentized Residual
minor

Fig. 30

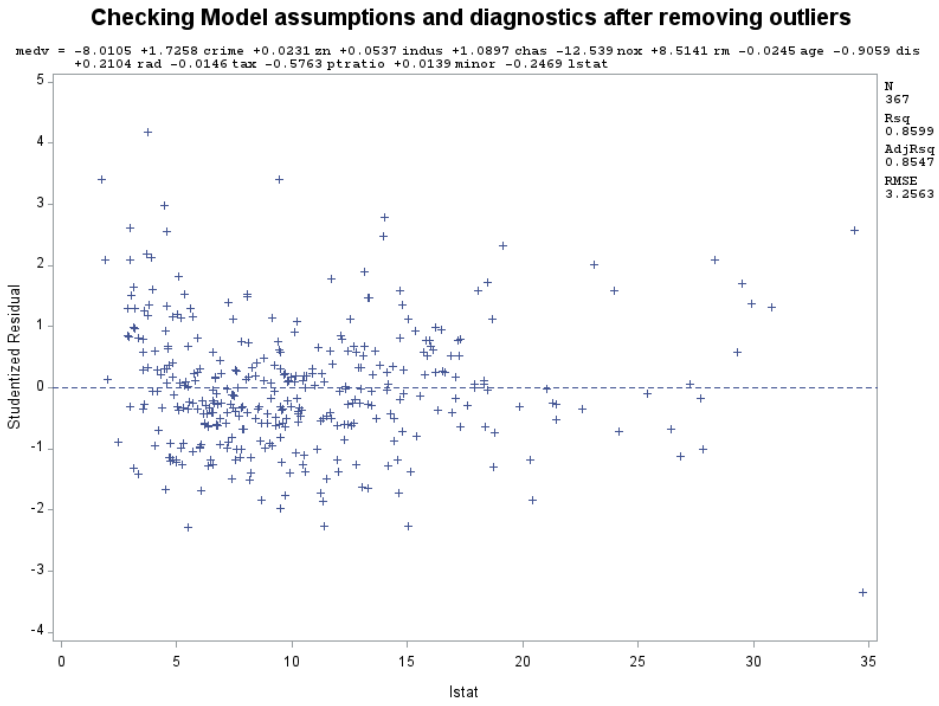

Checking Model assumptions and diagnostics after removing outliers
medv = -8.0105 +1.7258 crime +0.0231 zn +0.0537 indus +1.0897 chas -12.539 nox +8.5141 rm -0.0245 age -0.9059 dis
+0.2104 rad -0.0146 tax -0.5763 ptratio +0.0139 minor -0.2469 lstat
N
367
Rsq
0.8599
AdjRsq
0.8547
RMSE
3.2563
Studentized Residual
lstat

Fig. 31

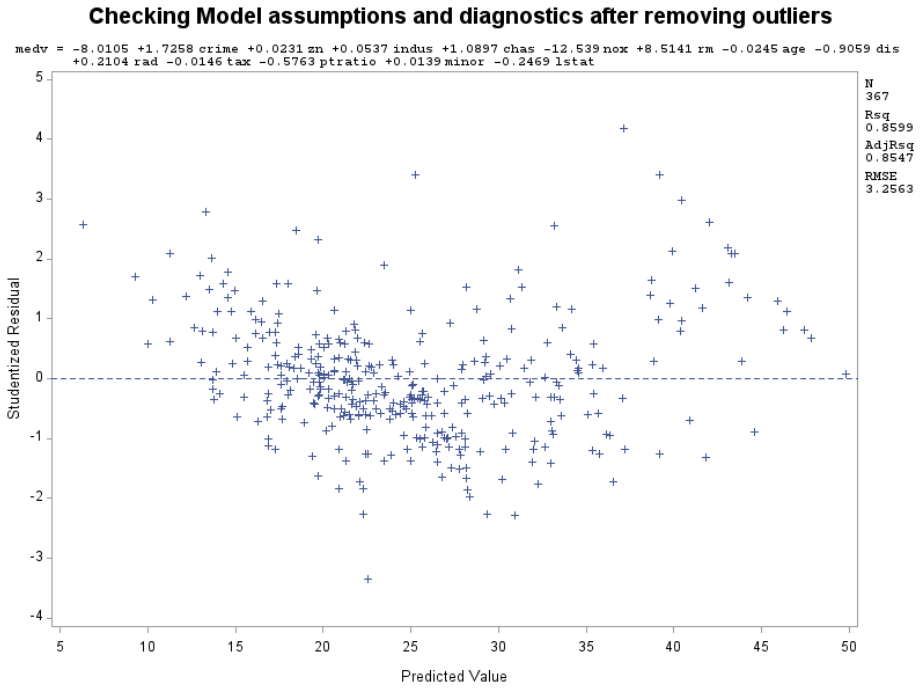
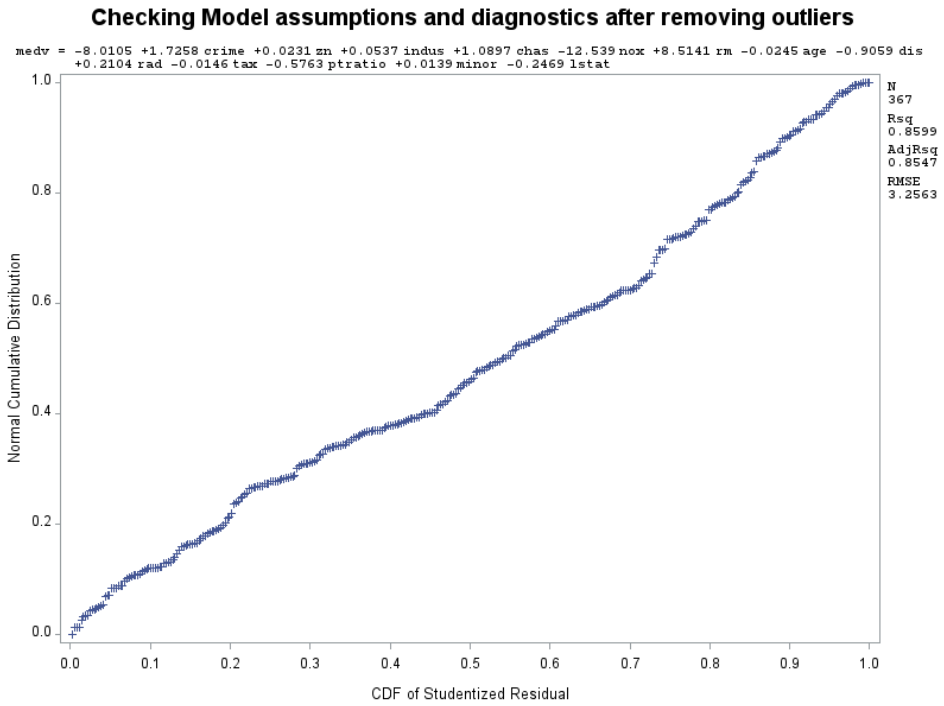

Checking Model assumptions and diagnostics after removing outliers
medv = -8.0105 +1.7258 crime +0.0231 zn +0.0537 indus +1.0897 chas -12.539 nox +8.5141 rm -0.0245 age -0.9059 dis
+0.2104 rad -0.0146 tax -0.5763 ptratio +0.0139 minor -0.2469 lstat
N
367
Rsq
0.8599
AdjRsq
0.8547
RMSE
3.2563
Studentized Residual
Predicted Value

Fig. 32



APPENDIX C – Log transformation

Table 1

Checking Model assumptions and diagnostics after transformation

The REG Procedure
Model: MODEL1
Dependent Variable: logmedv

Number of Observations Read	367
Number of Observations Used	367

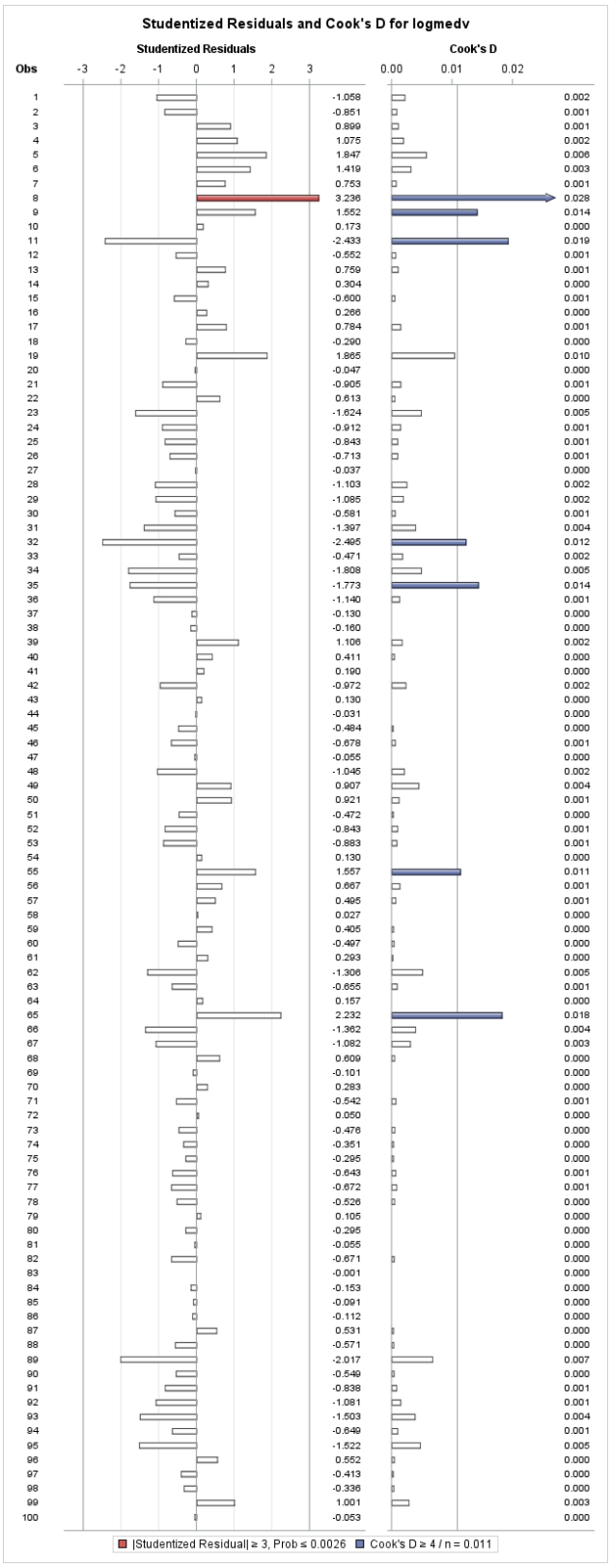
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	13	30.73803	2.36446	194.96	<.0001
Error	353	4.28116	0.01213		
Corrected Total	366	35.01918			

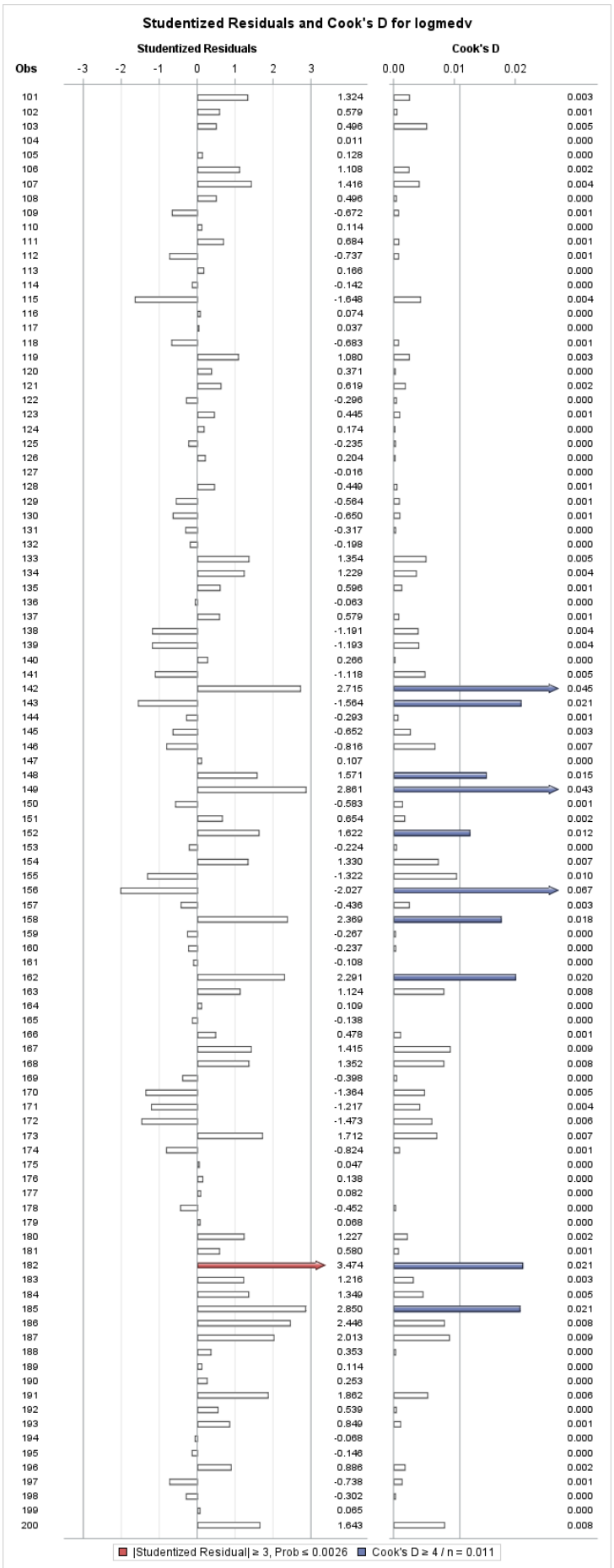
Root MSE	0.11013	**R-Square**	0.8777
Dependent Mean	3.17054	**Adj R-Sq**	0.8732

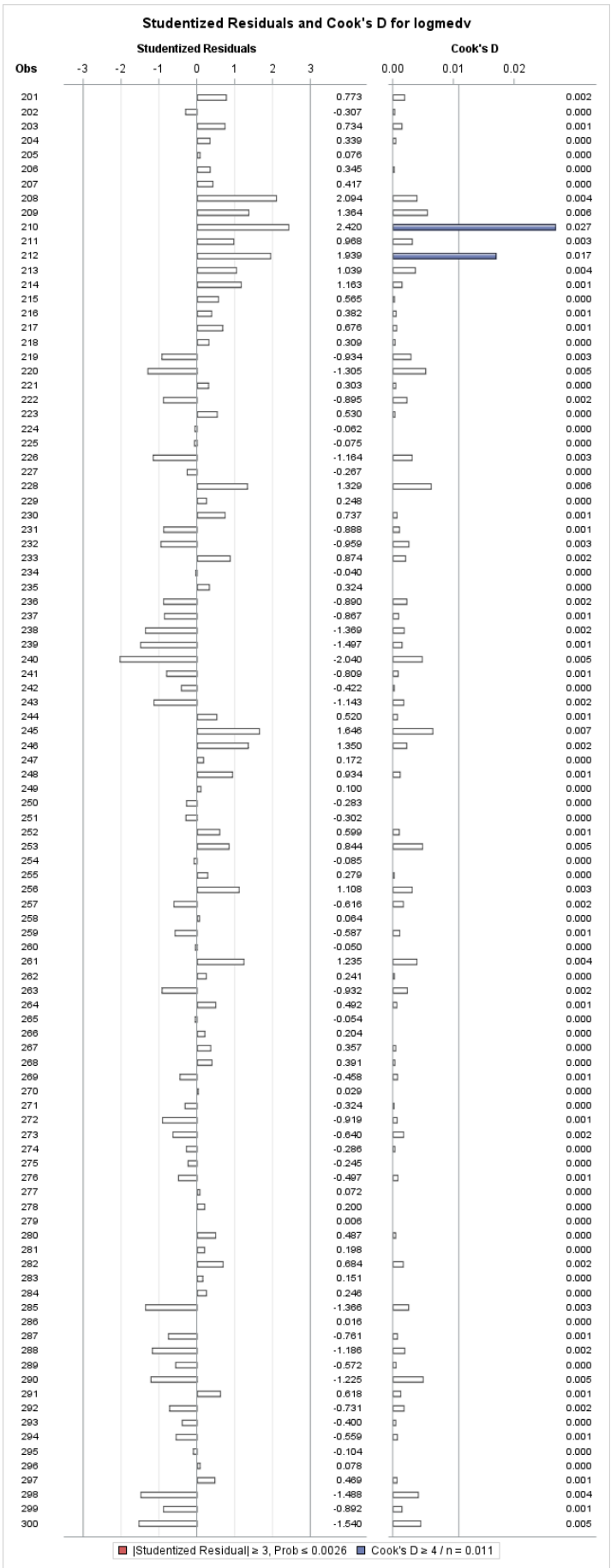
Coeff Var	3.47344		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standardized Estimate**	**Variance Inflation**
Intercept	**1**	2.51620	0.15482	16.25	<.0001	0	0
crime	**1**	0.04567	0.00828	5.51	<.0001	0.18916	3.39914
zn	**1**	0.00074642	0.00032864	2.27	0.0237	0.06306	2.22582
indus	**1**	0.00227	0.00151	1.50	0.1353	0.04577	2.69961
chas	**1**	0.03477	0.02156	1.61	0.1077	0.03175	1.11915
nox	**1**	-0.57152	0.10726	-5.33	<.0001	-0.19976	4.05816
rm	**1**	0.25083	0.01321	18.99	<.0001	0.55760	2.48915
age	**1**	-0.00092026	0.00035193	-2.61	0.0093	-0.08486	3.04143
dis	**1**	-0.03512	0.00499	-7.04	<.0001	-0.24076	3.38009
rad	**1**	0.01240	0.00276	4.49	<.0001	0.15311	3.35379
tax	**1**	-0.00057326	0.00010264	-5.58	<.0001	-0.17051	2.69156
ptratio	**1**	-0.02534	0.00326	-7.77	<.0001	-0.18166	1.57765
minor	**1**	0.00060553	0.00016140	3.75	0.0002	0.08190	1.37604
lstat	**1**	-0.01516	0.00170	-8.92	<.0001	-0.28803	3.00784

Fig. 1


Studentized Residuals and Cook's D for logmedv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02
1 -1.058 0.002
2 -0.851 0.001
3 0.899 0.001
4 1.075 0.002
5 1.847 0.006
6 1.419 0.003
7 0.753 0.001
8 3.236 0.028
9 1.552 0.014
10 0.173 0.000
11 -2.433 0.019
12 -0.552 0.001
13 0.759 0.001
14 0.304 0.000
15 -0.600 0.001
16 0.266 0.000
17 0.784 0.001
18 -0.290 0.000
19 1.865 0.010
20 -0.047 0.000
21 -0.905 0.001
22 0.613 0.000
23 -1.624 0.005
24 -0.912 0.001
25 -0.843 0.001
26 -0.713 0.001
27 -0.037 0.000
28 -1.103 0.002
29 -1.085 0.002
30 -0.581 0.001
31 -1.397 0.004
32 -2.495 0.012
33 -0.471 0.002
34 -1.808 0.005
35 -1.773 0.014
36 -1.140 0.001
37 -0.130 0.000
38 -0.160 0.000
39 1.106 0.002
40 0.411 0.000
41 0.190 0.000
42 -0.972 0.002
43 0.130 0.000
44 -0.031 0.000
45 -0.484 0.000
46 -0.678 0.001
47 -0.055 0.000
48 -1.045 0.002
49 0.907 0.004
50 0.921 0.001
51 -0.472 0.000
52 -0.843 0.001
53 -0.883 0.001
54 0.130 0.000
55 1.557 0.011
56 0.667 0.001
57 0.495 0.001
58 0.027 0.000
59 0.405 0.000
60 -0.497 0.000
61 0.293 0.000
62 -1.306 0.005
63 -0.655 0.001
64 0.157 0.000
65 2.232 0.018
66 -1.362 0.004
67 -1.082 0.003
68 0.609 0.000
69 -0.101 0.000
70 0.283 0.000
71 -0.542 0.001
72 0.050 0.000
73 -0.476 0.000
74 -0.351 0.000
75 -0.295 0.000
76 -0.643 0.001
77 -0.672 0.001
78 -0.526 0.000
79 0.105 0.000
80 -0.295 0.000
81 -0.055 0.000
82 -0.671 0.000
83 -0.001 0.000
84 -0.153 0.000
85 -0.091 0.000
86 -0.112 0.000
87 0.531 0.000
88 -0.571 0.000
89 -2.017 0.007
90 -0.549 0.000
91 -0.838 0.001
92 -1.081 0.001
93 -1.503 0.004
94 -0.649 0.001
95 -1.522 0.005
96 0.552 0.000
97 -0.413 0.000
98 -0.336 0.000
99 1.001 0.003
100 -0.053 0.000
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.011


Studentized Residuals and Cook's D for logmedv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02
101 1.324 0.003
102 0.579 0.001
103 0.496 0.005
104 0.011 0.000
105 0.128 0.000
106 1.108 0.002
107 1.416 0.004
108 0.496 0.000
109 -0.672 0.001
110 0.114 0.000
111 0.684 0.001
112 -0.737 0.001
113 0.166 0.000
114 -0.142 0.000
115 -1.648 0.004
116 0.074 0.000
117 0.037 0.000
118 -0.683 0.001
119 1.080 0.003
120 0.371 0.000
121 0.619 0.002
122 -0.296 0.000
123 0.445 0.001
124 0.174 0.000
125 -0.235 0.000
126 0.204 0.000
127 -0.016 0.000
128 0.449 0.001
129 -0.564 0.001
130 -0.650 0.001
131 -0.317 0.000
132 -0.198 0.000
133 1.354 0.005
134 1.229 0.004
135 0.596 0.001
136 -0.063 0.000
137 0.579 0.001
138 -1.191 0.004
139 -1.193 0.004
140 0.266 0.000
141 -1.118 0.005
142 2.715 0.045
143 -1.564 0.021
144 -0.293 0.001
145 -0.652 0.003
146 -0.816 0.007
147 0.107 0.000
148 1.571 0.015
149 2.861 0.043
150 -0.583 0.001
151 0.654 0.002
152 1.622 0.012
153 -0.224 0.000
154 1.330 0.007
155 -1.322 0.010
156 -2.027 0.067
157 -0.436 0.003
158 2.369 0.018
159 -0.267 0.000
160 -0.237 0.000
161 -0.108 0.000
162 2.291 0.020
163 1.124 0.008
164 0.109 0.000
165 -0.138 0.000
166 0.478 0.001
167 1.415 0.009
168 1.352 0.008
169 -0.398 0.000
170 -1.364 0.005
171 -1.217 0.004
172 -1.473 0.006
173 1.712 0.007
174 -0.824 0.001
175 0.047 0.000
176 0.138 0.000
177 0.082 0.000
178 -0.452 0.000
179 0.068 0.000
180 1.227 0.002
181 0.580 0.001
182 3.474 0.021
183 1.216 0.003
184 1.349 0.005
185 2.850 0.021
186 2.446 0.008
187 2.013 0.009
188 0.353 0.000
189 0.114 0.000
190 0.253 0.000
191 1.862 0.006
192 0.539 0.000
193 0.849 0.001
194 -0.068 0.000
195 -0.146 0.000
196 0.886 0.002
197 -0.738 0.001
198 -0.302 0.000
199 0.065 0.000
200 1.643 0.008
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.011


Studentized Residuals and Cook's D for logmedv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02
201 0.773 0.002
202 -0.307 0.000
203 0.734 0.001
204 0.339 0.000
205 0.076 0.000
206 0.345 0.000
207 0.417 0.000
208 2.094 0.004
209 1.364 0.006
210 2.420 0.027
211 0.968 0.003
212 1.939 0.017
213 1.039 0.004
214 1.163 0.001
215 0.565 0.000
216 0.382 0.001
217 0.676 0.001
218 0.309 0.000
219 -0.934 0.003
220 -1.305 0.005
221 0.303 0.000
222 -0.895 0.002
223 0.530 0.000
224 -0.062 0.000
225 -0.075 0.000
226 -1.164 0.003
227 -0.267 0.000
228 1.329 0.006
229 0.248 0.000
230 0.737 0.001
231 -0.888 0.001
232 -0.959 0.003
233 0.874 0.002
234 -0.040 0.000
235 0.324 0.000
236 -0.890 0.002
237 -0.867 0.001
238 -1.369 0.002
239 -1.497 0.001
240 -2.040 0.005
241 -0.809 0.001
242 -0.422 0.000
243 -1.143 0.002
244 0.520 0.001
245 1.646 0.007
246 1.350 0.002
247 0.172 0.000
248 0.934 0.001
249 0.100 0.000
250 -0.283 0.000
251 -0.302 0.000
252 0.599 0.001
253 0.844 0.005
254 -0.085 0.000
255 0.279 0.000
256 1.108 0.003
257 -0.616 0.002
258 0.064 0.000
259 -0.587 0.001
260 -0.050 0.000
261 1.235 0.004
262 0.241 0.000
263 -0.932 0.002
264 0.492 0.001
265 -0.054 0.000
266 0.204 0.000
267 0.357 0.000
268 0.391 0.000
269 -0.458 0.001
270 0.029 0.000
271 -0.324 0.000
272 -0.919 0.001
273 -0.640 0.002
274 -0.286 0.000
275 -0.245 0.000
276 -0.497 0.001
277 0.072 0.000
278 0.200 0.000
279 0.006 0.000
280 0.487 0.000
281 0.198 0.000
282 0.684 0.002
283 0.151 0.000
284 0.246 0.000
285 -1.366 0.003
286 0.016 0.000
287 -0.761 0.001
288 -1.186 0.002
289 -0.572 0.000
290 -1.225 0.005
291 0.618 0.001
292 -0.731 0.002
293 -0.400 0.000
294 -0.559 0.001
295 -0.104 0.000
296 0.078 0.000
297 0.469 0.001
298 -1.488 0.004
299 -0.892 0.001
300 -1.540 0.005
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.011

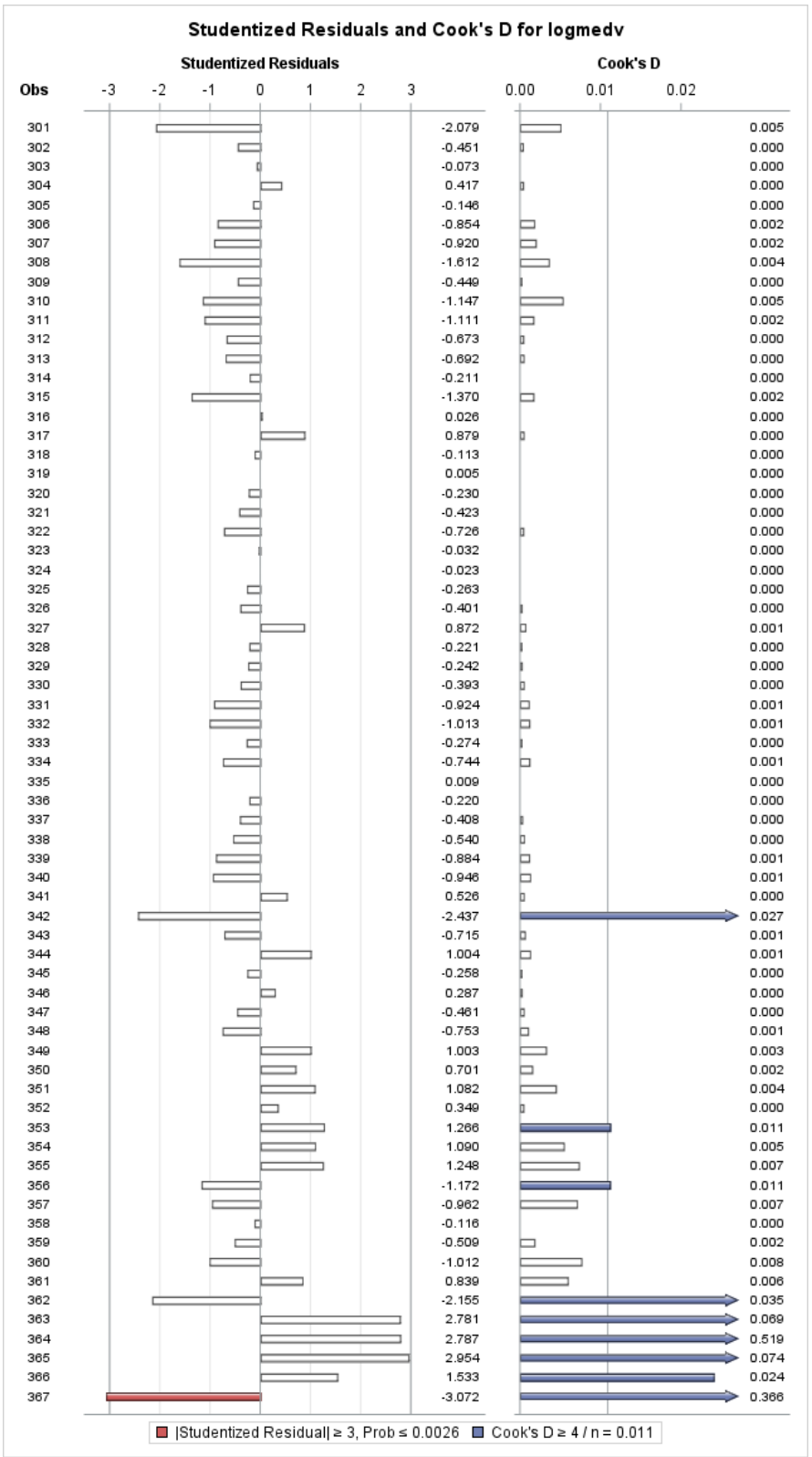

Studentized Residuals and Cook's D for logmedv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02
301 -2.079 0.005
302 -0.451 0.000
303 -0.073 0.000
304 0.417 0.000
305 -0.146 0.000
306 -0.854 0.002
307 -0.920 0.002
308 -1.612 0.004
309 -0.449 0.000
310 -1.147 0.005
311 -1.111 0.002
312 -0.673 0.000
313 -0.692 0.000
314 -0.211 0.000
315 -1.370 0.002
316 0.026 0.000
317 0.879 0.000
318 -0.113 0.000
319 0.005 0.000
320 -0.230 0.000
321 -0.423 0.000
322 -0.726 0.000
323 -0.032 0.000
324 -0.023 0.000
325 -0.263 0.000
326 -0.401 0.000
327 0.872 0.001
328 -0.221 0.000
329 -0.242 0.000
330 -0.393 0.000
331 -0.924 0.001
332 -1.013 0.001
333 -0.274 0.000
334 -0.744 0.001
335 0.009 0.000
336 -0.220 0.000
337 -0.408 0.000
338 -0.540 0.000
339 -0.884 0.001
340 -0.946 0.001
341 0.526 0.000
342 -2.437 0.027
343 -0.715 0.001
344 1.004 0.001
345 -0.258 0.000
346 0.287 0.000
347 -0.461 0.000
348 -0.753 0.001
349 1.003 0.003
350 0.701 0.002
351 1.082 0.004
352 0.349 0.000
353 1.266 0.011
354 1.090 0.005
355 1.248 0.007
356 -1.172 0.011
357 -0.962 0.007
358 -0.116 0.000
359 -0.509 0.002
360 -1.012 0.008
361 0.839 0.006
362 -2.155 0.035
363 2.781 0.069
364 2.787 0.519
365 2.954 0.074
366 1.533 0.024
367 -3.072 0.366
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.011

Fig. 2

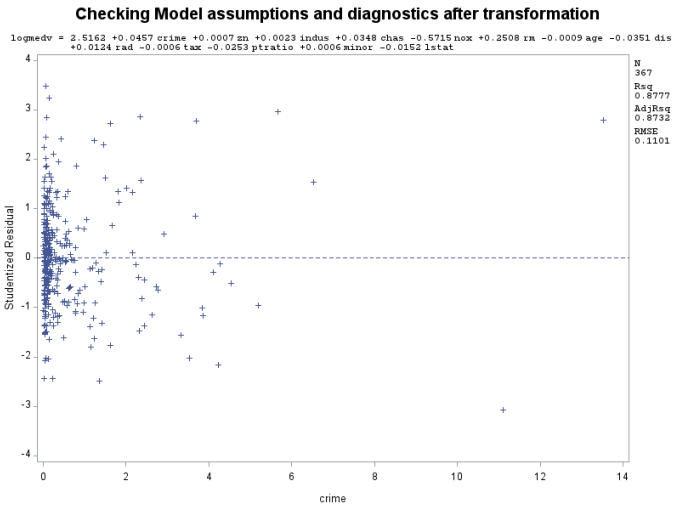

Checking Model assumptions and diagnostics after transformation
logmedv = 2.5162 +0.0457 crime +0.0007 zn +0.0023 indus +0.0348 chas -0.5715 nox +0.2508 rm -0.0009 age -0.0351 dis
+0.0124 rad -0.0006 tax -0.0253 ptratio +0.0006 minor -0.0152 lstat
N
367
Rsq
0.8777
AdjRsq
0.8732
RMSE
0.1101
Studentized Residual
crime

Fig. 3

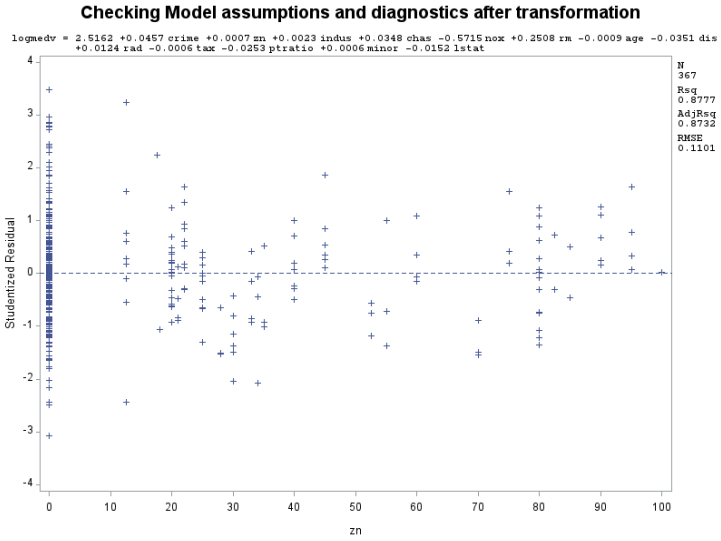

Checking Model assumptions and diagnostics after transformation
logmedv = 2.5162 +0.0457 crime +0.0007 zn +0.0023 indus +0.0348 chas -0.5715 nox +0.2508 rm -0.0009 age -0.0351 dis
+0.0124 rad -0.0006 tax -0.0253 ptratio +0.0006 minor -0.0152 lstat
N
367
Rsq
0.8777
AdjRsq
0.8732
RMSE
0.1101
Studentized Residual
zn

Fig. 4

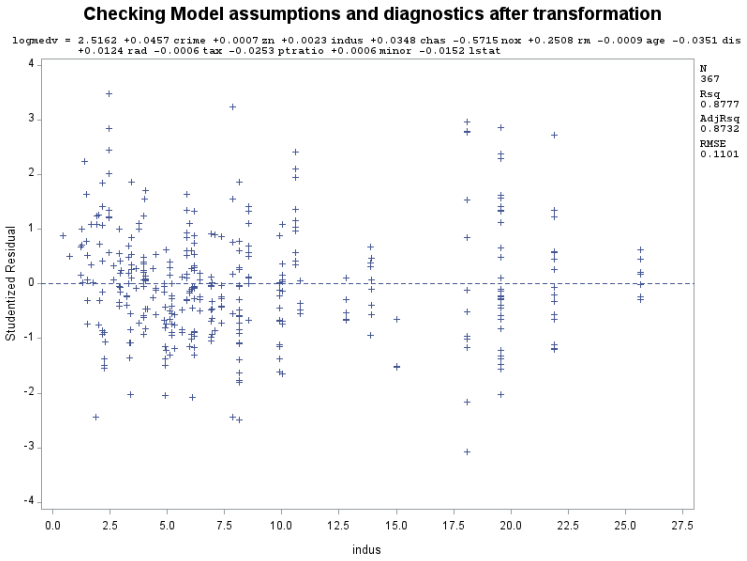

Checking Model assumptions and diagnostics after transformation
logmedv = 2.5162 +0.0457 crime +0.0007 zn +0.0023 indus +0.0348 chas -0.5715 nox +0.2508 rm -0.0009 age -0.0351 dis
+0.0124 rad -0.0006 tax -0.0253 ptratio +0.0006 minor -0.0152 lstat
N
367
Rsq
0.8777
AdjRsq
0.8732
RMSE
0.1101
Studentized Residual
indus

Fig. 5

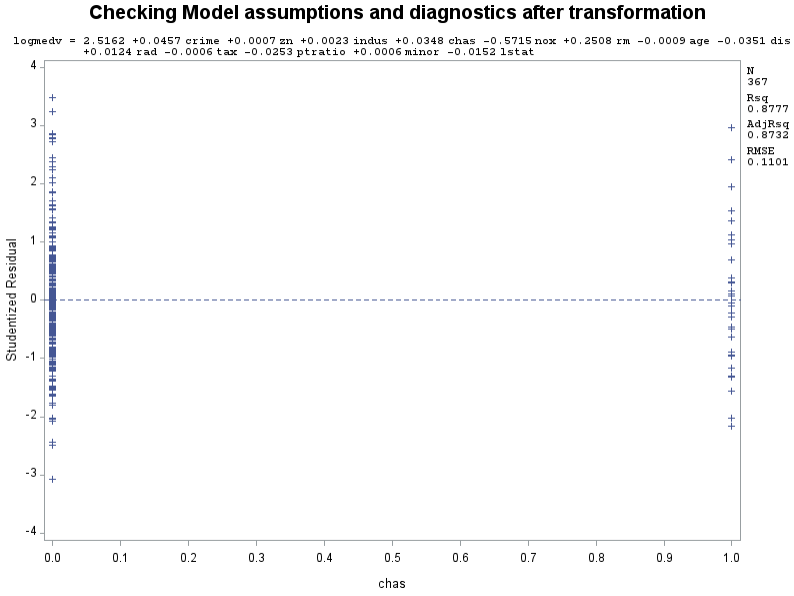

Checking Model assumptions and diagnostics after transformation
logmedv = 2.5162 +0.0457 crime +0.0007 zn +0.0023 indus +0.0348 chas -0.5715 nox +0.2508 rm -0.0009 age -0.0351 dis
+0.0124 rad -0.0006 tax -0.0253 ptratio +0.0006 minor -0.0152 lstat
N
367
Rsq
0.8777
AdjRsq
0.8732
RMSE
0.1101
Studentized Residual
chas

Fig. 6

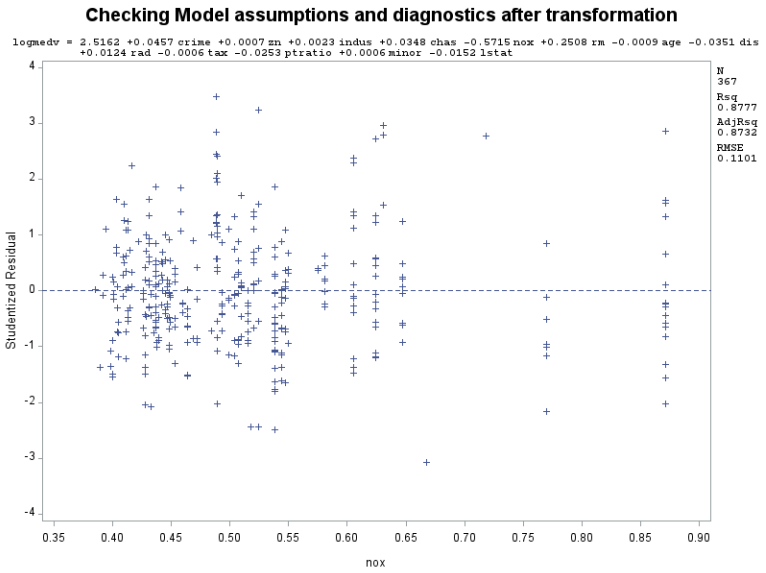

Checking Model assumptions and diagnostics after transformation
logmedv = 2.5162 +0.0457 crime +0.0007 zn +0.0023 indus +0.0348 chas -0.5715 nox +0.2508 rm -0.0009 age -0.0351 dis
+0.0124 rad -0.0006 tax -0.0253 ptratio +0.0006 minor -0.0152 lstat
N
367
Rsq
0.8777
AdjRsq
0.8732
RMSE
0.1101
Studentized Residual
nox

Fig. 7

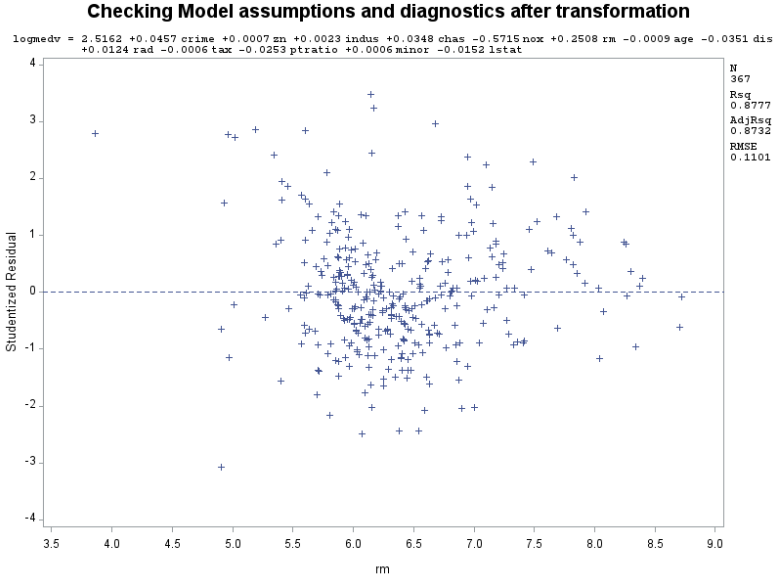

Checking Model assumptions and diagnostics after transformation
logmedv = 2.5162 +0.0457 crime +0.0007 zn +0.0023 indus +0.0348 chas -0.5715 nox +0.2508 rm -0.0009 age -0.0351 dis
+0.0124 rad -0.0006 tax -0.0253 ptratio +0.0006 minor -0.0152 lstat
N
367
Rsq
0.8777
AdjRsq
0.8732
RMSE
0.1101
Studentized Residual
rm

