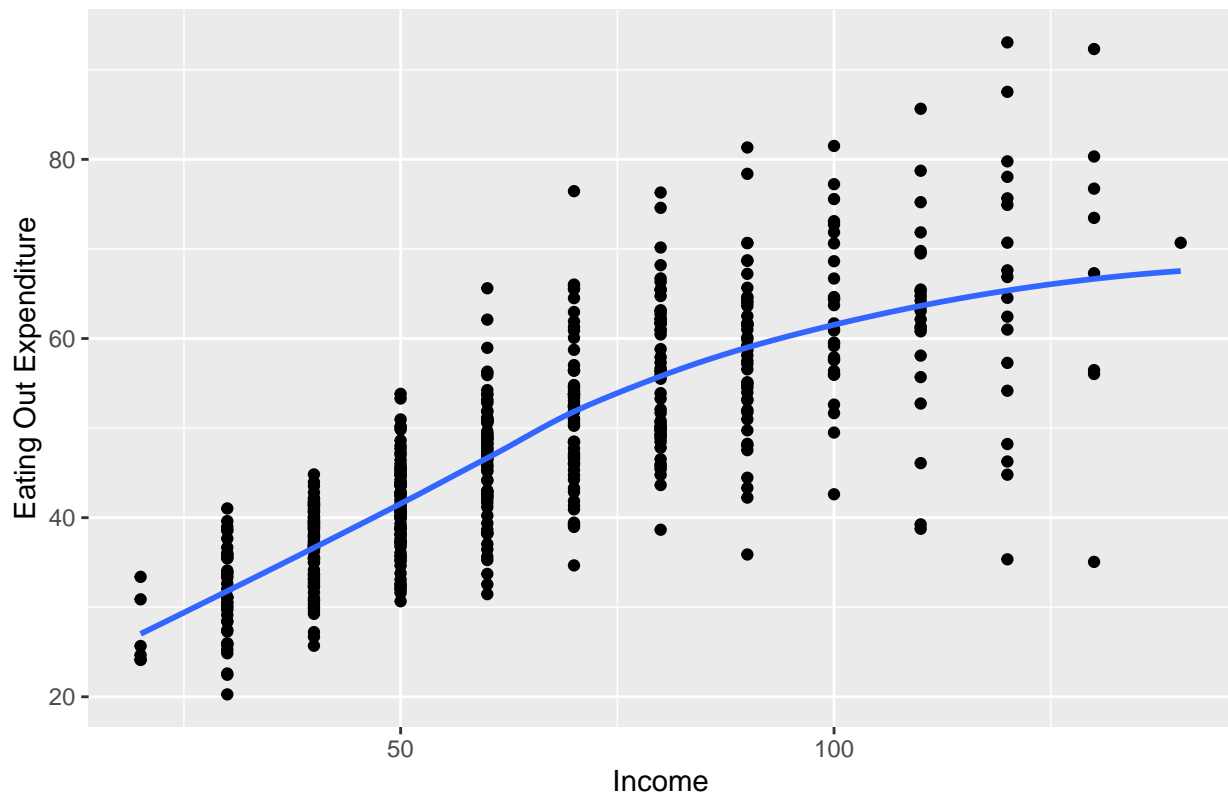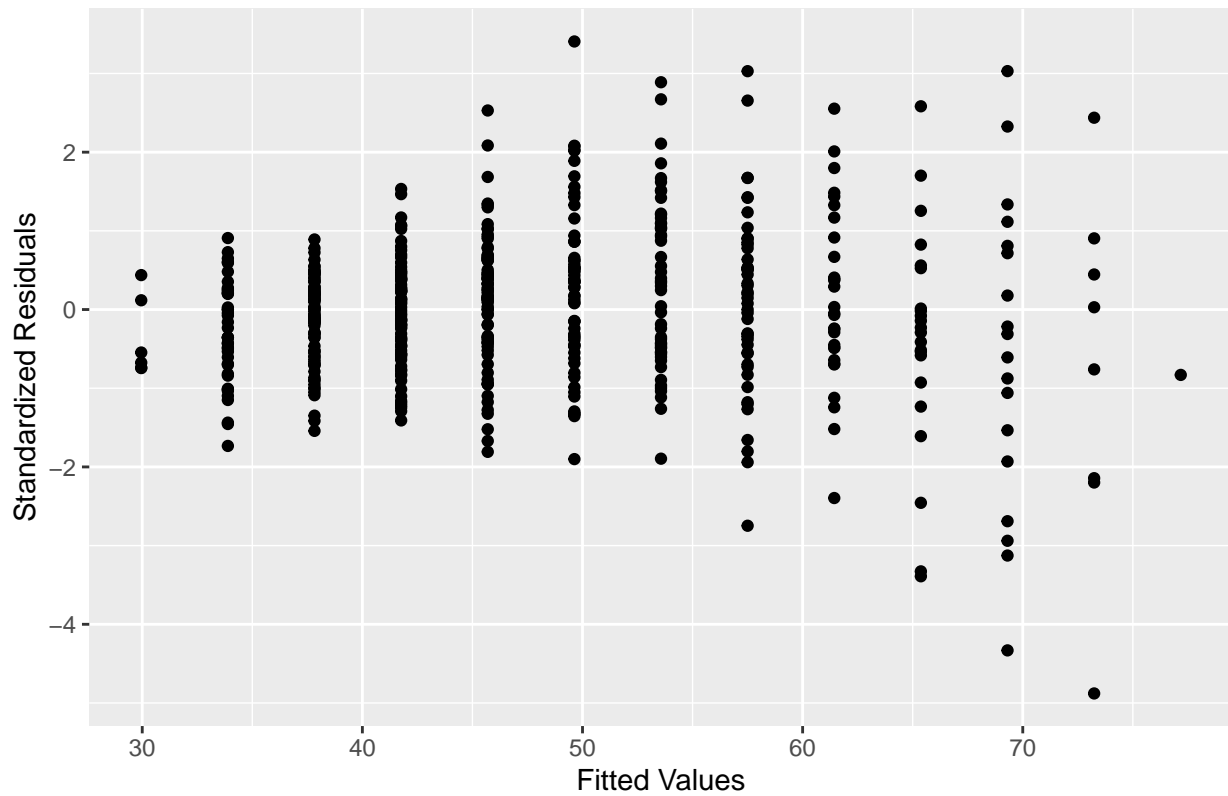# Food Expenditures

