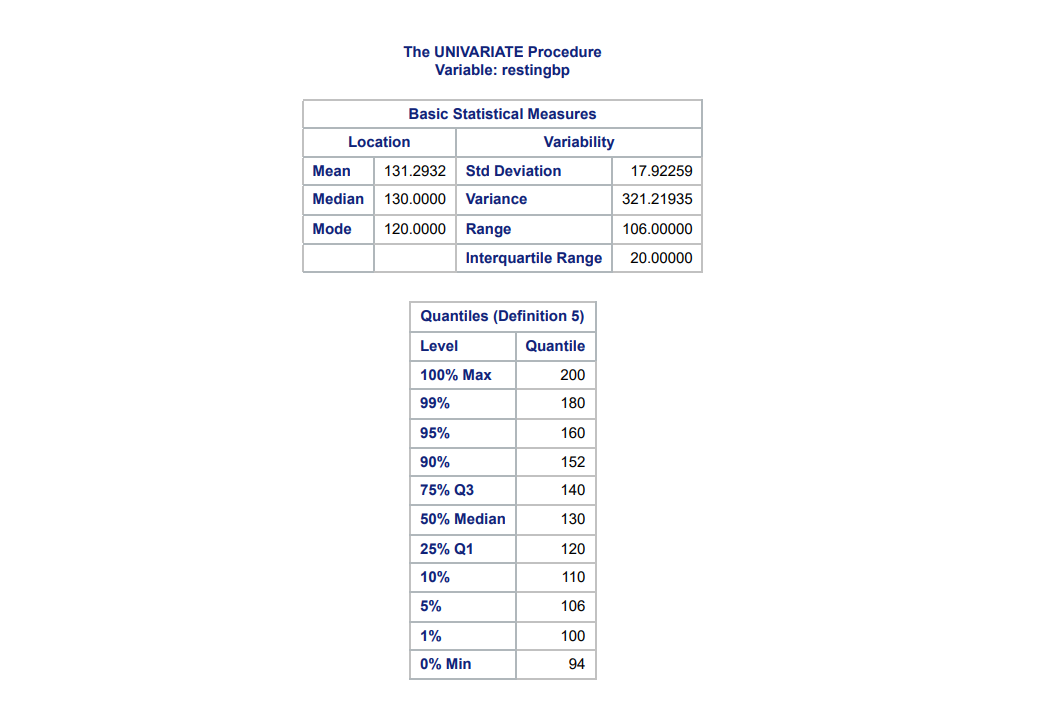
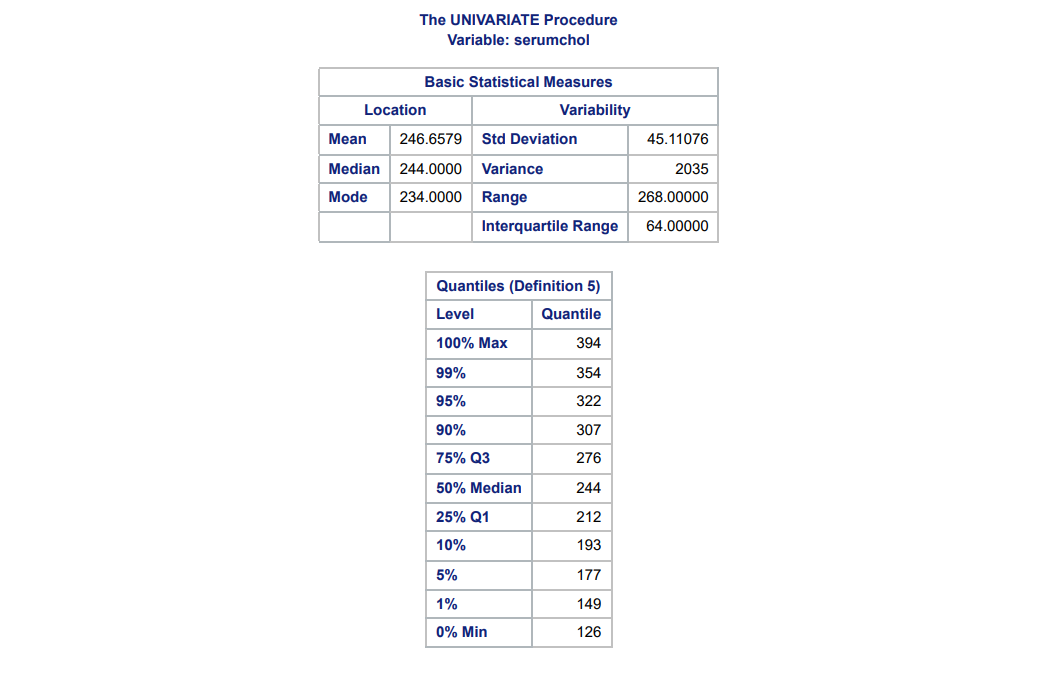
