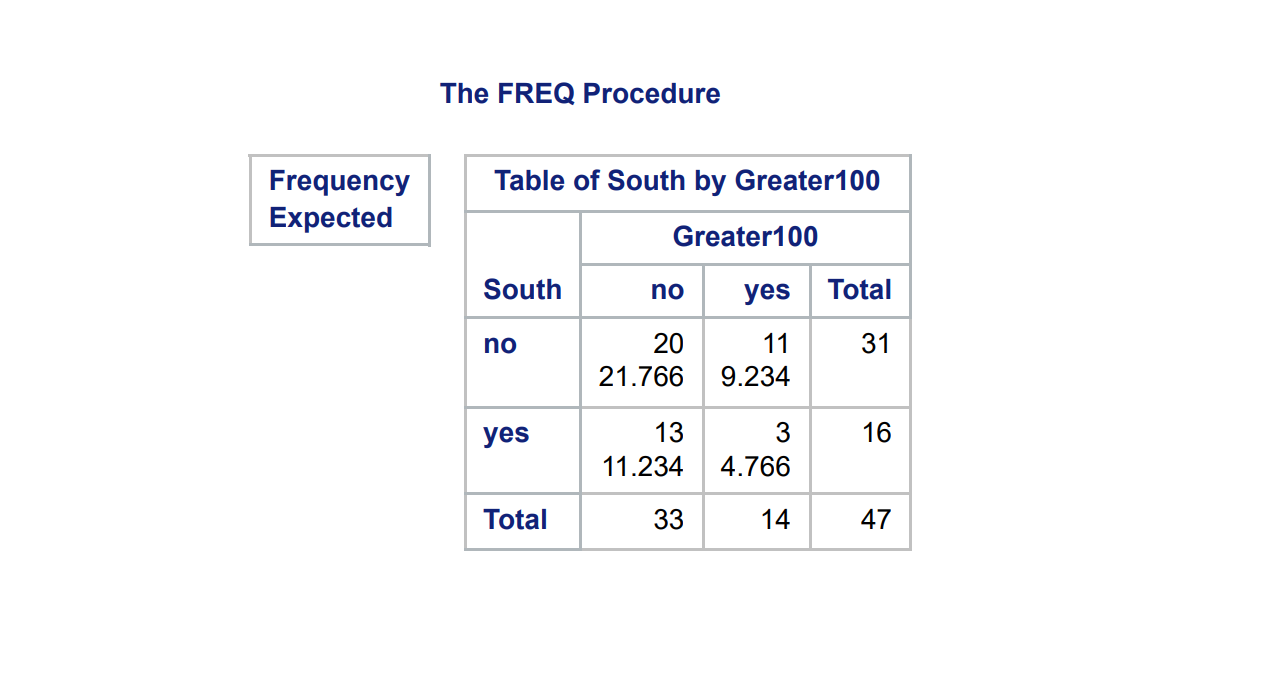
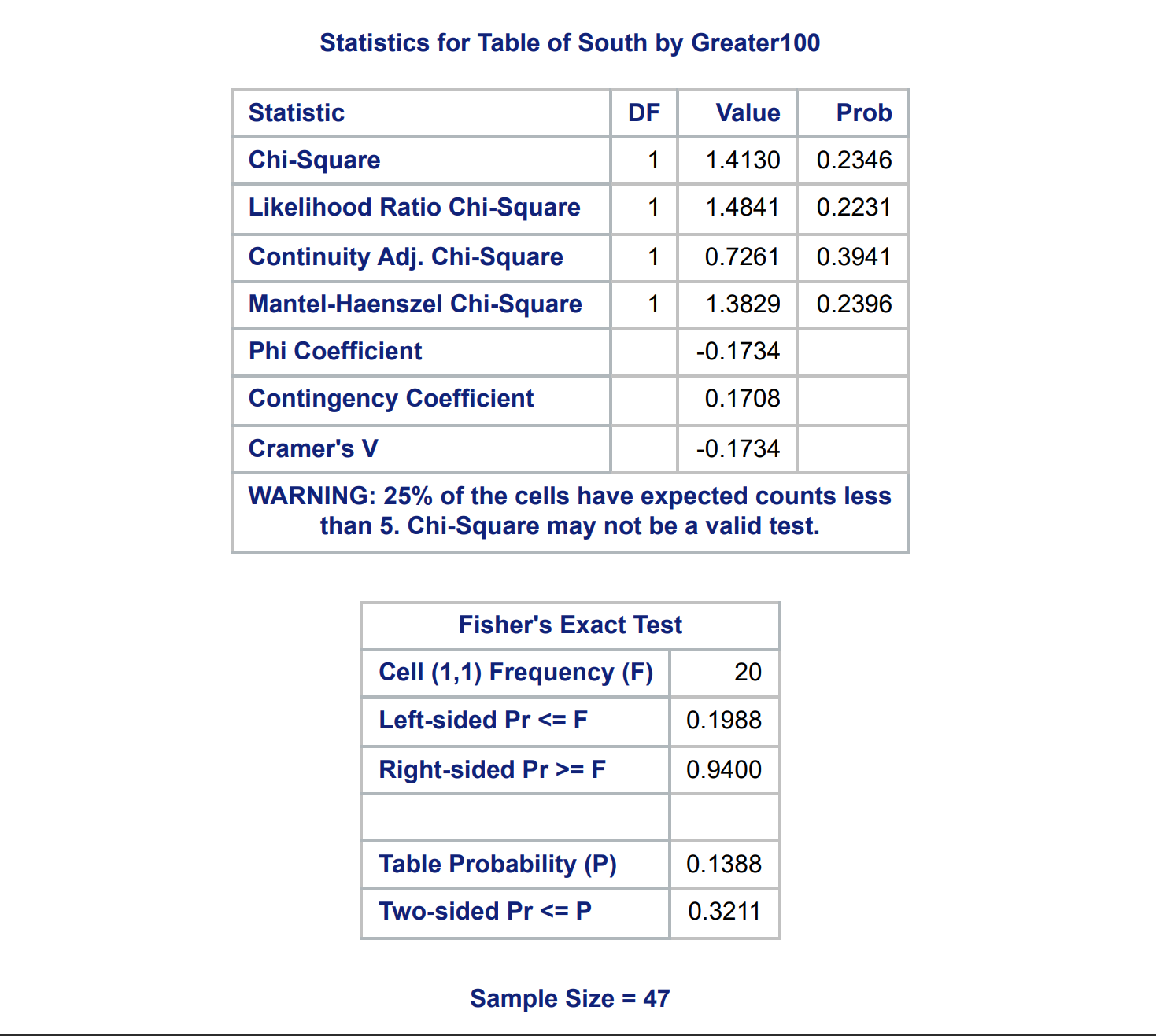
HW 2