Travis Andersen

1.

Scatterplot of eating out expenditures vs income

From the plot above, it seems that as income increases, eating out expenditures increase. The correlation between income and eating out expenditure is 0.7911.

2.

Scatterplot of fitted values against standardized residuals

First, we fit a homoskedastic linear model, and evaluate if the assumption of equal variance is met. Since the variance of the residuals increases as fitted values increase, the assumption is not met. A Breusch-Pagan test for equal variance gives a p-value very close to zero. If we continue to use this model, our confidence and prediction intervals will not work as they should. Therefore, we need to account for heteroskedasticity.

3.

$$\mathbf{y} \sim MVN(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{D}(\theta))$$
$$d_{ii} = e^{2log(X_i)\theta}$$

Above is the heteroskedastic linear regression model in matrix form. $\boldsymbol{\beta}$ is the vector of the intercept and the coefficient of income. σ^2 is the estimate of the residual variance. θ is the exponential variance parameter, which controls how much variance any particular observation is assigned. We can see the impact that income has on food expenditures by constructing a confidence interval for β_1, the coefficient of income.

4.

Scatterplot of income against eating out
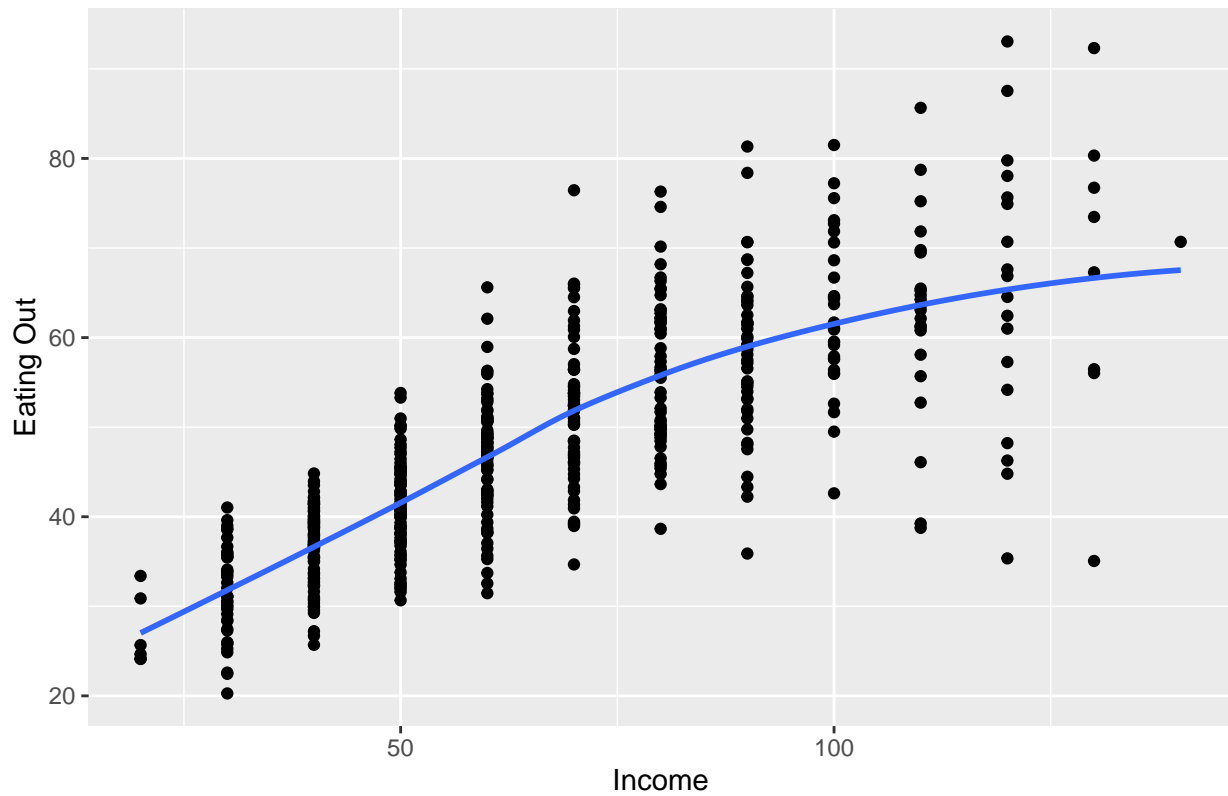
In order to use our model, we must check that it satisfies the assumptions of a linear regression model. First, we draw a scatterplot of the data, and confirm that it seems to be a linear relationship. It is possible that it slightly curves downward, but it is close enough to meet the assumption.

Second, the observations must be independent. The amount one family spends eating out should not significantly impact the amount another family spends eating out, so this assumption is met.

Histogram of standardized residuals
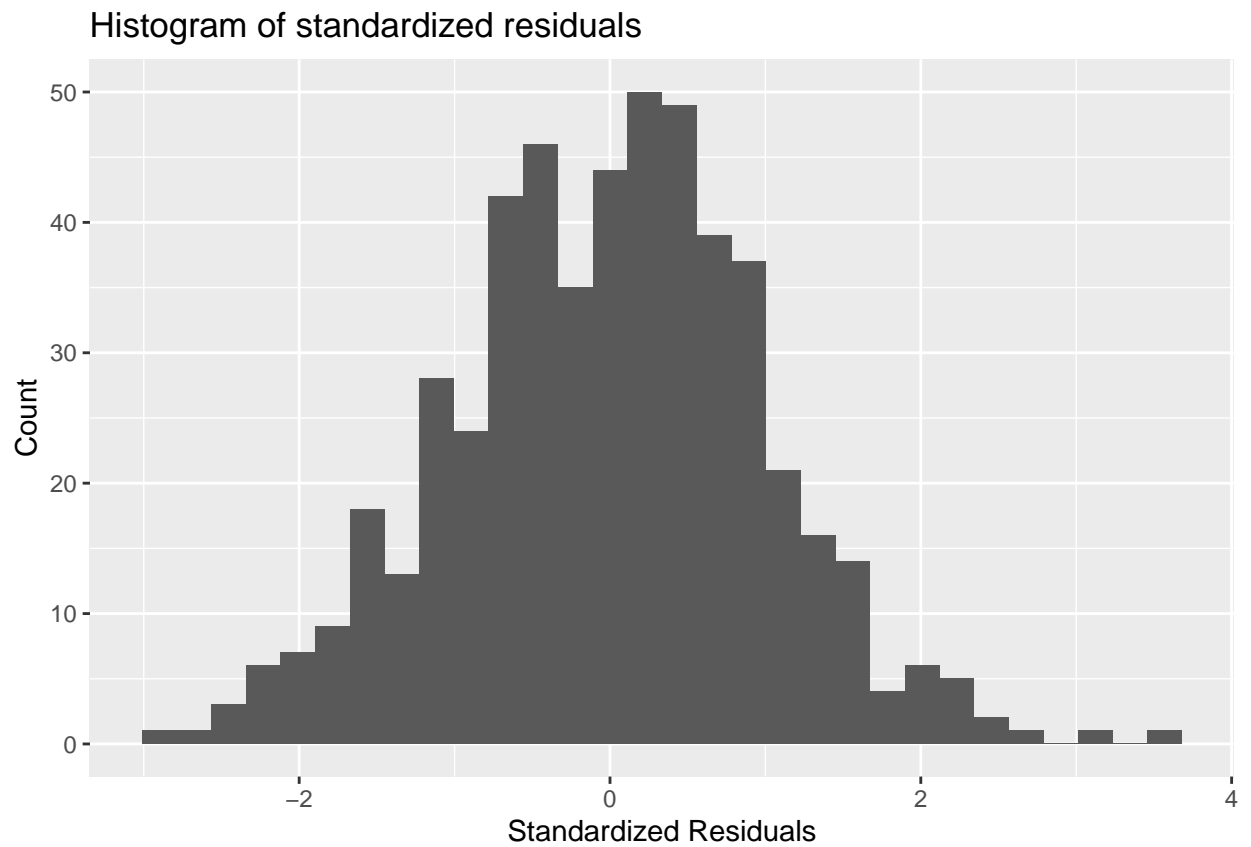
To check that the residuals follow a normal distribution, we check to see if a histogram of the residuals looks fairly normal. The histogram above looks fairly normal, so this condition is met.