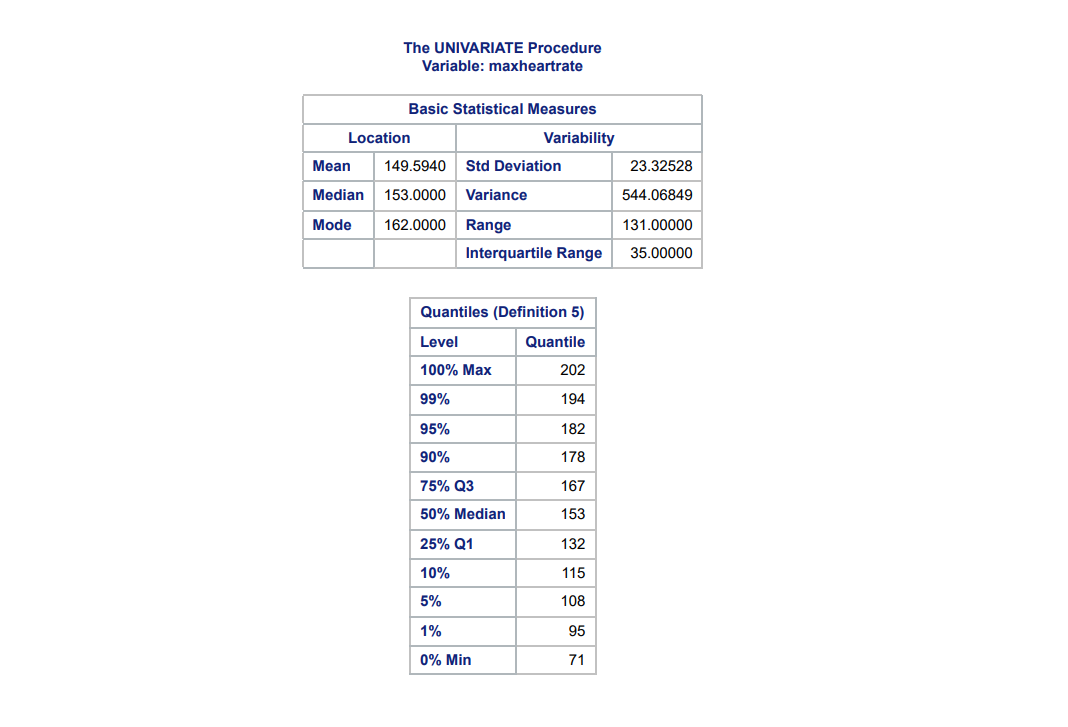
HW1