## Problem 1

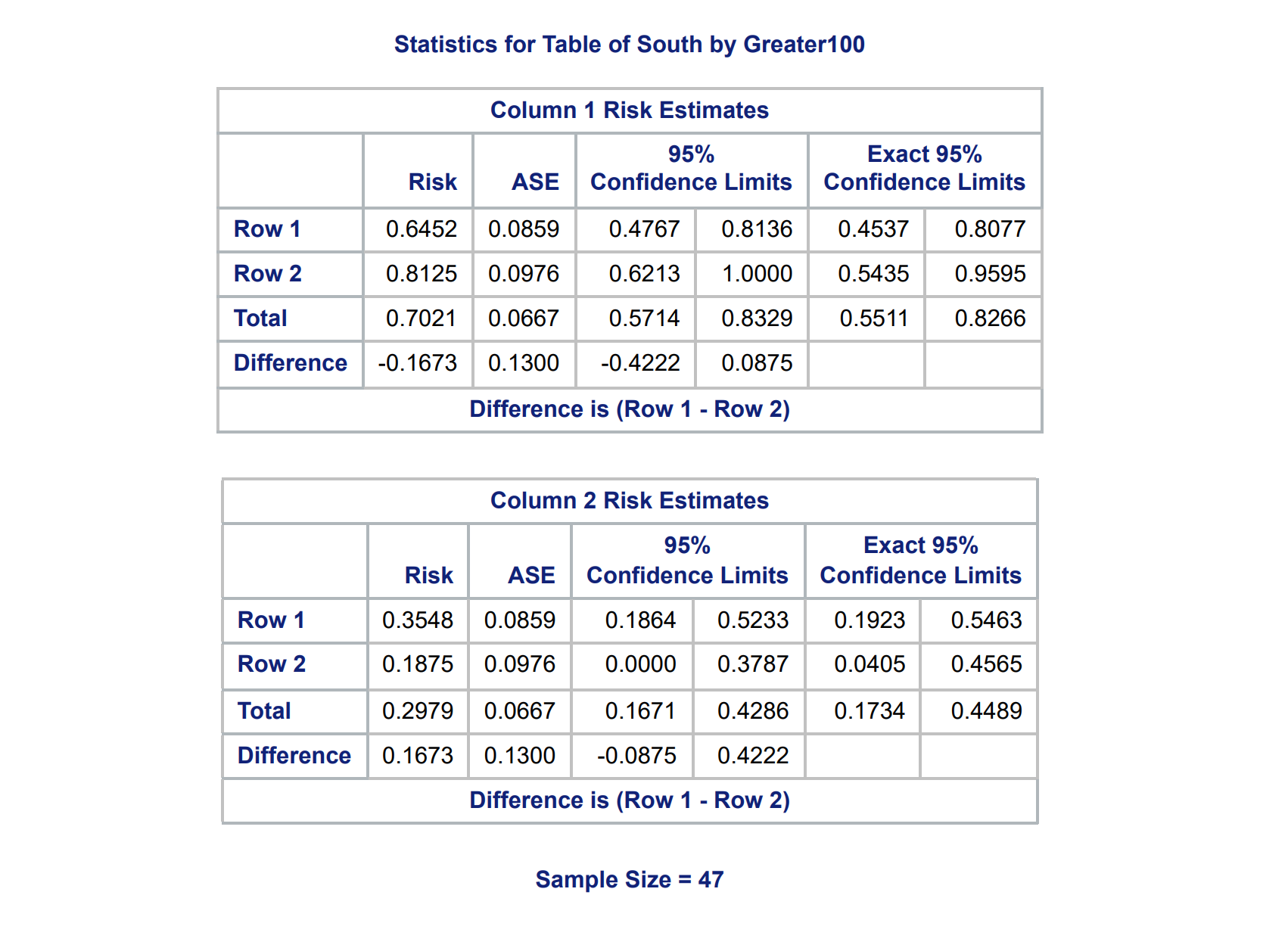
### 1a.)

