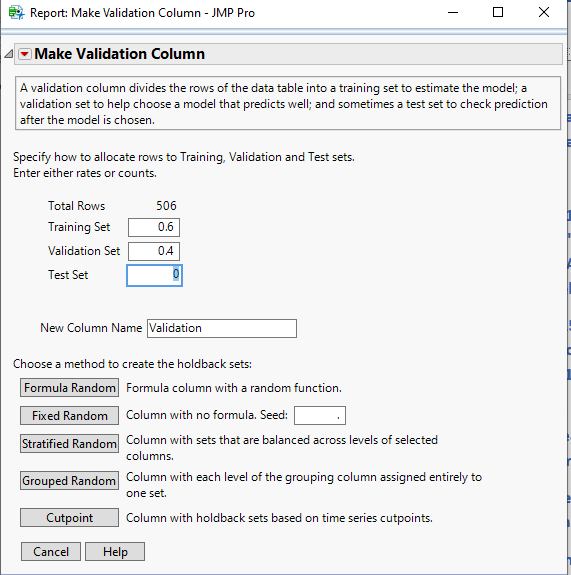
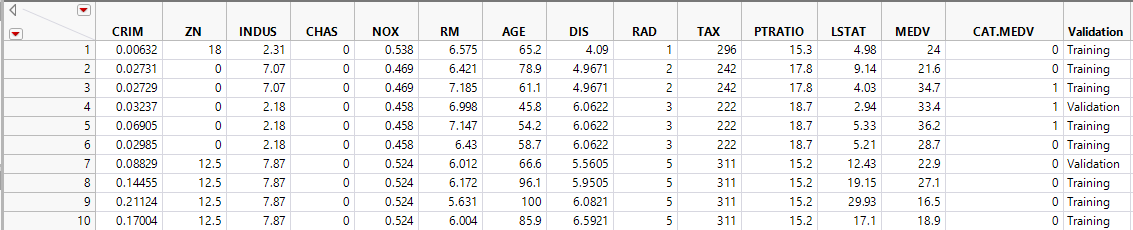
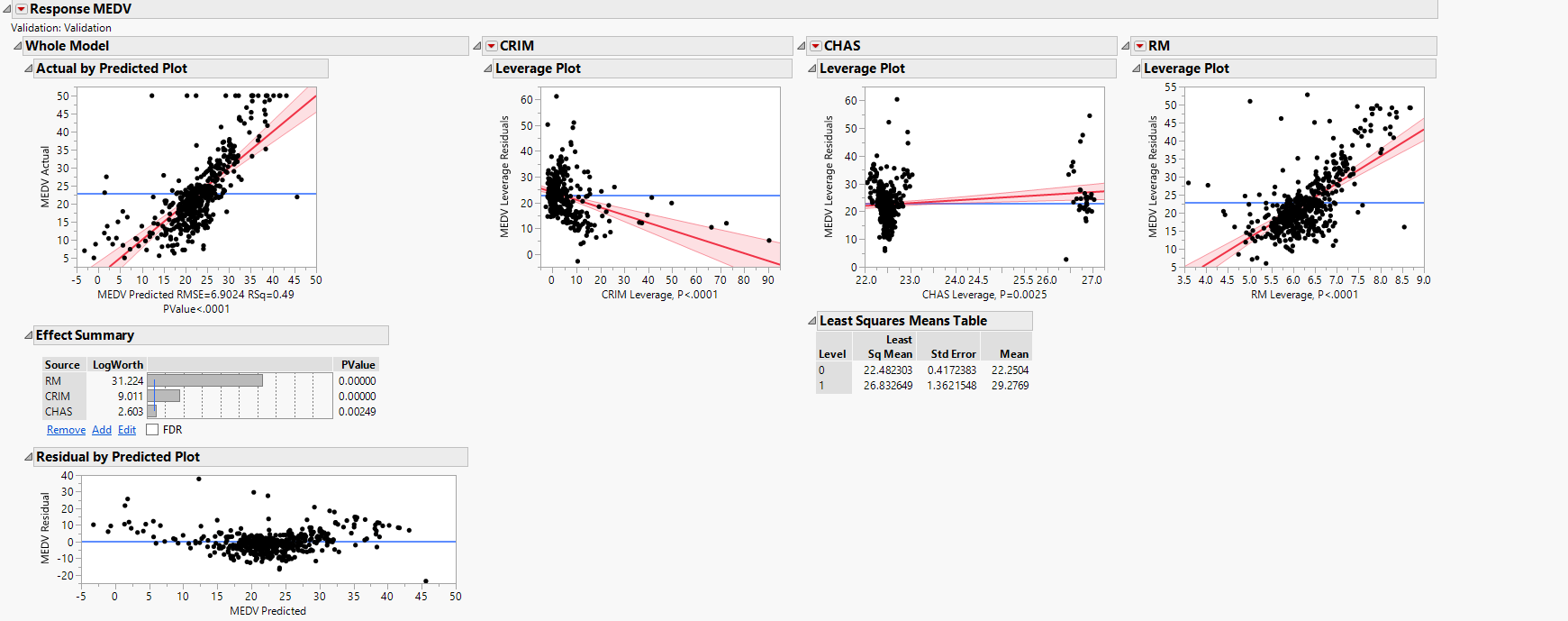
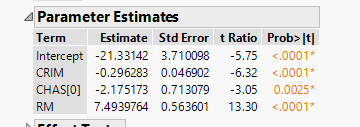
1. **Randomly partition the data in BostonHousing.jmp so that it is 60% Training and 40% Validation.  Include a screen shot of the first 10 rows of data.**