Fig. 8

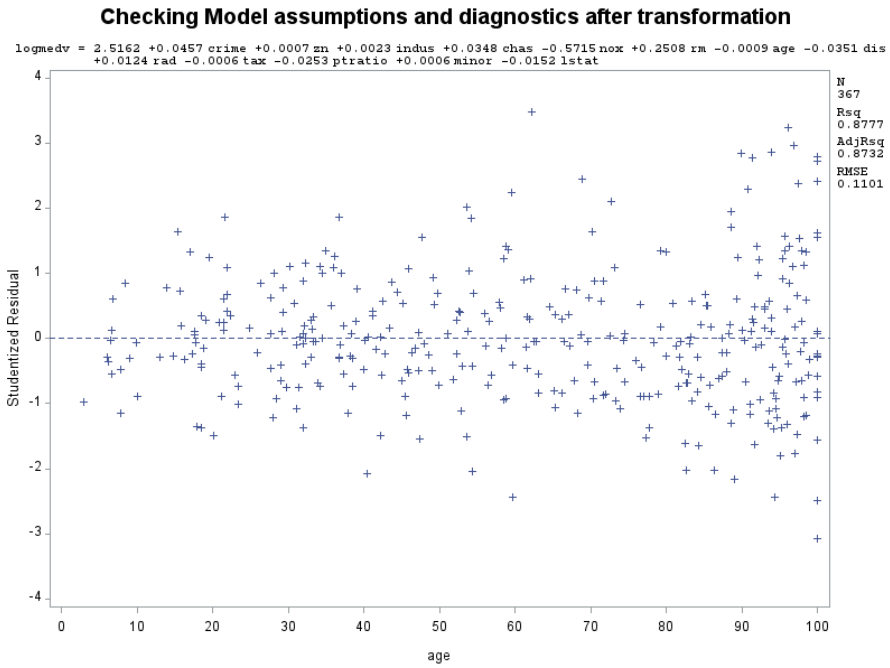

Checking Model assumptions and diagnostics after transformation
logmedv = 2.5162 +0.0457 crime +0.0007 zn +0.0023 indus +0.0348 chas -0.5715 nox +0.2508 rm -0.0009 age -0.0351 dis
+0.0124 rad -0.0006 tax -0.0253 ptratio +0.0006 minor -0.0152 lstat
N
367
Rsq
0.8777
AdjRsq
0.8732
RMSE
0.1101
Studentized Residual
age

Fig. 9

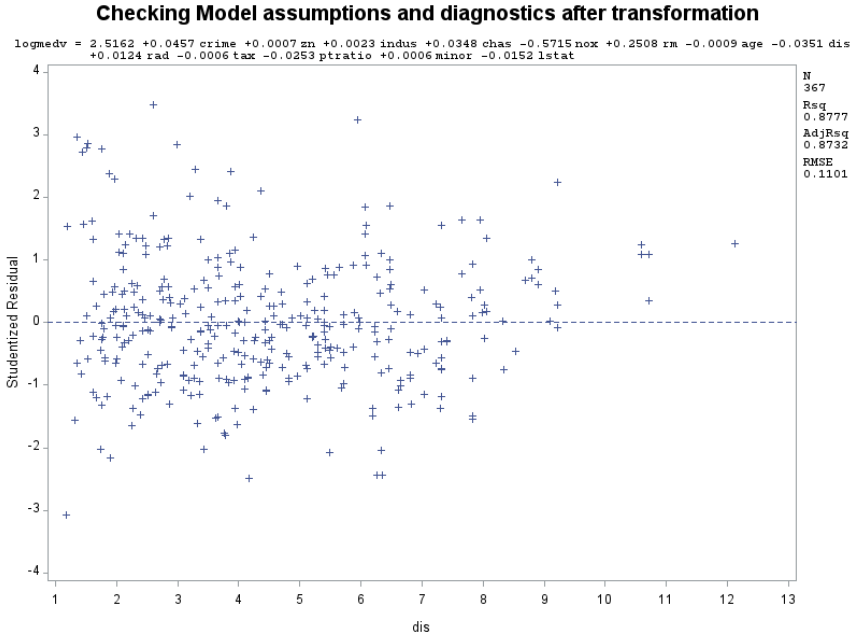

Checking Model assumptions and diagnostics after transformation
logmedv = 2.5162 +0.0457 crime +0.0007 zn +0.0023 indus +0.0348 chas -0.5715 nox +0.2508 rm -0.0009 age -0.0351 dis
+0.0124 rad -0.0006 tax -0.0253 ptratio +0.0006 minor -0.0152 lstat
N
367
Rsq
0.8777
AdjRsq
0.8732
RMSE
0.1101
Studentized Residual
dis

Fig. 10

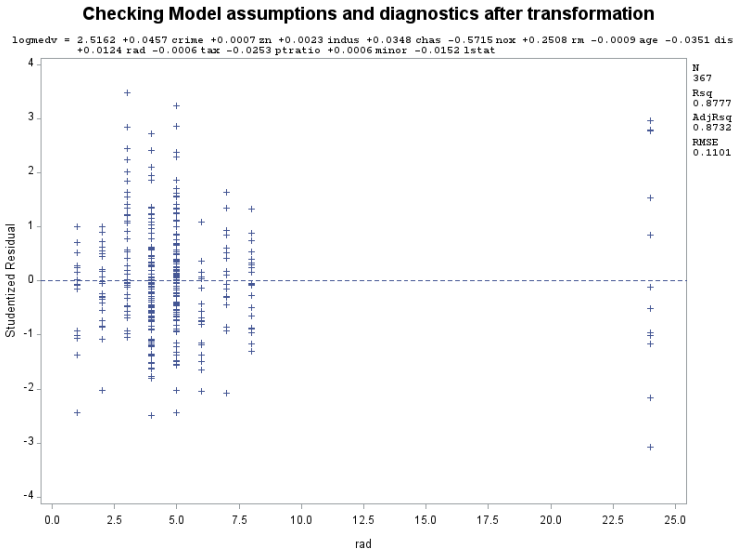

Checking Model assumptions and diagnostics after transformation
logmedv = 2.5162 +0.0457 crime +0.0007 zn +0.0023 indus +0.0348 chas -0.5715 nox +0.2508 rm -0.0009 age -0.0351 dis +0.0124 rad -0.0006 tax -0.0253 ptratio +0.0006 minor -0.0152 lstat
N 367
Rsq 0.8777
AdjRsq 0.8732
RMSE 0.1101
Studentized Residual
rad

Fig. 11

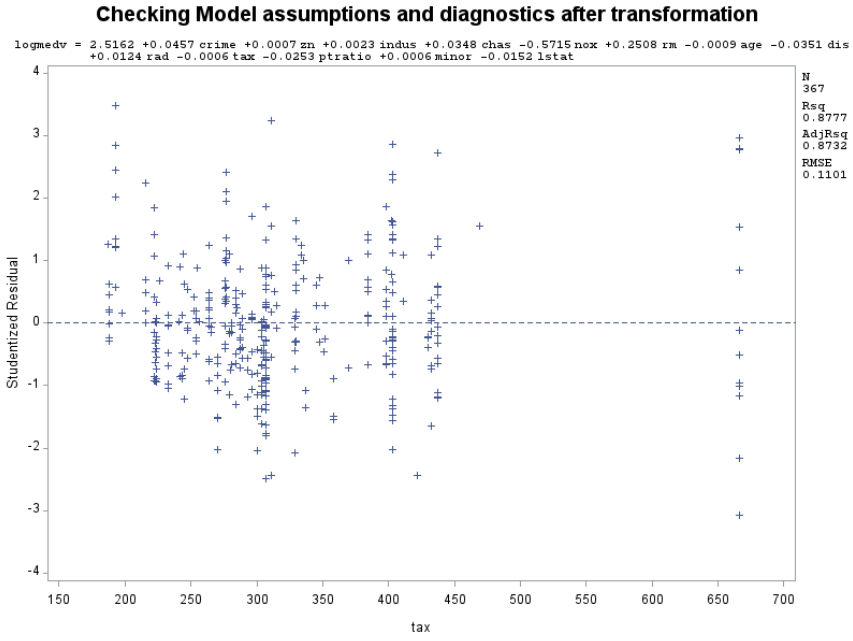

Checking Model assumptions and diagnostics after transformation
logmedv = 2.5162 +0.0457 crime +0.0007 zn +0.0023 indus +0.0348 chas -0.5715 nox +0.2508 rm -0.0009 age -0.0351 dis +0.0124 rad -0.0006 tax -0.0253 ptratio +0.0006 minor -0.0152 lstat
N 367
Rsq 0.8777
AdjRsq 0.8732
RMSE 0.1101
Studentized Residual
tax

Fig. 12

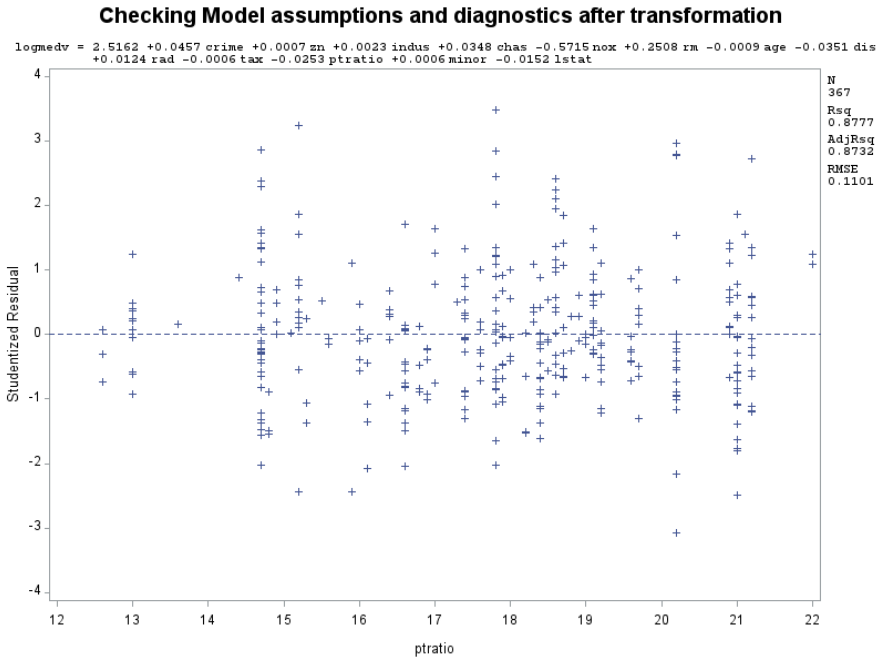

Checking Model assumptions and diagnostics after transformation
logmedv = 2.5162 +0.0457 crime +0.0007 zn +0.0023 indus +0.0348 chas -0.5715 nox +0.2508 rm -0.0009 age -0.0351 dis
+0.0124 rad -0.0006 tax -0.0253 ptratio +0.0006 minor -0.0152 lstat
N
367
Rsq
0.8777
AdjRsq
0.8732
RMSE
0.1101
Studentized Residual
ptratio

Fig. 13

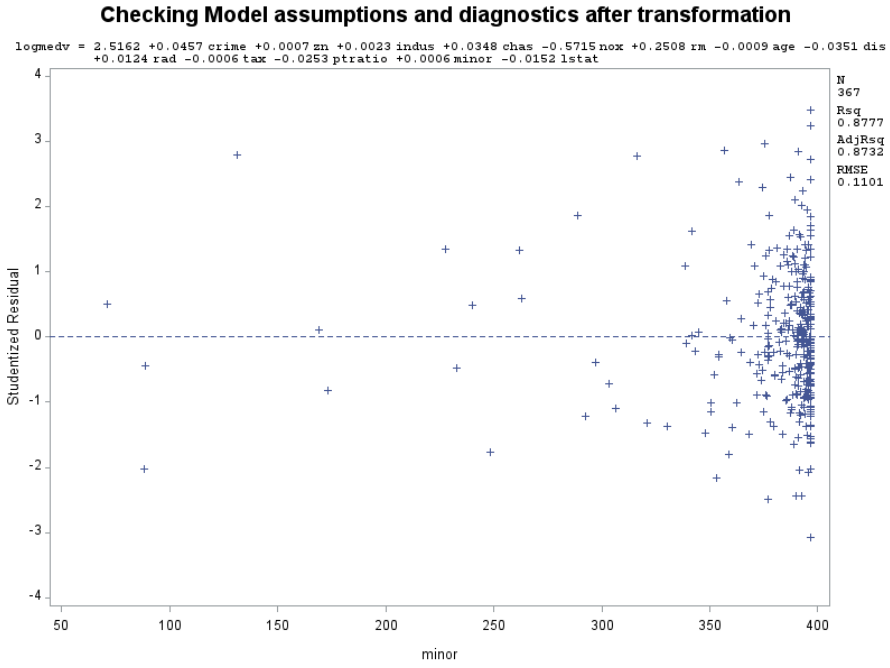

Checking Model assumptions and diagnostics after transformation
logmedv = 2.5162 +0.0457 crime +0.0007 zn +0.0023 indus +0.0348 chas -0.5715 nox +0.2508 rm -0.0009 age -0.0351 dis
+0.0124 rad -0.0006 tax -0.0253 ptratio +0.0006 minor -0.0152 lstat
N
367
Rsq
0.8777
AdjRsq
0.8732
RMSE
0.1101
Studentized Residual
minor

Fig. 14

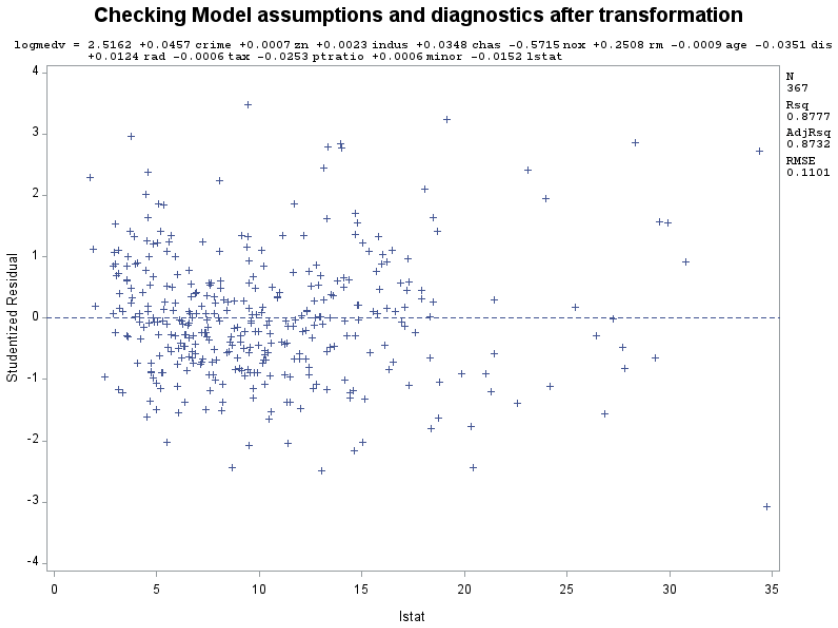

Checking Model assumptions and diagnostics after transformation
logmedv = 2.5162 +0.0457 crime +0.0007 zn +0.0023 indus +0.0348 chas -0.5715 nox +0.2508 rm -0.0009 age -0.0351 dis
+0.0124 rad -0.0006 tax -0.0253 ptratio +0.0006 minor -0.0152 lstat
N
367
Rsq
0.8777
AdjRsq
0.8732
RMSE
0.1101
Studentized Residual
lstat

Fig. 15

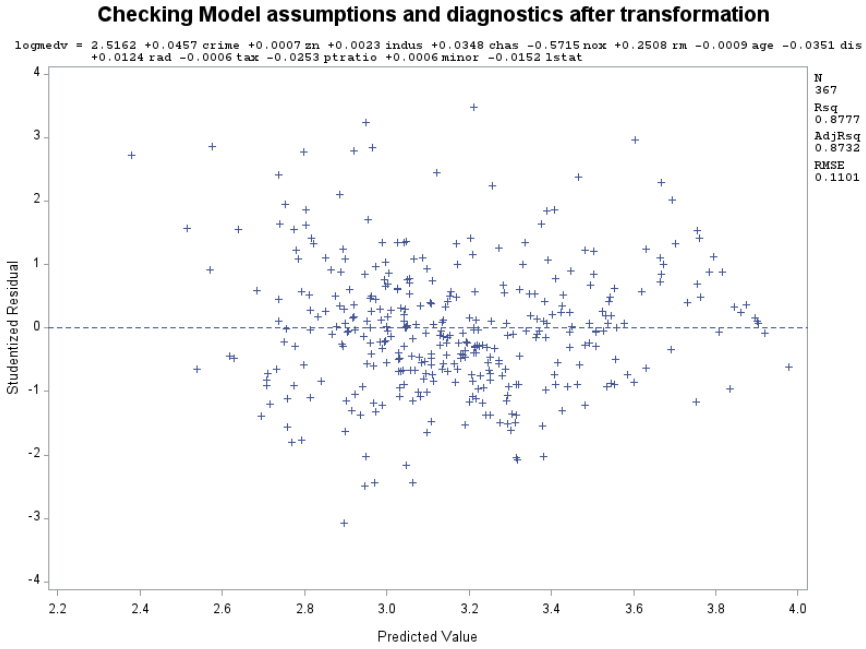

Checking Model assumptions and diagnostics after transformation
logmedv = 2.5162 +0.0457 crime +0.0007 zn +0.0023 indus +0.0348 chas -0.5715 nox +0.2508 rm -0.0009 age -0.0351 dis
+0.0124 rad -0.0006 tax -0.0253 ptratio +0.0006 minor -0.0152 lstat
N
367
Rsq
0.8777
AdjRsq
0.8732
RMSE
0.1101
Studentized Residual
Predicted Value

Fig. 16

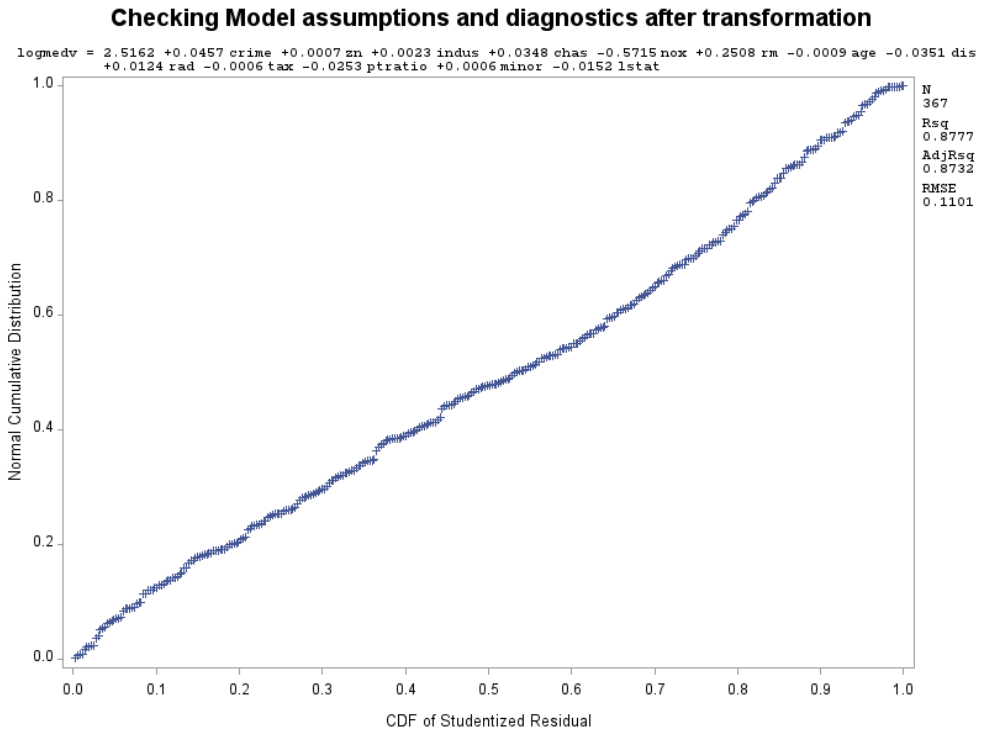


Table 2

Model 1

The REG Procedure
Model: MODEL1
Dependent Variable: new_medv

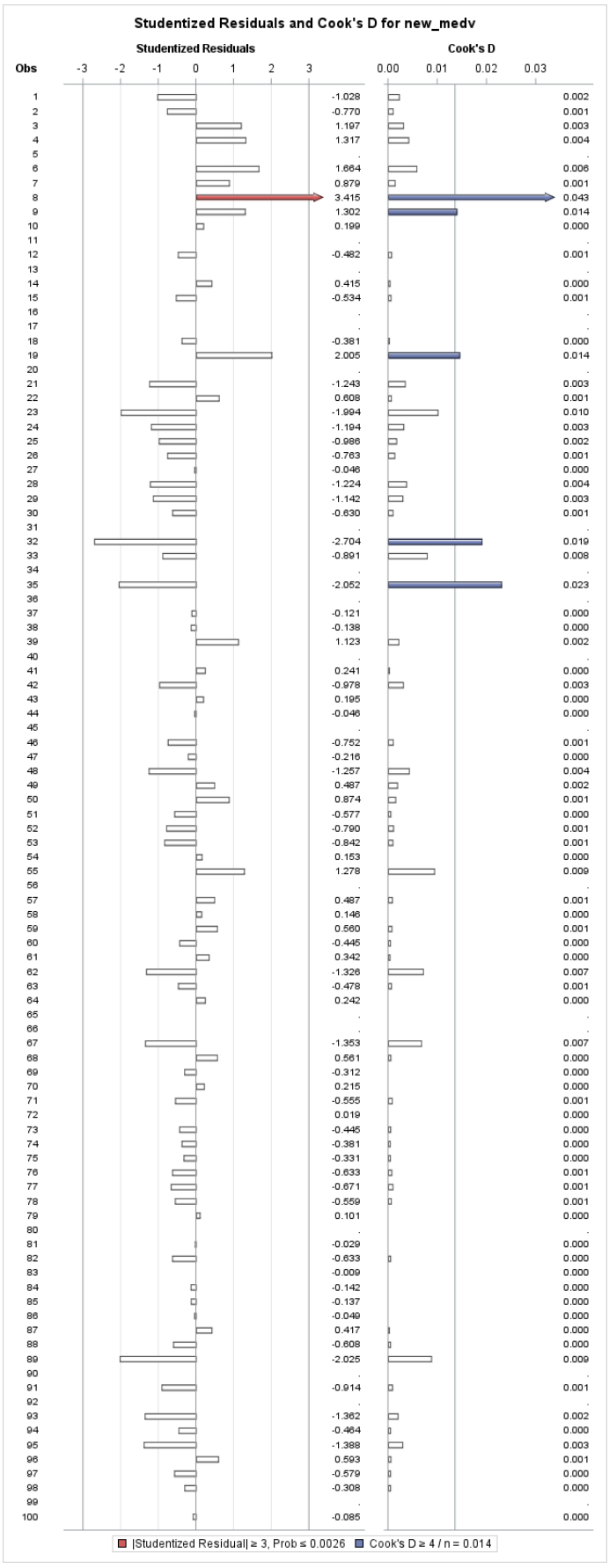
Number of Observations Read	367
Number of Observations Used	294
Number of Observations with Missing Values	73

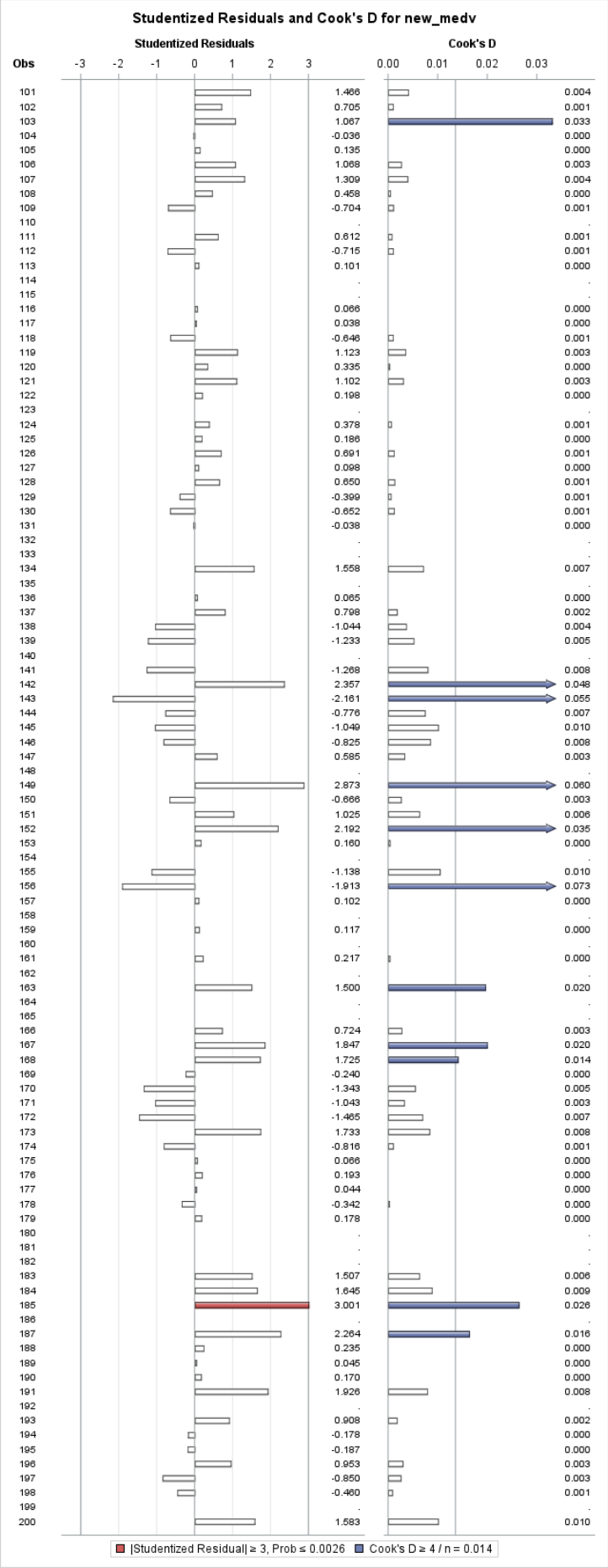
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	12	23.18421	1.93202	180.98	<.0001
Error	281	2.99983	0.01068		
Corrected Total	293	26.18404			

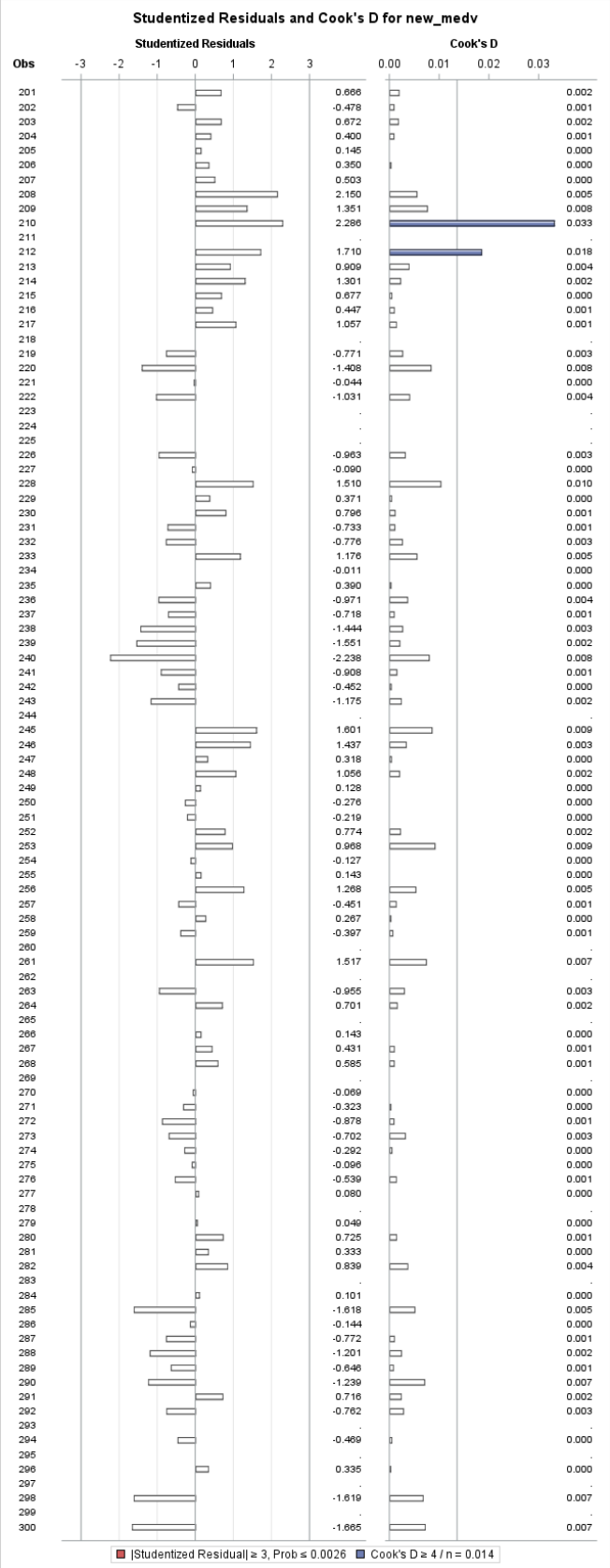
Root MSE	0.10332	**R-Square**	0.8854
Dependent Mean	3.16658	**Adj R-Sq**	0.8805
Coeff Var	3.26291		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standardized Estimate**	**Variance Inflation**
Intercept	**1**	2.47906	0.16384	15.13	<.0001	0	0
crime	**1**	0.05996	0.00955	6.28	<.0001	0.24271	3.66359
zn	**1**	0.00076665	0.00035056	2.19	0.0296	0.06692	2.29668
chas	**1**	0.04569	0.02259	2.02	0.0440	0.04347	1.13248
nox	**1**	-0.66316	0.10826	-6.13	<.0001	-0.23820	3.70877
rm	**1**	0.25313	0.01476	17.15	<.0001	0.55179	2.53851
age	**1**	-0.00123	0.00035849	-3.42	0.0007	-0.11843	2.94335
dis	**1**	-0.03781	0.00520	-7.27	<.0001	-0.26690	3.30945
rad	**1**	0.01079	0.00304	3.55	0.0005	0.12819	3.20278
tax	**1**	-0.00048448	0.00010412	-4.65	<.0001	-0.14233	2.29480
ptratio	**1**	-0.02516	0.00341	-7.37	<.0001	-0.18400	1.52954
minor	**1**	0.00074009	0.00016305	4.54	<.0001	0.11146	1.47913
lstat	**1**	-0.01197	0.00182	-6.57	<.0001	-0.23457	3.12529

Fig. 17


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
1 -1.028 0.002
2 -0.770 0.001
3 1.197 0.003
4 1.317 0.004
5 . .
6 1.664 0.006
7 0.879 0.001
8 3.415 0.043
9 1.302 0.014
10 0.199 0.000
11 . .
12 -0.482 0.001
13 . .
14 0.415 0.000
15 -0.534 0.001
16 . .
17 . .
18 -0.381 0.000
19 2.005 0.014
20 . .
21 -1.243 0.003
22 0.608 0.001
23 -1.994 0.010
24 -1.194 0.003
25 -0.986 0.002
26 -0.763 0.001
27 -0.046 0.000
28 -1.224 0.004
29 -1.142 0.003
30 -0.630 0.001
31 . .
32 -2.704 0.019
33 -0.891 0.008
34 . .
35 -2.052 0.023
36 . .
37 -0.121 0.000
38 -0.138 0.000
39 1.123 0.002
40 . .
41 0.241 0.000
42 -0.978 0.003
43 0.195 0.000
44 -0.046 0.000
45 . .
46 -0.752 0.001
47 -0.216 0.000
48 -1.257 0.004
49 0.487 0.002
50 0.874 0.001
51 -0.577 0.000
52 -0.790 0.001
53 -0.842 0.001
54 0.153 0.000
55 1.278 0.009
56 . .
57 0.487 0.001
58 0.146 0.000
59 0.560 0.001
60 -0.445 0.000
61 0.342 0.000
62 -1.326 0.007
63 -0.478 0.001
64 0.242 0.000
65 . .
66 . .
67 -1.353 0.007
68 0.561 0.000
69 -0.312 0.000
70 0.215 0.000
71 -0.555 0.001
72 0.019 0.000
73 -0.445 0.000
74 -0.381 0.000
75 -0.331 0.000
76 -0.633 0.001
77 -0.671 0.001
78 -0.559 0.001
79 0.101 0.000
80 . .
81 -0.029 0.000
82 -0.633 0.000
83 -0.009 0.000
84 -0.142 0.000
85 -0.137 0.000
86 -0.049 0.000
87 0.417 0.000
88 -0.608 0.000
89 -2.025 0.009
90 . .
91 -0.914 0.001
92 . .
93 -1.362 0.002
94 -0.464 0.000
95 -1.388 0.003
96 0.593 0.001
97 -0.579 0.000
98 -0.308 0.000
99 . .
100 -0.085 0.000
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
101 1.466 0.004
102 0.705 0.001
103 1.067 0.033
104 -0.036 0.000
105 0.135 0.000
106 1.068 0.003
107 1.309 0.004
108 0.458 0.000
109 -0.704 0.001
110 . .
111 0.612 0.001
112 -0.715 0.001
113 0.101 0.000
114 . .
115 . .
116 0.066 0.000
117 0.038 0.000
118 -0.646 0.001
119 1.123 0.003
120 0.335 0.000
121 1.102 0.003
122 0.198 0.000
123 . .
124 0.378 0.001
125 0.186 0.000
126 0.691 0.001
127 0.098 0.000
128 0.650 0.001
129 -0.399 0.001
130 -0.652 0.001
131 -0.038 0.000
132 . .
133 . .
134 1.558 0.007
135 . .
136 0.065 0.000
137 0.798 0.002
138 -1.044 0.004
139 -1.233 0.005
140 . .
141 -1.268 0.008
142 2.357 0.048
143 -2.161 0.055
144 -0.776 0.007
145 -1.049 0.010
146 -0.825 0.008
147 0.585 0.003
148 . .
149 2.873 0.060
150 -0.666 0.003
151 1.025 0.006
152 2.192 0.035
153 0.160 0.000
154 . .
155 -1.138 0.010
156 -1.913 0.073
157 0.102 0.000
158 . .
159 0.117 0.000
160 . .
161 0.217 0.000
162 . .
163 1.500 0.020
164 . .
165 . .
166 0.724 0.003
167 1.847 0.020
168 1.725 0.014
169 -0.240 0.000
170 -1.343 0.005
171 -1.043 0.003
172 -1.465 0.007
173 1.733 0.008
174 -0.816 0.001
175 0.066 0.000
176 0.193 0.000
177 0.044 0.000
178 -0.342 0.000
179 0.178 0.000
180 . .
181 . .
182 . .
183 1.507 0.006
184 1.645 0.009
185 3.001 0.026
186 . .
187 2.264 0.016
188 0.235 0.000
189 0.045 0.000
190 0.170 0.000
191 1.926 0.008
192 . .
193 0.908 0.002
194 -0.178 0.000
195 -0.187 0.000
196 0.953 0.003
197 -0.850 0.003
198 -0.460 0.001
199 . .
200 1.583 0.010
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014

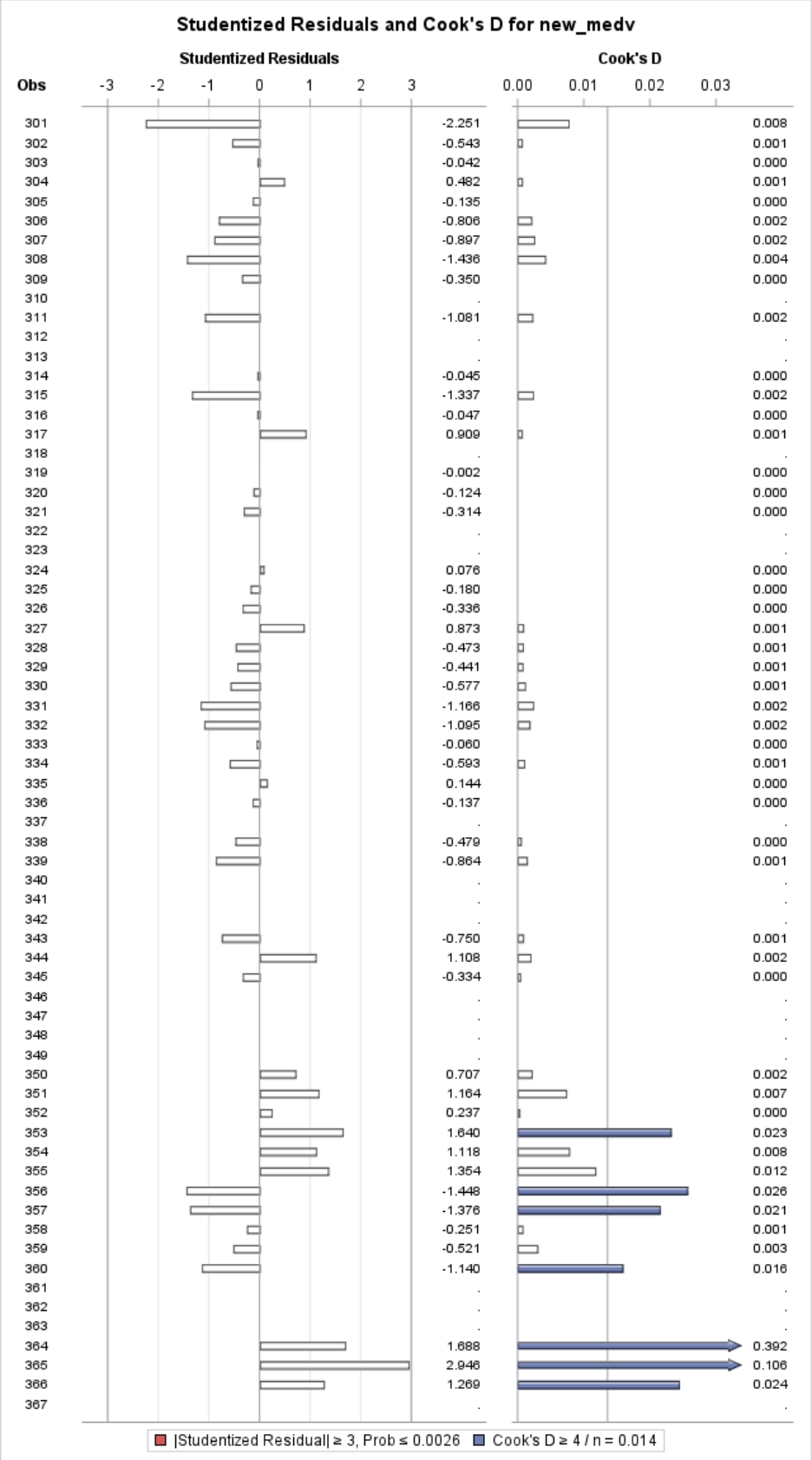

Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
301 -2.251 0.008
302 -0.543 0.001
303 -0.042 0.000
304 0.482 0.001
305 -0.135 0.000
306 -0.806 0.002
307 -0.897 0.002
308 -1.436 0.004
309 -0.350 0.000
310 . .
311 -1.081 0.002
312 . .
313 . .
314 -0.045 0.000
315 -1.337 0.002
316 -0.047 0.000
317 0.909 0.001
318 . .
319 -0.002 0.000
320 -0.124 0.000
321 -0.314 0.000
322 . .
323 . .
324 0.076 0.000
325 -0.180 0.000
326 -0.336 0.000
327 0.873 0.001
328 -0.473 0.001
329 -0.441 0.001
330 -0.577 0.001
331 -1.166 0.002
332 -1.095 0.002
333 -0.060 0.000
334 -0.593 0.001
335 0.144 0.000
336 -0.137 0.000
337 . .
338 -0.479 0.000
339 -0.864 0.001
340 . .
341 . .
342 . .
343 -0.750 0.001
344 1.108 0.002
345 -0.334 0.000
346 . .
347 . .
348 . .
349 . .
350 0.707 0.002
351 1.164 0.007
352 0.237 0.000
353 1.640 0.023
354 1.118 0.008
355 1.354 0.012
356 -1.448 0.026
357 -1.376 0.021
358 -0.251 0.001
359 -0.521 0.003
360 -1.140 0.016
361 . .
362 . .
363 . .
364 1.688 0.392
365 2.946 0.106
366 1.269 0.024
367 .
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014

Fig. 18

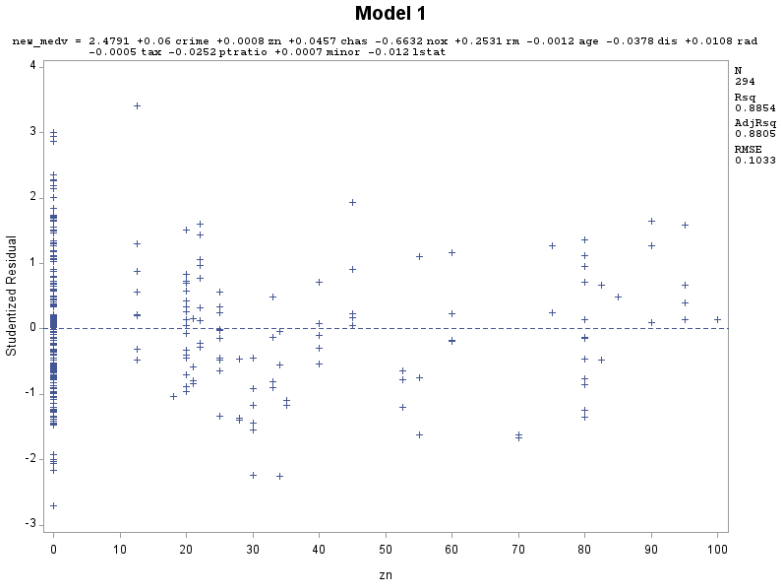


Fig. 19

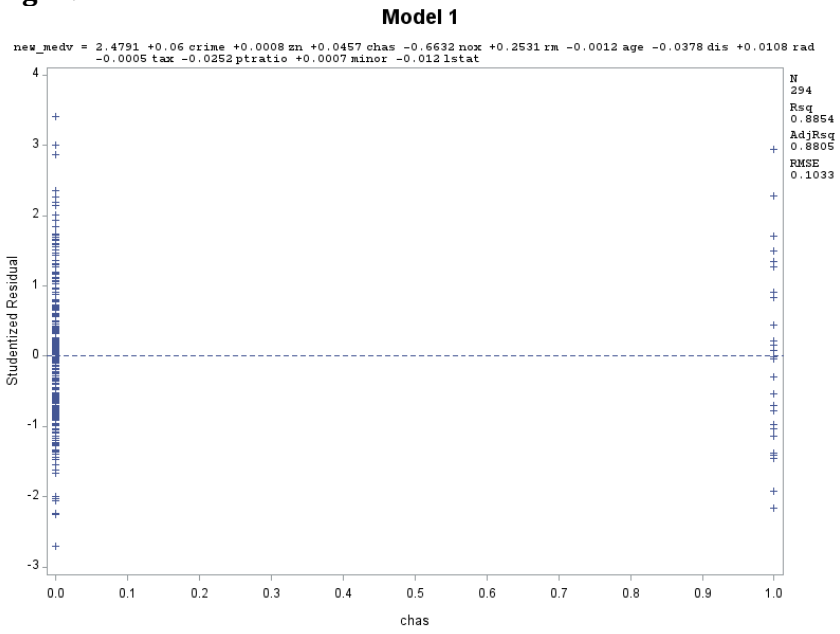


Fig. 20

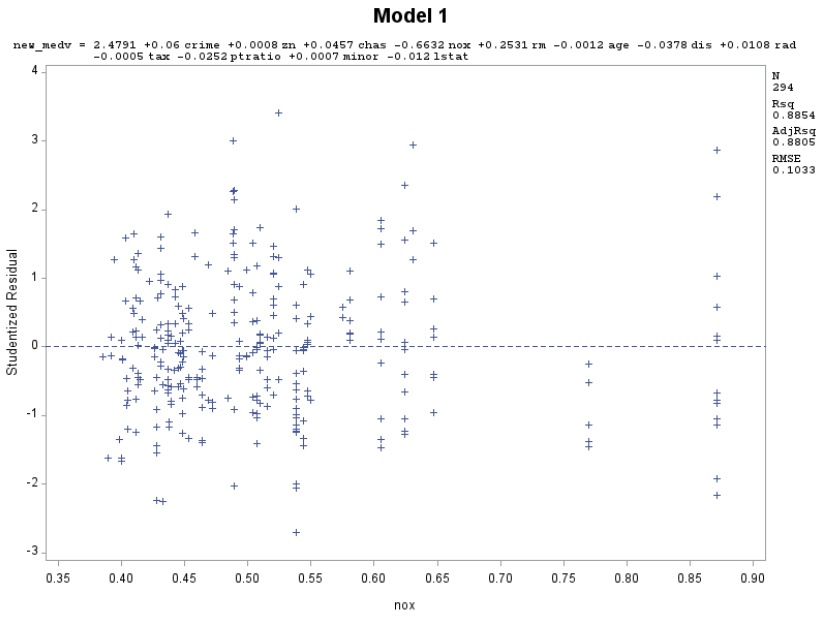

Model 1
new_medv = 2.4791 +0.06 crime +0.0008 zn +0.0457 chas -0.6632 nox +0.2531 rm -0.0012 age -0.0378 dis +0.0108 rad
-0.0005 tax -0.0252 ptratio +0.0007 minor -0.012 lstat
N
294
Rsq
0.8854
AdjRsq
0.8805
RMSE
0.1033
Studentized Residual
nox