  
There appears to be a very slight association between the two variables since we see that the expected values don't match with the actual counts. We see that Left to Right Diagonal has values that are both smaller than the expected which indicates a negative association. The phi coefficient confirms this as for these two variables it is -0.17 indicating a negative association.

### 1b.)

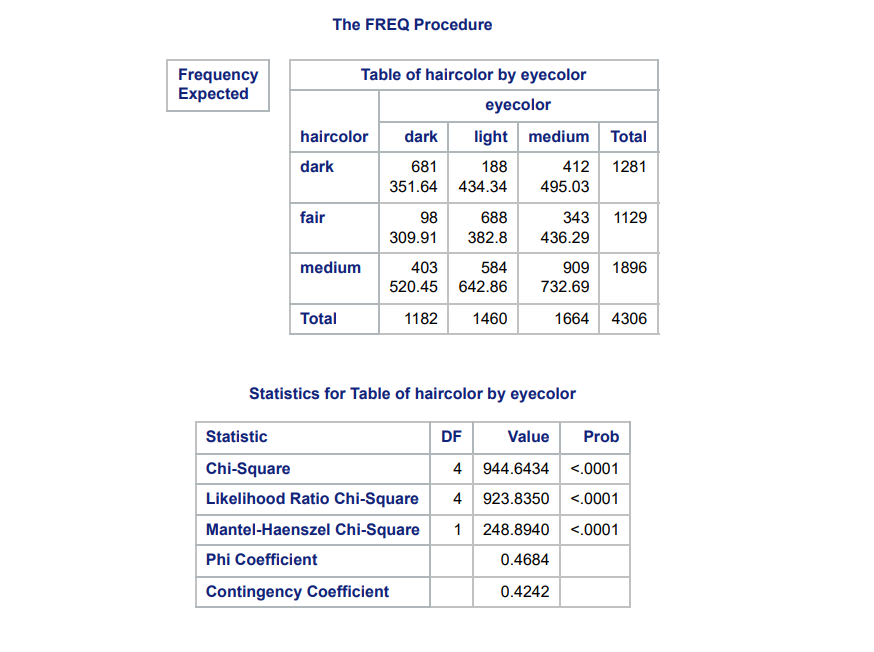
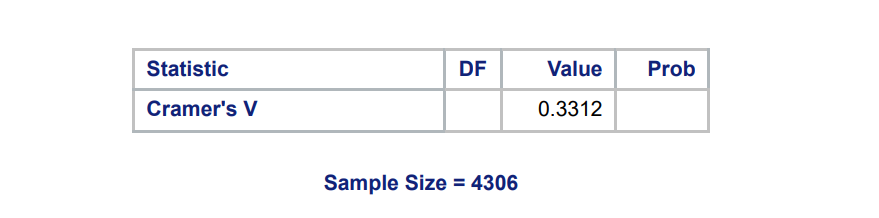
  
Since one of the cells has an expected count less than 5, we will need to use Fisher's Exact Test as opposed to the tests based on the chi-squared statistics. Our Null Hypothesis is that there is no association between Crime Rates greater than 100 per 100 million and whether or not there is a state is in the south. The p-value from the Fisher Exact Test is large, so we fail to reject the Null hypothesis and conclude that there isn't statistically significant evidence for an association between the two variables.