## Scatterplot of fitted values against standardized residuals
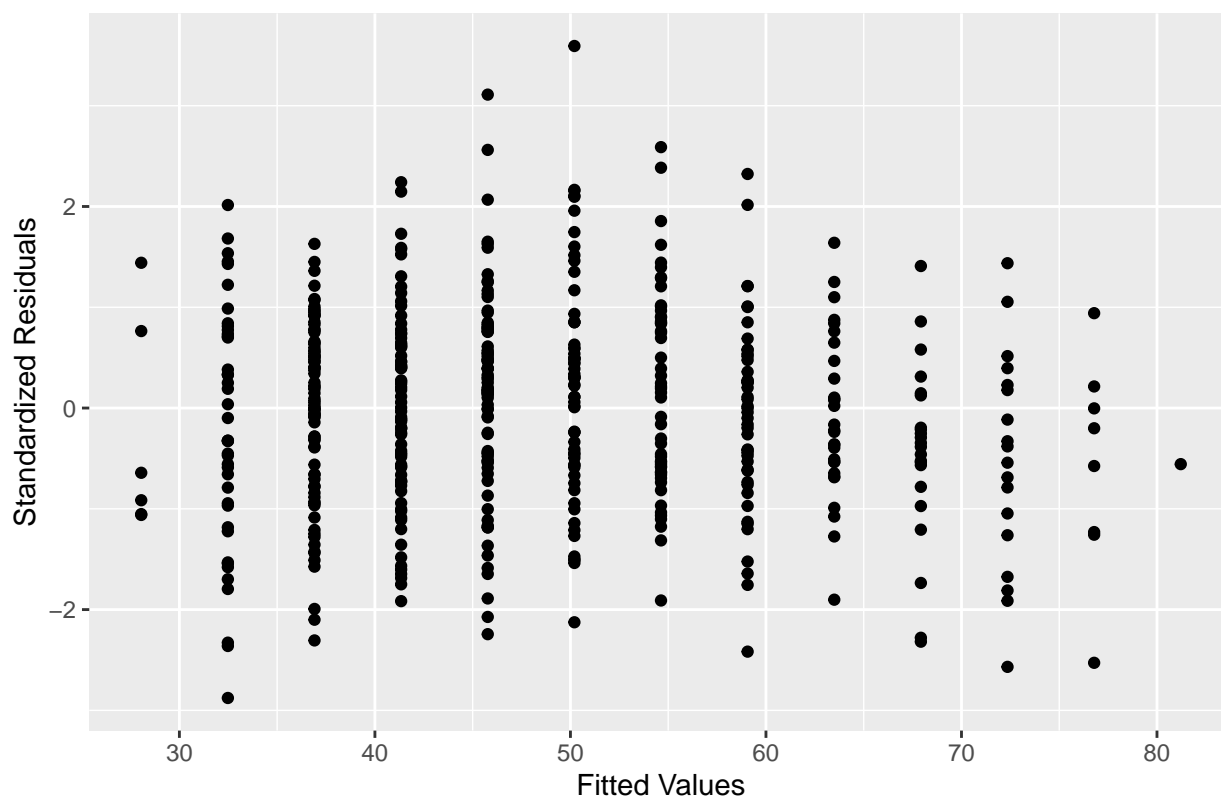
To check for equal variance, we plot fitted values against standardized residuals. Since the variance of the standardized residuals does not change when fitted values change, the assumption of equal variance is validated.