Fig. 21

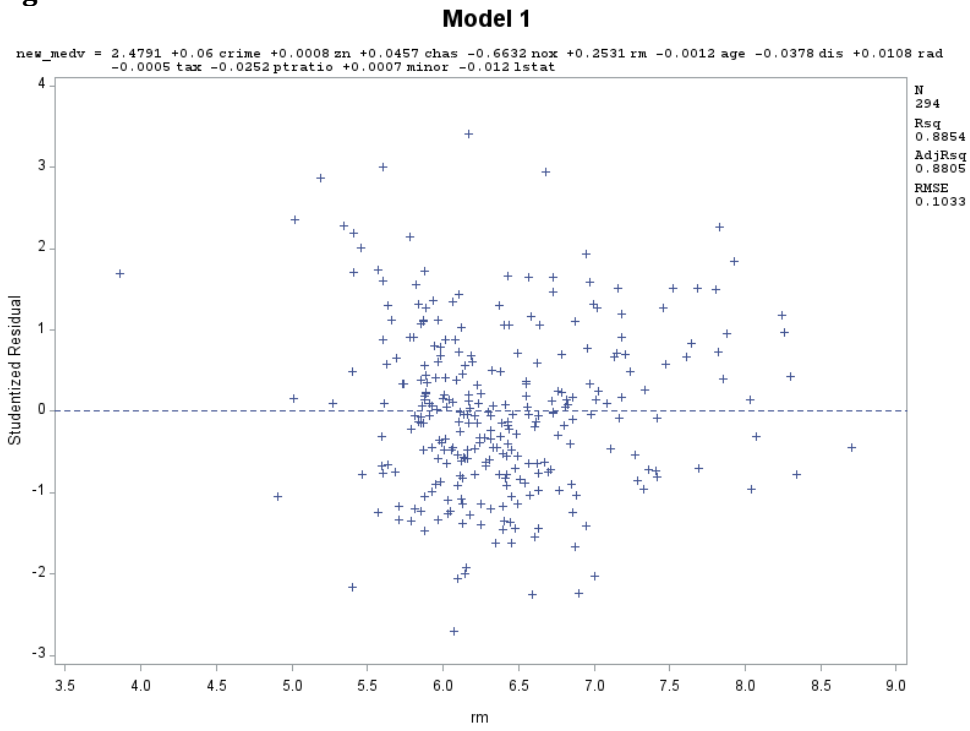

Model 1
new_medv = 2.4791 +0.06 crime +0.0008 zn +0.0457 chas -0.6632 nox +0.2531 rm -0.0012 age -0.0378 dis +0.0108 rad
-0.0005 tax -0.0252 ptratio +0.0007 minor -0.012 lstat
N
294
Rsq
0.8854
AdjRsq
0.8805
RMSE
0.1033
Studentized Residual
rm

Fig. 22

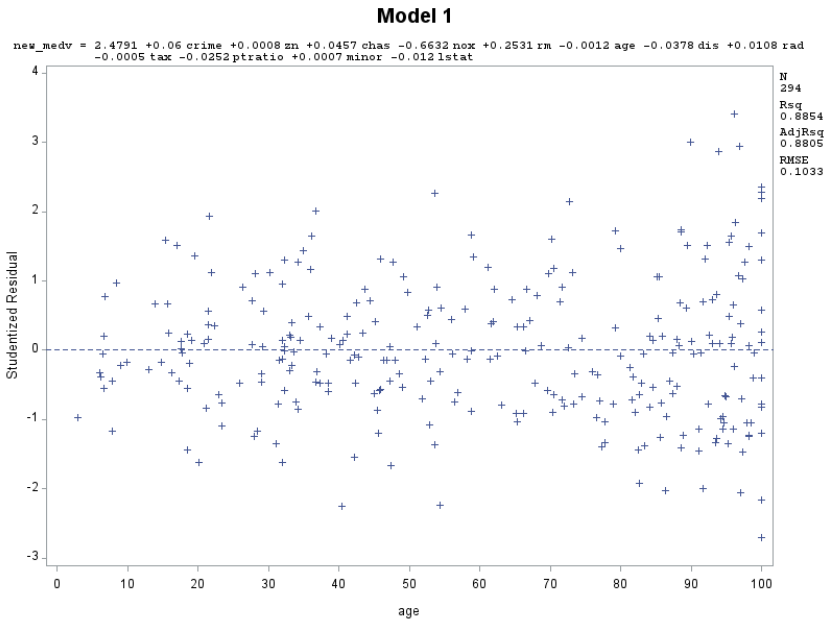

Model 1
new_medv = 2.4791 +0.06 crime +0.0008 zn +0.0457 chas -0.6632 nox +0.2531 rm -0.0012 age -0.0378 dis +0.0108 rad
-0.0005 tax -0.0252 ptratio +0.0007 minor -0.012 lstat
N
294
Rsq
0.8854
AdjRsq
0.8805
RMSE
0.1033
Studentized Residual
age

Fig. 23

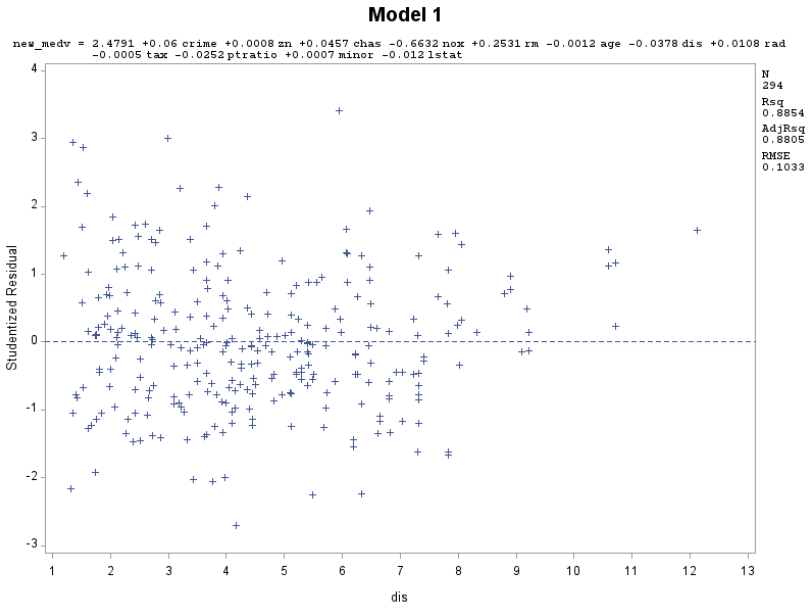

Model 1
new_medv = 2.4791 +0.06 crime +0.0008 zn +0.0457 chas -0.6632 nox +0.2531 rm -0.0012 age -0.0378 dis +0.0108 rad
-0.0005 tax -0.0252 ptratio +0.0007 minor -0.012 lstat
N
294
Rsq
0.8854
AdjRsq
0.8805
RMSE
0.1033
Studentized Residual
dis

Fig. 24

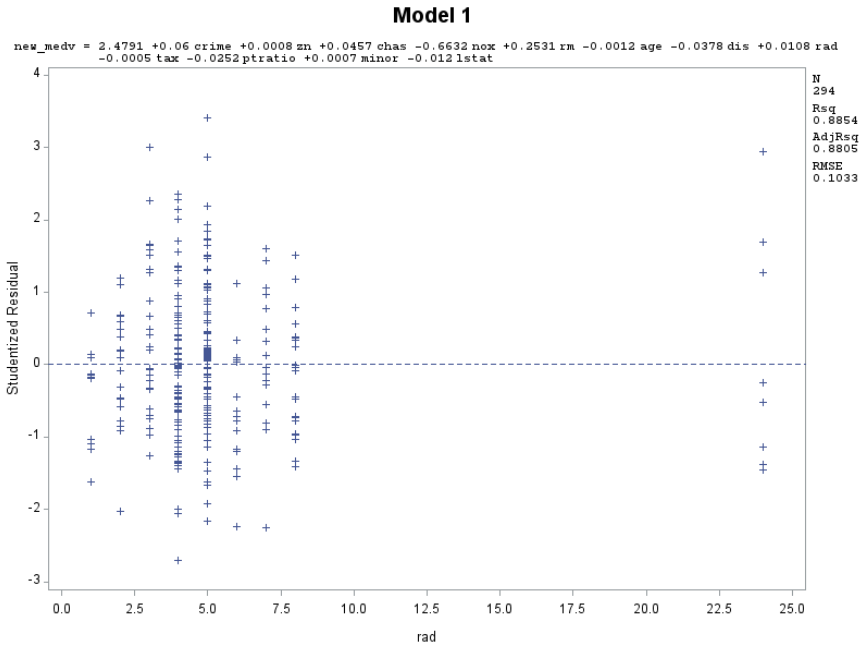

Model 1
new_medv = 2.4791 +0.06 crime +0.0008 zn +0.0457 chas -0.6632 nox +0.2531 rm -0.0012 age -0.0378 dis +0.0108 rad
-0.0005 tax -0.0252 ptratio +0.0007 minor -0.012 lstat
N
294
Rsq
0.8854
AdjRsq
0.8805
RMSE
0.1033
Studentized Residual
rad

Fig. 25

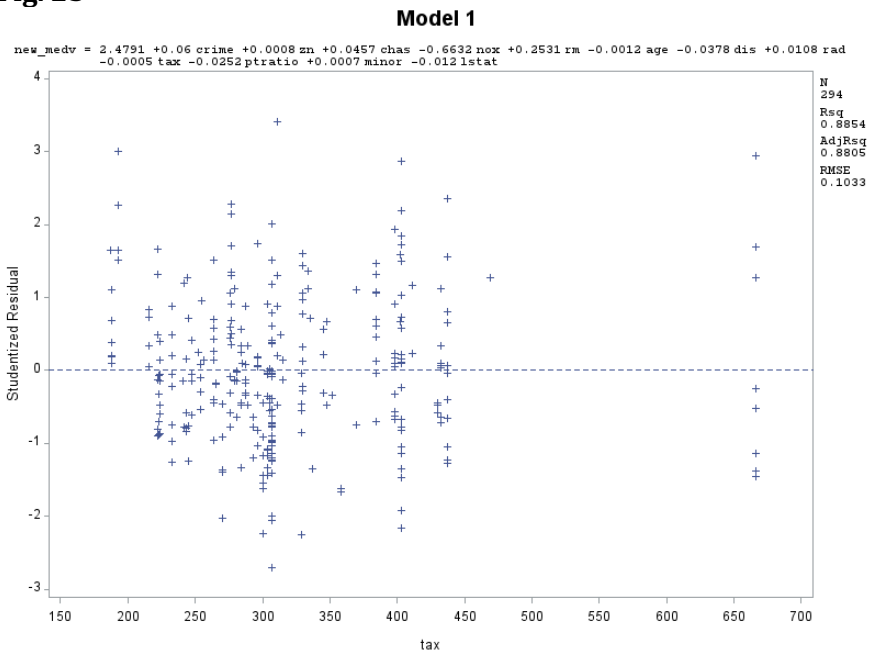

Model 1
new_medv = 2.4791 +0.06 crime +0.0008 zn +0.0457 chas -0.6632 nox +0.2531 rm -0.0012 age -0.0378 dis +0.0108 rad
-0.0005 tax -0.0252 ptratio +0.0007 minor -0.012 lstat
N
294
Rsq
0.8854
AdjRsq
0.8805
RMSE
0.1033
Studentized Residual
tax

Fig. 26

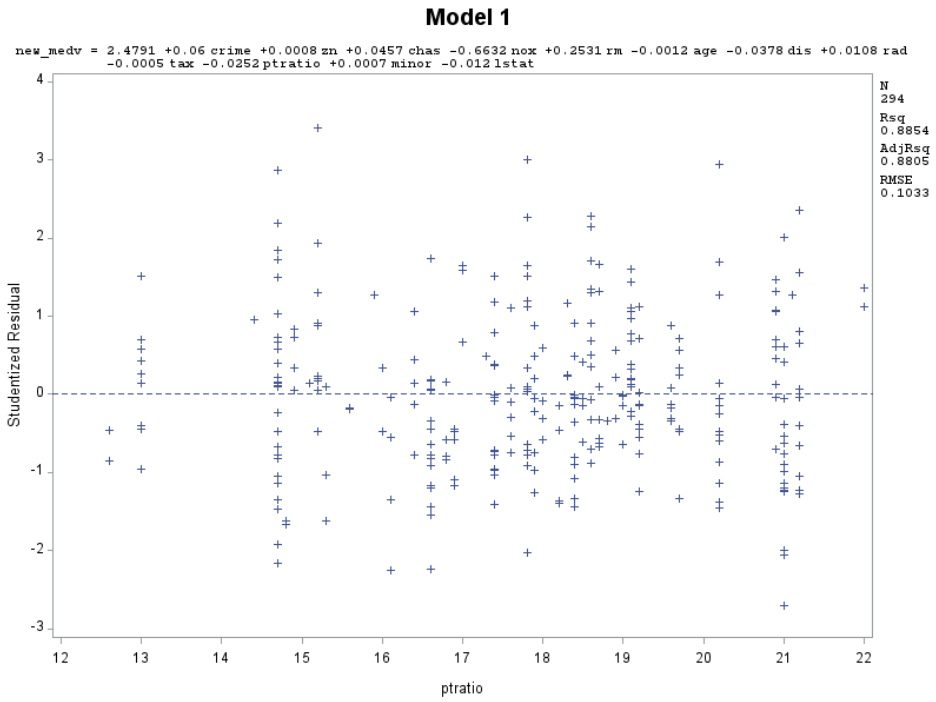

Model 1
new_medv = 2.4791 +0.06 crime +0.0008 zn +0.0457 chas -0.6632 nox +0.2531 rm -0.0012 age -0.0378 dis +0.0108 rad
-0.0005 tax -0.0252 ptratio +0.0007 minor -0.012 lstat
N
294
Rsq
0.8854
AdjRsq
0.8805
RMSE
0.1033
Studentized Residual
ptratio

Fig. 27

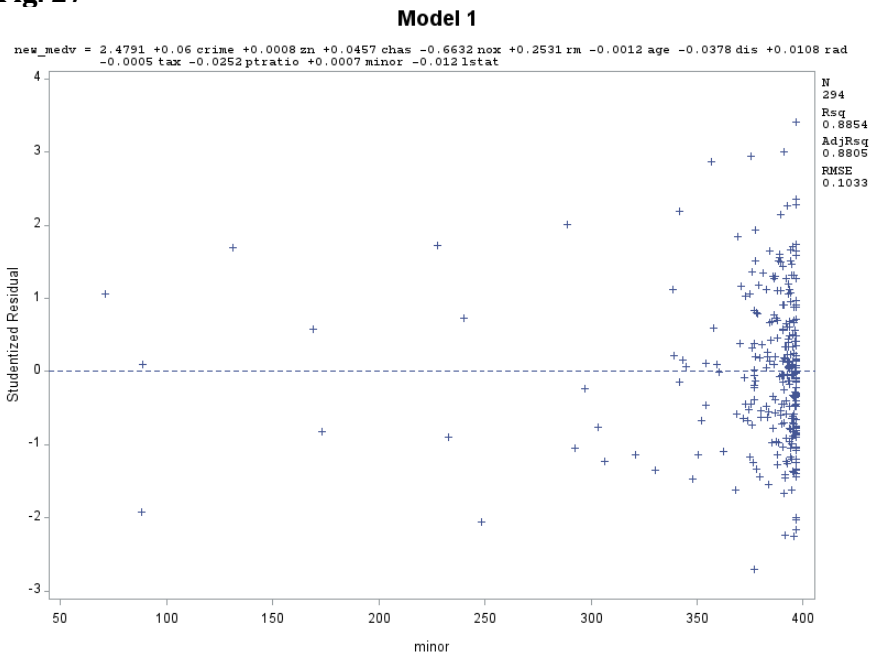

Model 1
new_medv = 2.4791 +0.06 crime +0.0008 zn +0.0457 chas -0.6632 nox +0.2531 rm -0.0012 age -0.0378 dis +0.0108 rad
-0.0005 tax -0.0252 ptratio +0.0007 minor -0.012 lstat
N
294
Rsq
0.8854
AdjRsq
0.8805
RMSE
0.1033
Studentized Residual
minor

Fig. 28

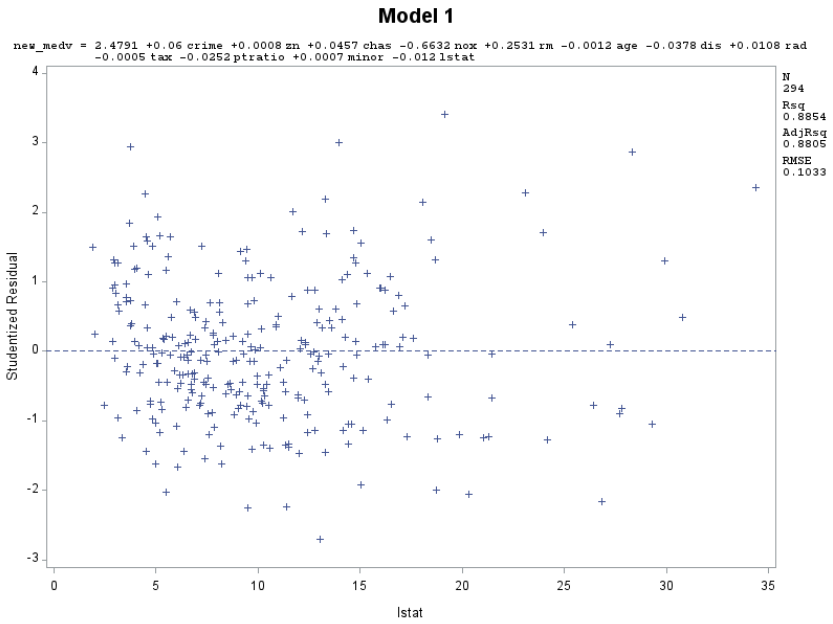

Model 1
new_medv = 2.4791 +0.06 crime +0.0008 zn +0.0457 chas -0.6632 nox +0.2531 rm -0.0012 age -0.0378 dis +0.0108 rad
-0.0005 tax -0.0252 ptratio +0.0007 minor -0.012 lstat
N
294
Rsq
0.8854
AdjRsq
0.8805
RMSE
0.1033
Studentized Residual
lstat

Fig. 29

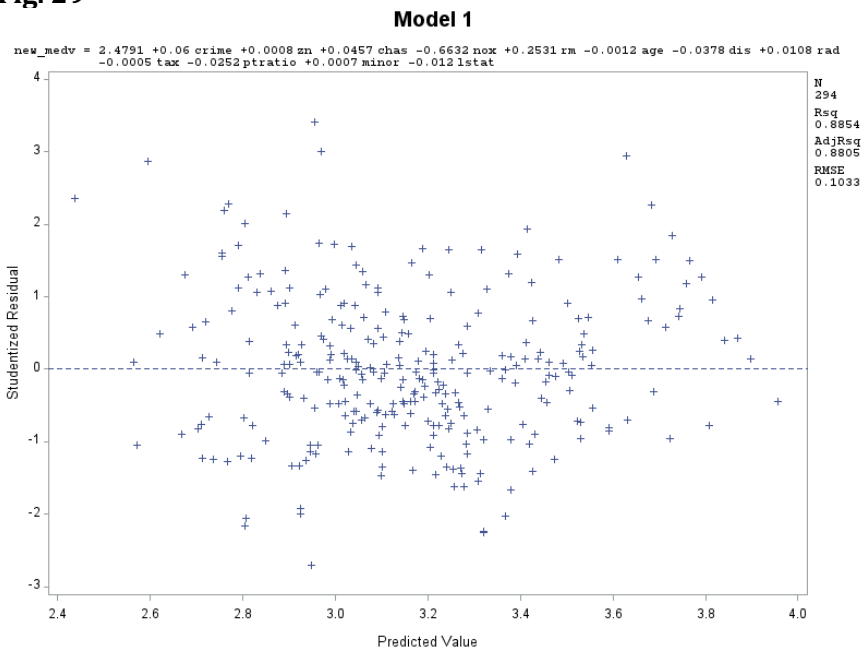

Model 1
new_medv = 2.4791 +0.06 crime +0.0008 zn +0.0457 chas -0.6632 nox +0.2531 rm -0.0012 age -0.0378 dis +0.0108 rad
-0.0005 tax -0.0252 ptratio +0.0007 minor -0.012 lstat
N
294
Rsq
0.8854
AdjRsq
0.8805
RMSE
0.1033
Studentized Residual
Predicted Value

Fig. 30

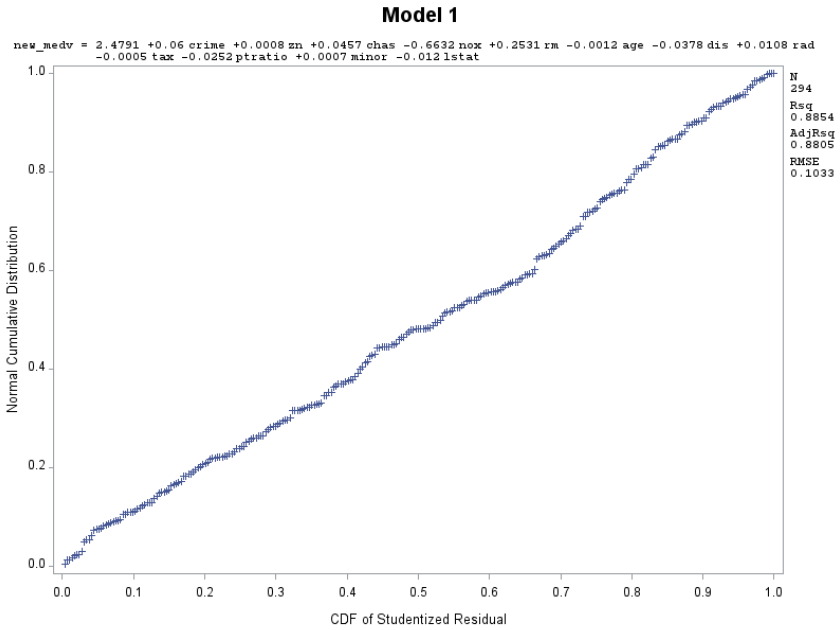


Table 3

Model 1 after removing outliers

The REG Procedure
Model: MODEL1
Dependent Variable: new_medv

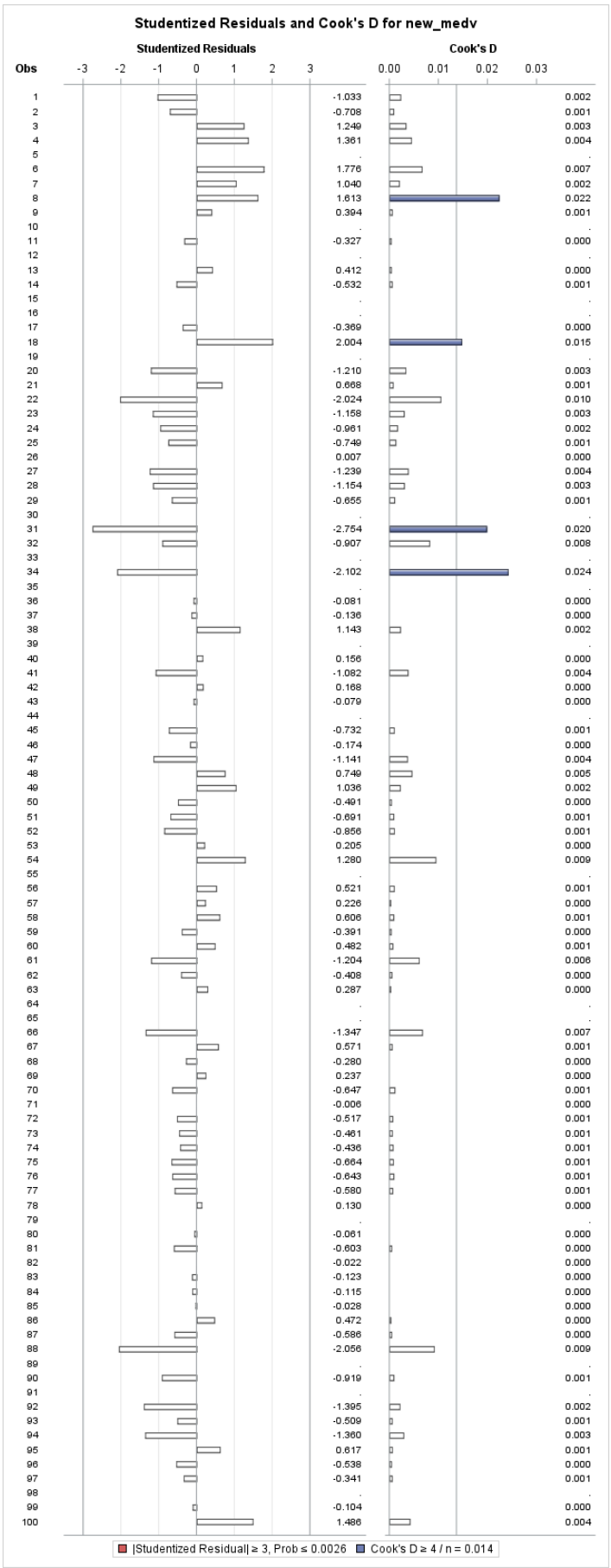
Number of Observations Read	365
Number of Observations Used	292
Number of Observations with Missing Values	73

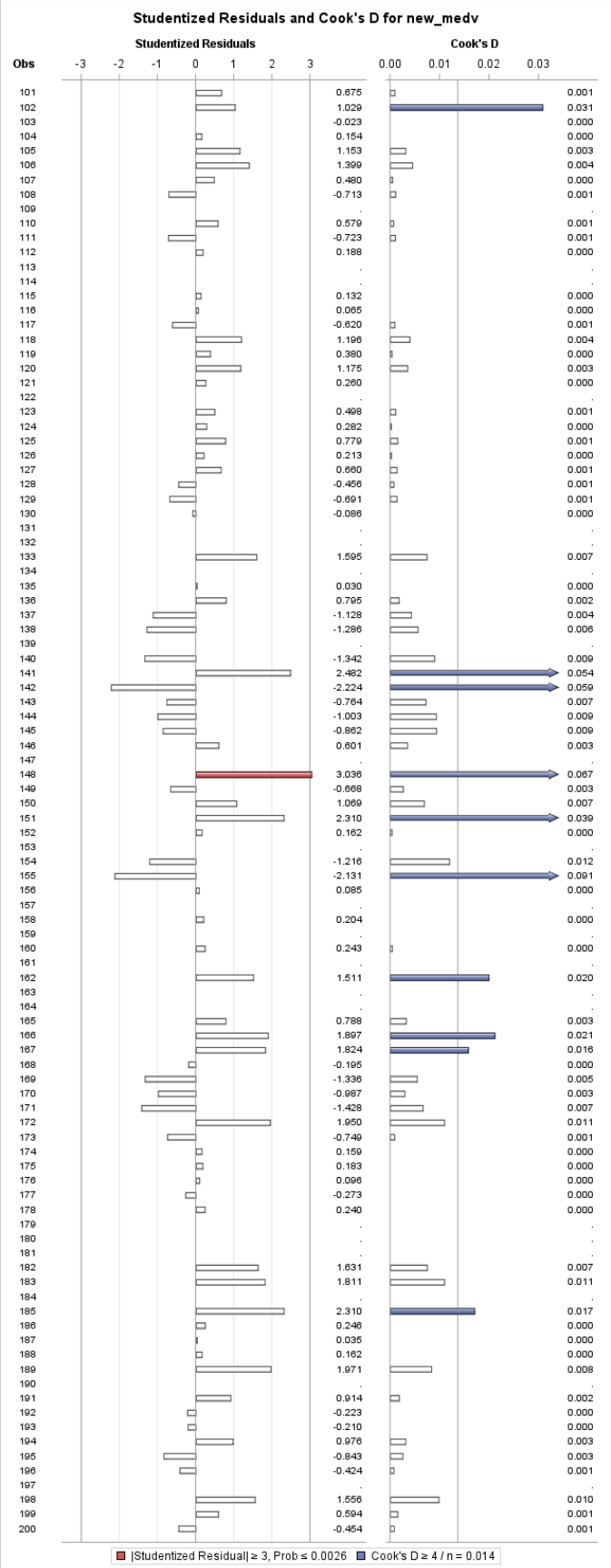
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	12	23.37772	1.94814	195.72	<.0001
Error	279	2.77704	0.00995		
Corrected Total	291	26.15476			

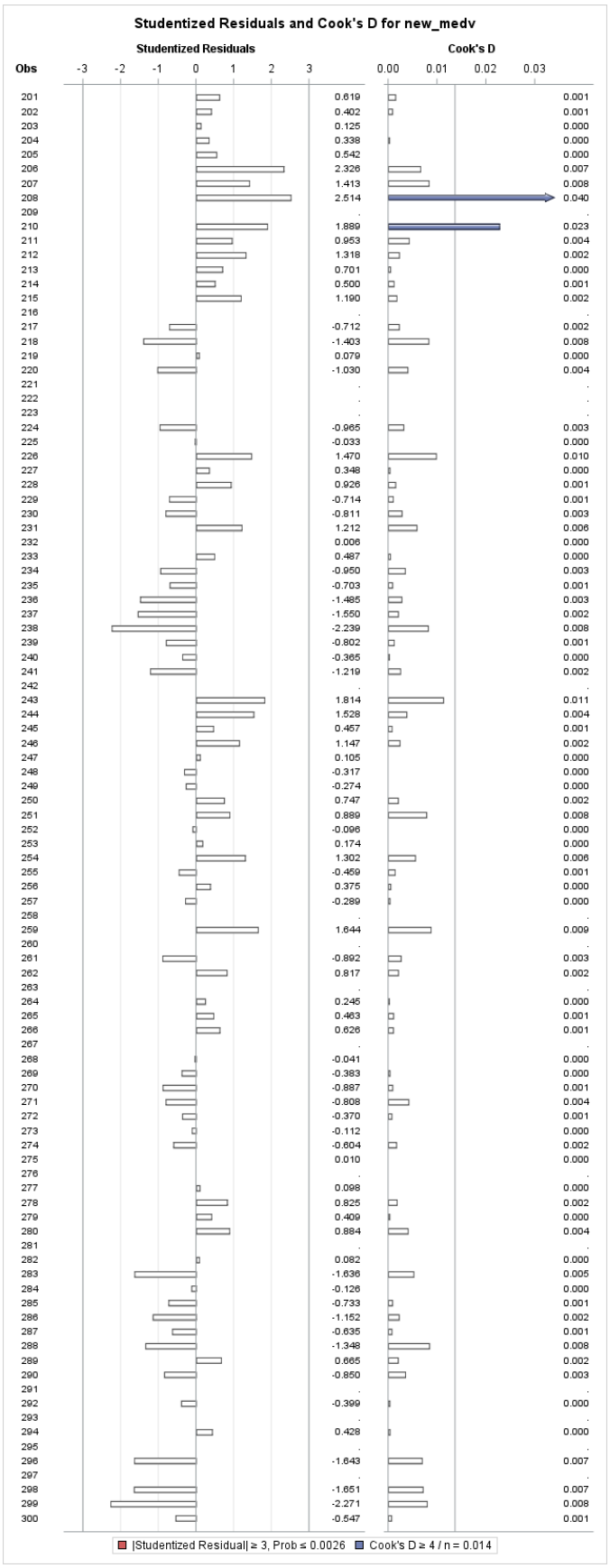
Root MSE	0.09977	**R-Square**	0.8938
Dependent Mean	3.16576	**Adj R-Sq**	0.8893
Coeff Var	3.15146		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standardized Estimate**	**Variance Inflation**
Intercept	**1**	2.42099	0.15932	15.20	<.0001	0	0
crime	**1**	0.06120	0.00922	6.64	<.0001	0.24779	3.66380
zn	**1**	0.00080050	0.00033922	2.36	0.0190	0.06986	2.30320
chas	**1**	0.04938	0.02182	2.26	0.0244	0.04699	1.13319
nox	**1**	-0.63361	0.10475	-6.05	<.0001	-0.22768	3.72284
rm	**1**	0.25741	0.01436	17.93	<.0001	0.55996	2.56263
age	**1**	-0.00143	0.00034877	-4.09	<.0001	-0.13738	2.96147
dis	**1**	-0.03883	0.00506	-7.67	<.0001	-0.27379	3.34696
rad	**1**	0.00988	0.00294	3.36	0.0009	0.11737	3.21295
tax	**1**	-0.00045523	0.00010106	-4.50	<.0001	-0.13333	2.30206
ptratio	**1**	-0.02333	0.00332	-7.02	<.0001	-0.17029	1.54708
minor	**1**	0.00072210	0.00015749	4.58	<.0001	0.10877	1.47888
lstat	**1**	-0.01217	0.00177	-6.89	<.0001	-0.23753	3.12320

Fig. 31


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
1 -1.033 0.002
2 -0.708 0.001
3 1.249 0.003
4 1.361 0.004
5 . .
6 1.776 0.007
7 1.040 0.002
8 1.613 0.022
9 0.394 0.001
10 . .
11 -0.327 0.000
12 . .
13 0.412 0.000
14 -0.532 0.001
15 . .
16 . .
17 -0.369 0.000
18 2.004 0.015
19 . .
20 -1.210 0.003
21 0.668 0.001
22 -2.024 0.010
23 -1.158 0.003
24 -0.961 0.002
25 -0.749 0.001
26 0.007 0.000
27 -1.239 0.004
28 -1.154 0.003
29 -0.655 0.001
30 . .
31 -2.754 0.020
32 -0.907 0.008
33 . .
34 -2.102 0.024
35 . .
36 -0.081 0.000
37 -0.136 0.000
38 1.143 0.002
39 . .
40 0.156 0.000
41 -1.082 0.004
42 0.168 0.000
43 -0.079 0.000
44 . .
45 -0.732 0.001
46 -0.174 0.000
47 -1.141 0.004
48 0.749 0.005
49 1.036 0.002
50 -0.491 0.000
51 -0.691 0.001
52 -0.856 0.001
53 0.205 0.000
54 1.280 0.009
55 . .
56 0.521 0.001
57 0.226 0.000
58 0.606 0.001
59 -0.391 0.000
60 0.482 0.001
61 -1.204 0.006
62 -0.408 0.000
63 0.287 0.000
64 . .
65 . .
66 -1.347 0.007
67 0.571 0.001
68 -0.280 0.000
69 0.237 0.000
70 -0.647 0.001
71 -0.006 0.000
72 -0.517 0.001
73 -0.461 0.001
74 -0.436 0.001
75 -0.664 0.001
76 -0.643 0.001
77 -0.580 0.001
78 0.130 0.000
79 . .
80 -0.061 0.000
81 -0.603 0.000
82 -0.022 0.000
83 -0.123 0.000
84 -0.115 0.000
85 -0.028 0.000
86 0.472 0.000
87 -0.586 0.000
88 -2.056 0.009
89 . .
90 -0.919 0.001
91 . .
92 -1.395 0.002
93 -0.509 0.001
94 -1.360 0.003
95 0.617 0.001
96 -0.538 0.000
97 -0.341 0.001
98 . .
99 -0.104 0.000
100 1.486 0.004
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
101 0.675 0.001
102 1.029 0.031
103 -0.023 0.000
104 0.154 0.000
105 1.153 0.003
106 1.399 0.004
107 0.480 0.000
108 -0.713 0.001
109 . .
110 0.579 0.001
111 -0.723 0.001
112 0.188 0.000
113 . .
114 . .
115 0.132 0.000
116 0.065 0.000
117 -0.620 0.001
118 1.196 0.004
119 0.380 0.000
120 1.175 0.003
121 0.260 0.000
122 . .
123 0.498 0.001
124 0.282 0.000
125 0.779 0.001
126 0.213 0.000
127 0.660 0.001
128 -0.456 0.001
129 -0.691 0.001
130 -0.086 0.000
131 . .
132 . .
133 1.595 0.007
134 . .
135 0.030 0.000
136 0.795 0.002
137 -1.128 0.004
138 -1.286 0.006
139 . .
140 -1.342 0.009
141 2.482 0.054
142 -2.224 0.059
143 -0.764 0.007
144 -1.003 0.009
145 -0.862 0.009
146 0.601 0.003
147 . .
148 3.036 0.067
149 -0.668 0.003
150 1.069 0.007
151 2.310 0.039
152 0.162 0.000
153 . .
154 -1.216 0.012
155 -2.131 0.091
156 0.085 0.000
157 . .
158 0.204 0.000
159 . .
160 0.243 0.000
161 . .
162 1.511 0.020
163 . .
164 . .
165 0.788 0.003
166 1.897 0.021
167 1.824 0.016
168 -0.195 0.000
169 -1.336 0.005
170 -0.987 0.003
171 -1.428 0.007
172 1.950 0.011
173 -0.749 0.001
174 0.159 0.000
175 0.183 0.000
176 0.096 0.000
177 -0.273 0.000
178 0.240 0.000
179 . .
180 . .
181 . .
182 1.631 0.007
183 1.811 0.011
184 . .
185 2.310 0.017
186 0.246 0.000
187 0.035 0.000
188 0.162 0.000
189 1.971 0.008
190 . .
191 0.914 0.002
192 -0.223 0.000
193 -0.210 0.000
194 0.976 0.003
195 -0.843 0.003
196 -0.424 0.001
197 . .
198 1.556 0.010
199 0.594 0.001
200 -0.454 0.001
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
201 0.619 0.001
202 0.402 0.001
203 0.125 0.000
204 0.338 0.000
205 0.542 0.000
206 2.326 0.007
207 1.413 0.008
208 2.514 0.040
209 . .
210 1.889 0.023
211 0.953 0.004
212 1.318 0.002
213 0.701 0.000
214 0.500 0.001
215 1.190 0.002
216 . .
217 -0.712 0.002
218 -1.403 0.008
219 0.079 0.000
220 -1.030 0.004
221 . .
222 . .
223 . .
224 -0.965 0.003
225 -0.033 0.000
226 1.470 0.010
227 0.348 0.000
228 0.926 0.001
229 -0.714 0.001
230 -0.811 0.003
231 1.212 0.006
232 0.006 0.000
233 0.487 0.000
234 -0.950 0.003
235 -0.703 0.001
236 -1.485 0.003
237 -1.550 0.002
238 -2.239 0.008
239 -0.802 0.001
240 -0.365 0.000
241 -1.219 0.002
242 . .
243 1.814 0.011
244 1.528 0.004
245 0.457 0.001
246 1.147 0.002
247 0.105 0.000
248 -0.317 0.000
249 -0.274 0.000
250 0.747 0.002
251 0.889 0.008
252 -0.096 0.000
253 0.174 0.000
254 1.302 0.006
255 -0.459 0.001
256 0.375 0.000
257 -0.289 0.000
258 . .
259 1.644 0.009
260 . .
261 -0.892 0.003
262 0.817 0.002
263 . .
264 0.245 0.000
265 0.463 0.001
266 0.626 0.001
267 . .
268 -0.041 0.000
269 -0.383 0.000
270 -0.887 0.001
271 -0.808 0.004
272 -0.370 0.001
273 -0.112 0.000
274 -0.604 0.002
275 0.010 0.000
276 . .
277 0.098 0.000
278 0.825 0.002
279 0.409 0.000
280 0.884 0.004
281 . .
282 0.082 0.000
283 -1.636 0.005
284 -0.126 0.000
285 -0.733 0.001
286 -1.152 0.002
287 -0.635 0.001
288 -1.348 0.008
289 0.665 0.002
290 -0.850 0.003
291 . .
292 -0.399 0.000
293 . .
294 0.428 0.000
295 . .
296 -1.643 0.007
297 . .
298 -1.651 0.007
299 -2.271 0.008
300 -0.547 0.001
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014