### 1c.)

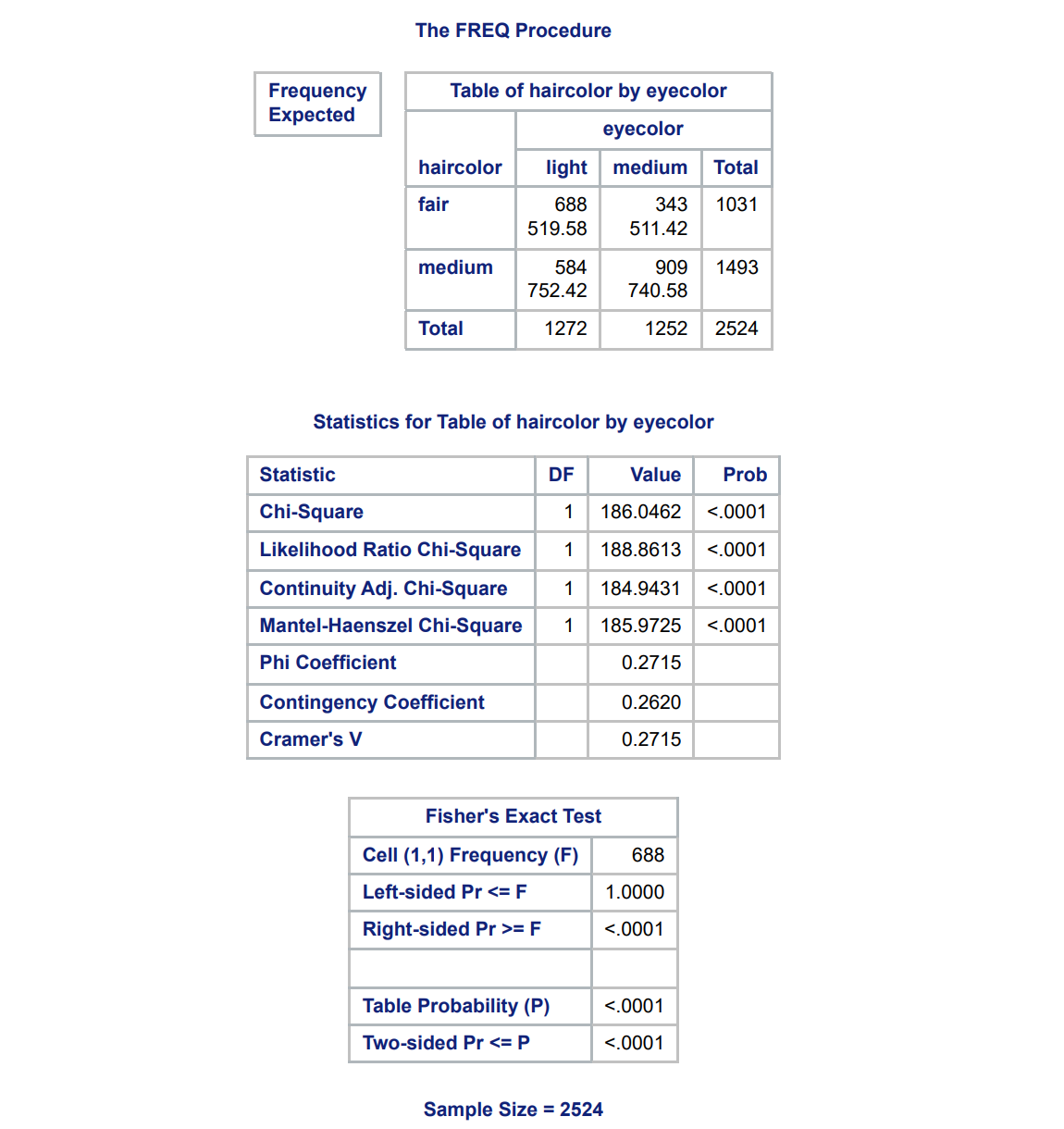
  
Our Null hypothesis for this test is that there is not a statistically significant higher risk of crime rates greater than 100 per million people for states in the south. We see that the confidence intervals for the difference contain zero, so because of this we fail to reject the null hypothesis, and we don't have any statistically significant evidence for southern states having a higher risk of crime.