## Problem 1

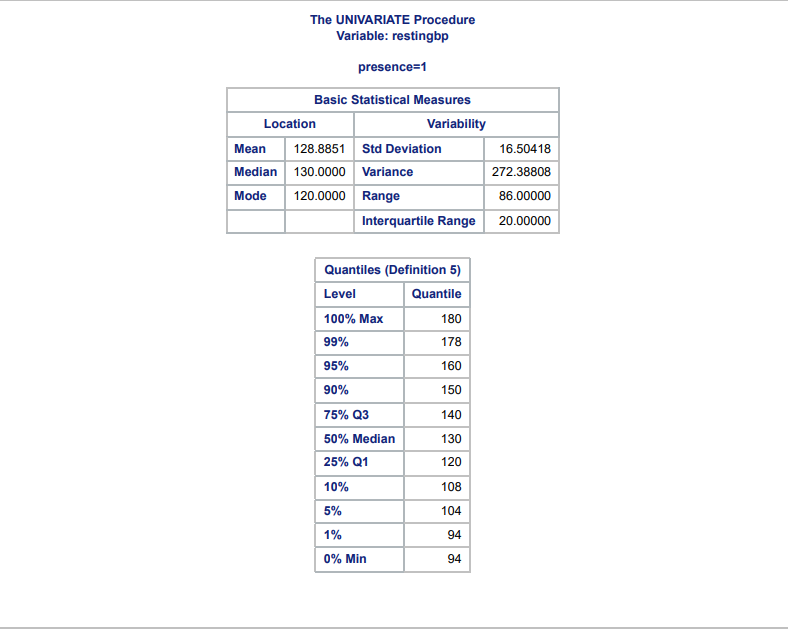
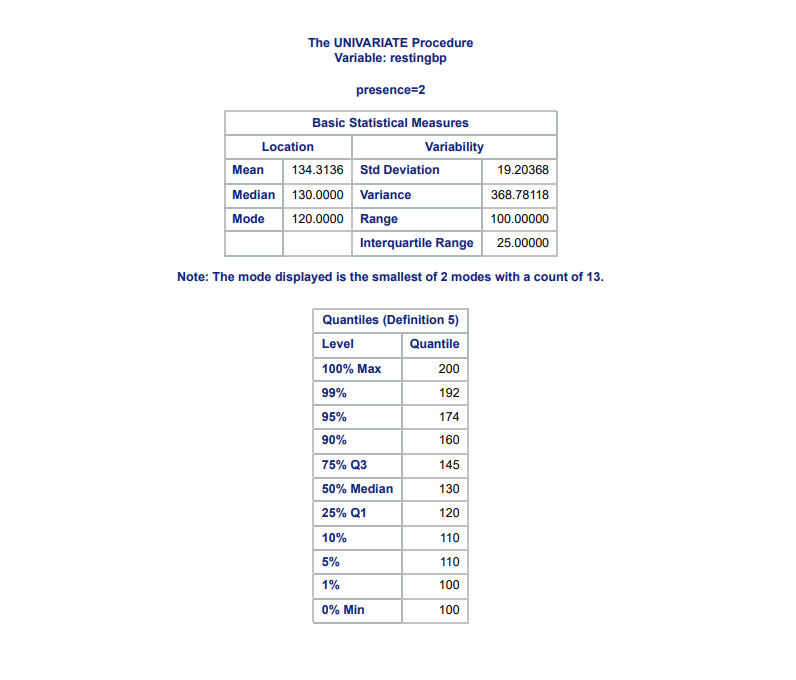
### 1a.)