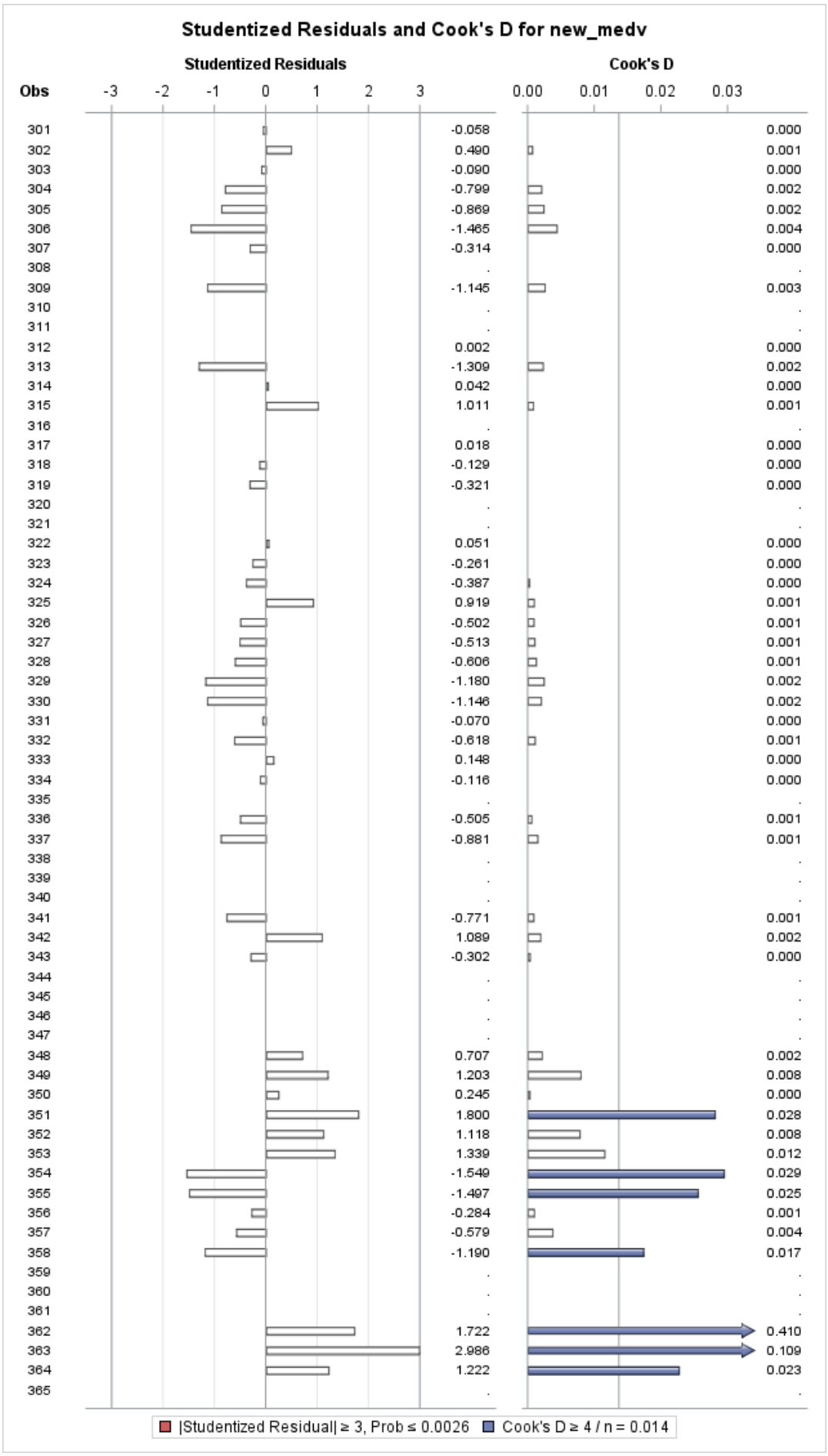

Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
301 -0.058 0.000
302 0.490 0.001
303 -0.090 0.000
304 -0.799 0.002
305 -0.869 0.002
306 -1.465 0.004
307 -0.314 0.000
308 . .
309 -1.145 0.003
310 . .
311 . .
312 0.002 0.000
313 -1.309 0.002
314 0.042 0.000
315 1.011 0.001
316 . .
317 0.018 0.000
318 -0.129 0.000
319 -0.321 0.000
320 . .
321 . .
322 0.051 0.000
323 -0.261 0.000
324 -0.387 0.000
325 0.919 0.001
326 -0.502 0.001
327 -0.513 0.001
328 -0.606 0.001
329 -1.180 0.002
330 -1.146 0.002
331 -0.070 0.000
332 -0.618 0.001
333 0.148 0.000
334 -0.116 0.000
335 . .
336 -0.505 0.001
337 -0.881 0.001
338 . .
339 . .
340 . .
341 -0.771 0.001
342 1.089 0.002
343 -0.302 0.000
344 . .
345 . .
346 . .
347 . .
348 0.707 0.002
349 1.203 0.008
350 0.245 0.000
351 1.800 0.028
352 1.118 0.008
353 1.339 0.012
354 -1.549 0.029
355 -1.497 0.025
356 -0.284 0.001
357 -0.579 0.004
358 -1.190 0.017
359 . .
360 . .
361 . .
362 1.722 0.410
363 2.986 0.109
364 1.222 0.023
365 . .
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014

Fig. 32

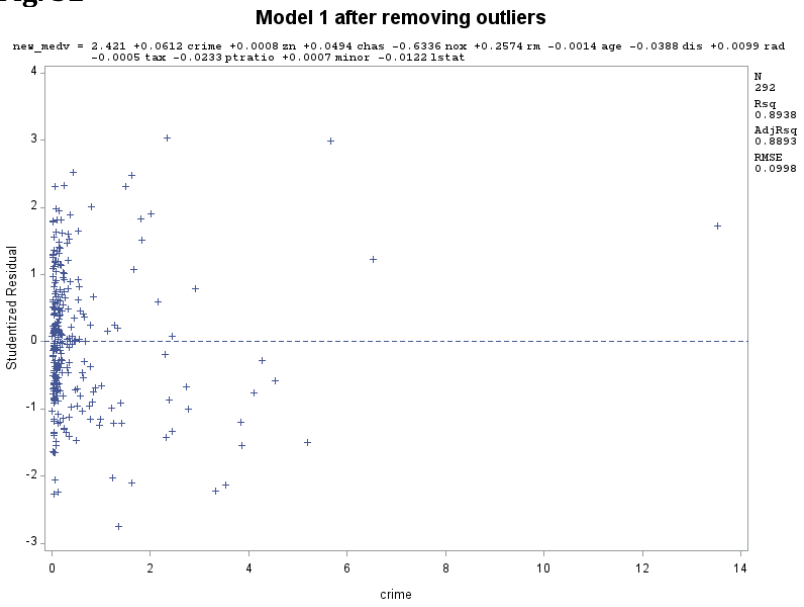


Fig. 33

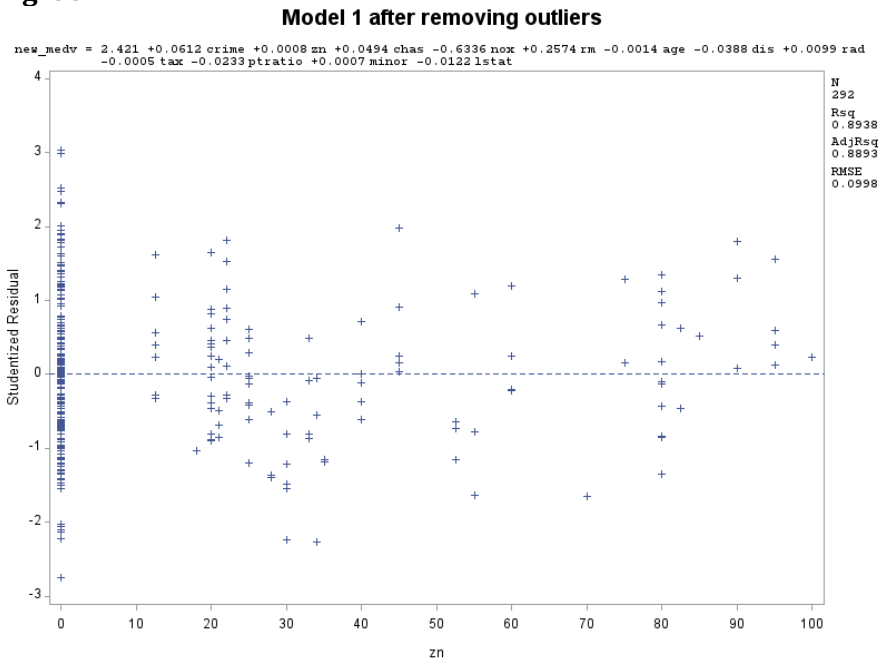


Fig. 34

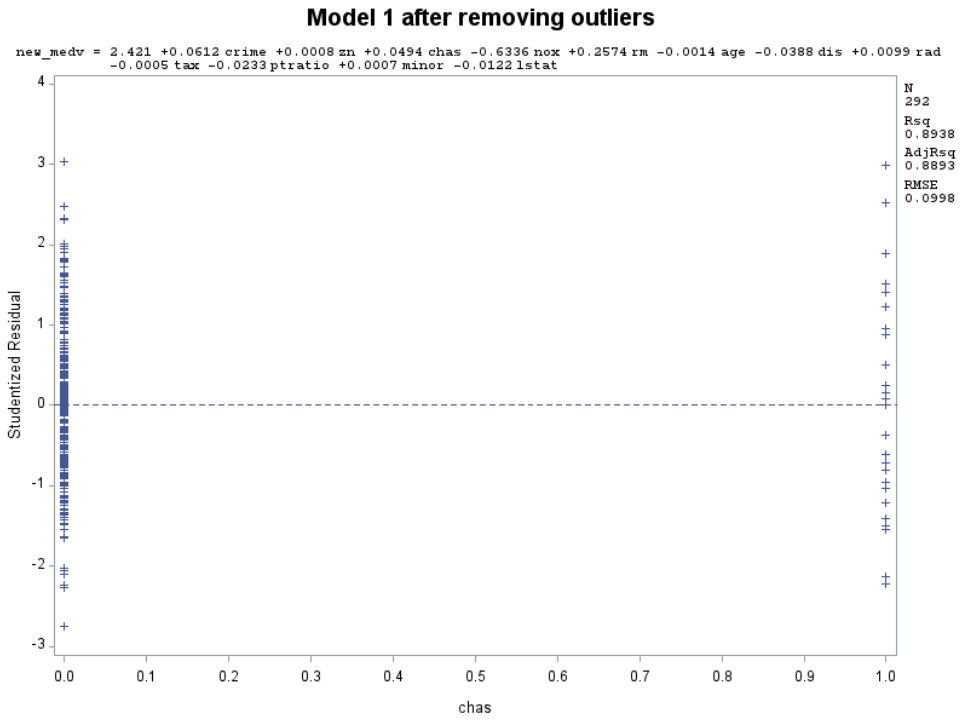

Model 1 after removing outliers
new_medv = 2.421 +0.0612 crime +0.0008 zn +0.0494 chas -0.6336 nox +0.2574 rm -0.0014 age -0.0388 dis +0.0099 rad
-0.0005 tax -0.0233 ptratio +0.0007 minor -0.0122 lstat
N
292
Rsq
0.8938
AdjRsq
0.8893
RMSE
0.0998
Studentized Residual
chas

Fig. 35

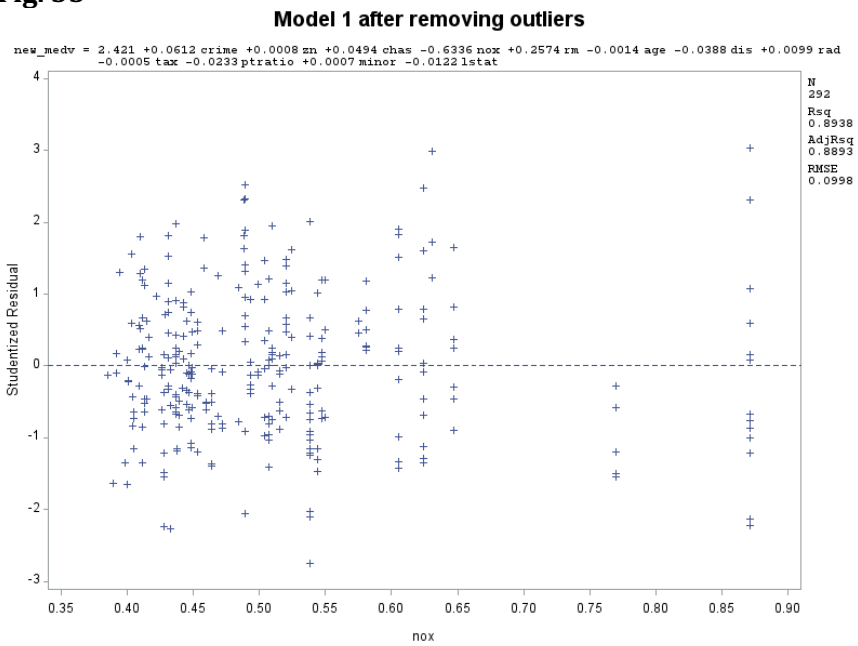

Model 1 after removing outliers
new_medv = 2.421 +0.0612 crime +0.0008 zn +0.0494 chas -0.6336 nox +0.2574 rm -0.0014 age -0.0388 dis +0.0099 rad
-0.0005 tax -0.0233 ptratio +0.0007 minor -0.0122 lstat
N
292
Rsq
0.8938
AdjRsq
0.8893
RMSE
0.0998
Studentized Residual
nox

Fig. 36

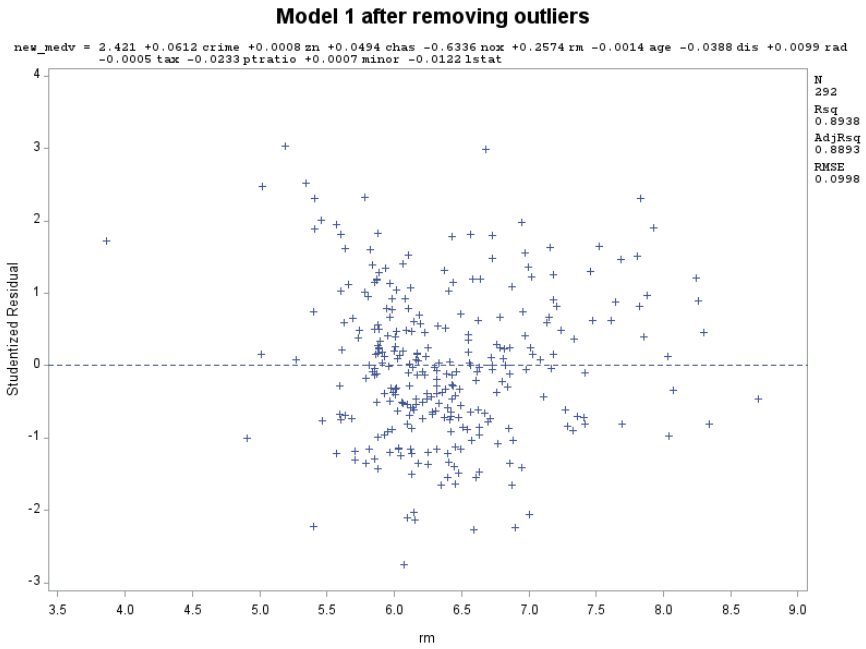

Model 1 after removing outliers
new_medv = 2.421 +0.0612 crime +0.0008 zn +0.0494 chas -0.6336 nox +0.2574 rm -0.0014 age -0.0388 dis +0.0099 rad
-0.0005 tax -0.0233 ptratio +0.0007 minor -0.0122 lstat
N
292
Rsq
0.8938
AdjRsq
0.8893
RMSE
0.0998
Studentized Residual
rm

Fig. 37

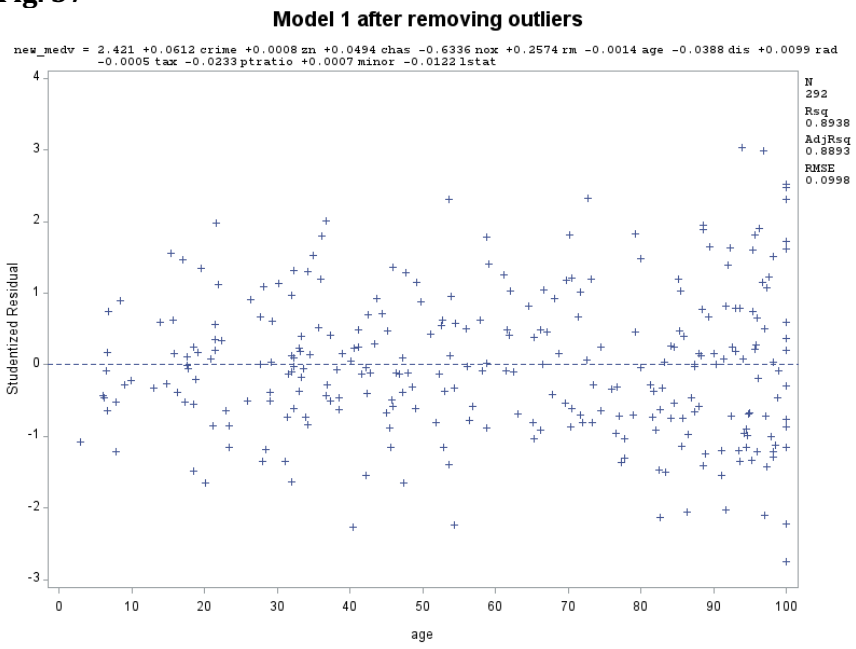

Model 1 after removing outliers
new_medv = 2.421 +0.0612 crime +0.0008 zn +0.0494 chas -0.6336 nox +0.2574 rm -0.0014 age -0.0388 dis +0.0099 rad
-0.0005 tax -0.0233 ptratio +0.0007 minor -0.0122 lstat
N
292
Rsq
0.8938
AdjRsq
0.8893
RMSE
0.0998
Studentized Residual
age

Fig. 38

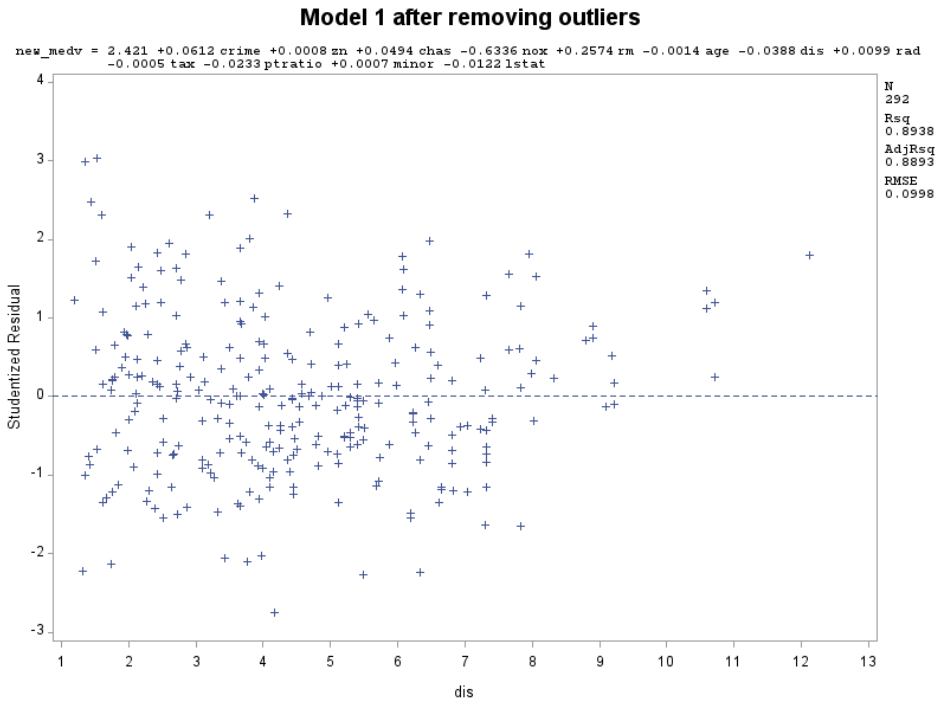

Model 1 after removing outliers
new_medv = 2.421 +0.0612 crime +0.0008 zn +0.0494 chas -0.6336 nox +0.2574 rm -0.0014 age -0.0388 dis +0.0099 rad
-0.0005 tax -0.0233 ptratio +0.0007 minor -0.0122 lstat
N
292
Rsq
0.8938
AdjRsq
0.8893
RMSE
0.0998
Studentized Residual
dis

Fig. 39

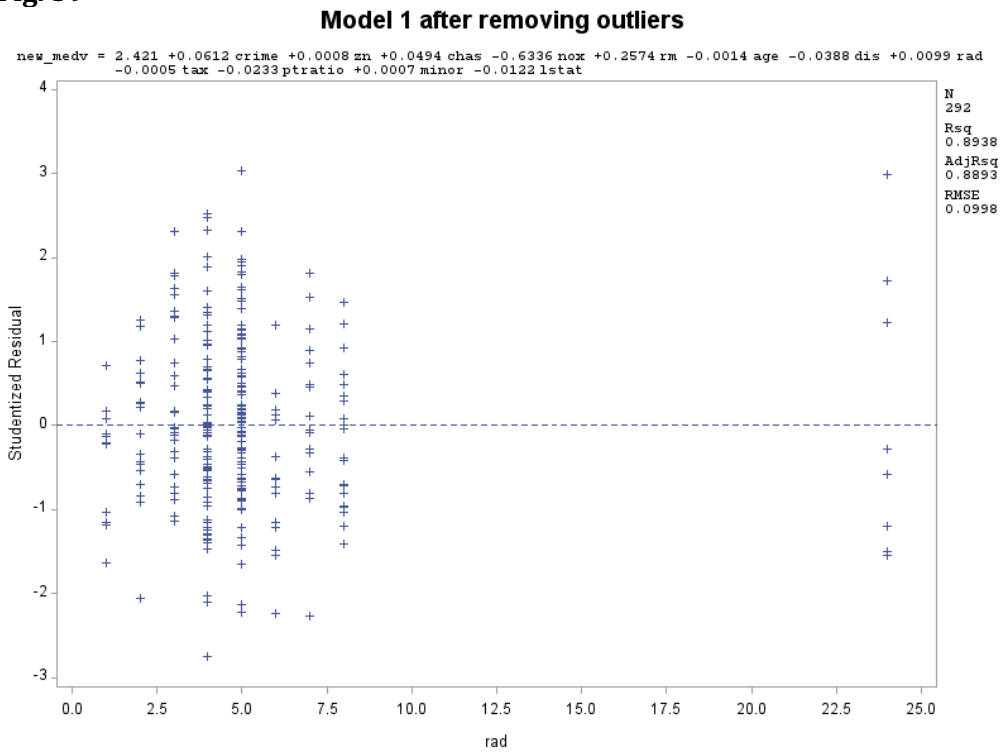

Model 1 after removing outliers
new_medv = 2.421 +0.0612 crime +0.0008 zn +0.0494 chas -0.6336 nox +0.2574 rm -0.0014 age -0.0388 dis +0.0099 rad
-0.0005 tax -0.0233 ptratio +0.0007 minor -0.0122 lstat
N
292
Rsq
0.8938
AdjRsq
0.8893
RMSE
0.0998
Studentized Residual
rad

Fig. 40

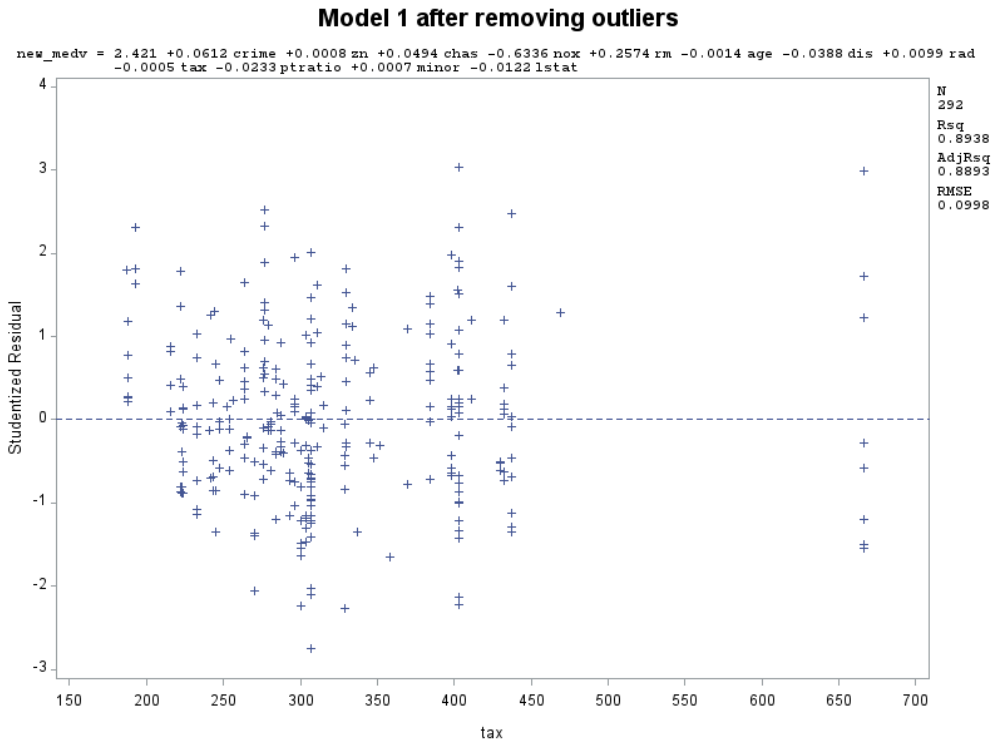

Model 1 after removing outliers
new_medv = 2.421 +0.0612 crime +0.0008 zn +0.0494 chas -0.6336 nox +0.2574 rm -0.0014 age -0.0388 dis +0.0099 rad
-0.0005 tax -0.0233 ptratio +0.0007 minor -0.0122 lstat
N
292
Rsq
0.8938
AdjRsq
0.8893
RMSE
0.0998
Studentized Residual
tax

Fig. 41

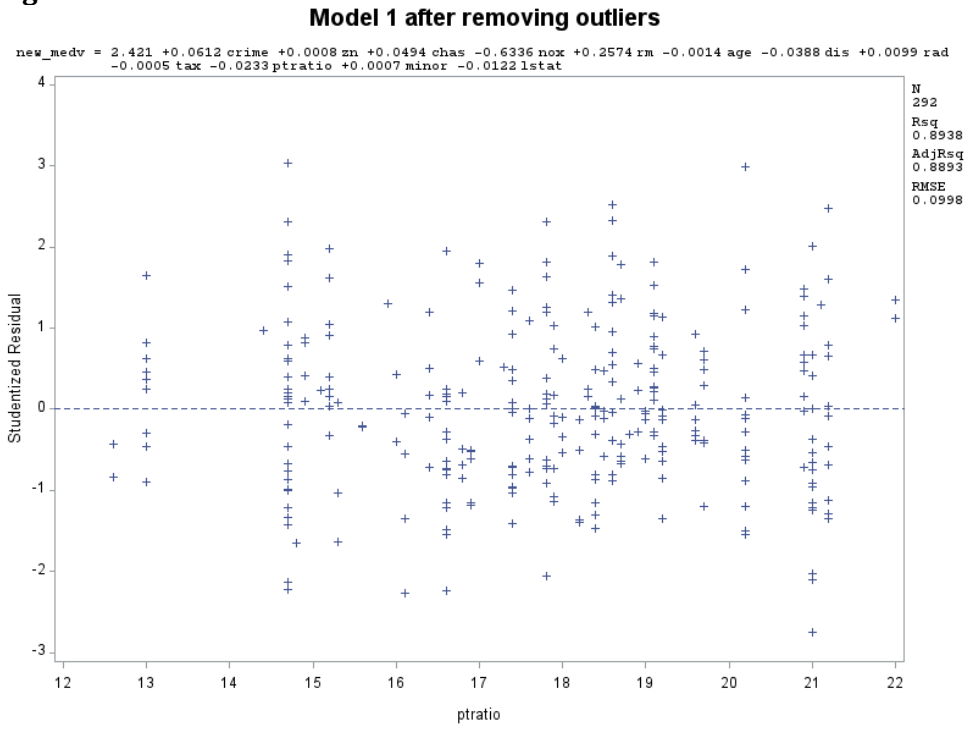

Model 1 after removing outliers
new_medv = 2.421 +0.0612 crime +0.0008 zn +0.0494 chas -0.6336 nox +0.2574 rm -0.0014 age -0.0388 dis +0.0099 rad
-0.0005 tax -0.0233 ptratio +0.0007 minor -0.0122 lstat
N
292
Rsq
0.8938
AdjRsq
0.8893
RMSE
0.0998
Studentized Residual
ptratio

Fig. 42

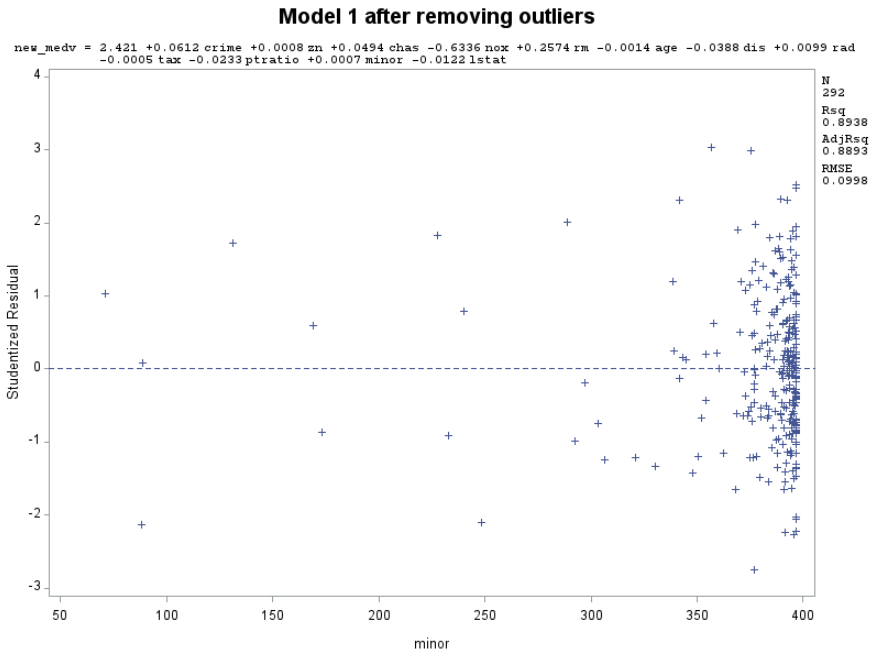

Model 1 after removing outliers
new_medv = 2.421 +0.0612 crime +0.0008 zn +0.0494 chas -0.6336 nox +0.2574 rm -0.0014 age -0.0388 dis +0.0099 rad
-0.0005 tax -0.0233 ptratio +0.0007 minor -0.0122 lstat
N
292
Rsq
0.8938
AdjRsq
0.8893
RMSE
0.0998
Studentized Residual
minor

Fig. 43

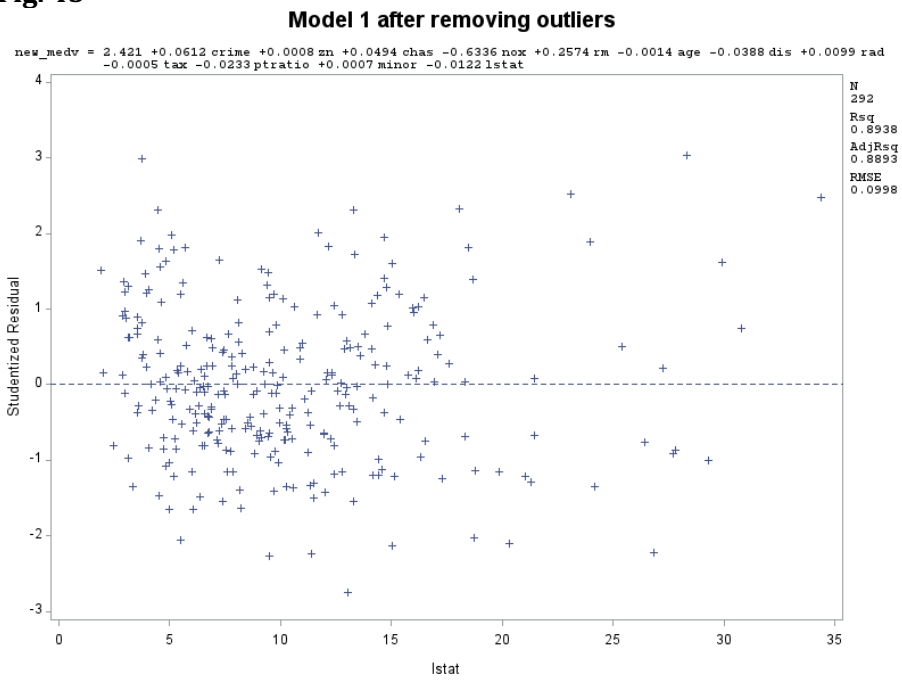

Model 1 after removing outliers
new_medv = 2.421 +0.0612 crime +0.0008 zn +0.0494 chas -0.6336 nox +0.2574 rm -0.0014 age -0.0388 dis +0.0099 rad
-0.0005 tax -0.0233 ptratio +0.0007 minor -0.0122 lstat
N
292
Rsq
0.8938
AdjRsq
0.8893
RMSE
0.0998
Studentized Residual
lstat

Fig. 44

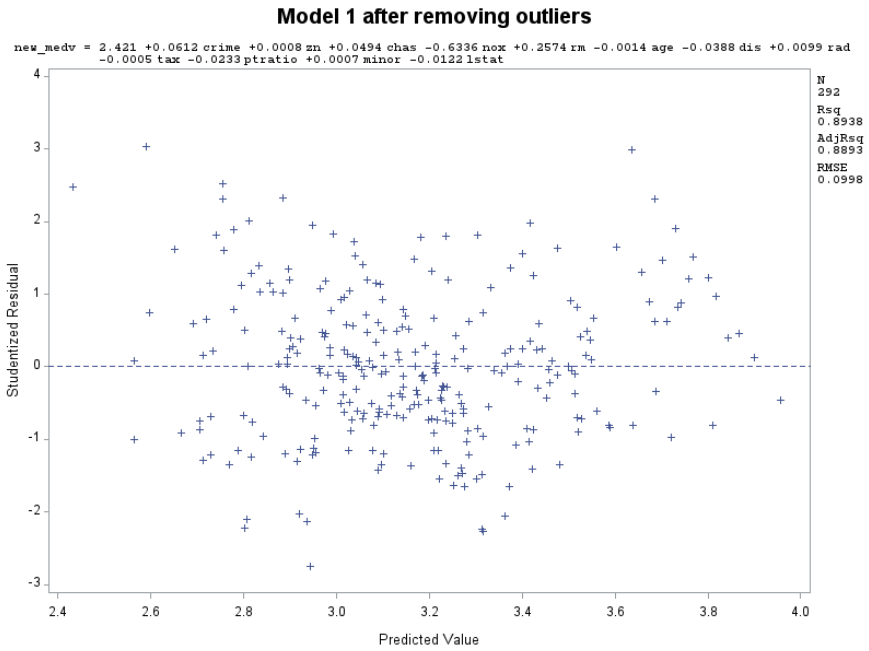


Fig. 45

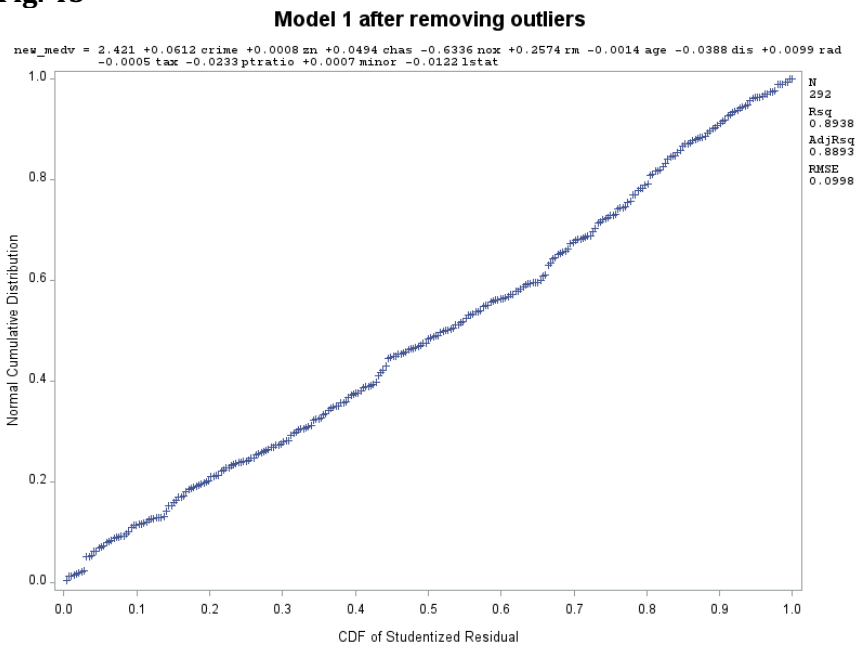


Table 4

Performance test parameters - Model 1

Obs	_TYPE_	_FREQ_	rmse	mae
1	0	73	9.57828	22.6199

Performance test parameters - Model 1

The CORR Procedure

2 Variables:	medv yhat

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**	**Label**
medv	73	25.79041	9.85777	1883	12.70000	50.00000	
yhat	73	3.17053	0.30402	231.44860	2.53534	3.91443	Predicted Value of new_medv

Pearson Correlation Coefficients, N = 73 Prob > |r| under H0: Rho=0		
	medv	**yhat**
medv	1.00000	0.92171 <.0001
yhat Predicted Value of new_medv	0.92171 <.0001	1.00000

APPENDIX D – Quadratic Polynomial Regression

Table 1

Quadratic Polynomial Regression

The GLMSELECT Procedure
Least Squares Model (No Selection)

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	91	25438	279.53963	60.20	<.0001
Error	275	1276.97106	4.64353		
Corrected Total	366	26715			

Root MSE	2.15489
Dependent Mean	25.04959
R-Square	0.9522
Adj R-Sq	0.9364
AIC	1010.60656
AICC	1074.65051
SBC	1000.89985

Table 2

Model 1- Checking Model assumptions and diagnostics

The REG Procedure
Model: MODEL1
Dependent Variable: new_medv

Number of Observations Read	367
Number of Observations Used	276
Number of Observations with Missing Values	91

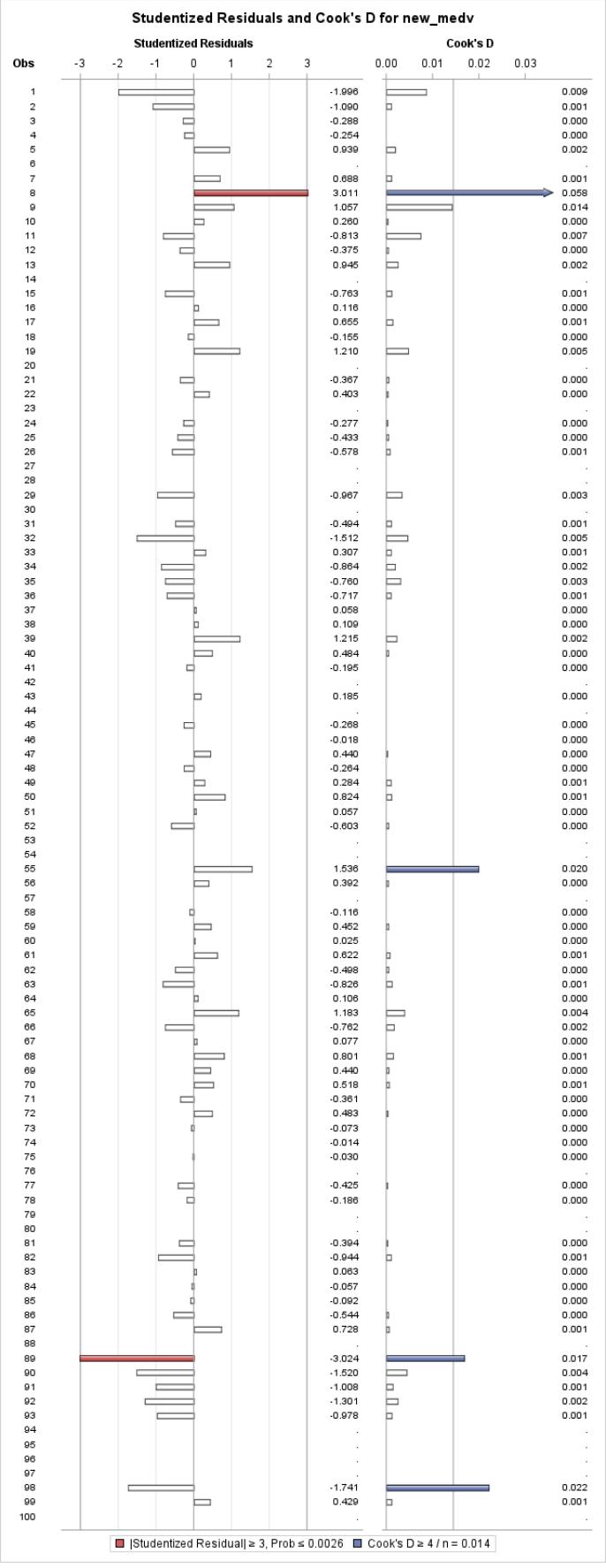
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	16519	3303.77401	341.57	<.0001
Error	270	2611.49196	9.67219		
Corrected Total	275	19130			

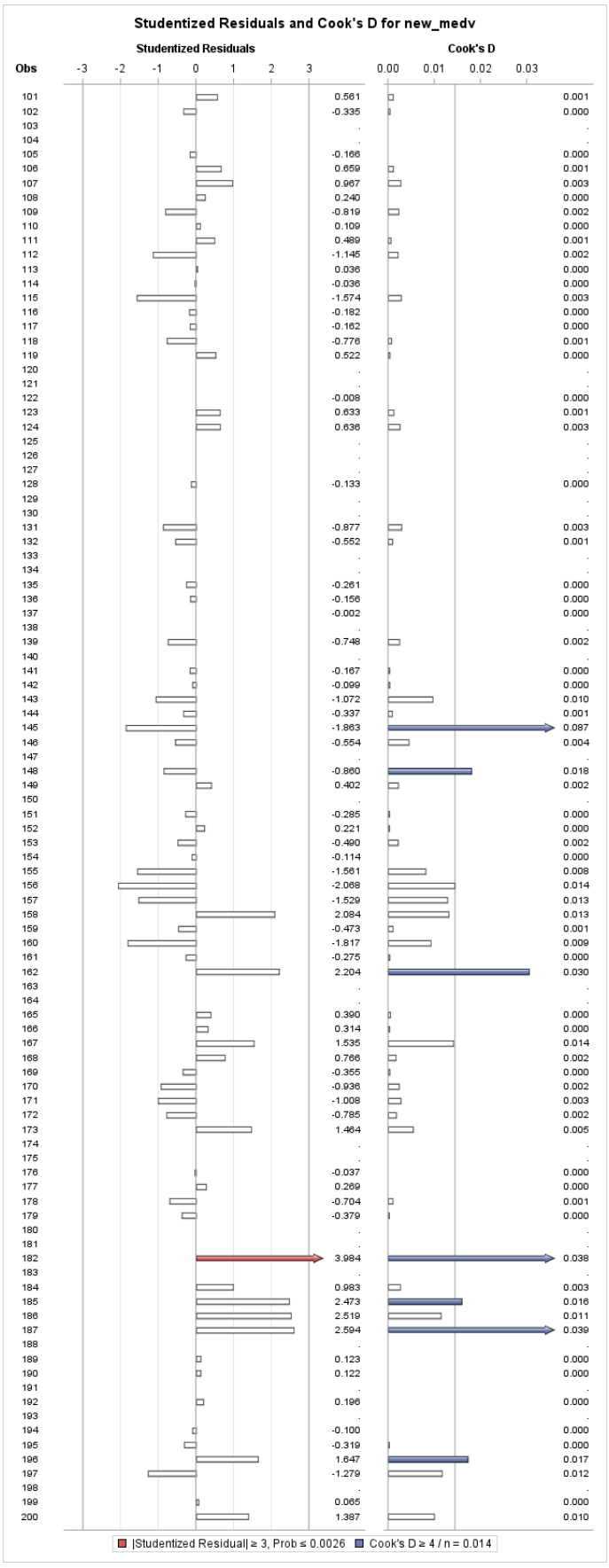
Root MSE	3.11001	**R-Square**	0.8635
Dependent Mean	24.78551	**Adj R-Sq**	0.8610
Coeff Var	12.54772		