## Problem 2

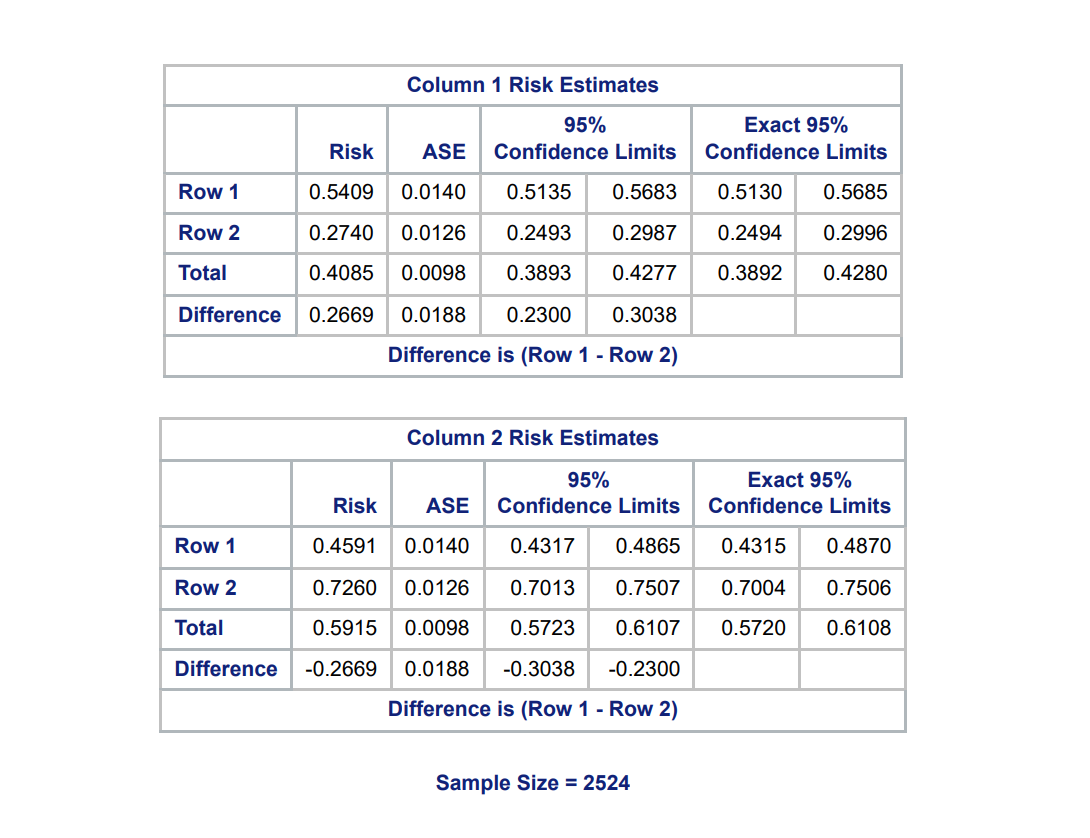
### 2a.)

  
  
All the values on the Left to Right Diagonal are larger than their expected values. Additionally the Top right and Bottom Left values are both smaller than their expected values which suggests a positive association between the two variables. So essentially we see that people with a certain hair color will tend to have the similar eye color. As in Dark hair and Dark eyes, or Medium hair and Medium eyes or Light Hair and Light eyes. The counts where the two variables differ tend to have counts lower than the expected number of counts which further reinforces this observation. The expected counts are large enough so we can also confirm from the Chi-Squared statistics for association that indeed there is an association between the two variables and the association is a positive one because of the positive phi coefficient. So our test would reject the Null hypothesis of their being no association as there is statistically significant evidence to suggest that there is an association between Hair color and eye color.