5.
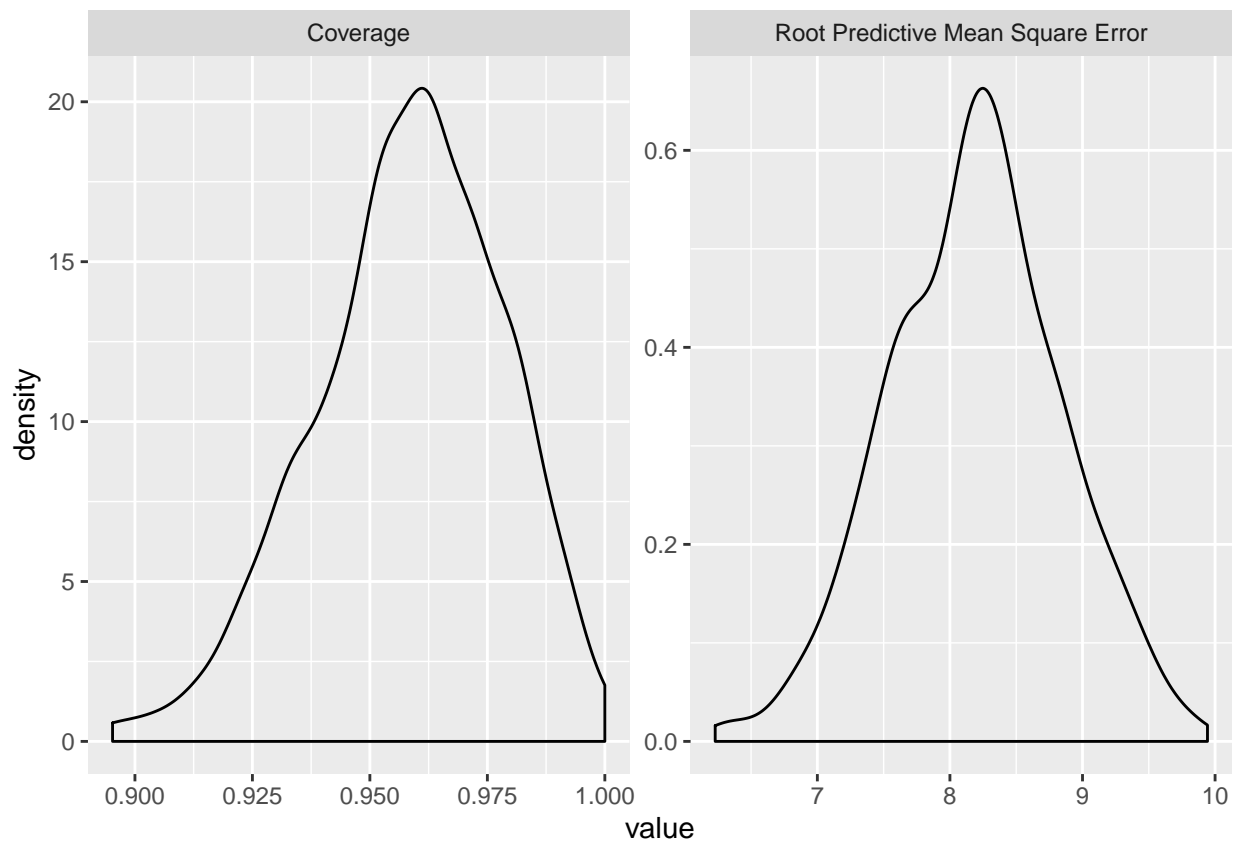
We ran 500 cross validations of our model, and calculated the following summary statistics. The coverage was very close to 0.95, where it should be. The root predictive mean square error was 8. This is pretty good because the standard deviation of food expenditure was 13.
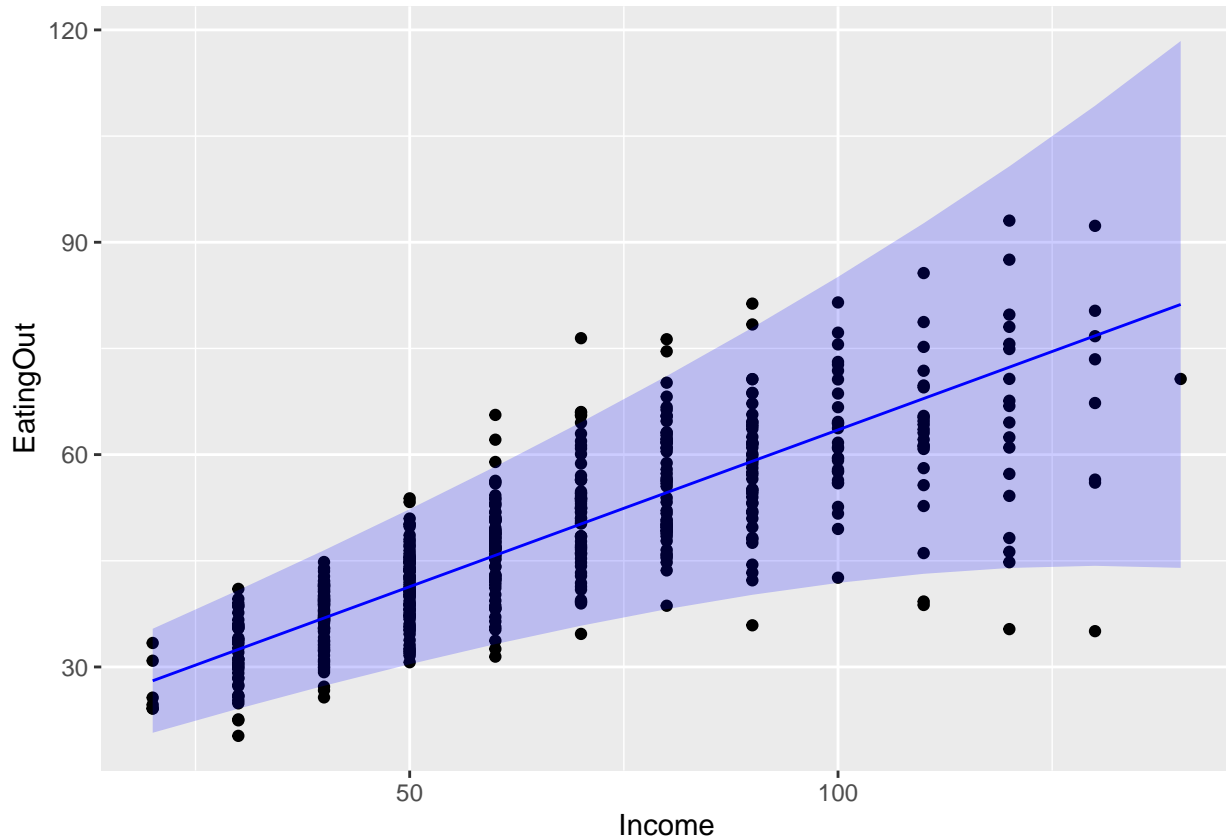
Above is a scatterplot of the data overlaid with model predictions and a 95% prediction interval.

6.

β_{Income} is estimated to be 0.4431, with a 95% confidence interval from 0.4165 to 0.4696. This means that when annual household income increases by one thousand, eating out expenditure increases by 0.4431 on average. θ is estimated to be 0.0135, with a 95% confidence interval from 0.01121 to 0.0159. This means that when salary increases, the variance in eating out expenditures increases.

7.

We want to test if when income increases by 1000, food expenditures increase by 0.5 or more. The null hypothesis is that $\beta_{Income} = 0.5$. The alternative hypothesis is that $\beta_{Income} < 0.5$. The p-value for this test is 0.000025, so we reject the null hypothesis and conclude that $\beta_{Income} = 0.5$. Therefore, the economy is not healthy for restaurant owners.

8.

With an income of 90000, I am predicted to spend 59.06 a week on eating out each week. We are 95% confident that I will spend between 40.20 and 77.93 on eating out each week.