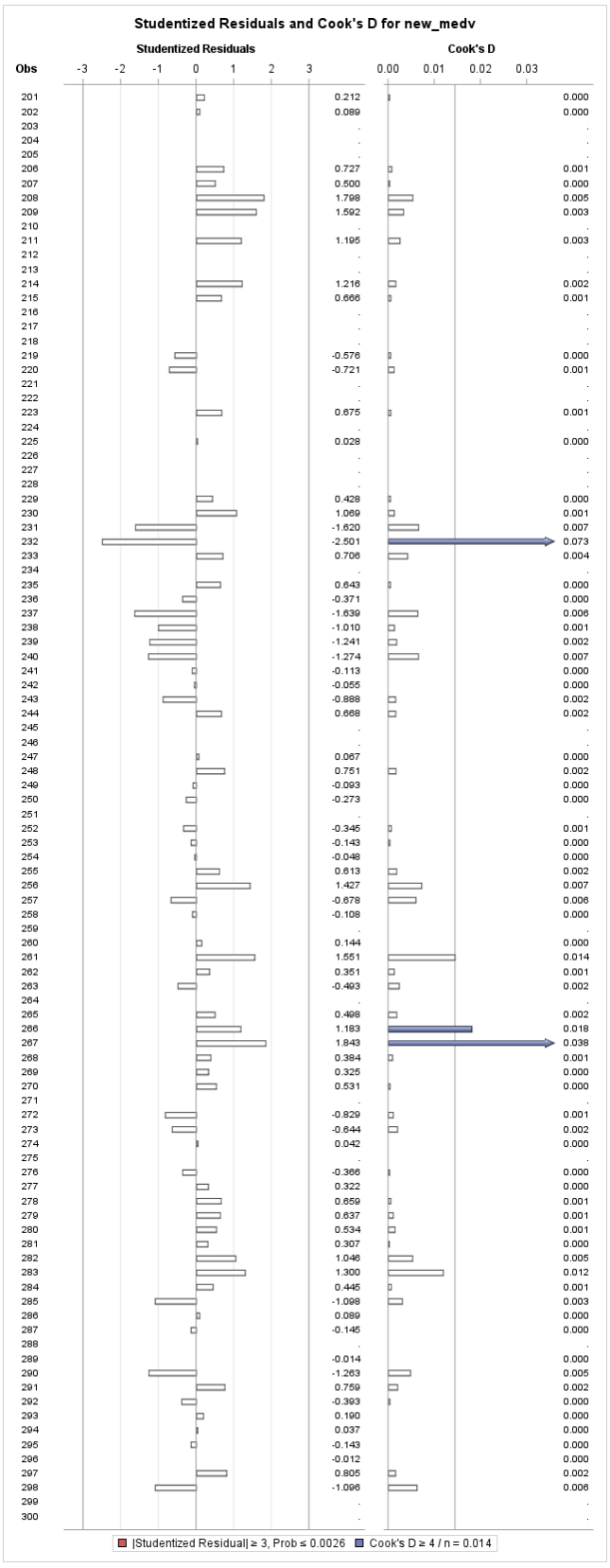
Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standardized Estimate**	**Variance Inflation**
Intercept	**1**	-40.38857	3.41815	-11.82	<.0001	0	0
rm	**1**	12.56605	0.48493	25.91	<.0001	1.00384	2.96823

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standardized Estimate**	**Variance Inflation**
dis_tax	**1**	-0.00151	0.00029123	-5.18	<.0001	-0.12445	1.14352
rm_ptratio	**1**	-0.06595	0.01397	-4.72	<.0001	-0.12075	1.29385
lstat	**1**	2.33246	0.28251	8.26	<.0001	1.67609	81.51468
rm_lstat	**1**	-0.46491	0.04915	-9.46	<.0001	-1.77338	69.53019

Fig. 1


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
1 -1.996 0.009
2 -1.090 0.001
3 -0.288 0.000
4 -0.254 0.000
5 0.939 0.002
6 . .
7 0.688 0.001
8 3.011 0.058
9 1.057 0.014
10 0.260 0.000
11 -0.813 0.007
12 -0.375 0.000
13 0.945 0.002
14 . .
15 -0.763 0.001
16 0.116 0.000
17 0.655 0.001
18 -0.155 0.000
19 1.210 0.005
20 . .
21 -0.367 0.000
22 0.403 0.000
23 . .
24 -0.277 0.000
25 -0.433 0.000
26 -0.578 0.001
27 . .
28 . .
29 -0.967 0.003
30 . .
31 -0.494 0.001
32 -1.512 0.005
33 0.307 0.001
34 -0.864 0.002
35 -0.760 0.003
36 -0.717 0.001
37 0.058 0.000
38 0.109 0.000
39 1.215 0.002
40 0.484 0.000
41 -0.195 0.000
42 . .
43 0.185 0.000
44 . .
45 -0.268 0.000
46 -0.018 0.000
47 0.440 0.000
48 -0.264 0.000
49 0.284 0.001
50 0.824 0.001
51 0.057 0.000
52 -0.603 0.000
53 . .
54 . .
55 1.536 0.020
56 0.392 0.000
57 . .
58 -0.116 0.000
59 0.452 0.000
60 0.025 0.000
61 0.622 0.001
62 -0.498 0.000
63 -0.826 0.001
64 0.106 0.000
65 1.183 0.004
66 -0.762 0.002
67 0.077 0.000
68 0.801 0.001
69 0.440 0.000
70 0.518 0.001
71 -0.361 0.000
72 0.483 0.000
73 -0.073 0.000
74 -0.014 0.000
75 -0.030 0.000
76 . .
77 -0.425 0.000
78 -0.186 0.000
79 . .
80 . .
81 -0.394 0.000
82 -0.944 0.001
83 0.063 0.000
84 -0.057 0.000
85 -0.092 0.000
86 -0.544 0.000
87 0.728 0.001
88 . .
89 -3.024 0.017
90 -1.520 0.004
91 -1.008 0.001
92 -1.301 0.002
93 -0.978 0.001
94 . .
95 . .
96 . .
97 . .
98 -1.741 0.022
99 0.429 0.001
100 . .
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
101 0.561 0.001
102 -0.335 0.000
103 . .
104 . .
105 -0.166 0.000
106 0.659 0.001
107 0.967 0.003
108 0.240 0.000
109 -0.819 0.002
110 0.109 0.000
111 0.489 0.001
112 -1.145 0.002
113 0.036 0.000
114 -0.036 0.000
115 -1.574 0.003
116 -0.182 0.000
117 -0.162 0.000
118 -0.776 0.001
119 0.522 0.000
120 . .
121 . .
122 -0.008 0.000
123 0.633 0.001
124 0.636 0.003
125 . .
126 . .
127 . .
128 -0.133 0.000
129 . .
130 . .
131 -0.877 0.003
132 -0.552 0.001
133 . .
134 . .
135 -0.261 0.000
136 -0.156 0.000
137 -0.002 0.000
138 . .
139 -0.748 0.002
140 . .
141 -0.167 0.000
142 -0.099 0.000
143 -1.072 0.010
144 -0.337 0.001
145 -1.863 0.087
146 -0.554 0.004
147 . .
148 -0.860 0.018
149 0.402 0.002
150 . .
151 -0.285 0.000
152 0.221 0.000
153 -0.490 0.002
154 -0.114 0.000
155 -1.561 0.008
156 -2.068 0.014
157 -1.529 0.013
158 2.084 0.013
159 -0.473 0.001
160 -1.817 0.009
161 -0.275 0.000
162 2.204 0.030
163 . .
164 . .
165 0.390 0.000
166 0.314 0.000
167 1.535 0.014
168 0.766 0.002
169 -0.355 0.000
170 -0.936 0.002
171 -1.008 0.003
172 -0.785 0.002
173 1.464 0.005
174 . .
175 . .
176 -0.037 0.000
177 0.269 0.000
178 -0.704 0.001
179 -0.379 0.000
180 . .
181 . .
182 3.984 0.038
183 . .
184 0.983 0.003
185 2.473 0.016
186 2.519 0.011
187 2.594 0.039
188 . .
189 0.123 0.000
190 0.122 0.000
191 . .
192 0.196 0.000
193 . .
194 -0.100 0.000
195 -0.319 0.000
196 1.647 0.017
197 -1.279 0.012
198 . .
199 0.065 0.000
200 1.387 0.010
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
201 0.212 0.000
202 0.089 0.000
206 0.727 0.001
207 0.500 0.000
208 1.798 0.005
209 1.592 0.003
211 1.195 0.003
214 1.216 0.002
215 0.666 0.001
219 -0.576 0.000
220 -0.721 0.001
223 0.675 0.001
225 0.028 0.000
229 0.428 0.000
230 1.069 0.001
231 -1.620 0.007
232 -2.501 0.073
233 0.706 0.004
235 0.643 0.000
236 -0.371 0.000
237 -1.639 0.006
238 -1.010 0.001
239 -1.241 0.002
240 -1.274 0.007
241 -0.113 0.000
242 -0.055 0.000
243 -0.888 0.002
244 0.668 0.002
247 0.067 0.000
248 0.751 0.002
249 -0.093 0.000
250 -0.273 0.000
252 -0.345 0.001
253 -0.143 0.000
254 -0.048 0.000
255 0.613 0.002
256 1.427 0.007
257 -0.678 0.006
258 -0.108 0.000
260 0.144 0.000
261 1.551 0.014
262 0.351 0.001
263 -0.493 0.002
265 0.498 0.002
266 1.183 0.018
267 1.843 0.038
268 0.384 0.001
269 0.325 0.000
270 0.531 0.000
272 -0.829 0.001
273 -0.644 0.002
274 0.042 0.000
276 -0.366 0.000
277 0.322 0.000
278 0.659 0.001
279 0.637 0.001
280 0.534 0.001
281 0.307 0.000
282 1.046 0.005
283 1.300 0.012
284 0.445 0.001
285 -1.098 0.003
286 0.089 0.000
287 -0.145 0.000
289 -0.014 0.000
290 -1.263 0.005
291 0.759 0.002
292 -0.393 0.000
293 0.190 0.000
294 0.037 0.000
295 -0.143 0.000
296 -0.012 0.000
297 0.805 0.002
298 -1.096 0.006
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014

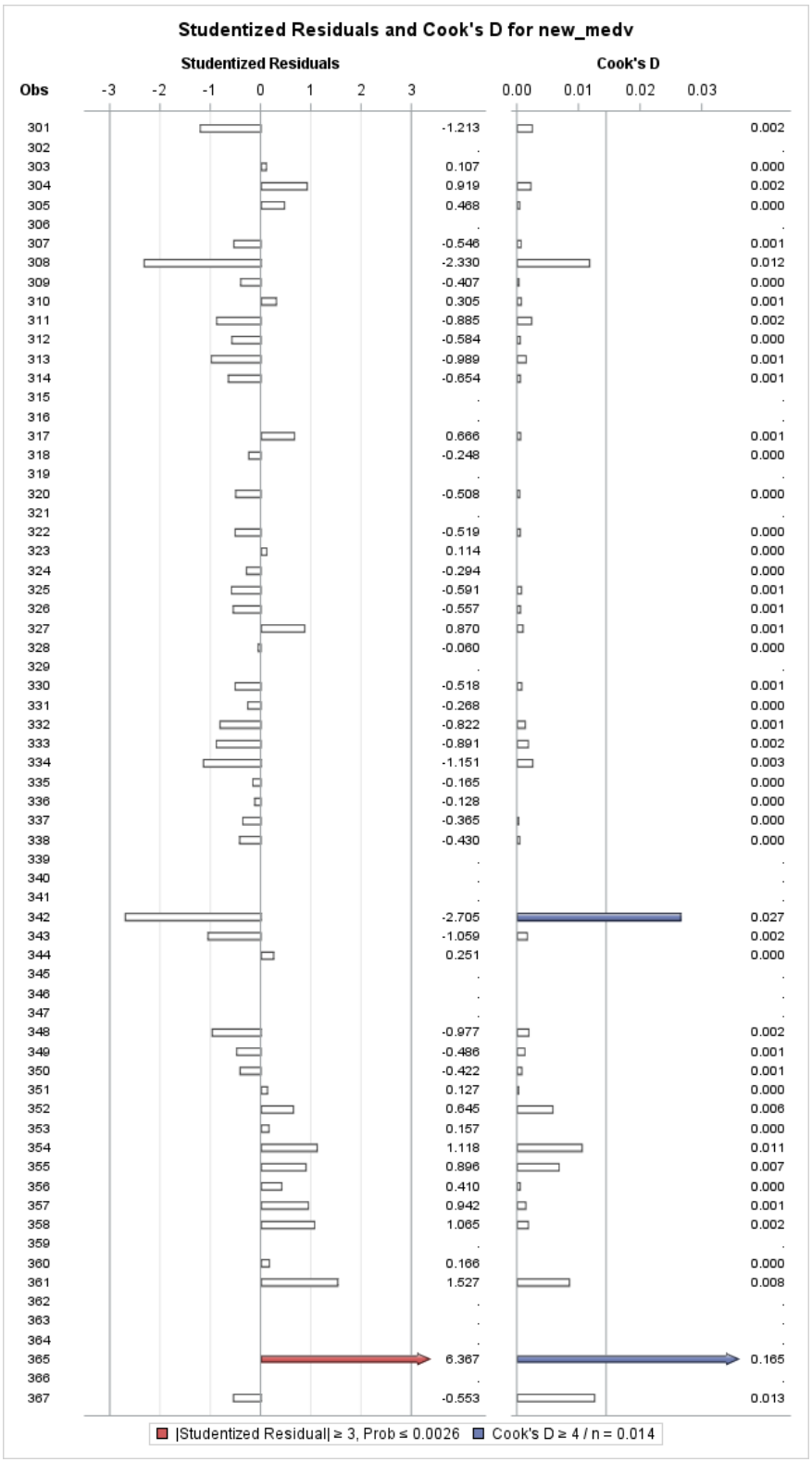

Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
301 -1.213 0.002
302 . .
303 0.107 0.000
304 0.919 0.002
305 0.468 0.000
306 . .
307 -0.546 0.001
308 -2.330 0.012
309 -0.407 0.000
310 0.305 0.001
311 -0.885 0.002
312 -0.584 0.000
313 -0.989 0.001
314 -0.654 0.001
315 . .
316 . .
317 0.666 0.001
318 -0.248 0.000
319 . .
320 -0.508 0.000
321 . .
322 -0.519 0.000
323 0.114 0.000
324 -0.294 0.000
325 -0.591 0.001
326 -0.557 0.001
327 0.870 0.001
328 -0.060 0.000
329 . .
330 -0.518 0.001
331 -0.268 0.000
332 -0.822 0.001
333 -0.891 0.002
334 -1.151 0.003
335 -0.165 0.000
336 -0.128 0.000
337 -0.365 0.000
338 -0.430 0.000
339 . .
340 . .
341 . .
342 -2.705 0.027
343 -1.059 0.002
344 0.251 0.000
345 . .
346 . .
347 . .
348 -0.977 0.002
349 -0.486 0.001
350 -0.422 0.001
351 0.127 0.000
352 0.645 0.006
353 0.157 0.000
354 1.118 0.011
355 0.896 0.007
356 0.410 0.000
357 0.942 0.001
358 1.065 0.002
359 . .
360 0.166 0.000
361 1.527 0.008
362 . .
363 . .
364 . .
365 6.367 0.165
366 . .
367 -0.553 0.013
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014

Fig. 2

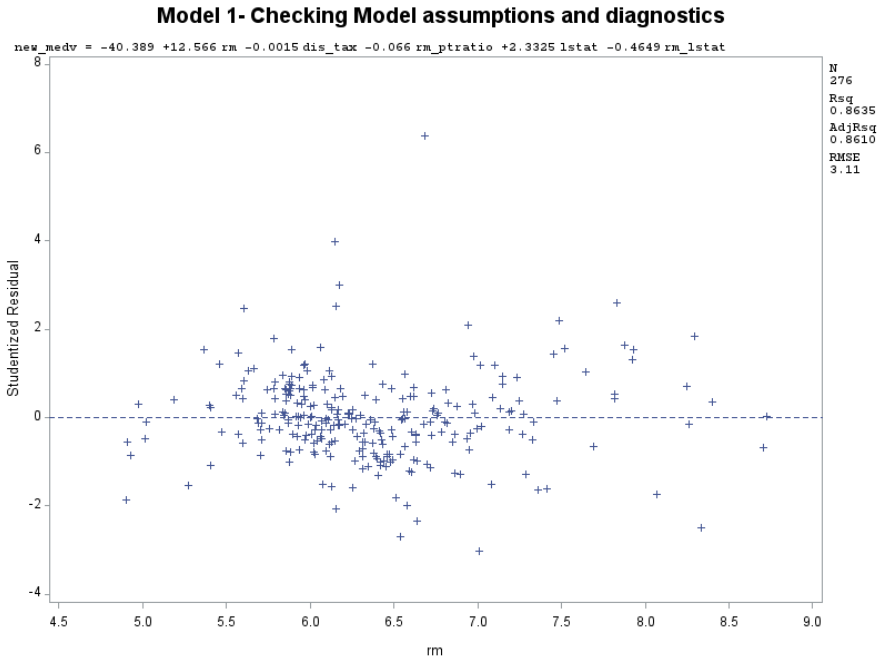

Model 1- Checking Model assumptions and diagnostics
new_medv = -40.389 +12.566 rm -0.0015 dis_tax -0.066 rm_ptratio +2.3325 lstat -0.4649 rm_lstat
N
276
Rsq
0.8635
AdjRsq
0.8610
RMSE
3.11
Studentized Residual
rm

Fig. 3

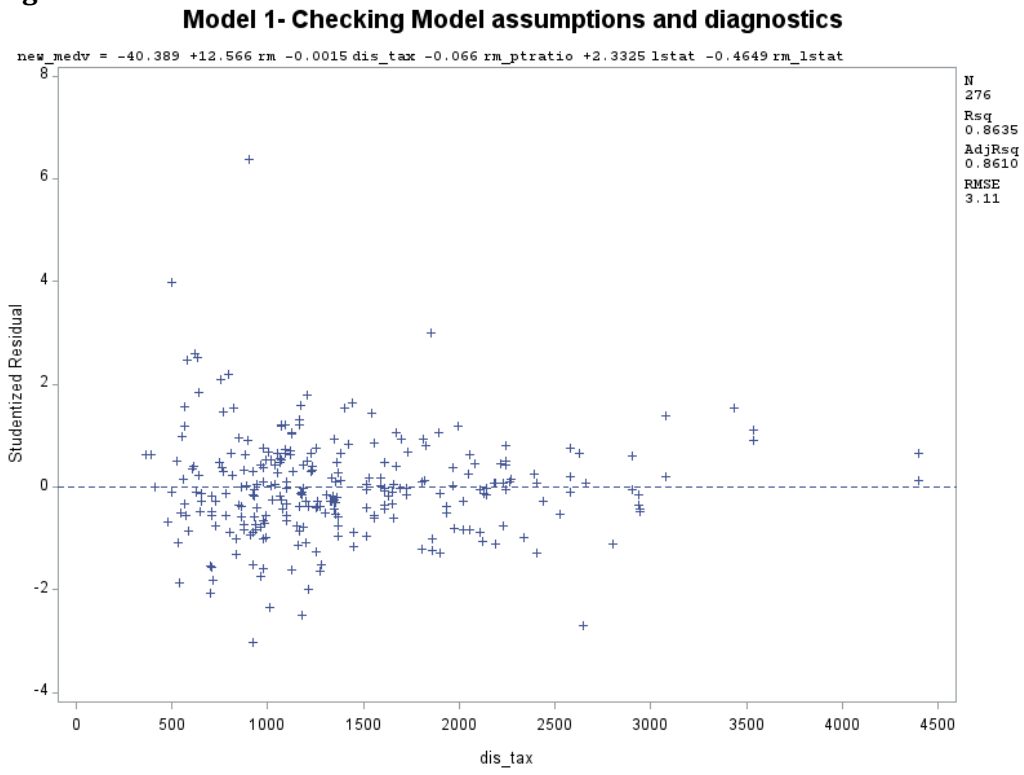

Model 1- Checking Model assumptions and diagnostics
new_medv = -40.389 +12.566 rm -0.0015 dis_tax -0.066 rm_ptratio +2.3325 lstat -0.4649 rm_lstat
N
276
Rsq
0.8635
AdjRsq
0.8610
RMSE
3.11
Studentized Residual
dis_tax

Fig. 4

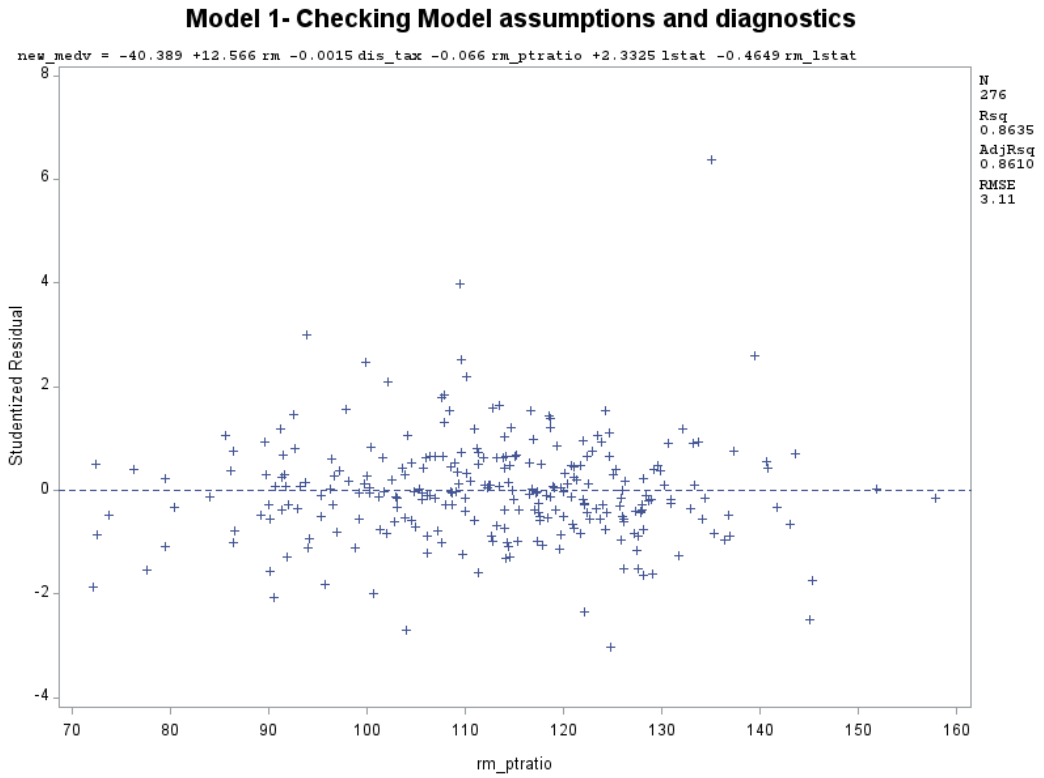

Model 1- Checking Model assumptions and diagnostics
new_medv = -40.389 +12.566 rm -0.0015 dis_tax -0.066 rm_ptratio +2.3325 lstat -0.4649 rm_lstat
N
276
Rsq
0.8635
AdjRsq
0.8610
RMSE
3.11
Studentized Residual
rm_ptratio

Fig. 5

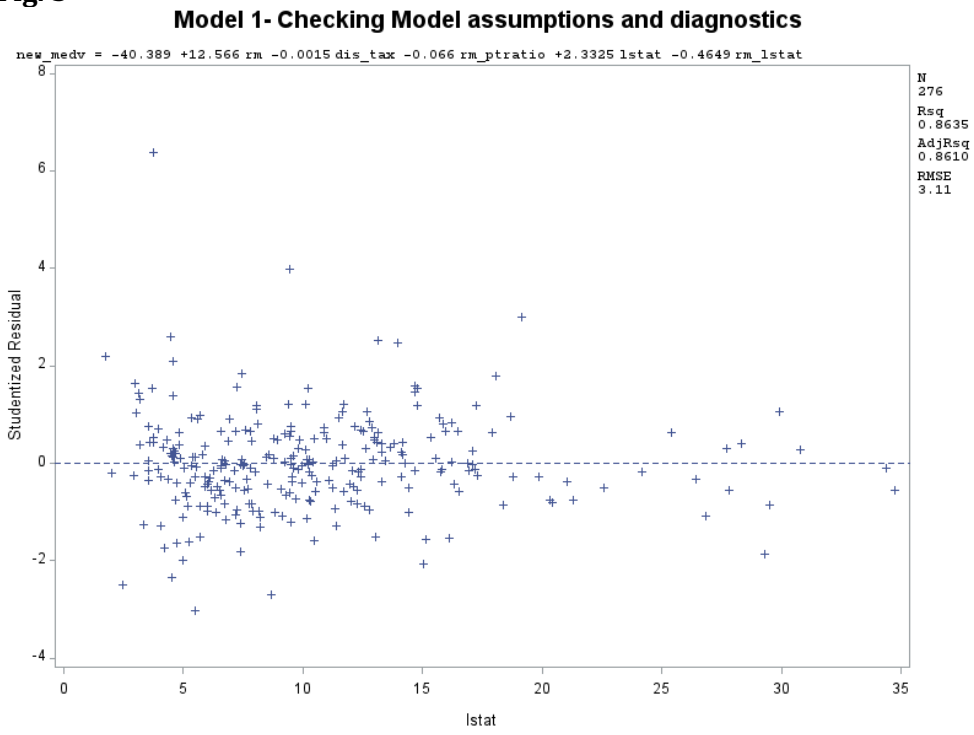

Model 1- Checking Model assumptions and diagnostics
new_medv = -40.389 +12.566 rm -0.0015 dis_tax -0.066 rm_ptratio +2.3325 lstat -0.4649 rm_lstat
N
276
Rsq
0.8635
AdjRsq
0.8610
RMSE
3.11
Studentized Residual
lstat

Fig. 6

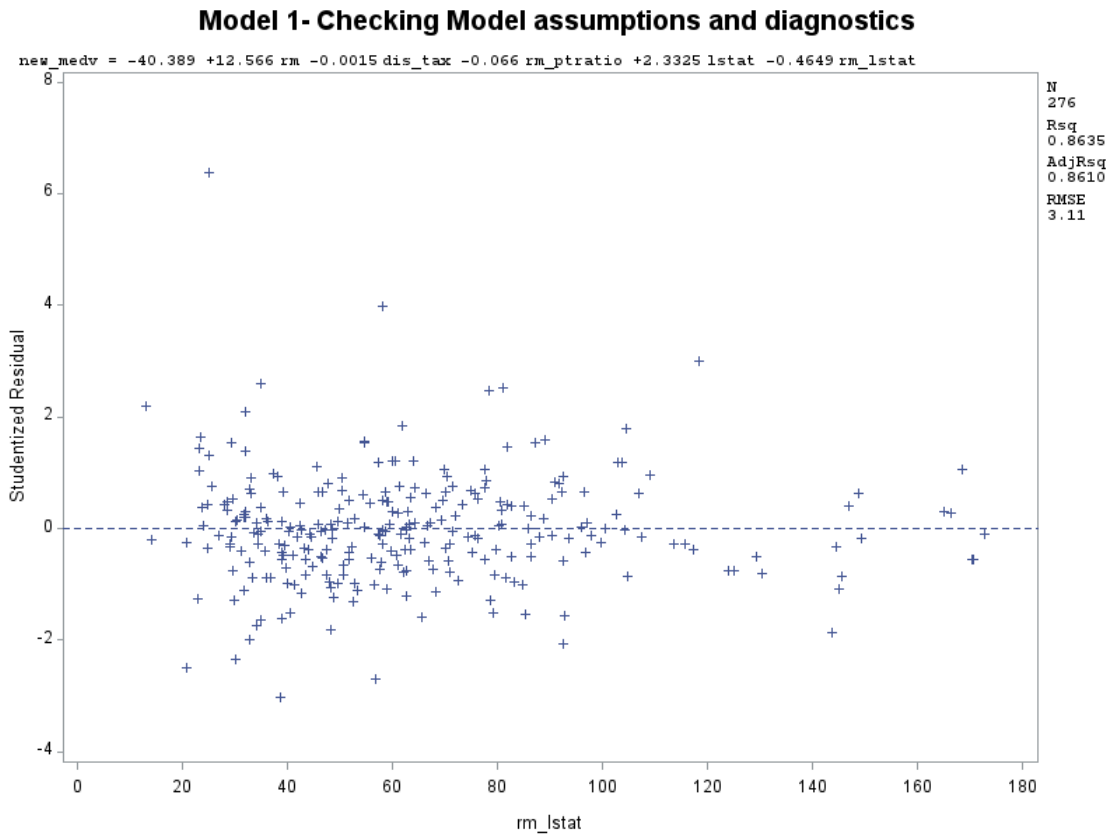

Model 1- Checking Model assumptions and diagnostics
new_medv = -40.389 +12.566 rm -0.0015 dis_tax -0.066 rm_ptratio +2.3325 lstat -0.4649 rm_lstat
N
276
Rsq
0.8635
AdjRsq
0.8610
RMSE
3.11
Studentized Residual
rm_lstat

Fig. 7

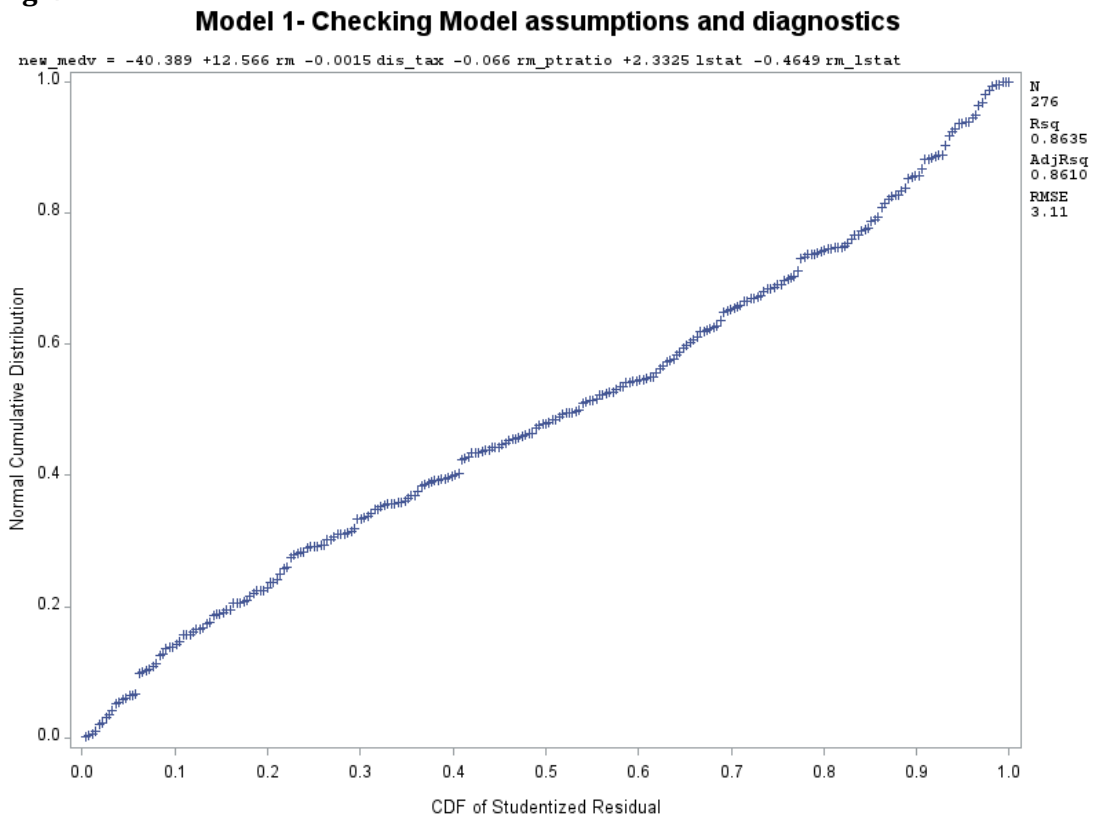

Model 1- Checking Model assumptions and diagnostics
new_medv = -40.389 +12.566 rm -0.0015 dis_tax -0.066 rm_ptratio +2.3325 lstat -0.4649 rm_lstat
N
276
Rsq
0.8635
AdjRsq
0.8610
RMSE
3.11
Normal Cumulative Distribution
CDF of Studentized Residual

Table 3

The REG Procedure
Model: MODEL1
Dependent Variable: new_medv

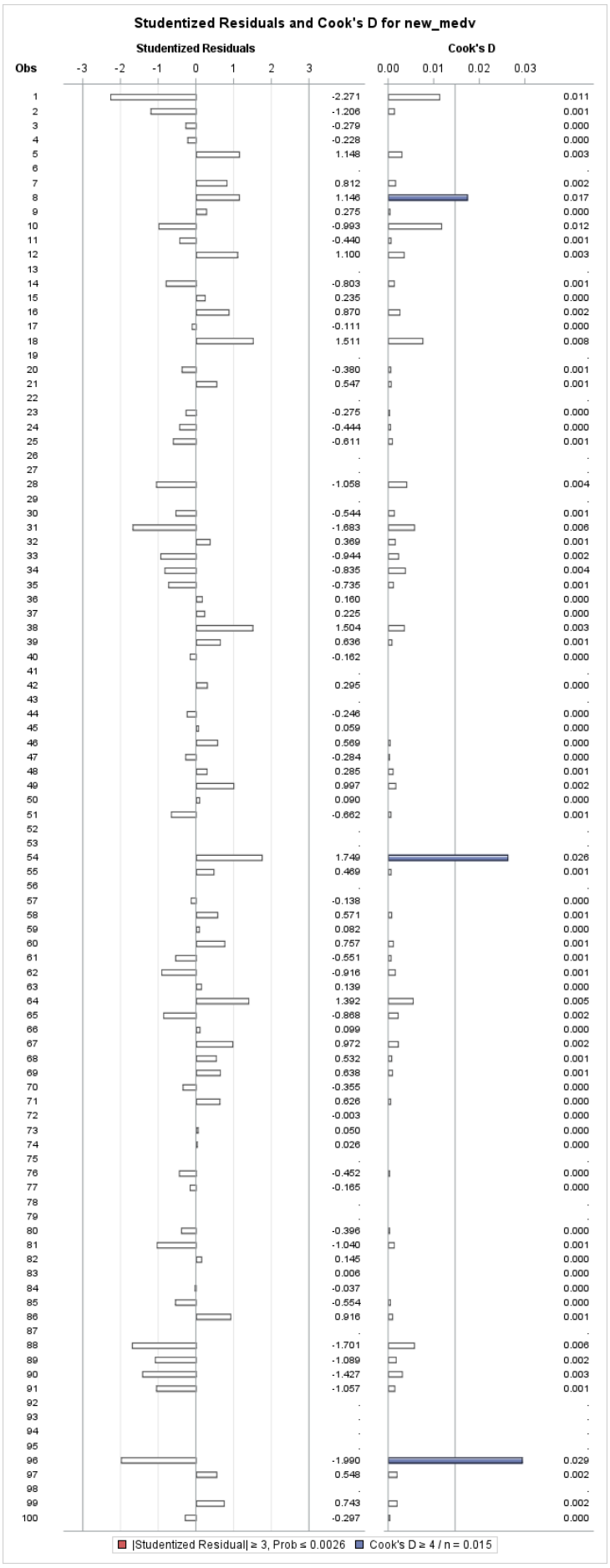
Number of Observations Read	363
Number of Observations Used	272
Number of Observations with Missing Values	91

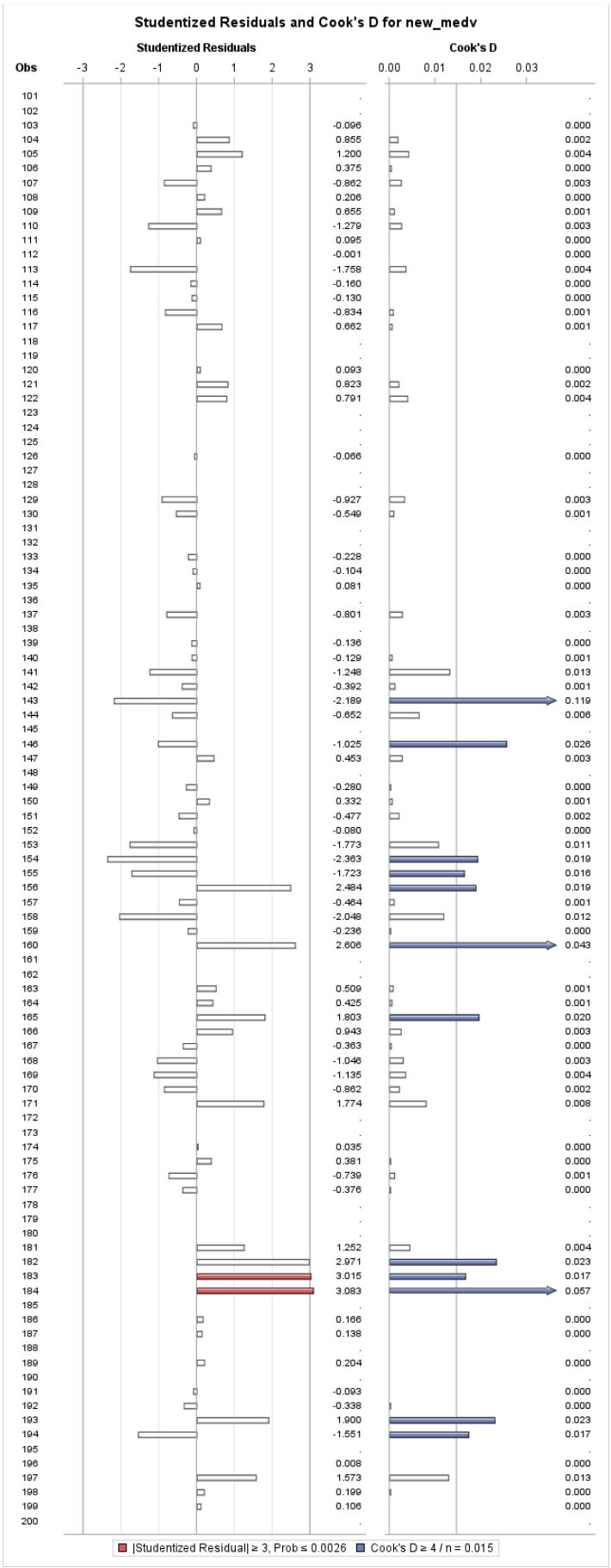
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	16454	3290.74414	461.05	<.0001
Error	266	1898.57631	7.13750		
Corrected Total	271	18352			

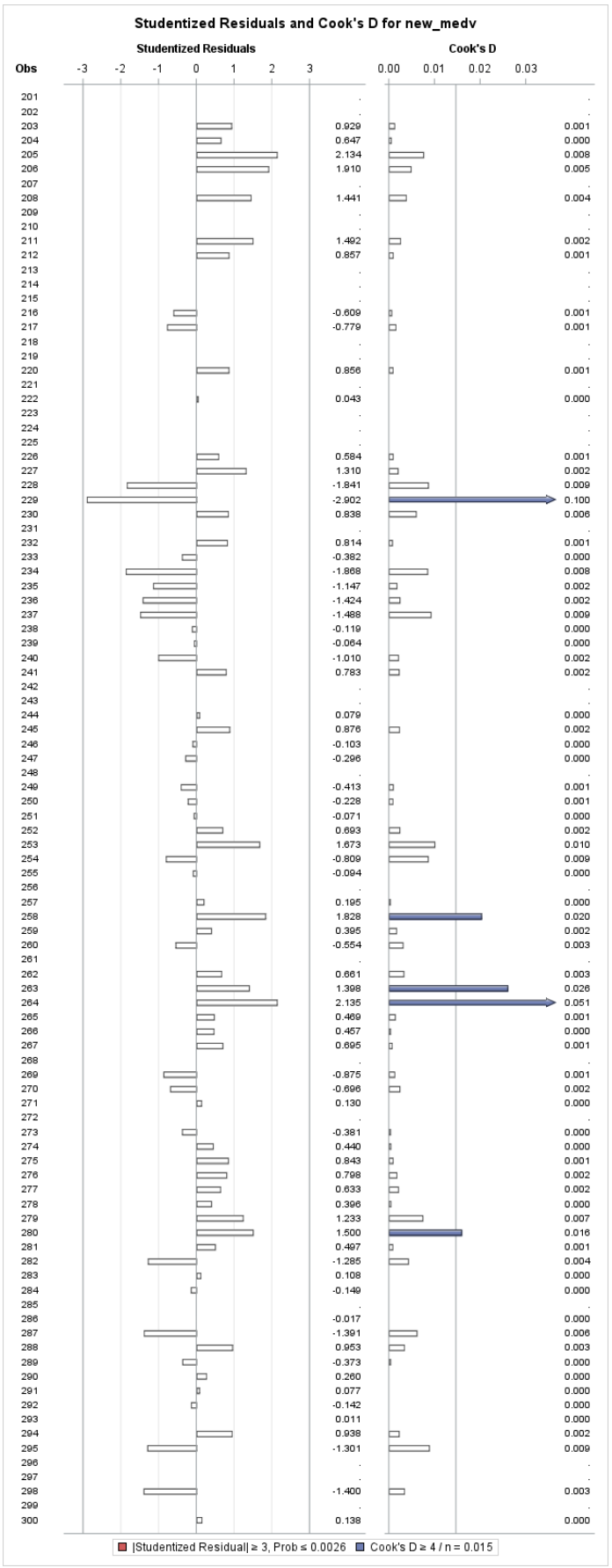
Root MSE	2.67161	**R-Square**	0.8965
Dependent Mean	24.64669	**Adj R-Sq**	0.8946
Coeff Var	10.83963		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standardized Estimate**	**Variance Inflation**
Intercept	**1**	-41.58632	2.94417	-14.12	<.0001	0	0
rm	**1**	12.74890	0.41748	30.54	<.0001	1.03728	2.96662
dis_tax	**1**	-0.00137	0.00025347	-5.41	<.0001	-0.11484	1.16055
rm_ptratio	**1**	-0.06953	0.01211	-5.74	<.0001	-0.12894	1.29691
lstat	**1**	2.37504	0.24389	9.74	<.0001	1.72954	81.10798
rm_lstat	**1**	-0.46988	0.04254	-11.05	<.0001	-1.81295	69.26354