### 2b.)

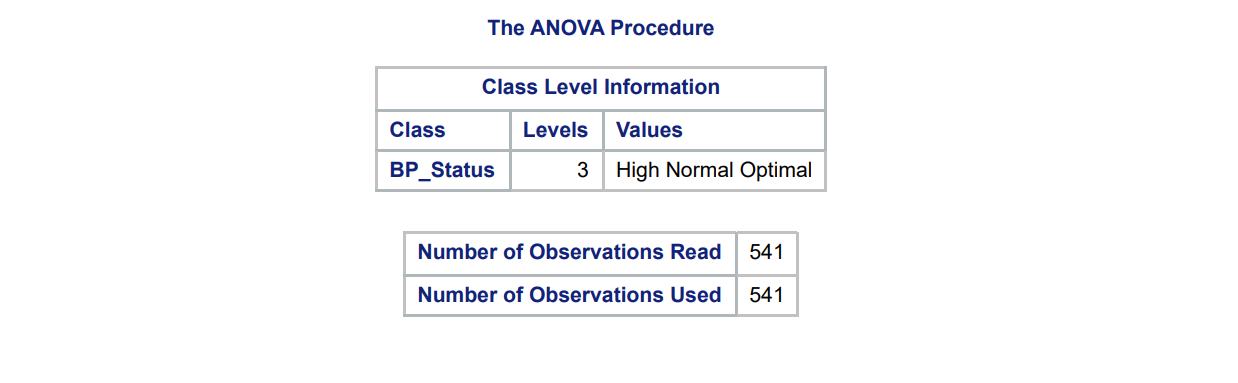
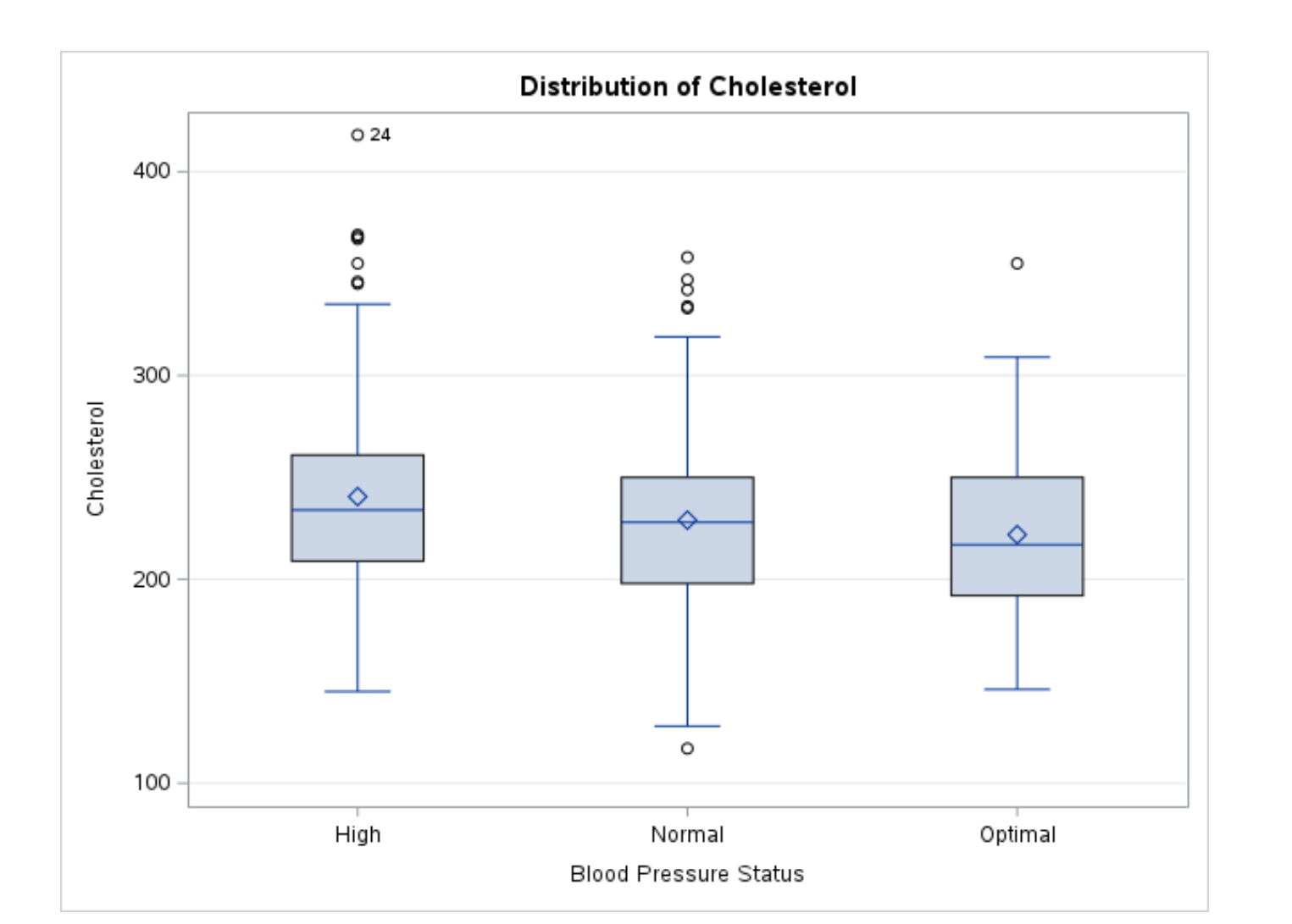
  