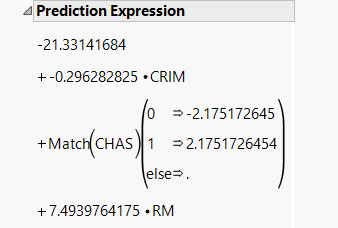




1. **Include screen shots of the Parameter Estimates and the Prediction Formula. "Fit a multiple linear regression model to the median house price (MEDV) as a function of CRIM, CHAS, and RM.  Write the equation for predicting the median house price from the predictors in the model."**







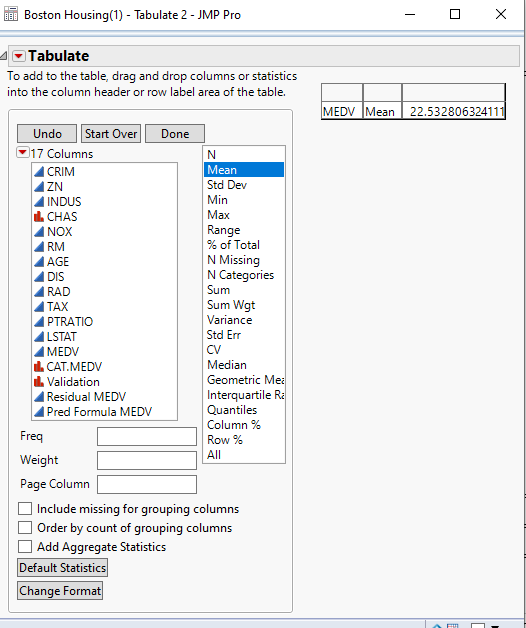
1. **"Using the estimated regression model, what median house price is predicted for a tract in the Boston area that does not bound the Charles River, has a crime rate of 0.1, and where the average number of rooms per house is 6? What is the prediction error?"**

MEDV= -21.33141684+(-0.296282825\*CRIM) +(-2.175172645\*CHAS) +(7.4939764175\*RM)

MEDV= -21.33141684 - 0.0296282825 + (-2.175172645\*0) + (7.4939764175\*6)