Fig. 8


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
1 -2.271 0.011
2 -1.206 0.001
3 -0.279 0.000
4 -0.228 0.000
5 1.148 0.003
6 . .
7 0.812 0.002
8 1.146 0.017
9 0.275 0.000
10 -0.993 0.012
11 -0.440 0.001
12 1.100 0.003
13 . .
14 -0.803 0.001
15 0.235 0.000
16 0.870 0.002
17 -0.111 0.000
18 1.511 0.008
19 . .
20 -0.380 0.001
21 0.547 0.001
22 . .
23 -0.275 0.000
24 -0.444 0.000
25 -0.611 0.001
26 . .
27 . .
28 -1.058 0.004
29 . .
30 -0.544 0.001
31 -1.683 0.006
32 0.369 0.001
33 -0.944 0.002
34 -0.835 0.004
35 -0.735 0.001
36 0.160 0.000
37 0.225 0.000
38 1.504 0.003
39 0.636 0.001
40 -0.162 0.000
41 . .
42 0.295 0.000
43 . .
44 -0.246 0.000
45 0.059 0.000
46 0.569 0.000
47 -0.284 0.000
48 0.285 0.001
49 0.997 0.002
50 0.090 0.000
51 -0.662 0.001
52 . .
53 . .
54 1.749 0.026
55 0.469 0.001
56 . .
57 -0.138 0.000
58 0.571 0.001
59 0.082 0.000
60 0.757 0.001
61 -0.551 0.001
62 -0.916 0.001
63 0.139 0.000
64 1.392 0.005
65 -0.868 0.002
66 0.099 0.000
67 0.972 0.002
68 0.532 0.001
69 0.638 0.001
70 -0.355 0.000
71 0.626 0.000
72 -0.003 0.000
73 0.050 0.000
74 0.026 0.000
75 . .
76 -0.452 0.000
77 -0.165 0.000
78 . .
79 . .
80 -0.396 0.000
81 -1.040 0.001
82 0.145 0.000
83 0.006 0.000
84 -0.037 0.000
85 -0.554 0.000
86 0.916 0.001
87 . .
88 -1.701 0.006
89 -1.089 0.002
90 -1.427 0.003
91 -1.057 0.001
92 . .
93 . .
94 . .
95 . .
96 -1.990 0.029
97 0.548 0.002
98 . .
99 0.743 0.002
100 -0.297 0.000
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.015


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
101 .
102 .
103 -0.096 0.000
104 0.855 0.002
105 1.200 0.004
106 0.375 0.000
107 -0.862 0.003
108 0.206 0.000
109 0.655 0.001
110 -1.279 0.003
111 0.095 0.000
112 -0.001 0.000
113 -1.758 0.004
114 -0.160 0.000
115 -0.130 0.000
116 -0.834 0.001
117 0.662 0.001
118 .
119 .
120 0.093 0.000
121 0.823 0.002
122 0.791 0.004
123 .
124 .
125 .
126 -0.066 0.000
127 .
128 .
129 -0.927 0.003
130 -0.549 0.001
131 .
132 .
133 -0.228 0.000
134 -0.104 0.000
135 0.081 0.000
136 .
137 -0.801 0.003
138 .
139 -0.136 0.000
140 -0.129 0.001
141 -1.248 0.013
142 -0.392 0.001
143 -2.189 0.119
144 -0.652 0.006
145 .
146 -1.025 0.026
147 0.453 0.003
148 .
149 -0.280 0.000
150 0.332 0.001
151 -0.477 0.002
152 -0.080 0.000
153 -1.773 0.011
154 -2.363 0.019
155 -1.723 0.016
156 2.484 0.019
157 -0.464 0.001
158 -2.048 0.012
159 -0.236 0.000
160 2.606 0.043
161 .
162 .
163 0.509 0.001
164 0.425 0.001
165 1.803 0.020
166 0.943 0.003
167 -0.363 0.000
168 -1.046 0.003
169 -1.135 0.004
170 -0.862 0.002
171 1.774 0.008
172 .
173 .
174 0.035 0.000
175 0.381 0.000
176 -0.739 0.001
177 -0.376 0.000
178 .
179 .
180 .
181 1.252 0.004
182 2.971 0.023
183 3.015 0.017
184 3.083 0.057
185 .
186 0.166 0.000
187 0.138 0.000
188 .
189 0.204 0.000
190 .
191 -0.093 0.000
192 -0.338 0.000
193 1.900 0.023
194 -1.551 0.017
195 .
196 0.008 0.000
197 1.573 0.013
198 0.199 0.000
199 0.106 0.000
200 .
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.015


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
201 .
202 .
203 0.929 0.001
204 0.647 0.000
205 2.134 0.008
206 1.910 0.005
207 .
208 1.441 0.004
209 .
210 .
211 1.492 0.002
212 0.857 0.001
213 .
214 .
215 .
216 -0.609 0.001
217 -0.779 0.001
218 .
219 .
220 0.856 0.001
221 .
222 0.043 0.000
223 .
224 .
225 .
226 0.584 0.001
227 1.310 0.002
228 -1.841 0.009
229 -2.902 0.100
230 0.838 0.006
231 .
232 0.814 0.001
233 -0.382 0.000
234 -1.868 0.008
235 -1.147 0.002
236 -1.424 0.002
237 -1.488 0.009
238 -0.119 0.000
239 -0.064 0.000
240 -1.010 0.002
241 0.783 0.002
242 .
243 .
244 0.079 0.000
245 0.876 0.002
246 -0.103 0.000
247 -0.296 0.000
248 .
249 -0.413 0.001
250 -0.228 0.001
251 -0.071 0.000
252 0.693 0.002
253 1.673 0.010
254 -0.809 0.009
255 -0.094 0.000
256 .
257 0.195 0.000
258 1.828 0.020
259 0.395 0.002
260 -0.554 0.003
261 .
262 0.661 0.003
263 1.398 0.026
264 2.135 0.051
265 0.469 0.001
266 0.457 0.000
267 0.695 0.001
268 .
269 -0.875 0.001
270 -0.696 0.002
271 0.130 0.000
272 .
273 -0.381 0.000
274 0.440 0.000
275 0.843 0.001
276 0.798 0.002
277 0.633 0.002
278 0.396 0.000
279 1.233 0.007
280 1.500 0.016
281 0.497 0.001
282 -1.285 0.004
283 0.108 0.000
284 -0.149 0.000
285 .
286 -0.017 0.000
287 -1.391 0.006
288 0.953 0.003
289 -0.373 0.000
290 0.260 0.000
291 0.077 0.000
292 -0.142 0.000
293 0.011 0.000
294 0.938 0.002
295 -1.301 0.009
296 .
297 .
298 -1.400 0.003
299 .
300 0.138 0.000
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.015

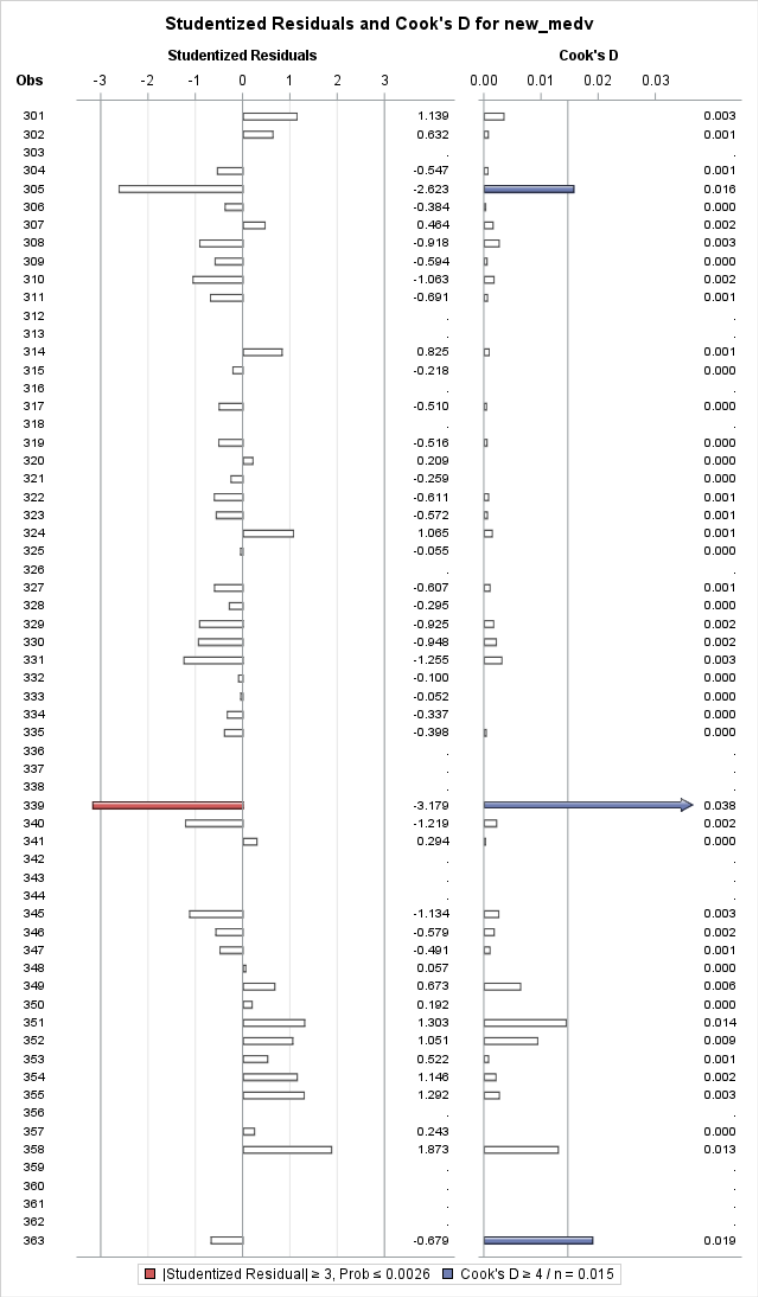

Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
301 1.139 0.003
302 0.632 0.001
303 . .
304 -0.547 0.001
305 -2.623 0.016
306 -0.384 0.000
307 0.464 0.002
308 -0.918 0.003
309 -0.594 0.000
310 -1.063 0.002
311 -0.691 0.001
312 . .
313 . .
314 0.825 0.001
315 -0.218 0.000
316 . .
317 -0.510 0.000
318 . .
319 -0.516 0.000
320 0.209 0.000
321 -0.259 0.000
322 -0.611 0.001
323 -0.572 0.001
324 1.065 0.001
325 -0.055 0.000
326 . .
327 -0.607 0.001
328 -0.295 0.000
329 -0.925 0.002
330 -0.948 0.002
331 -1.255 0.003
332 -0.100 0.000
333 -0.052 0.000
334 -0.337 0.000
335 -0.398 0.000
336 . .
337 . .
338 . .
339 -3.179 0.038
340 -1.219 0.002
341 0.294 0.000
342 . .
343 . .
344 . .
345 -1.134 0.003
346 -0.579 0.002
347 -0.491 0.001
348 0.057 0.000
349 0.673 0.006
350 0.192 0.000
351 1.303 0.014
352 1.051 0.009
353 0.522 0.001
354 1.146 0.002
355 1.292 0.003
356 . .
357 0.243 0.000
358 1.873 0.013
359 . .
360 . .
361 . .
362 . .
363 -0.679 0.019
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.015

Fig. 9


Model 1 after removing outliers
new_medv = -41.586 +12.749 rm -0.0014 dis_tax -0.0695 rm_ptratio +2.375 lstat -0.4699 rm_lstat
N
272
Rsq
0.8965
AdjRsq
0.8946
RMSE
2.6716
Studentized Residual
rm

Fig. 10

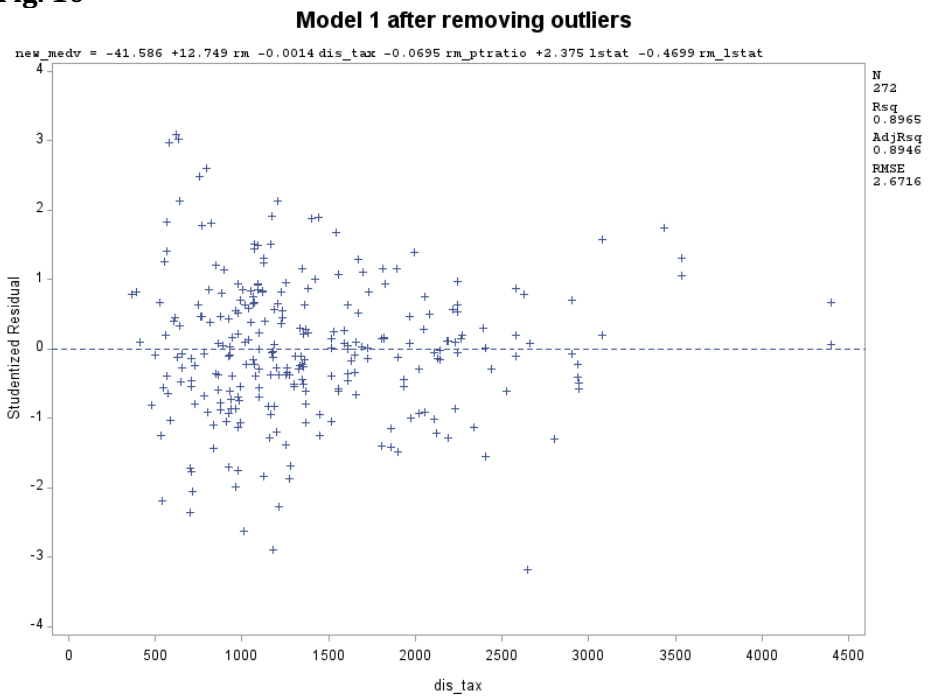

Model 1 after removing outliers
new_medv = -41.586 +12.749 rm -0.0014 dis_tax -0.0695 rm_ptratio +2.375 lstat -0.4699 rm_lstat
N
272
Rsq
0.8965
AdjRsq
0.8946
RMSE
2.6716
Studentized Residual
dis_tax

Fig. 11

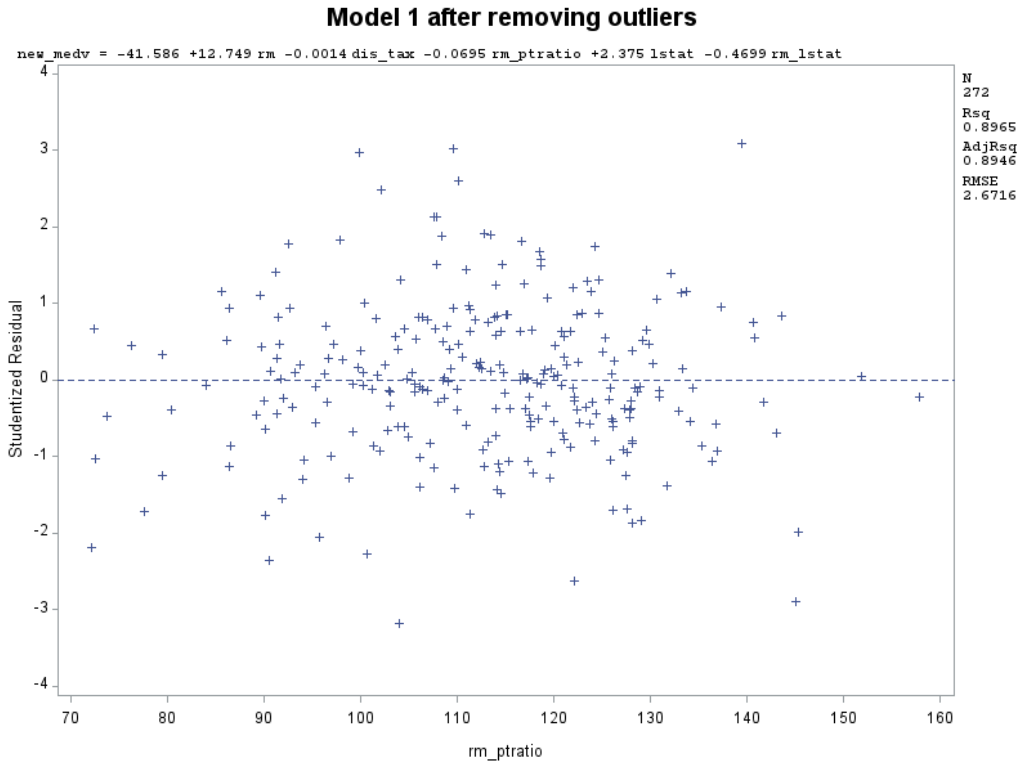

Model 1 after removing outliers
new_medv = -41.586 +12.749 rm -0.0014 dis_tax -0.0695 rm_ptratio +2.375 lstat -0.4699 rm_lstat
N
272
Rsq
0.8965
AdjRsq
0.8946
RMSE
2.6716
Studentized Residual
rm_ptratio

Fig. 12

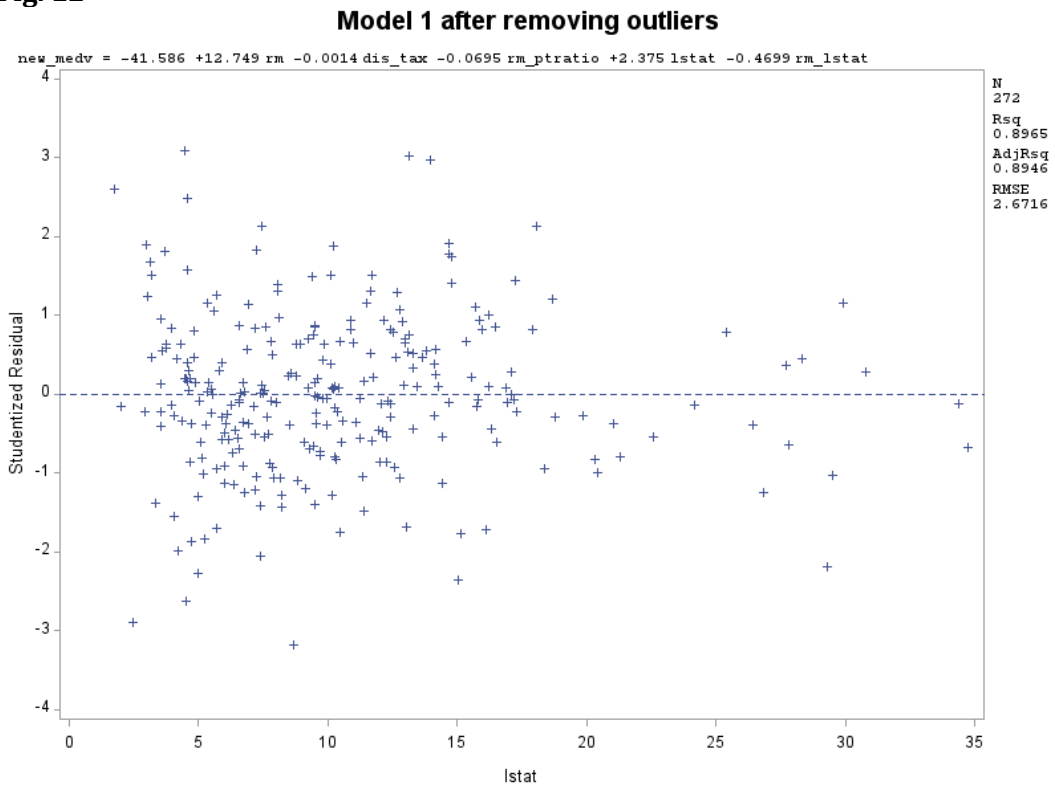

Model 1 after removing outliers
new_medv = -41.586 +12.749 rm -0.0014 dis_tax -0.0695 rm_ptratio +2.375 lstat -0.4699 rm_lstat
N
272
Rsq
0.8965
AdjRsq
0.8946
RMSE
2.6716
Studentized Residual
lstat

Fig. 13

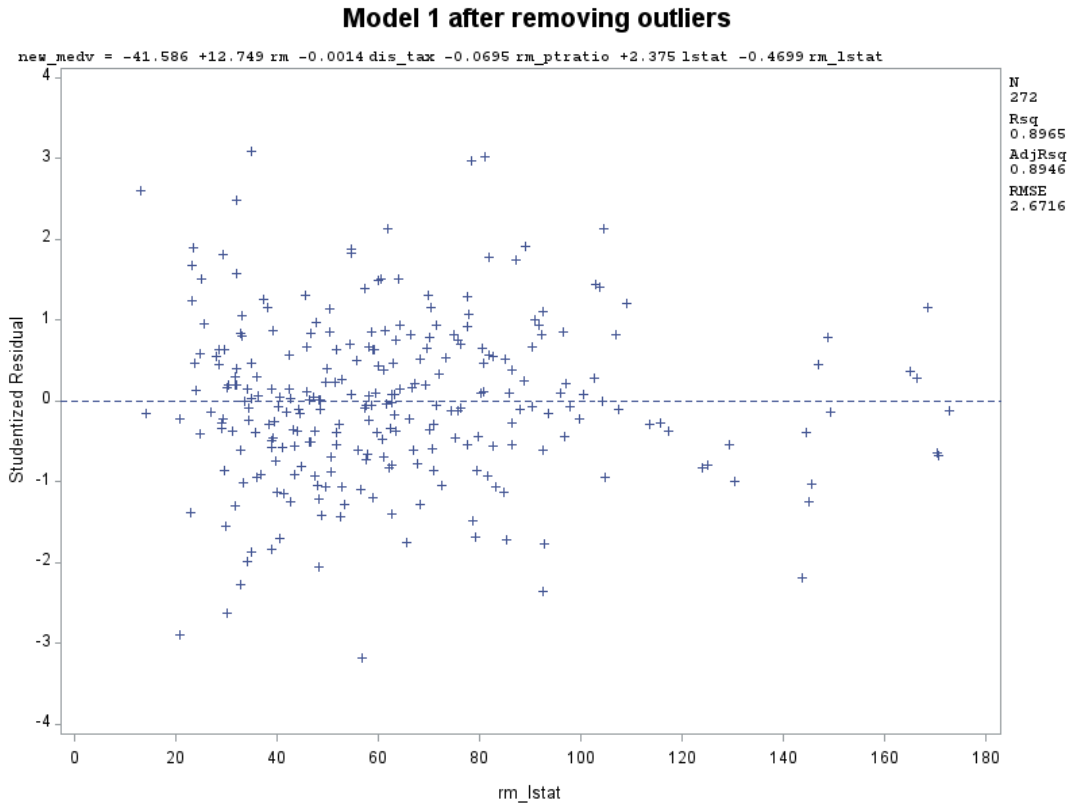

Model 1 after removing outliers
new_medv = -41.586 +12.749 rm -0.0014 dis_tax -0.0695 rm_ptratio +2.375 lstat -0.4699 rm_lstat
N
272
Rsq
0.8965
AdjRsq
0.8946
RMSE
2.6716
Studentized Residual
rm_lstat

Fig. 14

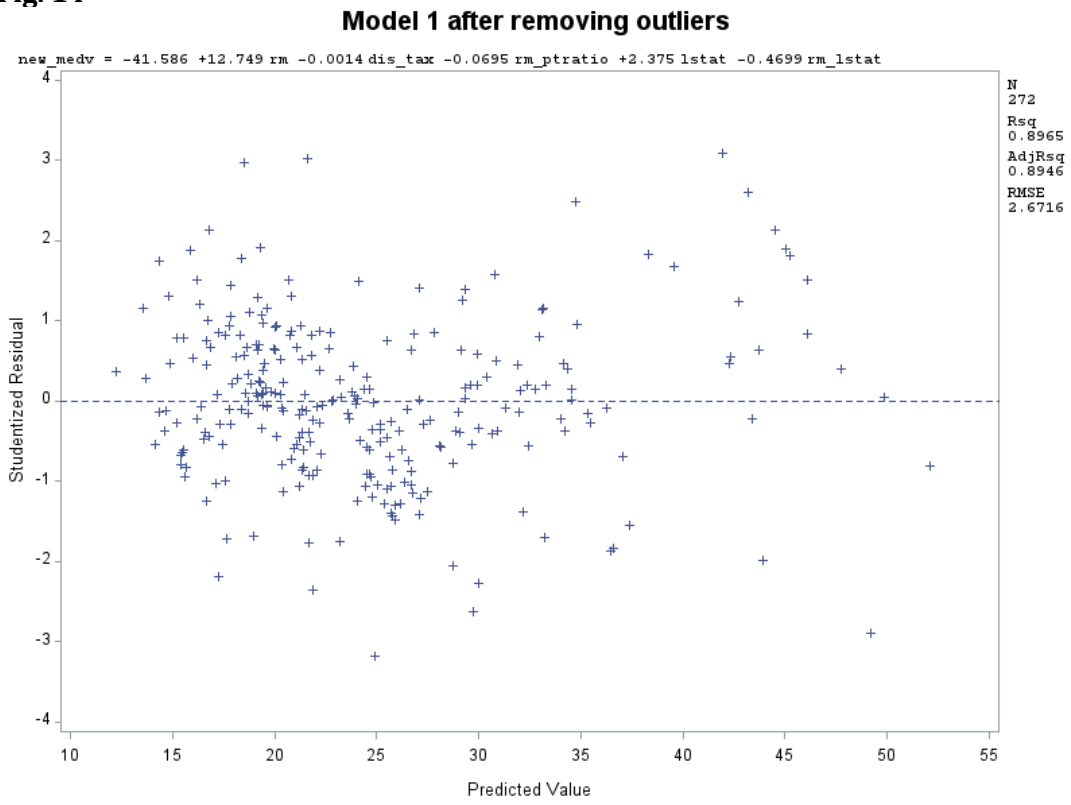

Model 1 after removing outliers
new_medv = -41.586 +12.749 rm -0.0014 dis_tax -0.0695 rm_ptratio +2.375 lstat -0.4699 rm_lstat
N
272
Rsq
0.8965
AdjRsq
0.8946
RMSE
2.6716
Studentized Residual
Predicted Value

Fig. 15

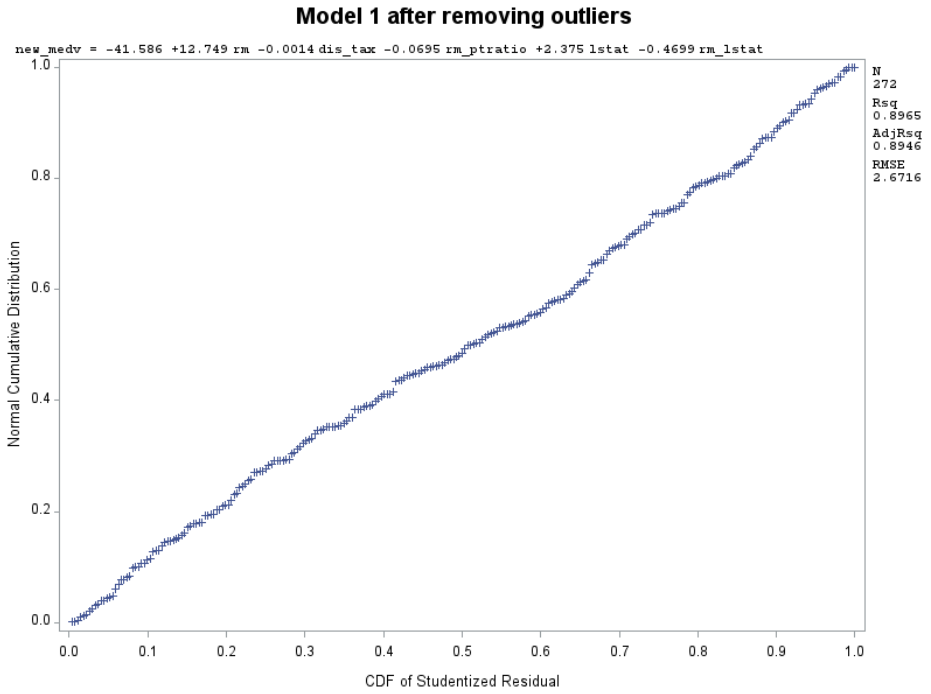


Table 4

Performance test parameters - Model 1

Obs	_TYPE_	_FREQ_	rmse	mae
1	0	91	3.46342	2.36930

Performance test parameters - Model 1

The CORR Procedure

2 Variables:	medv yhat

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**	**Label**
medv	91	25.85055	9.13302	2352	14.30000	50.00000	
yhat	91	25.34226	8.81136	2306	8.32210	50.25216	Predicted Value of new_medv

Pearson Correlation Coefficients, N = 91 Prob > |r| under H0: Rho=0		
	medv	**yhat**
medv	1.00000	0.92611 <.0001
yhat Predicted Value of new_medv	0.92611 <.0001	1.00000

Table 5

Model 2- Checking Model assumptions and diagnostics

The REG Procedure
Model: MODEL1
Dependent Variable: new_medv

Number of Observations Read	367
Number of Observations Used	276
Number of Observations with Missing Values	91

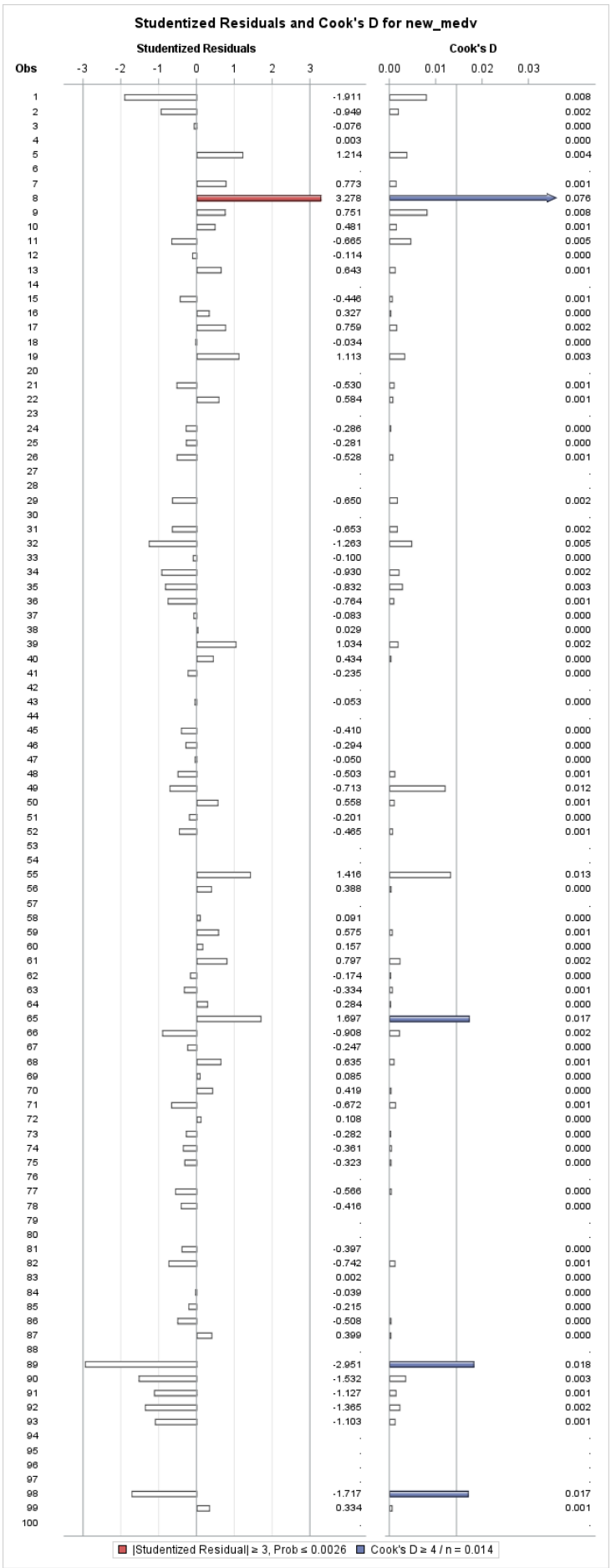
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	7	16688	2384.02414	261.62	<.0001
Error	268	2442.19304	9.11266		
Corrected Total	275	19130			

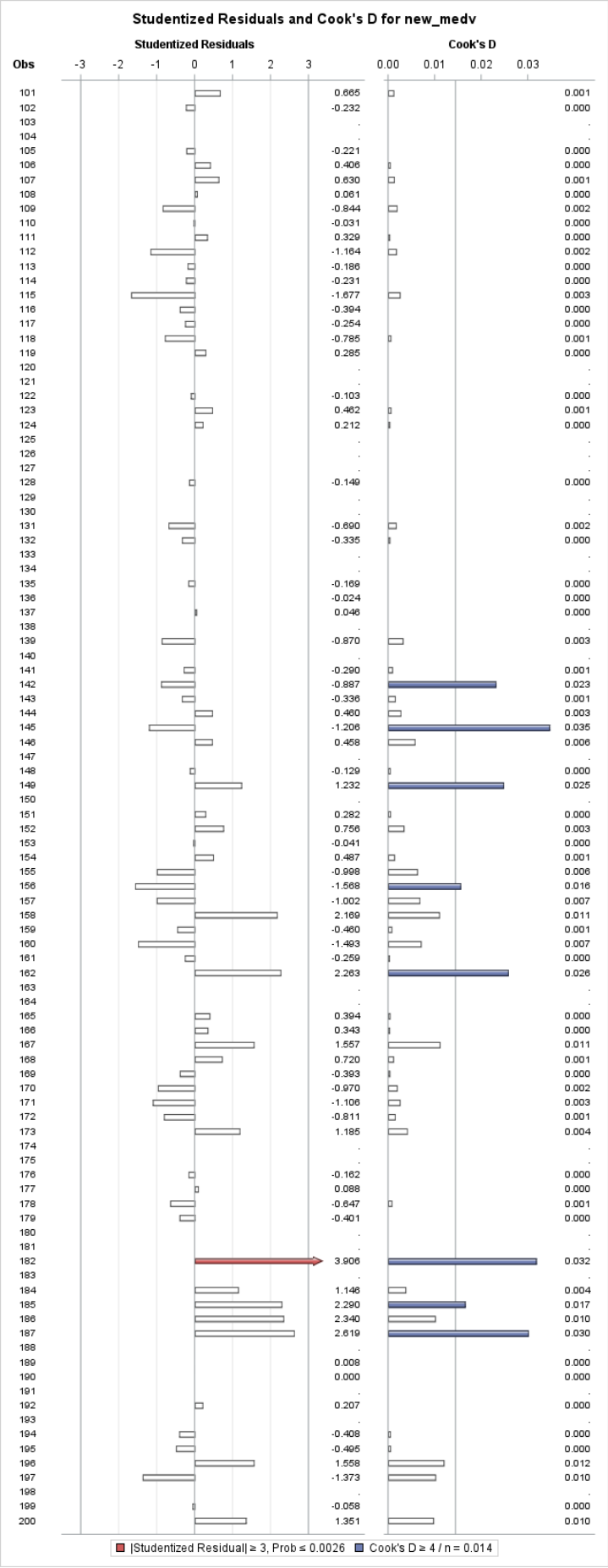
Root MSE	3.01872	**R-Square**	0.8723
Dependent Mean	24.78551	**Adj R-Sq**	0.8690
Coeff Var	12.17937		

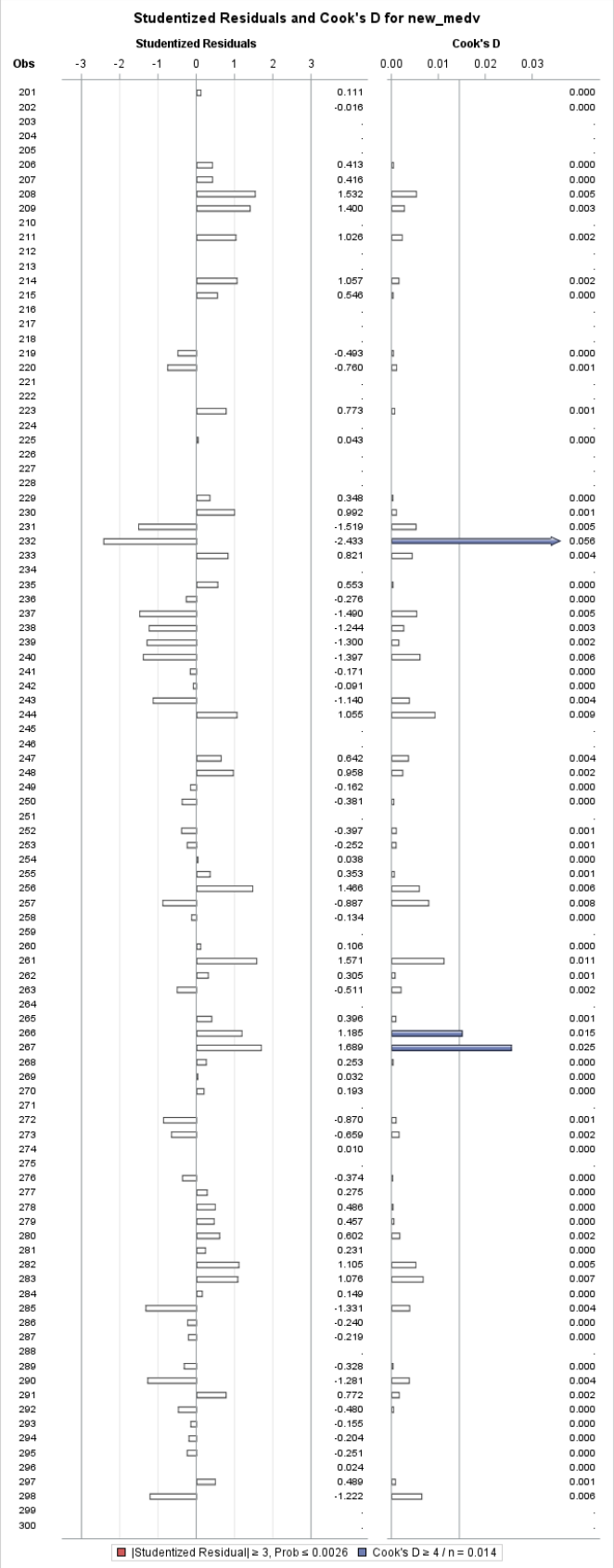
Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standardized Estimate**	**Variance Inflation**
Intercept	**1**	-42.33980	3.35135	-12.63	<.0001	0	0
rm	**1**	13.17032	0.49332	26.70	<.0001	1.05212	3.26034

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standardized Estimate**	**Variance Inflation**
age_dis	**1**	-0.00482	0.00198	-2.43	0.0159	-0.06428	1.47185
dis_tax	**1**	-0.00165	0.00030821	-5.34	<.0001	-0.13590	1.35933
rm_ptratio	**1**	-0.08043	0.01408	-5.71	<.0001	-0.14726	1.39503
lstat	**1**	2.99927	0.32916	9.11	<.0001	2.15526	117.45268
nox_lstat	**1**	-0.59786	0.14380	-4.16	<.0001	-0.31722	12.22035
rm_lstat	**1**	-0.50394	0.05085	-9.91	<.0001	-1.92226	78.97871

Fig. 16


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
1 -1.911 0.008
2 -0.949 0.002
3 -0.076 0.000
4 0.003 0.000
5 1.214 0.004
6 . .
7 0.773 0.001
8 3.278 0.076
9 0.751 0.008
10 0.481 0.001
11 -0.665 0.005
12 -0.114 0.000
13 0.643 0.001
14 . .
15 -0.446 0.001
16 0.327 0.000
17 0.759 0.002
18 -0.034 0.000
19 1.113 0.003
20 . .
21 -0.530 0.001
22 0.584 0.001
23 . .
24 -0.286 0.000
25 -0.281 0.000
26 -0.528 0.001
27 . .
28 . .
29 -0.650 0.002
30 . .
31 -0.653 0.002
32 -1.263 0.005
33 -0.100 0.000
34 -0.930 0.002
35 -0.832 0.003
36 -0.764 0.001
37 -0.083 0.000
38 0.029 0.000
39 1.034 0.002
40 0.434 0.000
41 -0.235 0.000
42 . .
43 -0.053 0.000
44 . .
45 -0.410 0.000
46 -0.294 0.000
47 -0.050 0.000
48 -0.503 0.001
49 -0.713 0.012
50 0.558 0.001
51 -0.201 0.000
52 -0.465 0.001
53 . .
54 . .
55 1.416 0.013
56 0.388 0.000
57 . .
58 0.091 0.000
59 0.575 0.001
60 0.157 0.000
61 0.797 0.002
62 -0.174 0.000
63 -0.334 0.001
64 0.284 0.000
65 1.697 0.017
66 -0.908 0.002
67 -0.247 0.000
68 0.635 0.001
69 0.085 0.000
70 0.419 0.000
71 -0.672 0.001
72 0.108 0.000
73 -0.282 0.000
74 -0.361 0.000
75 -0.323 0.000
76 . .
77 -0.566 0.000
78 -0.416 0.000
79 . .
80 . .
81 -0.397 0.000
82 -0.742 0.001
83 0.002 0.000
84 -0.039 0.000
85 -0.215 0.000
86 -0.508 0.000
87 0.399 0.000
88 . .
89 -2.951 0.018
90 -1.532 0.003
91 -1.127 0.001
92 -1.365 0.002
93 -1.103 0.001
94 . .
95 . .
96 . .
97 . .
98 -1.717 0.017
99 0.334 0.001
100 . .
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
101 0.665 0.001
102 -0.232 0.000
103 . .
104 . .
105 -0.221 0.000
106 0.406 0.000
107 0.630 0.001
108 0.061 0.000
109 -0.844 0.002
110 -0.031 0.000
111 0.329 0.000
112 -1.164 0.002
113 -0.186 0.000
114 -0.231 0.000
115 -1.677 0.003
116 -0.394 0.000
117 -0.254 0.000
118 -0.785 0.001
119 0.285 0.000
120 . .
121 . .
122 -0.103 0.000
123 0.462 0.001
124 0.212 0.000
125 . .
126 . .
127 . .
128 -0.149 0.000
129 . .
130 . .
131 -0.690 0.002
132 -0.335 0.000
133 . .
134 . .
135 -0.169 0.000
136 -0.024 0.000
137 0.046 0.000
138 . .
139 -0.870 0.003
140 . .
141 -0.290 0.001
142 -0.887 0.023
143 -0.336 0.001
144 0.460 0.003
145 -1.206 0.035
146 0.458 0.006
147 . .
148 -0.129 0.000
149 1.232 0.025
150 . .
151 0.282 0.000
152 0.756 0.003
153 -0.041 0.000
154 0.487 0.001
155 -0.998 0.006
156 -1.568 0.016
157 -1.002 0.007
158 2.169 0.011
159 -0.460 0.001
160 -1.493 0.007
161 -0.259 0.000
162 2.263 0.026
163 . .
164 . .
165 0.394 0.000
166 0.343 0.000
167 1.557 0.011
168 0.720 0.001
169 -0.393 0.000
170 -0.970 0.002
171 -1.106 0.003
172 -0.811 0.001
173 1.185 0.004
174 . .
175 . .
176 -0.162 0.000
177 0.088 0.000
178 -0.647 0.001
179 -0.401 0.000
180 . .
181 . .
182 3.906 0.032
183 . .
184 1.146 0.004
185 2.290 0.017
186 2.340 0.010
187 2.619 0.030
188 . .
189 0.008 0.000
190 0.000 0.000
191 . .
192 0.207 0.000
193 . .
194 -0.408 0.000
195 -0.495 0.000
196 1.558 0.012
197 -1.373 0.010
198 . .
199 -0.058 0.000
200 1.351 0.010
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
201 0.111 0.000
202 -0.016 0.000
203 . .
204 . .
205 . .
206 0.413 0.000
207 0.416 0.000
208 1.532 0.005
209 1.400 0.003
210 . .
211 1.026 0.002
212 . .
213 . .
214 1.057 0.002
215 0.546 0.000
216 . .
217 . .
218 . .
219 -0.493 0.000
220 -0.760 0.001
221 . .
222 . .
223 0.773 0.001
224 . .
225 0.043 0.000
226 . .
227 . .
228 . .
229 0.348 0.000
230 0.992 0.001
231 -1.519 0.005
232 -2.433 0.056
233 0.821 0.004
234 . .
235 0.553 0.000
236 -0.276 0.000
237 -1.490 0.005
238 -1.244 0.003
239 -1.300 0.002
240 -1.397 0.006
241 -0.171 0.000
242 -0.091 0.000
243 -1.140 0.004
244 1.055 0.009
245 . .
246 . .
247 0.642 0.004
248 0.958 0.002
249 -0.162 0.000
250 -0.381 0.000
251 . .
252 -0.397 0.001
253 -0.252 0.001
254 0.038 0.000
255 0.353 0.001
256 1.466 0.006
257 -0.887 0.008
258 -0.134 0.000
259 . .
260 0.106 0.000
261 1.571 0.011
262 0.305 0.001
263 -0.511 0.002
264 . .
265 0.396 0.001
266 1.185 0.015
267 1.689 0.025
268 0.253 0.000
269 0.032 0.000
270 0.193 0.000
271 . .
272 -0.870 0.001
273 -0.659 0.002
274 0.010 0.000
275 . .
276 -0.374 0.000
277 0.275 0.000
278 0.486 0.000
279 0.457 0.000
280 0.602 0.002
281 0.231 0.000
282 1.105 0.005
283 1.076 0.007
284 0.149 0.000
285 -1.331 0.004
286 -0.240 0.000
287 -0.219 0.000
288 . .
289 -0.328 0.000
290 -1.281 0.004
291 0.772 0.002
292 -0.480 0.000
293 -0.155 0.000
294 -0.204 0.000
295 -0.251 0.000
296 0.024 0.000
297 0.489 0.001
298 -1.222 0.006
299 . .
300 . .
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014