We see that there is also an association between hair color and eye color when we drop dark eyes & hair. We see that this association is positive meaning that people with medium hair are more likely to hair medium eye color as well. We see that there is a significant association as well since the p-value of the Chi-squared statistics are all very small and our expected counts are large enough to use these tests. So we have statistically significant evidence for an association between hair color and eye color and thus we reject the null hypothesis of their being no association.

### 2c.)

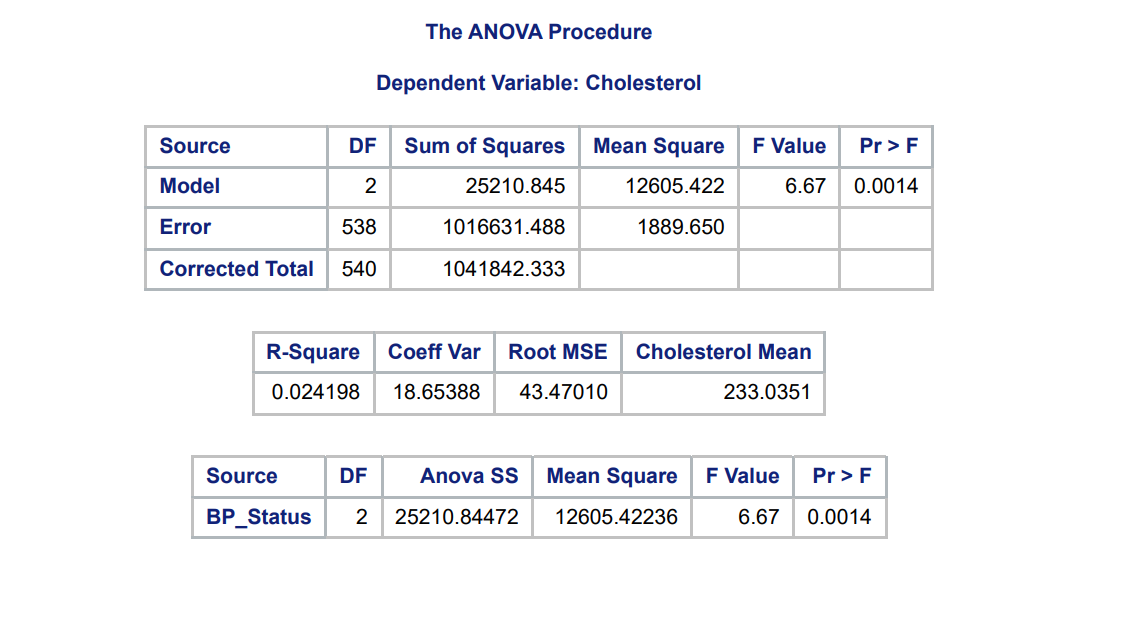
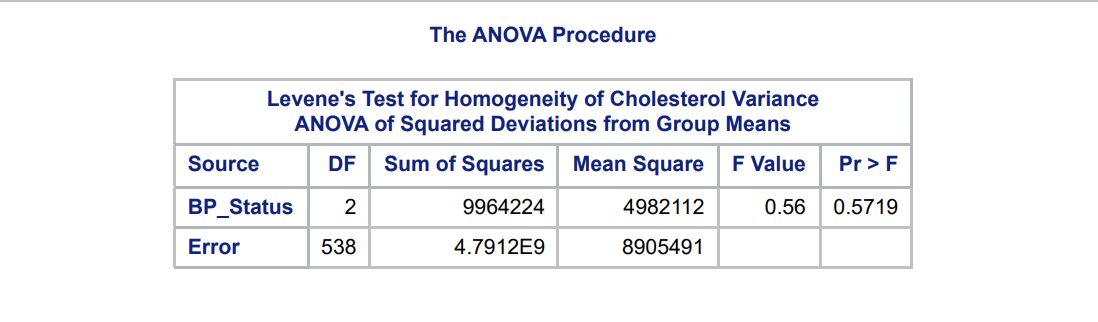
  
Our Null hypothesis will be that people with light eye color don't have a higher risk for having light eyes. However our confidence interval in our first table does not contain zero, so we have statistically significant evidence that those with light eye color are significantly more likely to have fair hair, thus we reject the null hypothesis.