MEDV= 23.6028133825

Median house price is $23,602.81

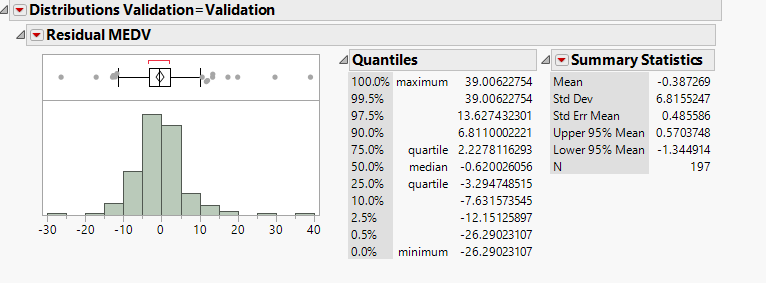


Prediction Error: mean (Actual Median Price) – Predicted Formula Price

= 23.53280 – 23.60281

= -0.07001

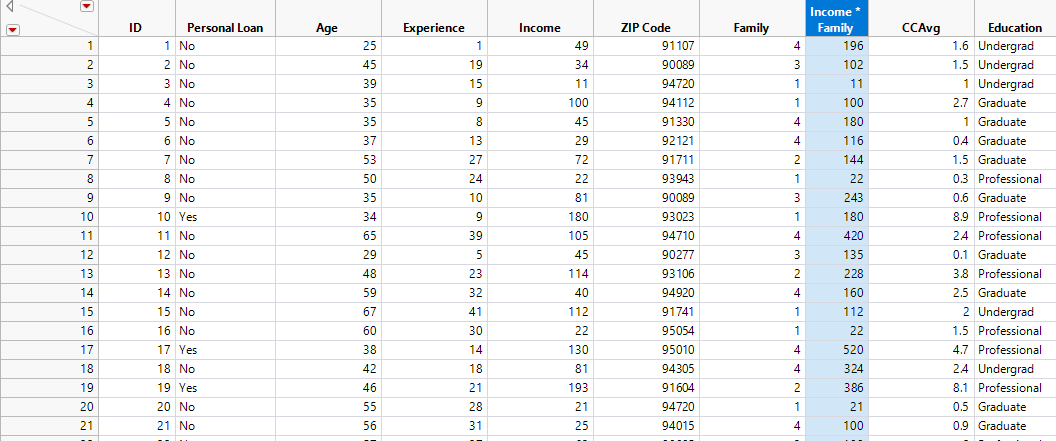
1. **Create a histogram, boxplot, and summary statistics for the model residuals that resulted from the model created in question 2.  What is the average error?  What does this tell us about the median values that the model predicts?**