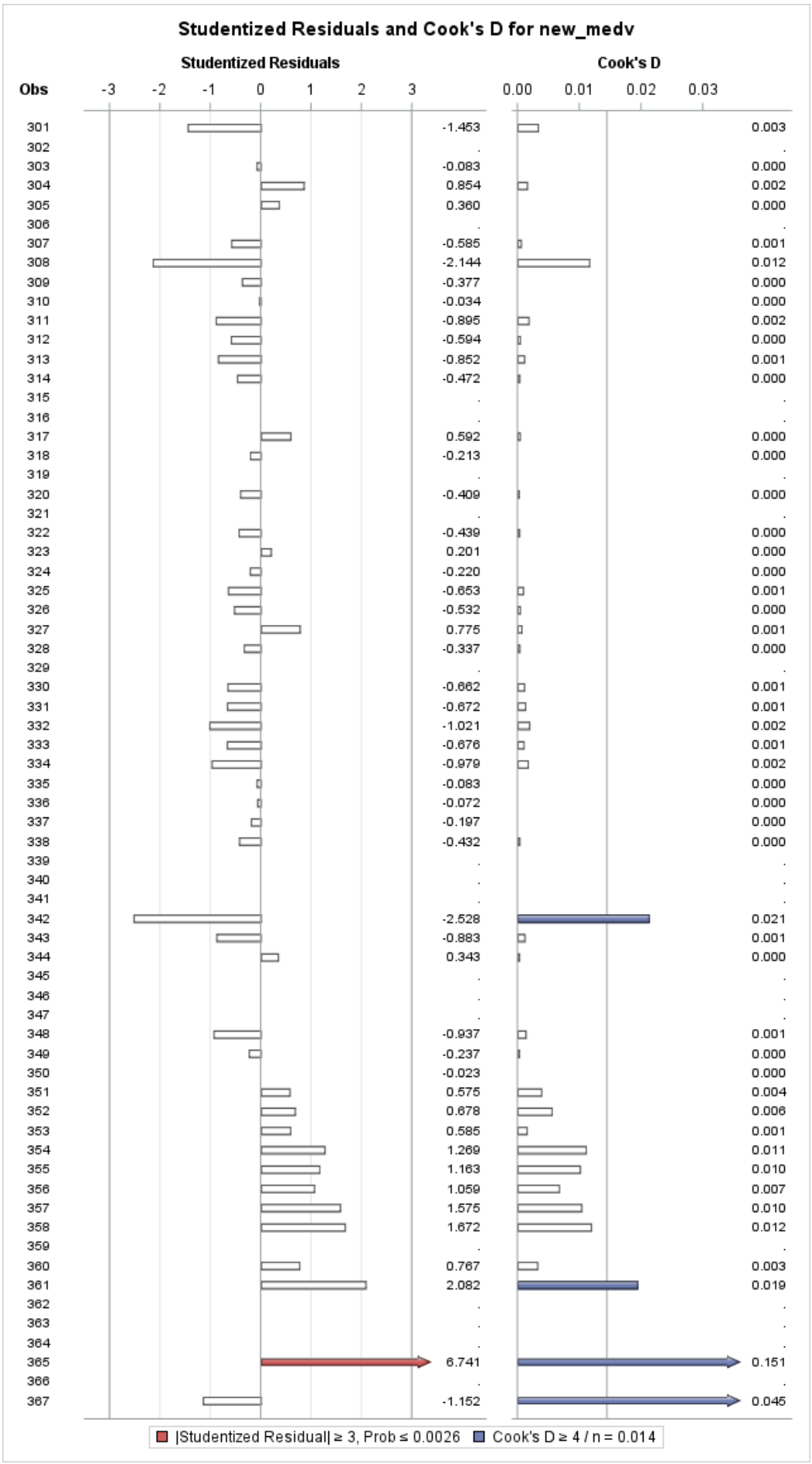

Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
301 -1.453 0.003
302 . .
303 -0.083 0.000
304 0.854 0.002
305 0.360 0.000
306 . .
307 -0.585 0.001
308 -2.144 0.012
309 -0.377 0.000
310 -0.034 0.000
311 -0.895 0.002
312 -0.594 0.000
313 -0.852 0.001
314 -0.472 0.000
315 . .
316 . .
317 0.592 0.000
318 -0.213 0.000
319 . .
320 -0.409 0.000
321 . .
322 -0.439 0.000
323 0.201 0.000
324 -0.220 0.000
325 -0.653 0.001
326 -0.532 0.000
327 0.775 0.001
328 -0.337 0.000
329 . .
330 -0.662 0.001
331 -0.672 0.001
332 -1.021 0.002
333 -0.676 0.001
334 -0.979 0.002
335 -0.083 0.000
336 -0.072 0.000
337 -0.197 0.000
338 -0.432 0.000
339 . .
340 . .
341 . .
342 -2.528 0.021
343 -0.883 0.001
344 0.343 0.000
345 . .
346 . .
347 . .
348 -0.937 0.001
349 -0.237 0.000
350 -0.023 0.000
351 0.575 0.004
352 0.678 0.006
353 0.585 0.001
354 1.269 0.011
355 1.163 0.010
356 1.059 0.007
357 1.575 0.010
358 1.672 0.012
359 . .
360 0.767 0.003
361 2.082 0.019
362 . .
363 . .
364 . .
365 6.741 0.151
366 . .
367 -1.152 0.045
|Studentized Residual| ≥ 3, Prob ≤ 0.0026
Cook's D ≥ 4 / n = 0.014

Fig. 17

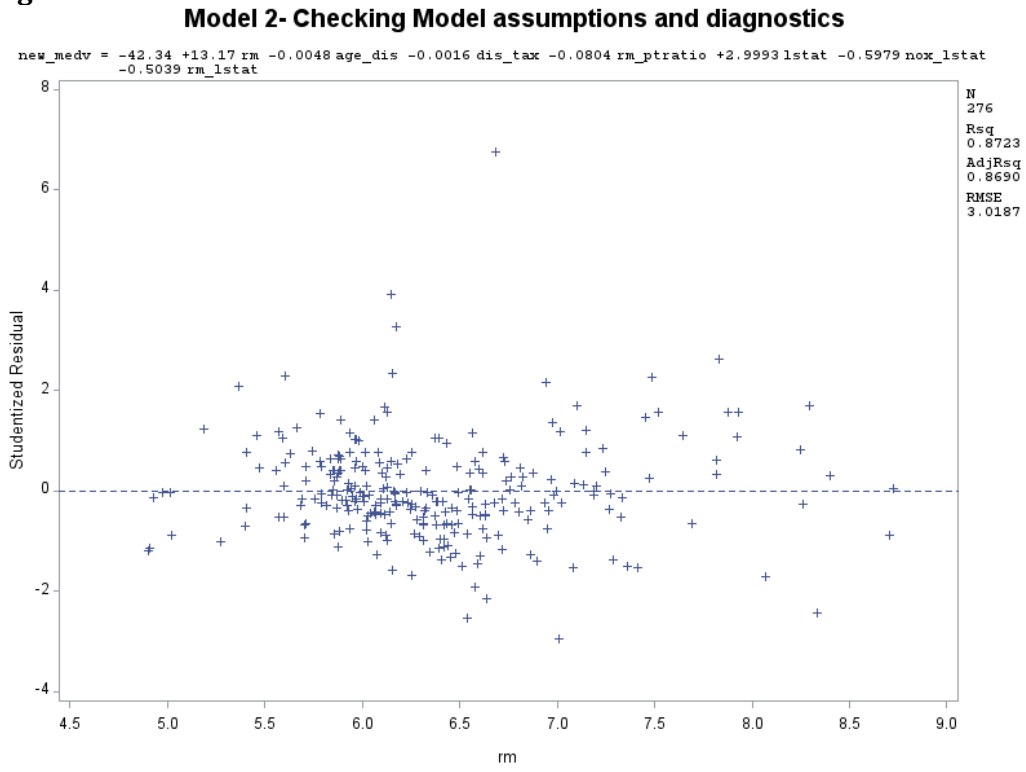


Fig. 18

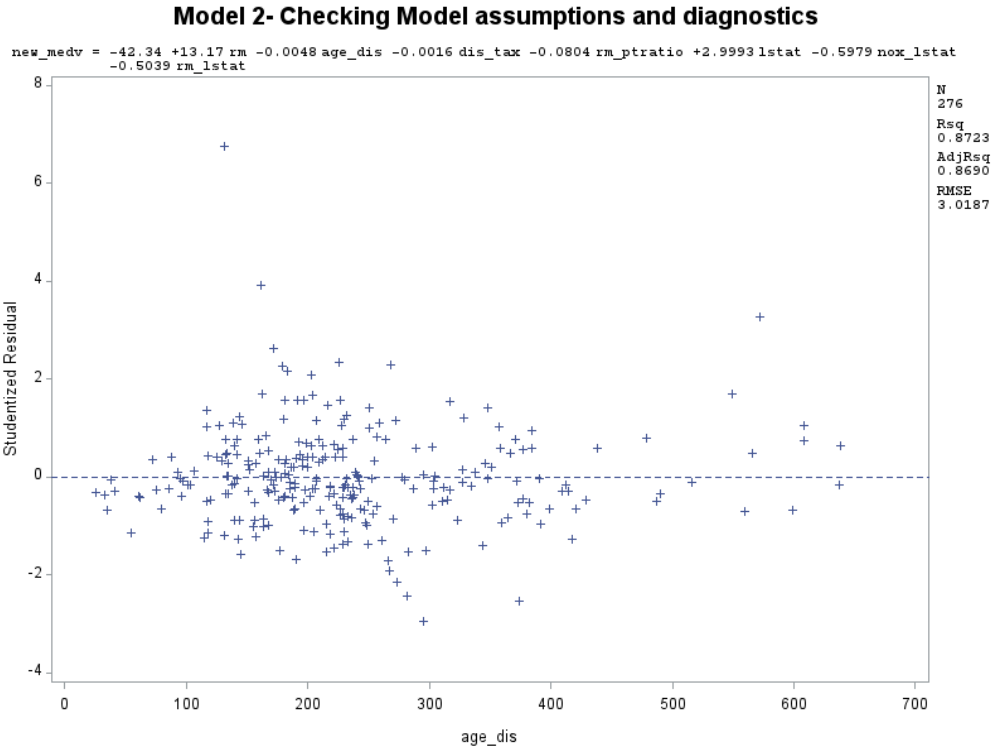


Fig. 19

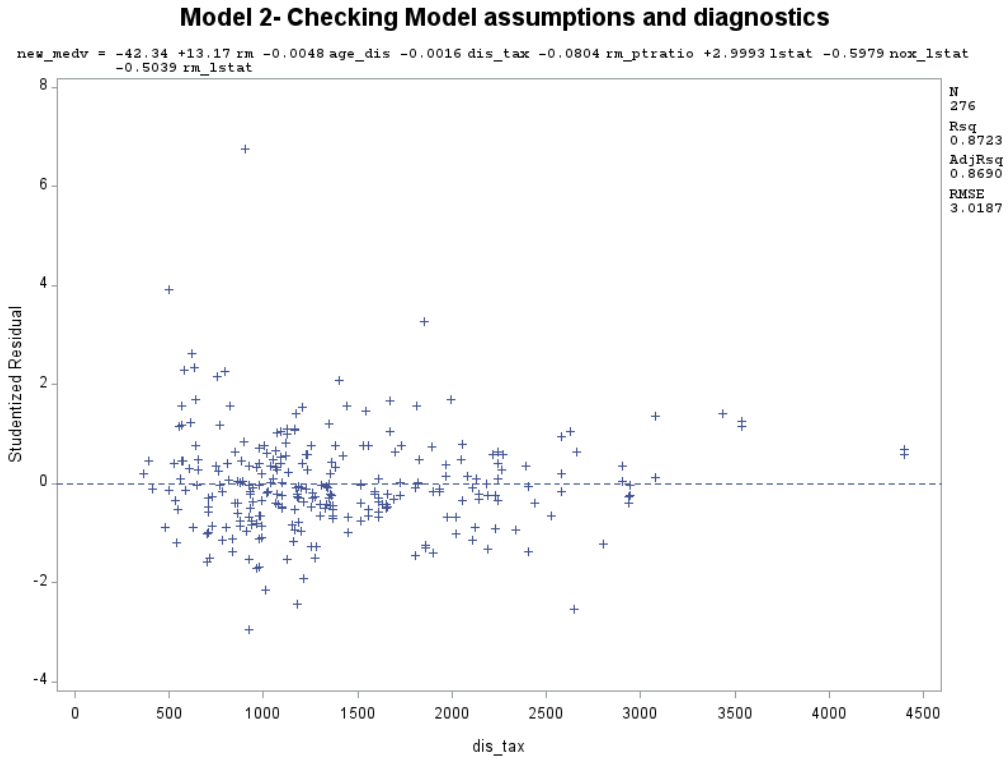

Model 2- Checking Model assumptions and diagnostics
new_medv = -42.34 +13.17 rm -0.0048 age_dis -0.0016 dis_tax -0.0804 rm_ptratio +2.9993 lstat -0.5979 nox_lstat -0.5039 rm_lstat
N 276
Rsq 0.8723
AdjRsq 0.8690
RMSE 3.0187
Studentized Residual
dis_tax

Fig. 20

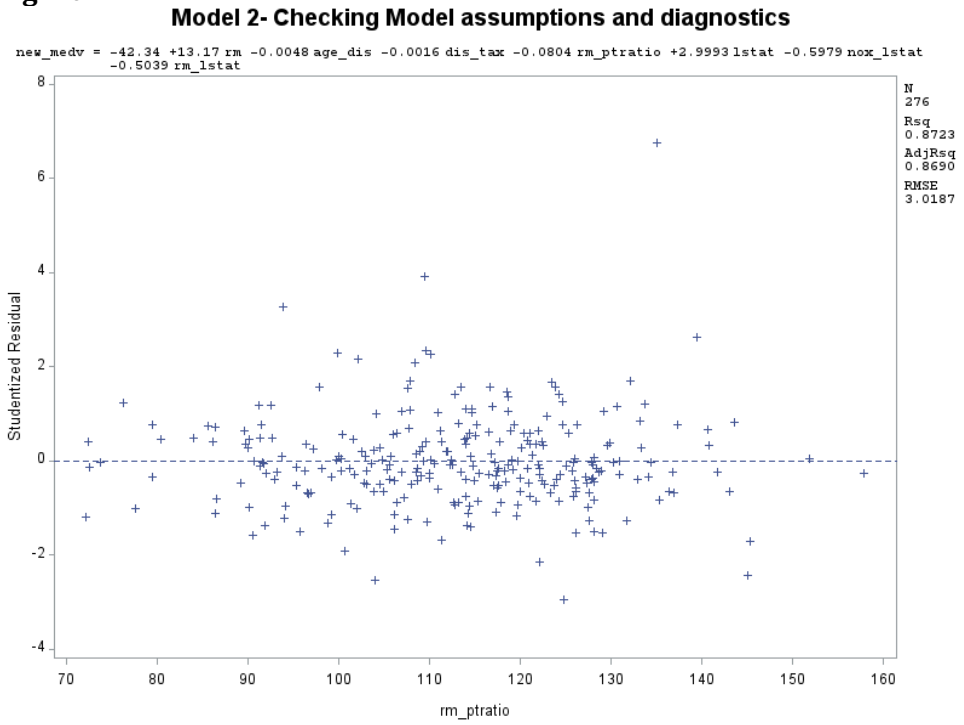

Model 2- Checking Model assumptions and diagnostics
new_medv = -42.34 +13.17 rm -0.0048 age_dis -0.0016 dis_tax -0.0804 rm_ptratio +2.9993 lstat -0.5979 nox_lstat -0.5039 rm_lstat
N 276
Rsq 0.8723
AdjRsq 0.8690
RMSE 3.0187
Studentized Residual
rm_ptratio

Fig. 21

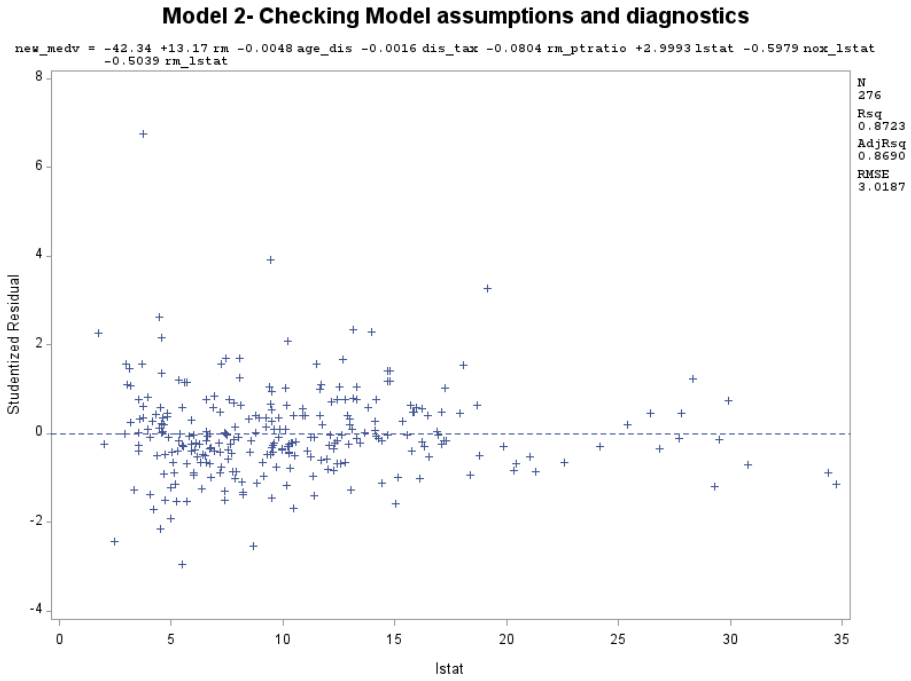

Model 2- Checking Model assumptions and diagnostics
new_medv = -42.34 +13.17 rm -0.0048 age_dis -0.0016 dis_tax -0.0804 rm_ptratio +2.9993 lstat -0.5979 nox_lstat -0.5039 rm_lstat
N 276
Rsq 0.8723
AdjRsq 0.8690
RMSE 3.0187
Studentized Residual
lstat

Fig. 22

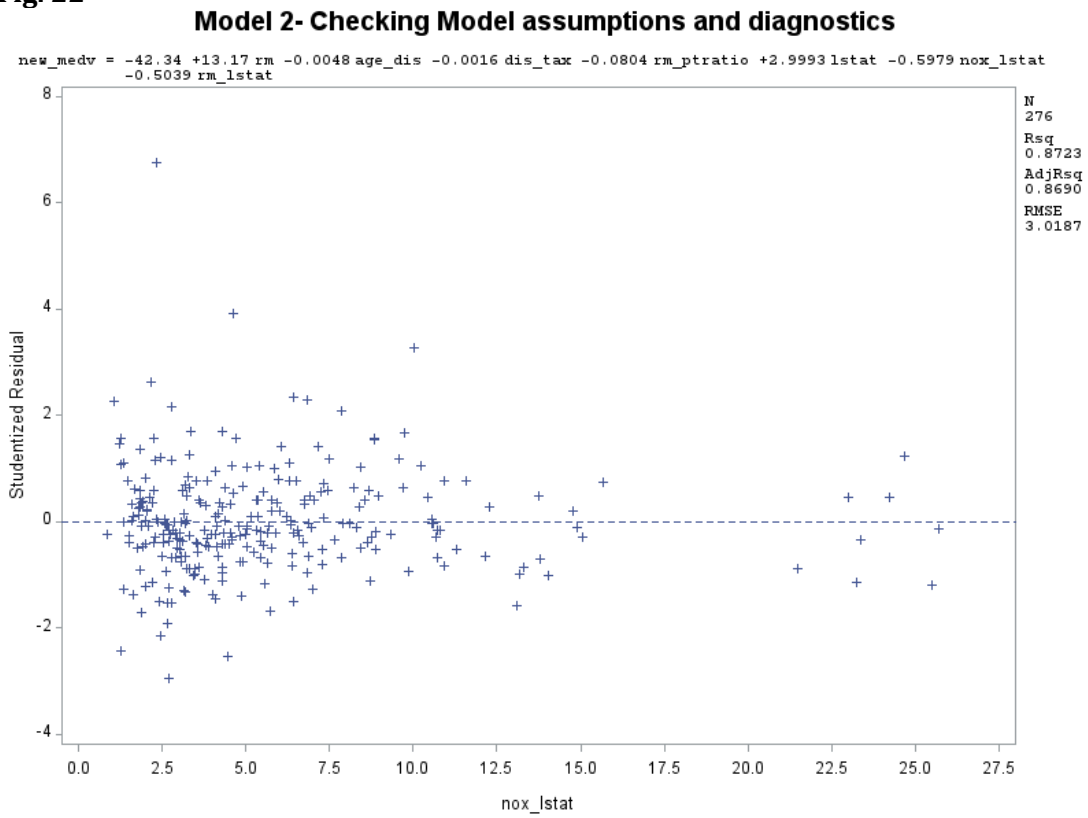

Model 2- Checking Model assumptions and diagnostics
new_medv = -42.34 +13.17 rm -0.0048 age_dis -0.0016 dis_tax -0.0804 rm_ptratio +2.9993 lstat -0.5979 nox_lstat -0.5039 rm_lstat
N 276
Rsq 0.8723
AdjRsq 0.8690
RMSE 3.0187
Studentized Residual
nox_lstat

Fig. 23

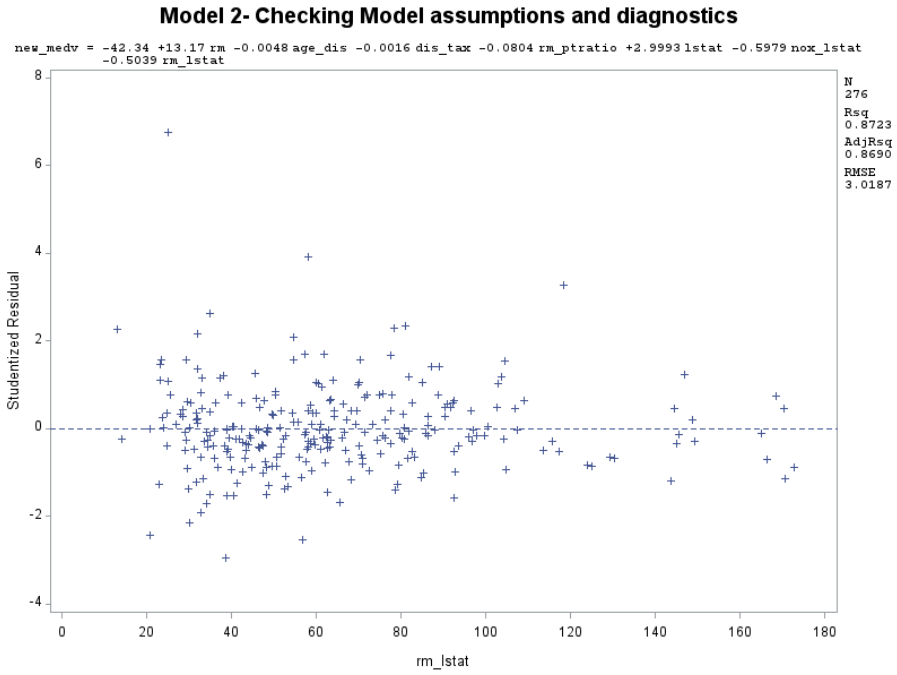

Model 2- Checking Model assumptions and diagnostics
new_medv = -42.34 +13.17 rm -0.0048 age_dis -0.0016 dis_tax -0.0804 rm_ptratio +2.9993 lstat -0.5979 nox_lstat -0.5039 rm_lstat
N 276
Rsq 0.8723
AdjRsq 0.8690
RMSE 3.0187
Studentized Residual
rm_lstat

Fig. 24

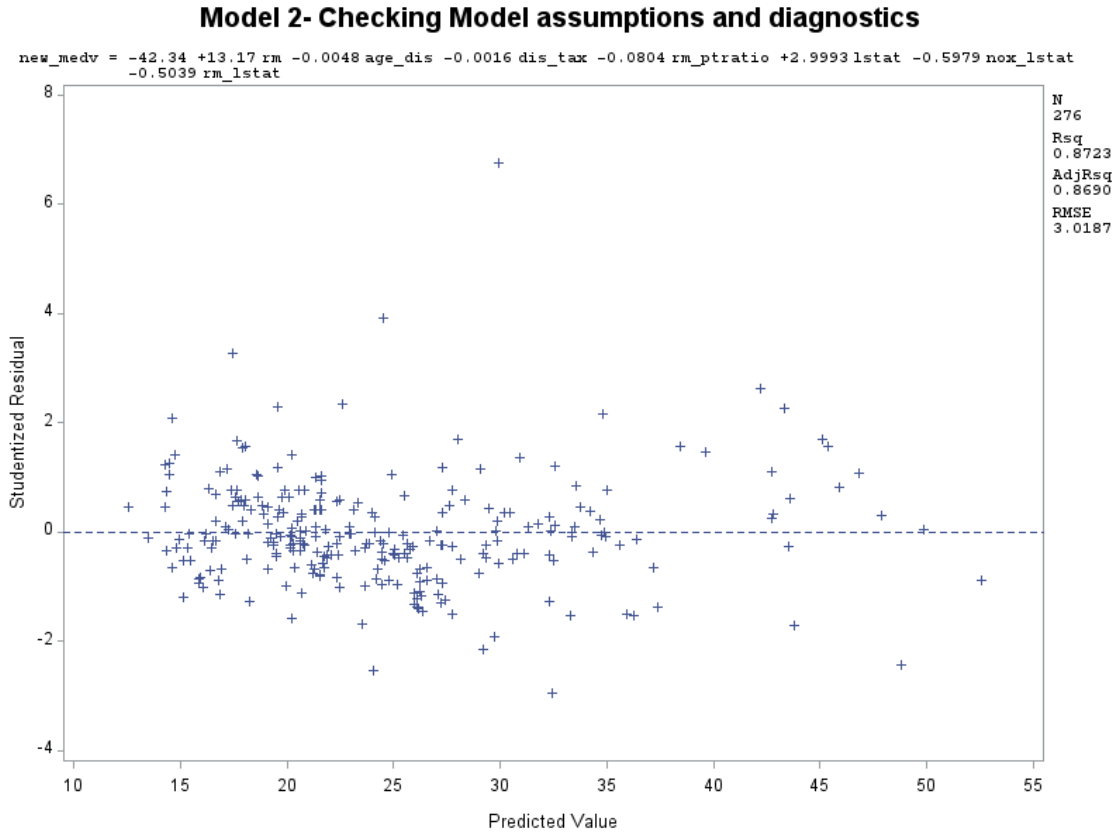

Model 2- Checking Model assumptions and diagnostics
new_medv = -42.34 +13.17 rm -0.0048 age_dis -0.0016 dis_tax -0.0804 rm_ptratio +2.9993 lstat -0.5979 nox_lstat -0.5039 rm_lstat
N 276
Rsq 0.8723
AdjRsq 0.8690
RMSE 3.0187
Studentized Residual
Predicted Value

Fig. 25

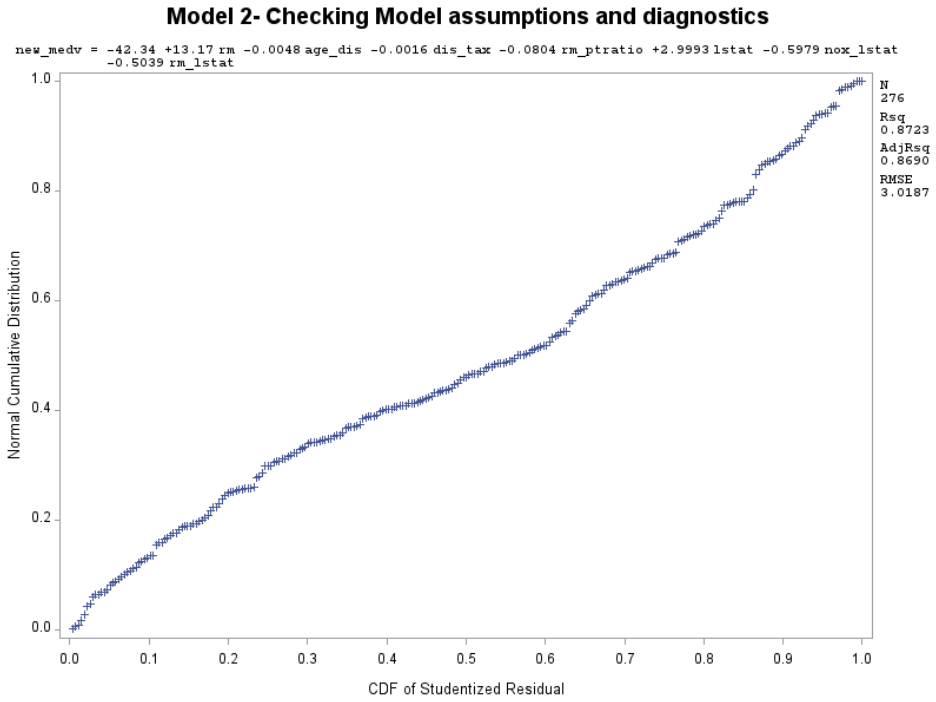


Table 5

Model 2 after removing outliers

The REG Procedure
Model: MODEL1
Dependent Variable: new_medv

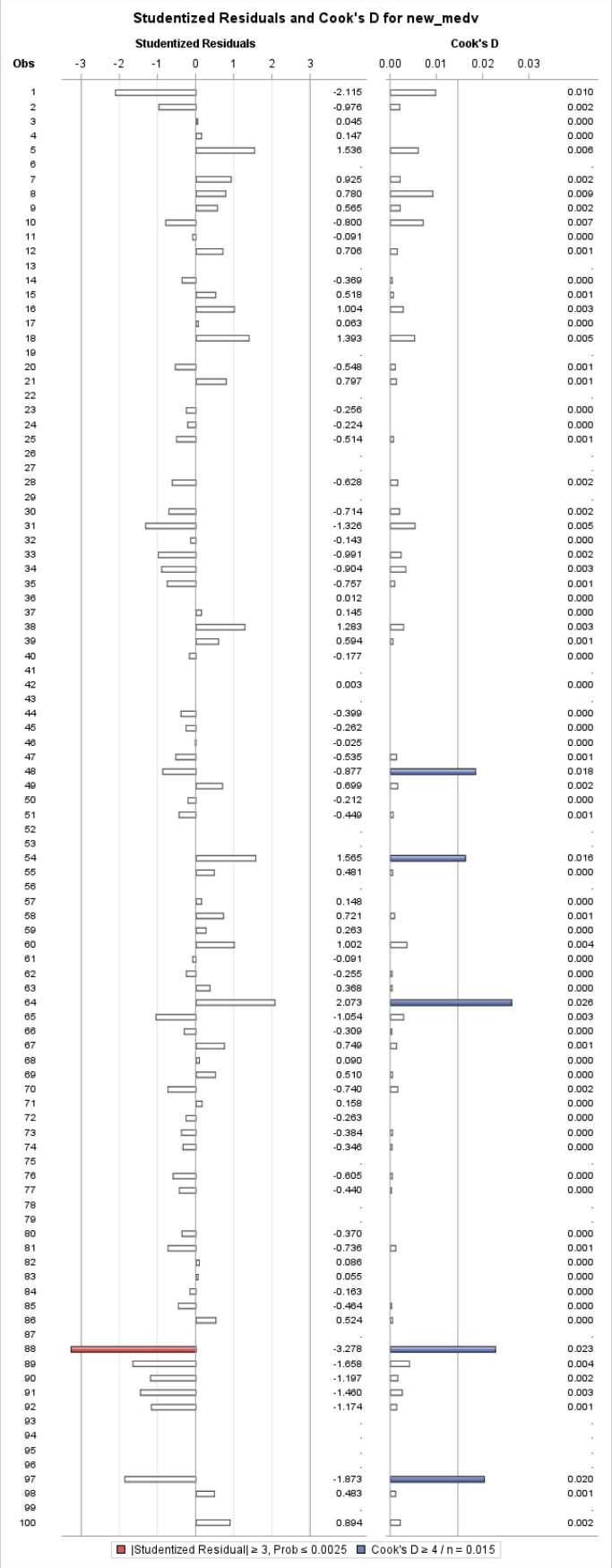
Number of Observations Read	364
Number of Observations Used	273
Number of Observations with Missing Values	91

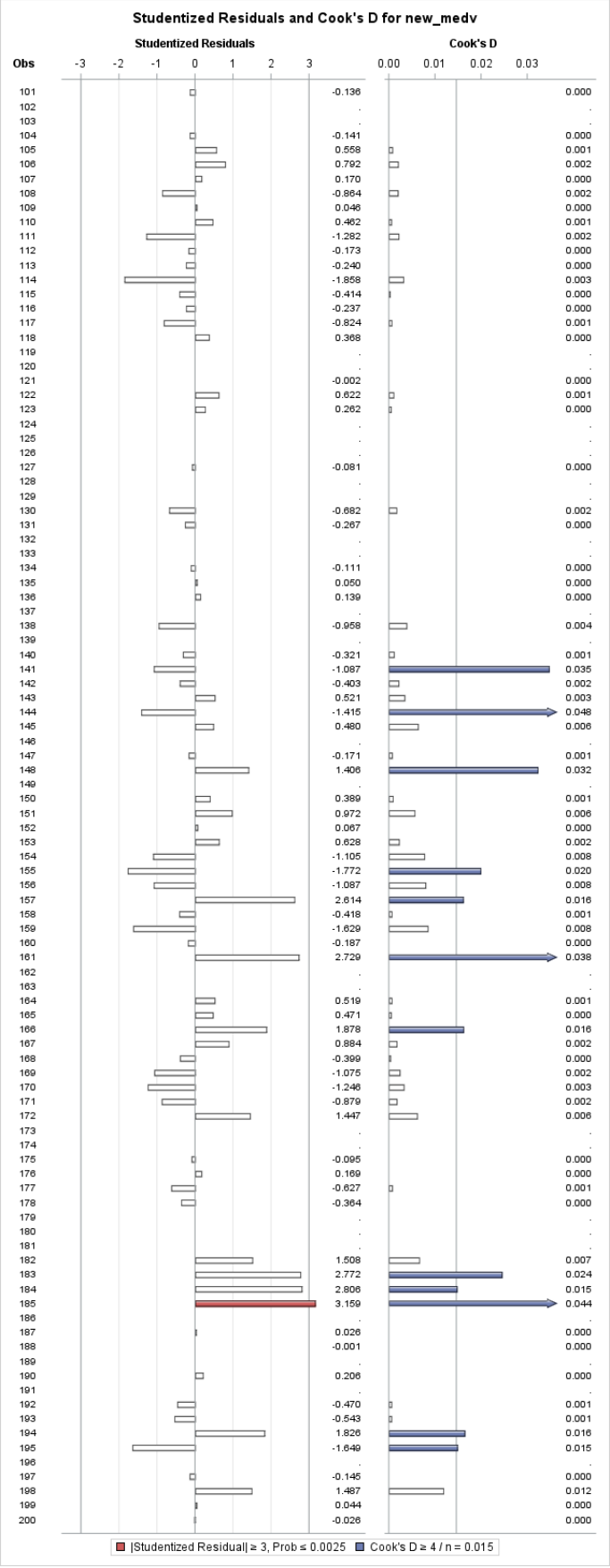
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	7	16556	2365.16956	348.75	<.0001
Error	265	1797.20164	6.78189		
Corrected Total	272	18353			

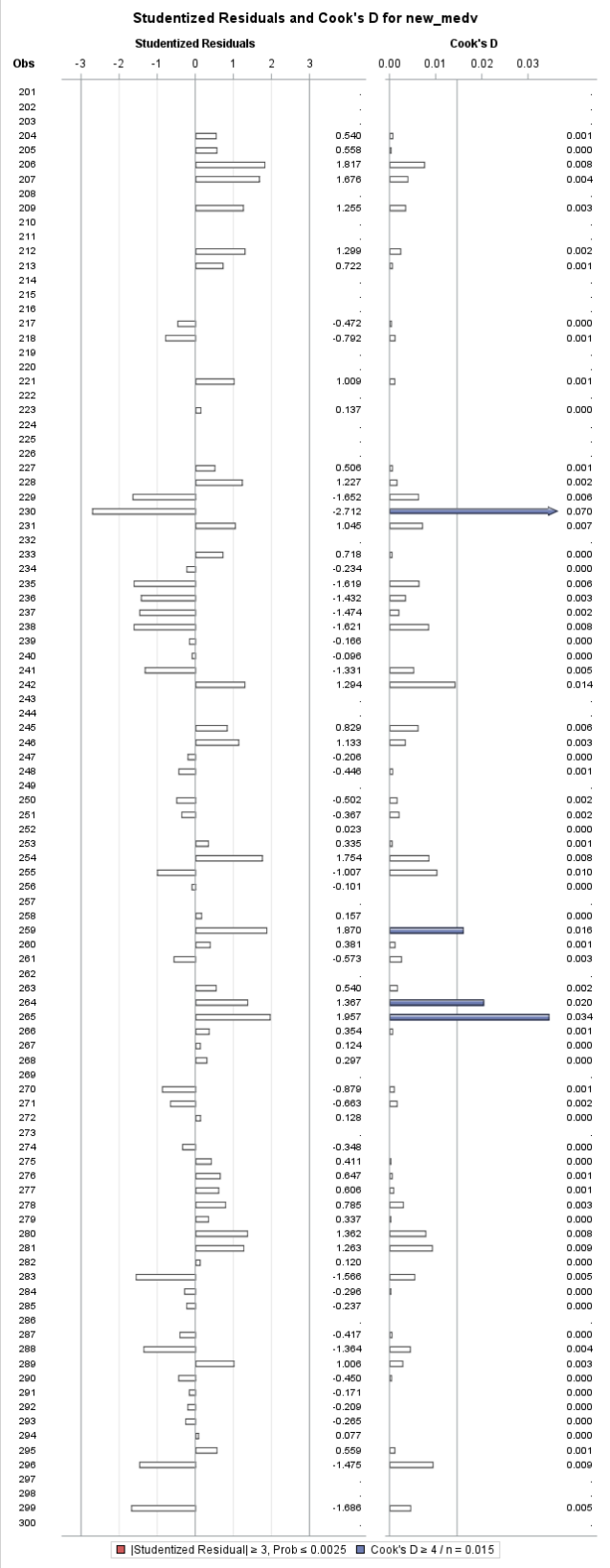
Root MSE	2.60421	**R-Square**	0.9021
Dependent Mean	24.64286	**Adj R-Sq**	0.8995
Coeff Var	10.56780		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standardized Estimate**	**Variance Inflation**
Intercept	**1**	-43.37581	2.89730	-14.97	<.0001	0	0
rm	**1**	13.33592	0.42628	31.28	<.0001	1.08682	3.26602
age_dis	**1**	-0.00549	0.00173	-3.18	0.0016	-0.07338	1.43868
dis_tax	**1**	-0.00144	0.00026773	-5.38	<.0001	-0.12082	1.36506
rm_ptratio	**1**	-0.08550	0.01227	-6.97	<.0001	-0.15874	1.40408
lstat	**1**	3.03364	0.28433	10.67	<.0001	2.21184	116.30657
nox_lstat	**1**	-0.61369	0.12431	-4.94	<.0001	-0.33147	12.20018
rm_lstat	**1**	-0.50391	0.04400	-11.45	<.0001	-1.94646	78.17365

Fig. 26


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
1 -2.115 0.010
2 -0.976 0.002
3 0.045 0.000
4 0.147 0.000
5 1.536 0.006
6 . .
7 0.925 0.002
8 0.780 0.009
9 0.565 0.002
10 -0.800 0.007
11 -0.091 0.000
12 0.706 0.001
13 . .
14 -0.369 0.000
15 0.518 0.001
16 1.004 0.003
17 0.063 0.000
18 1.393 0.005
19 . .
20 -0.548 0.001
21 0.797 0.001
22 . .
23 -0.256 0.000
24 -0.224 0.000
25 -0.514 0.001
26 . .
27 . .
28 -0.628 0.002
29 . .
30 -0.714 0.002
31 -1.326 0.005
32 -0.143 0.000
33 -0.991 0.002
34 -0.904 0.003
35 -0.757 0.001
36 0.012 0.000
37 0.145 0.000
38 1.283 0.003
39 0.594 0.001
40 -0.177 0.000
41 . .
42 0.003 0.000
43 . .
44 -0.399 0.000
45 -0.262 0.000
46 -0.025 0.000
47 -0.535 0.001
48 -0.877 0.018
49 0.699 0.002
50 -0.212 0.000
51 -0.449 0.001
52 . .
53 . .
54 1.565 0.016
55 0.481 0.000
56 . .
57 0.148 0.000
58 0.721 0.001
59 0.263 0.000
60 1.002 0.004
61 -0.091 0.000
62 -0.255 0.000
63 0.368 0.000
64 2.073 0.026
65 -1.054 0.003
66 -0.309 0.000
67 0.749 0.001
68 0.090 0.000
69 0.510 0.000
70 -0.740 0.002
71 0.158 0.000
72 -0.263 0.000
73 -0.384 0.000
74 -0.346 0.000
75 . .
76 -0.605 0.000
77 -0.440 0.000
78 . .
79 . .
80 -0.370 0.000
81 -0.736 0.001
82 0.086 0.000
83 0.055 0.000
84 -0.163 0.000
85 -0.464 0.000
86 0.524 0.000
87 . .
88 -3.278 0.023
89 -1.658 0.004
90 -1.197 0.002
91 -1.460 0.003
92 -1.174 0.001
93 . .
94 . .
95 . .
96 . .
97 -1.873 0.020
98 0.483 0.001
99 . .
100 0.894 0.002
|Studentized Residual| ≥ 3, Prob ≤ 0.0025
Cook's D ≥ 4 / n = 0.015