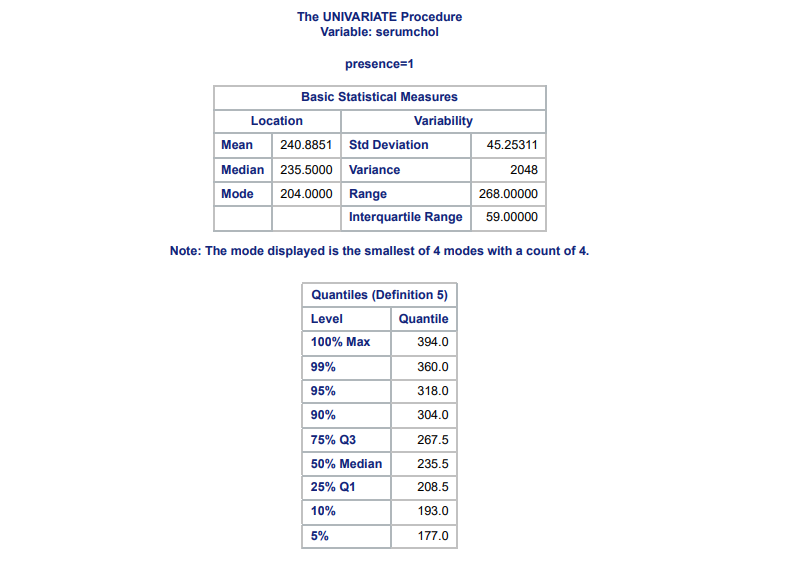
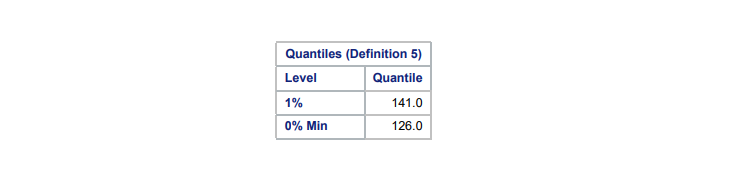
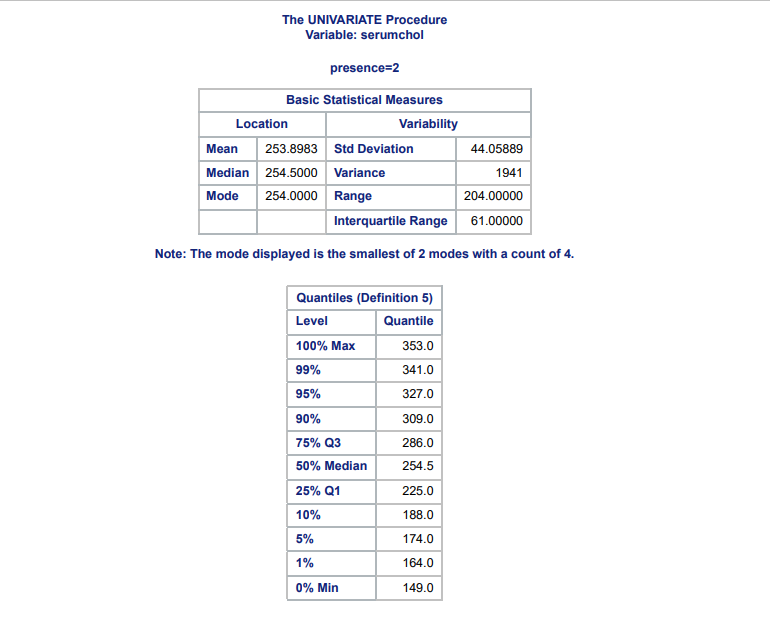
**Resting Blood Pressure**  
  
Resting Blood Pressure ranges from 94 to 200, with a Median Blood Pressure of 130. The most common blood pressure level was 120, and the mean was about 131.