## Problem 3

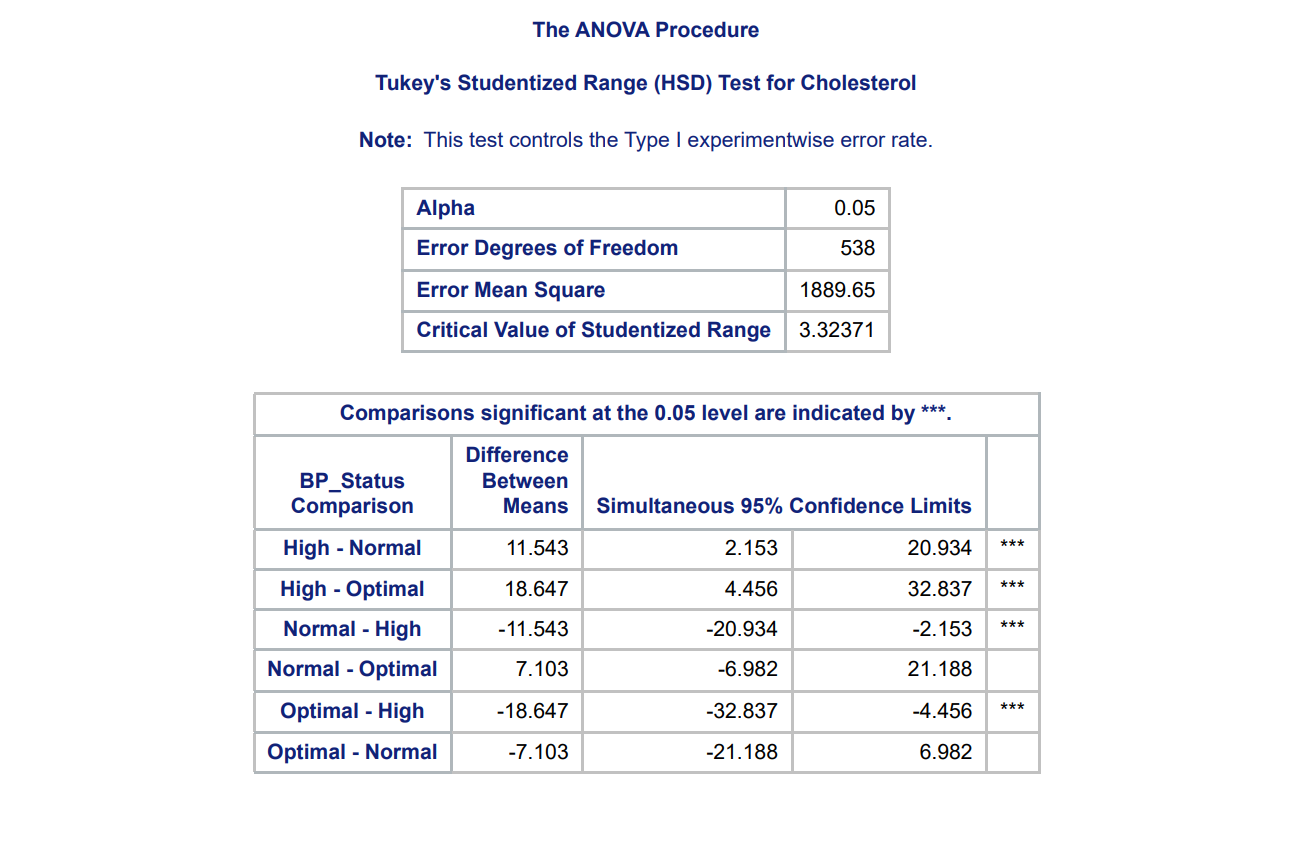
### 3a.)

### 3b.)

  
  
Based on the test for significance, we see that the P-value for the contribution of the model is less than 0.05. So we conclude that the model is significantly better at predicting Cholesterol level than the model without BP Status also know as the model with just the mean of the entire data. However, the R-Square is very small so not much of the variation in the data is explained by our model. We also see that from the test for homogeneity of Variance, we see that there is no statistically significant evidence for the variance in Cholesterol being different amongst the groups of BP status. So the variance assumption for our ANOVA model is appropriate.

### 3c.)

  
We see that the difference between High & Normal and High & Optimal are significantly different. Meaning that for those with High BP Status, the mean of the cholesterol level is significantly higher than mean of those with Optimal or Normal Blood Pressure. However there isn't a significant difference between Normal & Optimal BP Status since the confidence interval for difference in the means contains zero.