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
101 -0.136 0.000
102 . .
103 . .
104 -0.141 0.000
105 0.558 0.001
106 0.792 0.002
107 0.170 0.000
108 -0.864 0.002
109 0.046 0.000
110 0.462 0.001
111 -1.282 0.002
112 -0.173 0.000
113 -0.240 0.000
114 -1.858 0.003
115 -0.414 0.000
116 -0.237 0.000
117 -0.824 0.001
118 0.368 0.000
119 . .
120 . .
121 -0.002 0.000
122 0.622 0.001
123 0.262 0.000
124 . .
125 . .
126 . .
127 -0.081 0.000
128 . .
129 . .
130 -0.682 0.002
131 -0.267 0.000
132 . .
133 . .
134 -0.111 0.000
135 0.050 0.000
136 0.139 0.000
137 . .
138 -0.958 0.004
139 . .
140 -0.321 0.001
141 -1.087 0.035
142 -0.403 0.002
143 0.521 0.003
144 -1.415 0.048
145 0.480 0.006
146 . .
147 -0.171 0.001
148 1.406 0.032
149 . .
150 0.389 0.001
151 0.972 0.006
152 0.067 0.000
153 0.628 0.002
154 -1.105 0.008
155 -1.772 0.020
156 -1.087 0.008
157 2.614 0.016
158 -0.418 0.001
159 -1.629 0.008
160 -0.187 0.000
161 2.729 0.038
162 . .
163 . .
164 0.519 0.001
165 0.471 0.000
166 1.878 0.016
167 0.884 0.002
168 -0.399 0.000
169 -1.075 0.002
170 -1.246 0.003
171 -0.879 0.002
172 1.447 0.006
173 . .
174 . .
175 -0.095 0.000
176 0.169 0.000
177 -0.627 0.001
178 -0.364 0.000
179 . .
180 . .
181 . .
182 1.508 0.007
183 2.772 0.024
184 2.806 0.015
185 3.159 0.044
186 . .
187 0.026 0.000
188 -0.001 0.000
189 . .
190 0.206 0.000
191 . .
192 -0.470 0.001
193 -0.543 0.001
194 1.826 0.016
195 -1.649 0.015
196 . .
197 -0.146 0.000
198 1.487 0.012
199 0.044 0.000
200 -0.026 0.000
|Studentized Residual| ≥ 3, Prob ≤ 0.0025
Cook's D ≥ 4 / n = 0.015


Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
201 .
202 .
203 .
204 0.540 0.001
205 0.558 0.000
206 1.817 0.008
207 1.676 0.004
208 .
209 1.255 0.003
210 .
211 .
212 1.299 0.002
213 0.722 0.001
214 .
215 .
216 .
217 -0.472 0.000
218 -0.792 0.001
219 .
220 .
221 1.009 0.001
222 .
223 0.137 0.000
224 .
225 .
226 .
227 0.506 0.001
228 1.227 0.002
229 -1.652 0.006
230 -2.712 0.070
231 1.045 0.007
232 .
233 0.718 0.000
234 -0.234 0.000
235 -1.619 0.006
236 -1.432 0.003
237 -1.474 0.002
238 -1.621 0.008
239 -0.166 0.000
240 -0.096 0.000
241 -1.331 0.005
242 1.294 0.014
243 .
244 .
245 0.829 0.006
246 1.133 0.003
247 -0.206 0.000
248 -0.446 0.001
249 .
250 -0.502 0.002
251 -0.367 0.002
252 0.023 0.000
253 0.335 0.001
254 1.754 0.008
255 -1.007 0.010
256 -0.101 0.000
257 .
258 0.157 0.000
259 1.870 0.016
260 0.381 0.001
261 -0.573 0.003
262 .
263 0.540 0.002
264 1.367 0.020
265 1.957 0.034
266 0.354 0.001
267 0.124 0.000
268 0.297 0.000
269 .
270 -0.879 0.001
271 -0.663 0.002
272 0.128 0.000
273 .
274 -0.348 0.000
275 0.411 0.000
276 0.647 0.001
277 0.606 0.001
278 0.785 0.003
279 0.337 0.000
280 1.362 0.008
281 1.263 0.009
282 0.120 0.000
283 -1.566 0.005
284 -0.296 0.000
285 -0.237 0.000
286 .
287 -0.417 0.000
288 -1.364 0.004
289 1.006 0.003
290 -0.450 0.000
291 -0.171 0.000
292 -0.209 0.000
293 -0.265 0.000
294 0.077 0.000
295 0.559 0.001
296 -1.475 0.009
297 .
298 .
299 -1.686 0.005
300 .
|Studentized Residual| ≥ 3, Prob ≤ 0.0025
Cook's D ≥ 4 / n = 0.015

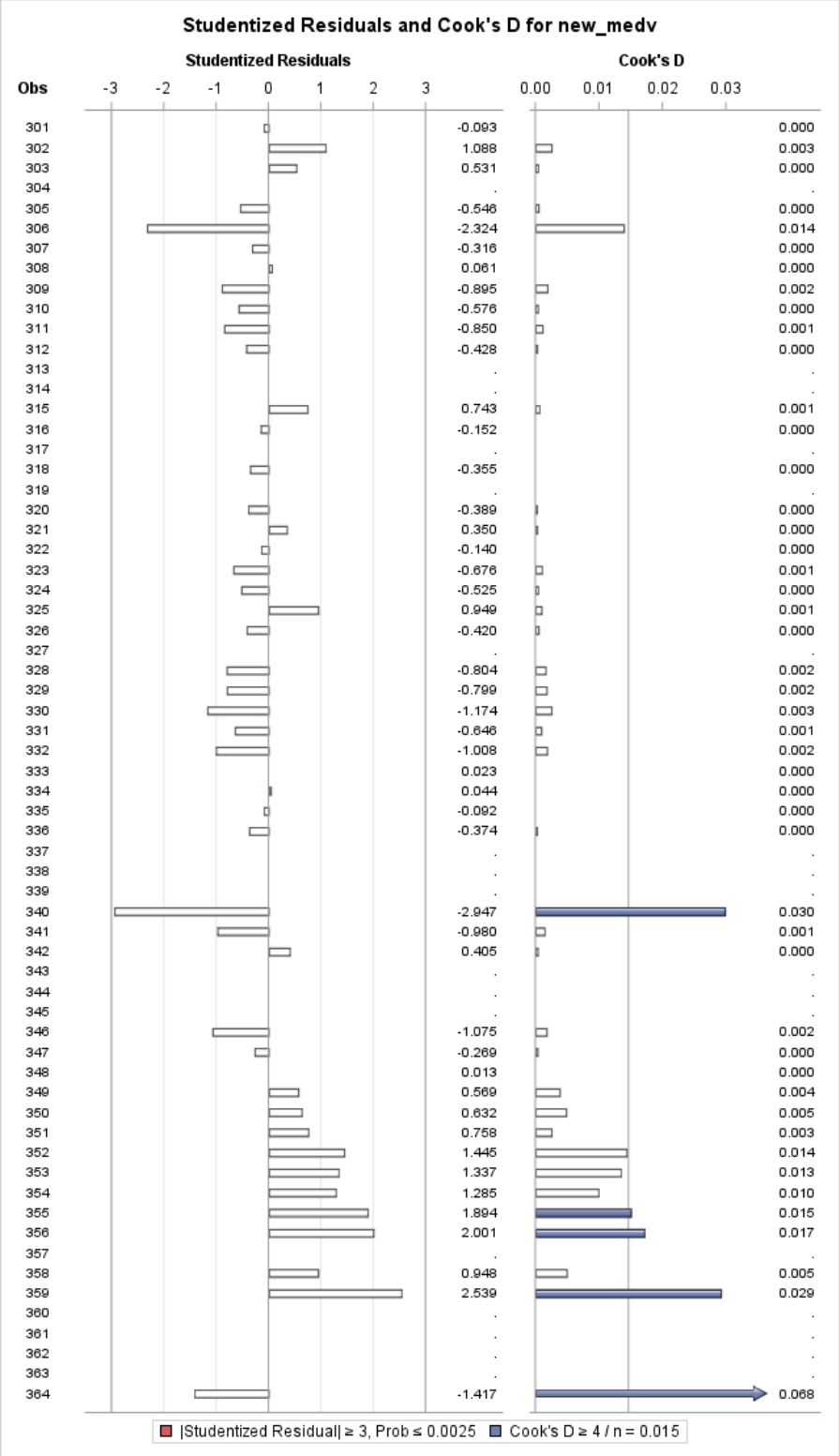

Studentized Residuals and Cook's D for new_medv
Studentized Residuals
Cook's D
Obs
-3 -2 -1 0 1 2 3
0.00 0.01 0.02 0.03
301 -0.093 0.000
302 1.088 0.003
303 0.531 0.000
304 . .
305 -0.546 0.000
306 -2.324 0.014
307 -0.316 0.000
308 0.061 0.000
309 -0.895 0.002
310 -0.576 0.000
311 -0.850 0.001
312 -0.428 0.000
313 . .
314 . .
315 0.743 0.001
316 -0.152 0.000
317 . .
318 -0.355 0.000
319 . .
320 -0.389 0.000
321 0.350 0.000
322 -0.140 0.000
323 -0.676 0.001
324 -0.525 0.000
325 0.949 0.001
326 -0.420 0.000
327 . .
328 -0.804 0.002
329 -0.799 0.002
330 -1.174 0.003
331 -0.646 0.001
332 -1.008 0.002
333 0.023 0.000
334 0.044 0.000
335 -0.092 0.000
336 -0.374 0.000
337 . .
338 . .
339 . .
340 -2.947 0.030
341 -0.980 0.001
342 0.405 0.000
343 . .
344 . .
345 . .
346 -1.075 0.002
347 -0.269 0.000
348 0.013 0.000
349 0.569 0.004
350 0.632 0.005
351 0.758 0.003
352 1.445 0.014
353 1.337 0.013
354 1.285 0.010
355 1.894 0.015
356 2.001 0.017
357 . .
358 0.948 0.005
359 2.539 0.029
360 . .
361 . .
362 . .
363 . .
364 -1.417 0.068
|Studentized Residual| ≥ 3, Prob ≤ 0.0025
Cook's D ≥ 4 / n = 0.015

Fig. 27

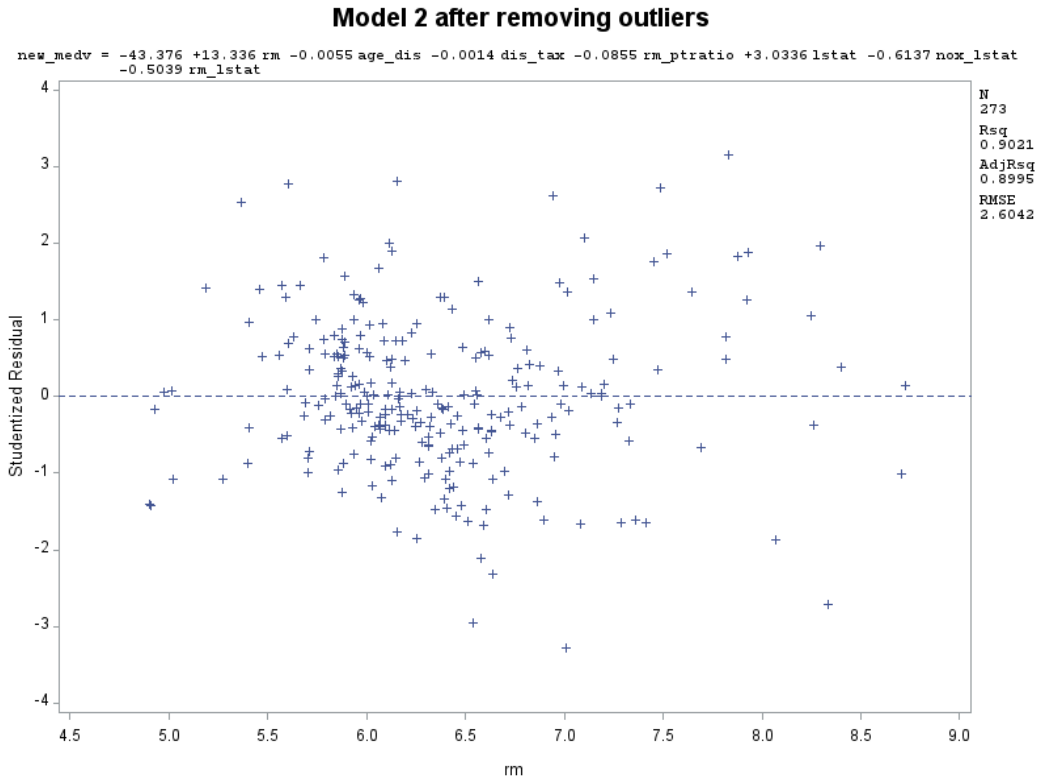

Model 2 after removing outliers
new_medv = -43.376 +13.336 rm -0.0055 age_dis -0.0014 dis_tax -0.0855 rm_ptratio +3.0336 lstat -0.6137 nox_lstat -0.5039 rm_lstat
N 273
Rsq 0.9021
AdjRsq 0.8995
RMSE 2.6042
Studentized Residual
rm

Fig. 28

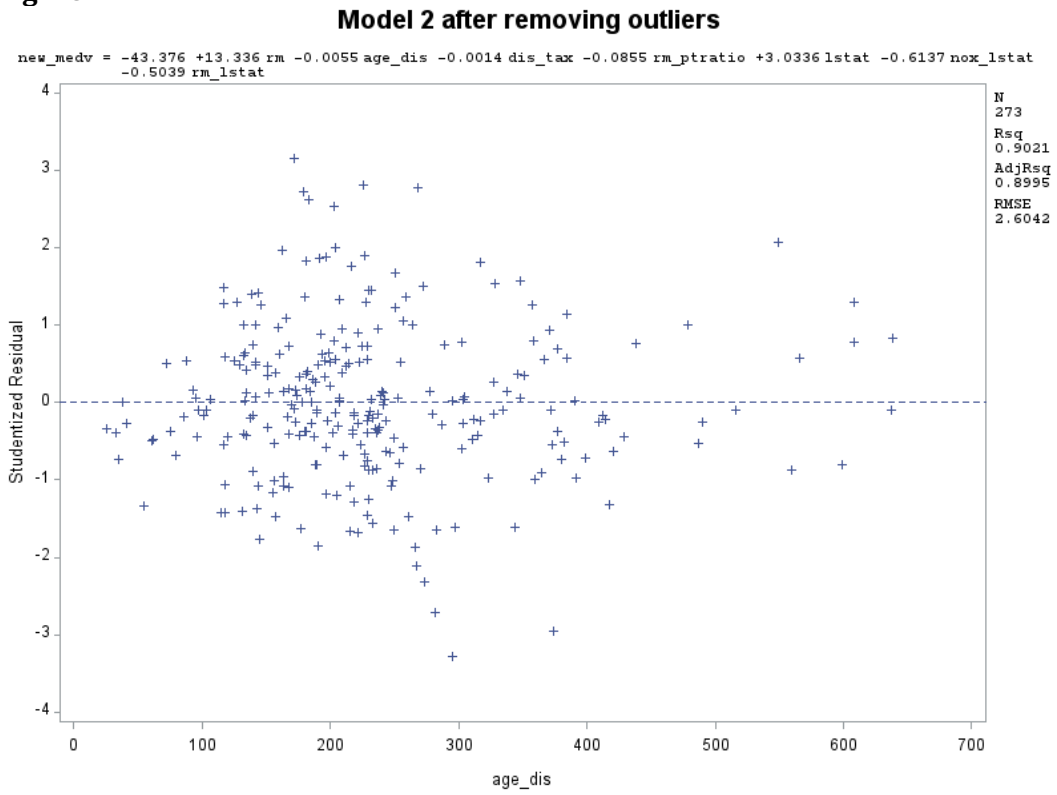

Model 2 after removing outliers
new_medv = -43.376 +13.336 rm -0.0055 age_dis -0.0014 dis_tax -0.0855 rm_ptratio +3.0336 lstat -0.6137 nox_lstat -0.5039 rm_lstat
N 273
Rsq 0.9021
AdjRsq 0.8995
RMSE 2.6042
Studentized Residual
age_dis

Fig. 29

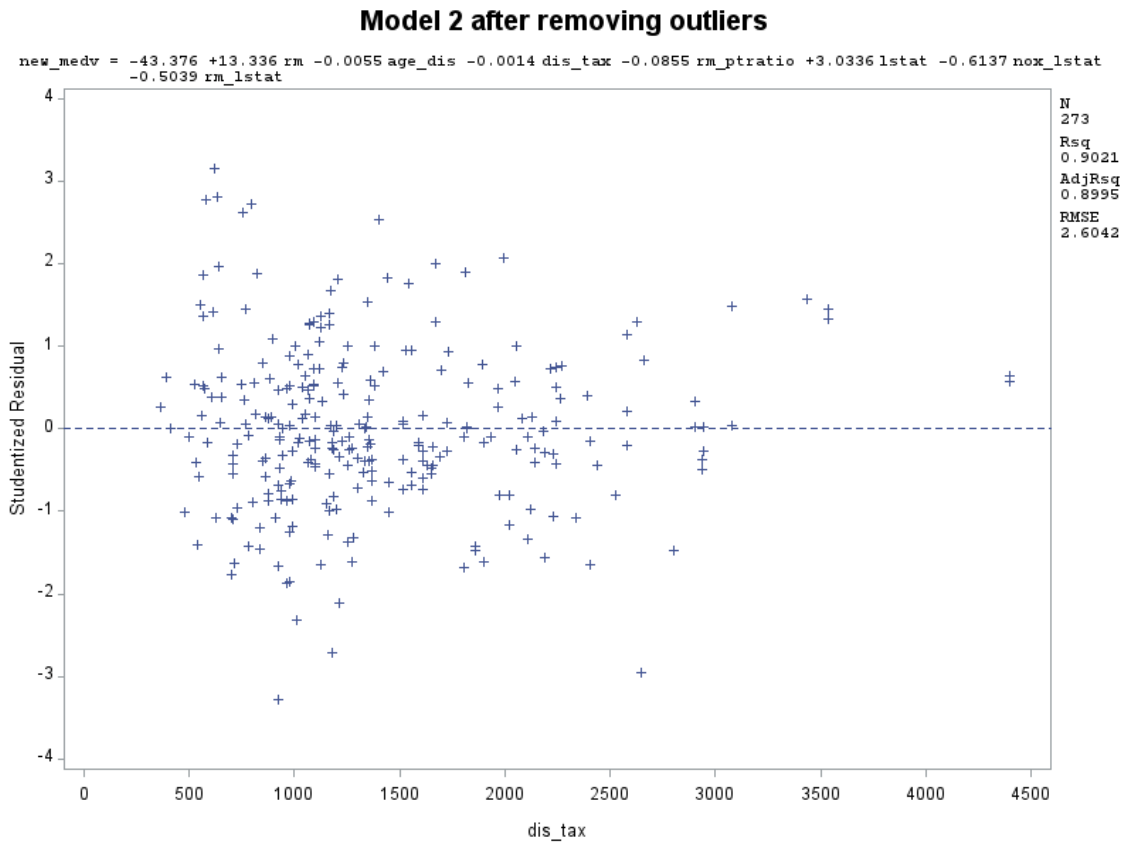

Model 2 after removing outliers
new_medv = -43.376 +13.336 rm -0.0055 age_dis -0.0014 dis_tax -0.0855 rm_ptratio +3.0336 lstat -0.6137 nox_lstat -0.5039 rm_lstat
N 273
Rsq 0.9021
AdjRsq 0.8995
RMSE 2.6042
Studentized Residual
dis_tax

Fig. 30

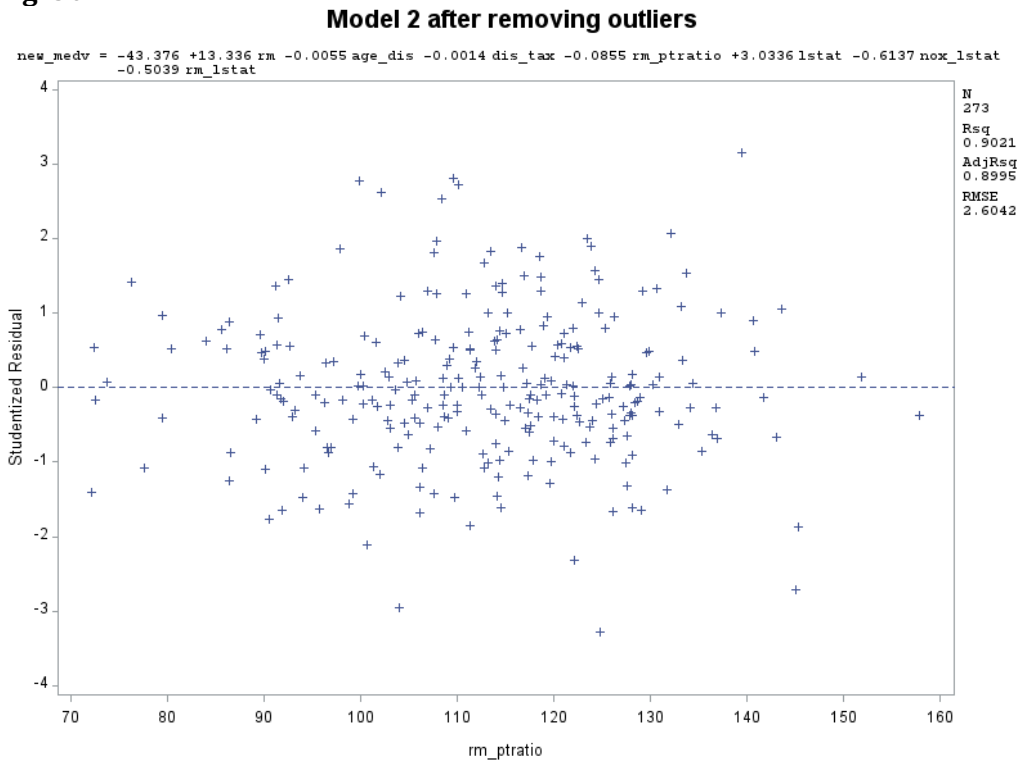

Model 2 after removing outliers
new_medv = -43.376 +13.336 rm -0.0055 age_dis -0.0014 dis_tax -0.0855 rm_ptratio +3.0336 lstat -0.6137 nox_lstat -0.5039 rm_lstat
N 273
Rsq 0.9021
AdjRsq 0.8995
RMSE 2.6042
Studentized Residual
rm_ptratio

Fig. 31

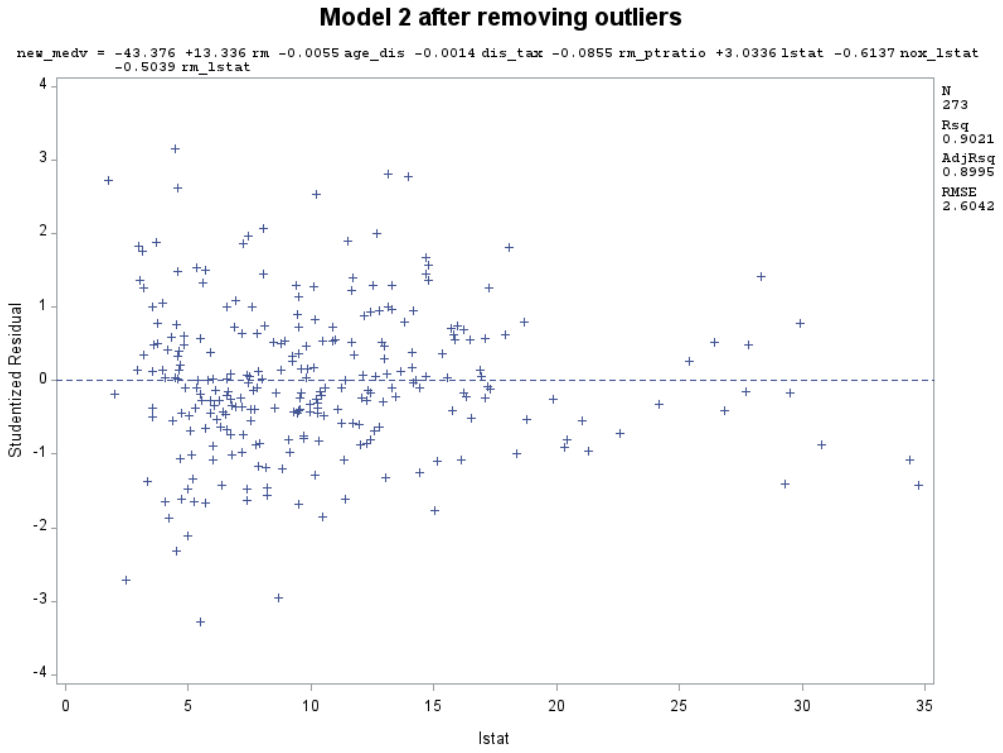

Model 2 after removing outliers
new_medv = -43.376 +13.336 rm -0.0055 age_dis -0.0014 dis_tax -0.0855 rm_ptratio +3.0336 lstat -0.6137 nox_lstat -0.5039 rm_lstat
N
273
Rsq
0.9021
AdjRsq
0.8995
RMSE
2.6042
Studentized Residual
lstat

Fig. 32

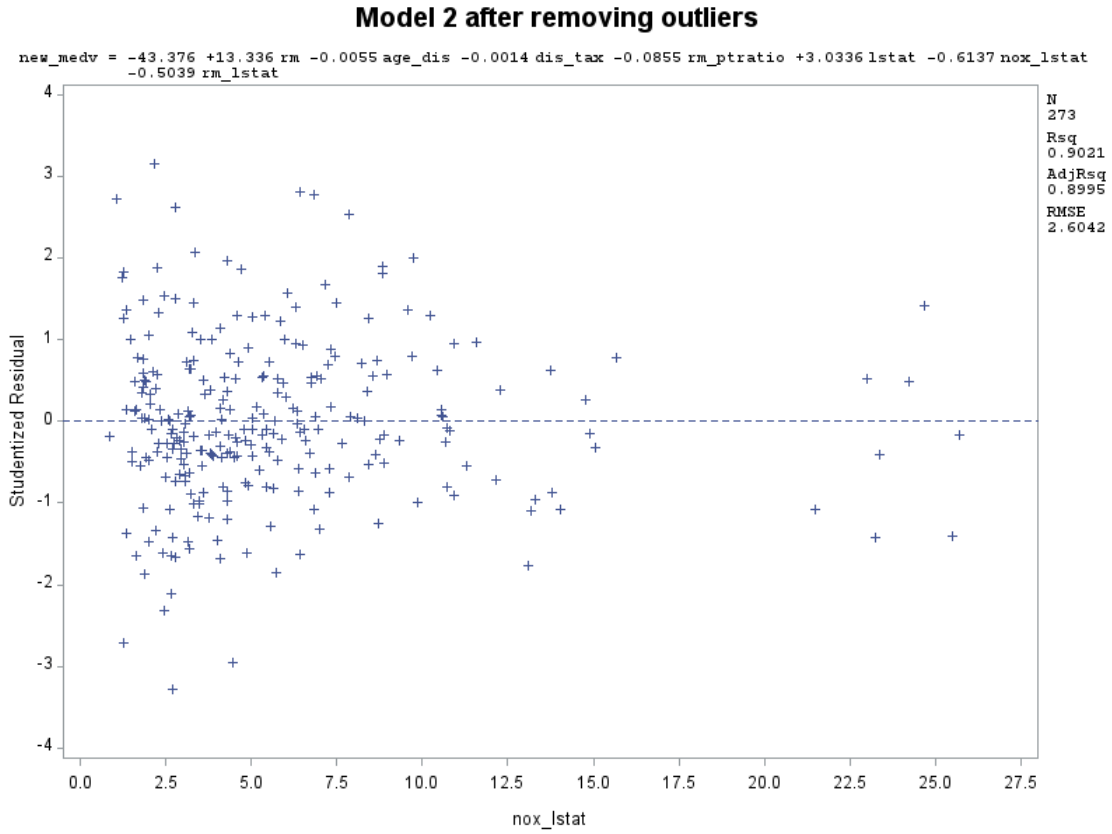

Model 2 after removing outliers
new_medv = -43.376 +13.336 rm -0.0055 age_dis -0.0014 dis_tax -0.0855 rm_ptratio +3.0336 lstat -0.6137 nox_lstat -0.5039 rm_lstat
N
273
Rsq
0.9021
AdjRsq
0.8995
RMSE
2.6042
Studentized Residual
nox_lstat

Fig. 33

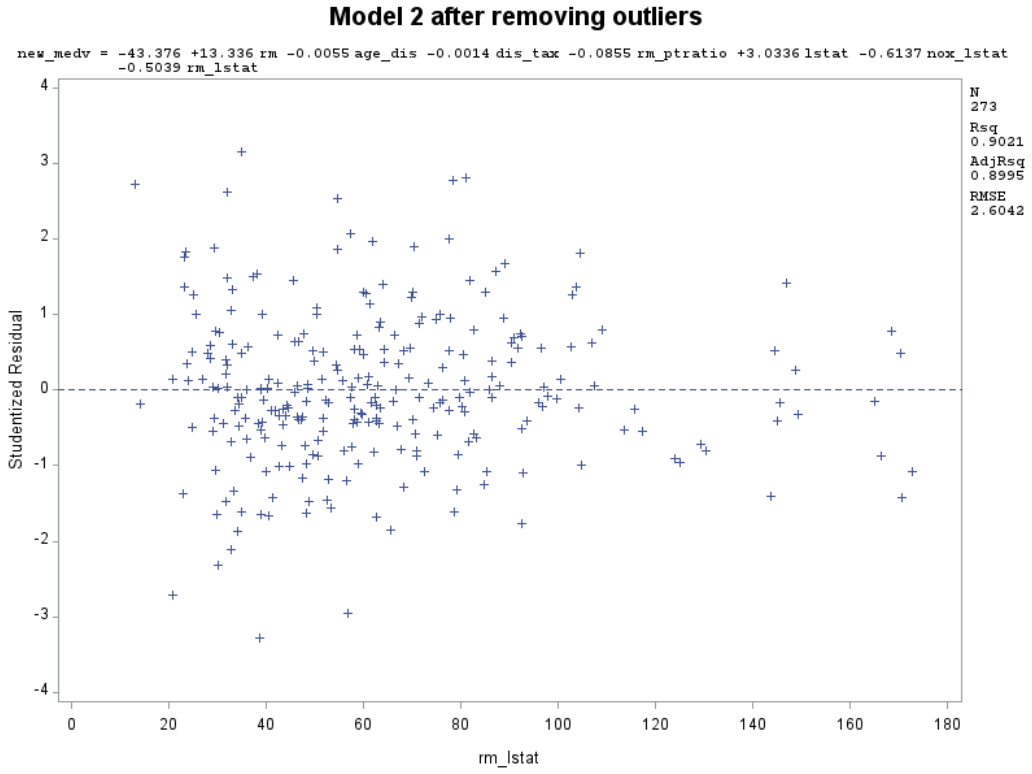

Model 2 after removing outliers
new_medv = -43.376 +13.336 rm -0.0055 age_dis -0.0014 dis_tax -0.0855 rm_ptratio +3.0336 lstat -0.6137 nox_lstat -0.5039 rm_lstat
N 273
Rsq 0.9021
AdjRsq 0.8995
RMSE 2.6042
Studentized Residual
rm_lstat

Fig. 34

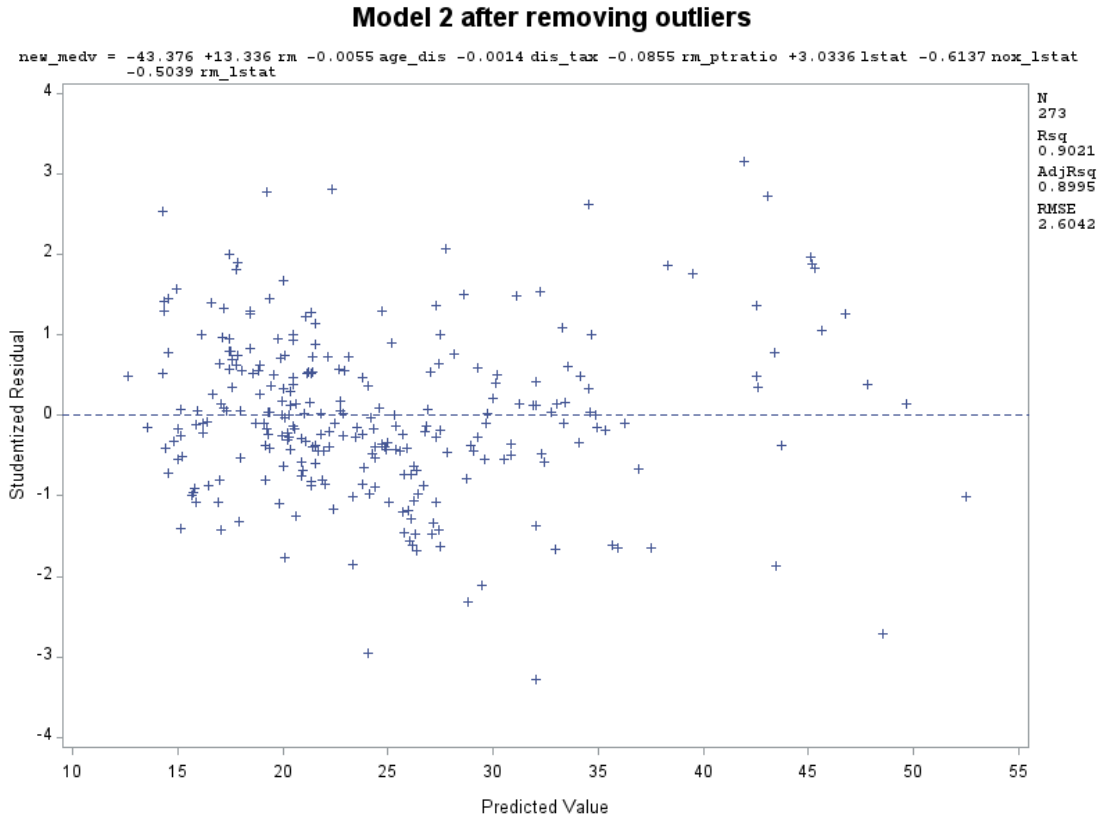

Model 2 after removing outliers
new_medv = -43.376 +13.336 rm -0.0055 age_dis -0.0014 dis_tax -0.0855 rm_ptratio +3.0336 lstat -0.6137 nox_lstat -0.5039 rm_lstat
N 273
Rsq 0.9021
AdjRsq 0.8995
RMSE 2.6042
Studentized Residual
Predicted Value

Fig. 35

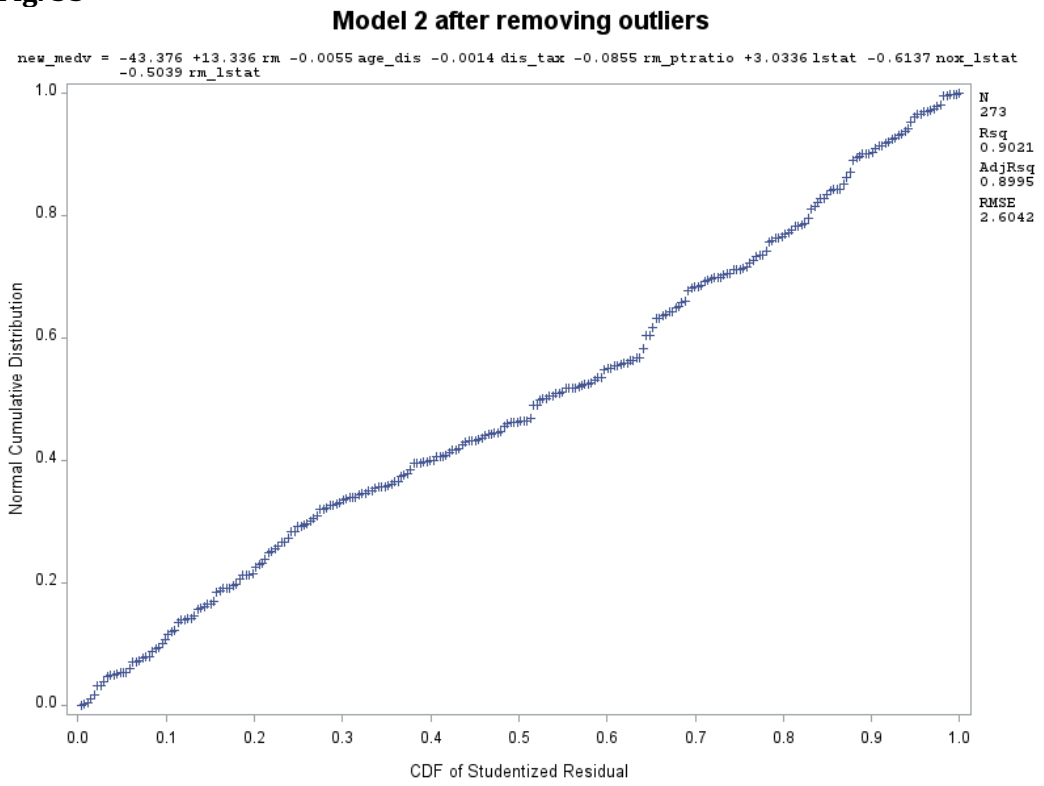


Table 6

Performance test parameters - Model 2

Obs	_TYPE_	_FREQ_	rmse	mae
1	0	91	3.36385	2.22168

Performance test parameters - Model 2

The CORR Procedure

2 Variables:	medv yhat

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**	**Label**
medv	91	25.85055	9.13302	2352	14.30000	50.00000	
yhat	91	25.38279	8.81051	2310	8.51833	50.24374	Predicted Value of new_medv

Pearson Correlation Coefficients, N = 91 Prob > |r| under H0: Rho=0		
	medv	**yhat**
medv	1.00000	0.93033 <.0001
yhat Predicted Value of new_medv	0.93033 <.0001	1.00000

APPENDIX E – FINAL MODEL & PREDICTIONS

Table 1

Final Model - The log transformation

The REG Procedure
Model: MODEL1
Dependent Variable: logmedv

Number of Observations Read	367
Number of Observations Used	367

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	12	30.71085	2.55924	210.28	<.0001
Error	354	4.30833	0.01217		
Corrected Total	366	35.01918			

Root MSE	0.11032	**R-Square**	0.8770
Dependent Mean	3.17054	**Adj R-Sq**	0.8728
Coeff Var	3.47952		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standardized Estimate**	**Variance Inflation**
Intercept	**1**	2.50502	0.15491	16.17	<.0001	0	0
crime	**1**	0.04650	0.00828	5.62	<.0001	0.19259	3.38398

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standardized Estimate**	**Variance Inflation**
zn	**1**	0.00070974	0.00032830	2.16	0.0313	0.05996	2.21345
chas	**1**	0.03864	0.02144	1.80	0.0723	0.03529	1.10302
nox	**1**	-0.52470	0.10277	-5.11	<.0001	-0.18340	3.71298
rm	**1**	0.24922	0.01319	18.90	<.0001	0.55401	2.47260
age	**1**	-0.00094060	0.00035229	-2.67	0.0079	-0.08674	3.03690
dis	**1**	-0.03668	0.00489	-7.50	<.0001	-0.25148	3.23198
rad	**1**	0.01153	0.00270	4.27	<.0001	0.14234	3.20443
tax	**1**	-0.00052892	0.00009845	-5.37	<.0001	-0.15733	2.46745
ptratio	**1**	-0.02459	0.00323	-7.62	<.0001	-0.17623	1.53953
minor	**1**	0.00060319	0.00016167	3.73	0.0002	0.08158	1.37591
lstat	**1**	-0.01494	0.00170	-8.81	<.0001	-0.28368	2.98341

Table 2

Final Model - Removing chas

The REG Procedure
Model: MODEL1
Dependent Variable: logmedv

Number of Observations Read	367
Number of Observations Used	367

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	11	30.67131	2.78830	227.66	<.0001
Error	355	4.34787	0.01225		
Corrected Total	366	35.01918			

Root MSE	0.11067	**R-Square**	0.8758
Dependent Mean	3.17054	**Adj R-Sq**	0.8720

Coeff Var	3.49053		

Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standardized Estimate**	**Variance Inflation**
Intercept	**1**	2.50664	0.15540	16.13	<.0001	0	0
crime	**1**	0.04657	0.00831	5.61	<.0001	0.19288	3.38391
zn	**1**	0.00071733	0.00032931	2.18	0.0300	0.06060	2.21308
nox	**1**	-0.52204	0.10309	-5.06	<.0001	-0.18247	3.71222
rm	**1**	0.25034	0.01321	18.95	<.0001	0.55651	2.46710
age	**1**	-0.00092748	0.00035333	-2.63	0.0090	-0.08553	3.03560
dis	**1**	-0.03722	0.00489	-7.60	<.0001	-0.25517	3.21991
rad	**1**	0.01231	0.00268	4.60	<.0001	0.15203	3.12132
tax	**1**	-0.00054065	0.00009854	-5.49	<.0001	-0.16082	2.45667
ptratio	**1**	-0.02506	0.00323	-7.77	<.0001	-0.17964	1.52922
minor	**1**	0.00060768	0.00016217	3.75	0.0002	0.08219	1.37558
lstat	**1**	-0.01482	0.00170	-8.72	<.0001	-0.28150	2.97921

Fig. 1

Predictions

The REG Procedure
Model: MODEL1
Dependent Variable: logmedv

Output Statistics								
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual
1	.	2.1043	0.0418	2.0220	2.1865	1.8716	2.3370	.
2	.	3.1416	0.1361	2.8739	3.4094	2.7966	3.4867	.
3	3.18	3.3085	0.0153	3.2784	3.3386	3.0888	3.5282	-0.1305
4	3.07	3.1659	0.0137	3.1390	3.1928	2.9466	3.3852	-0.0932
5	3.55	3.4469	0.0148	3.4178	3.4761	3.2273	3.6665	0.0998
6	3.51	3.3973	0.0164	3.3650	3.4296	3.1773	3.6174	0.1112
7	3.59	3.3945	0.0164	3.3623	3.4267	3.1745	3.6145	0.1945
8	3.36	3.2091	0.0157	3.1782	3.2400	2.9893	3.4289	0.1478
9	3.13	3.0512	0.0144	3.0228	3.0795	2.8317	3.2706	0.0800

X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X-X