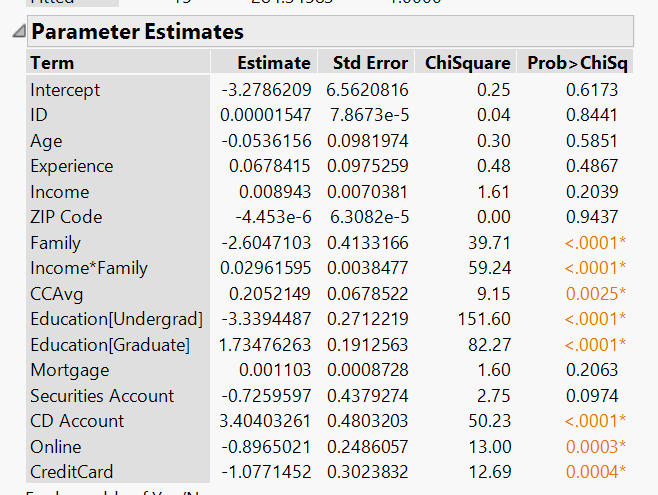


**Average Error = -0.38726**

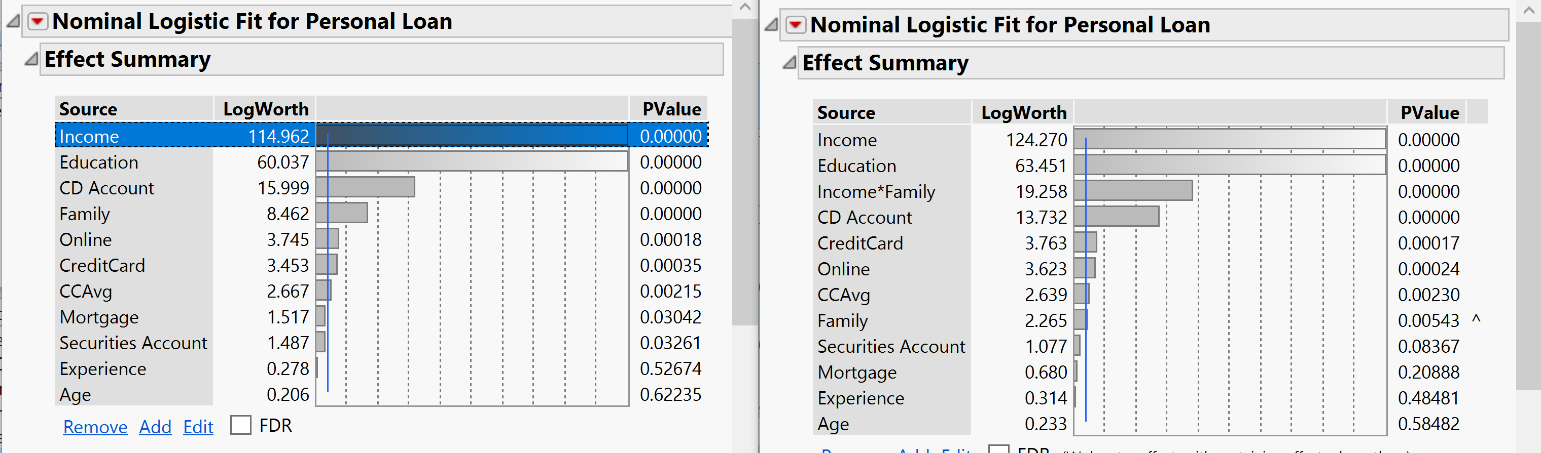
On an average, MEDV will have a deviation of -0.38726 from the original value. The median value of residual error is less than mean. From the boxplot, we can confirm that our data is right skewed.

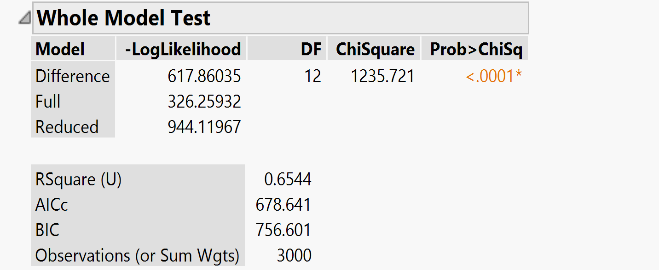
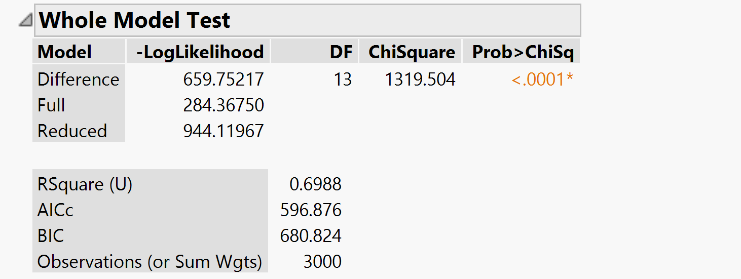
1. **For the UniversalBank.jmp data, modify the model that we created in class to include an interaction term of Income x Family.  Include a screen shot of the resulting Parameter Estimates.**





1. **Is the new interaction term significant?  How do you know?**



The LogWorth value of all parameters increased. The precision of accuracy increased when we added the new interaction term. First model recognized 4796 rows correctly out of 5000 but the Second model recognizes 4820 rows correctly. Therefore, we can conclude that the new parameter is significant.