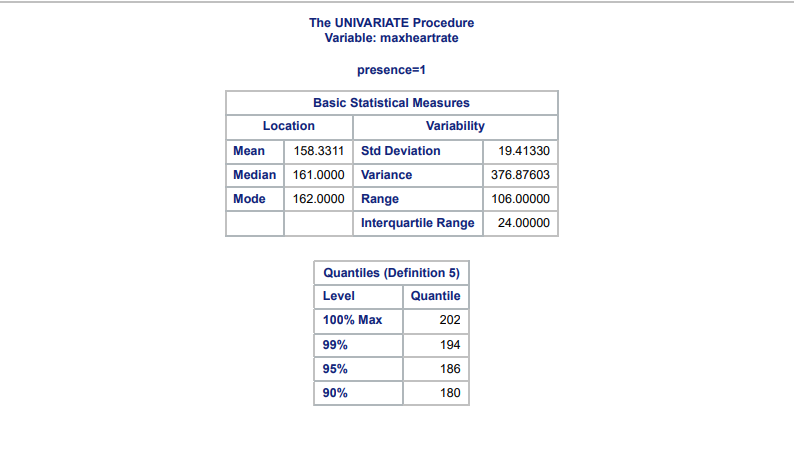
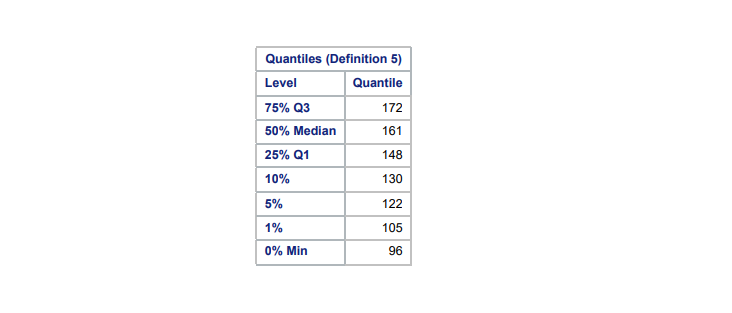
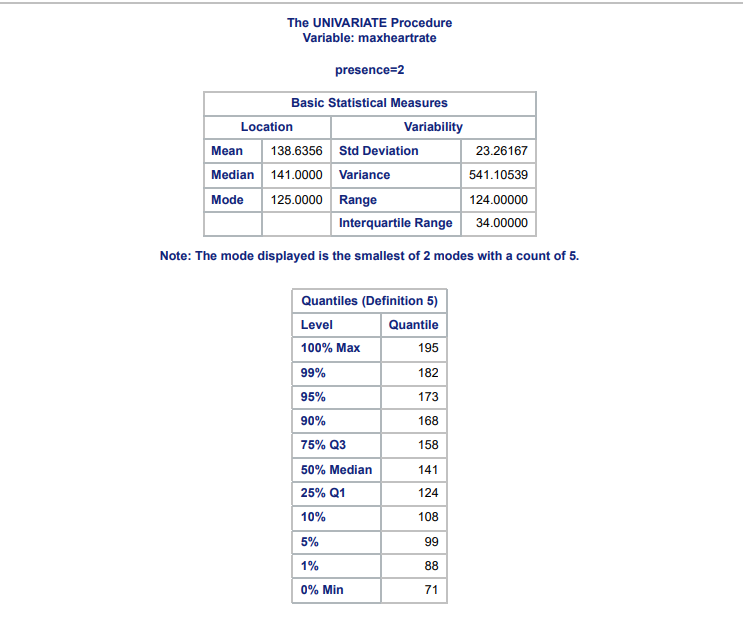
**Serum Cholesterols**  
  
Serum Cholesterol ranged from 126 to 394, with a Median of 244. The most common Serum Cholesterol value was 234, and the mean was about 246.

**Max Heart Rate**  
  
Max Heart Rate ranged from 71 to 202, with a Median of 153. The most common Max Heart Rate value was 162. and the mean was about 150.

### 1b.)

**Resting Blood Pressure**  
  
  
We see that patients with heart disease had a mean resting heart rate that was slightly higher than patients who did not have heart disease. The difference is about 5.5 mm HG (Or whichever blood pressure unit this dataset is using as it is not specified)

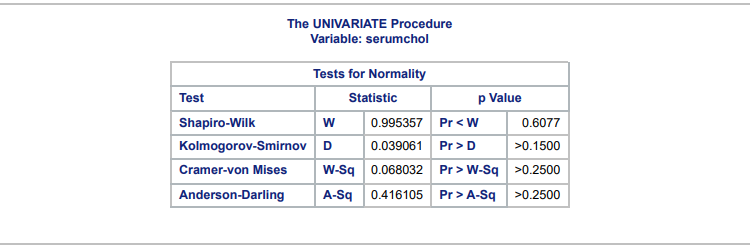
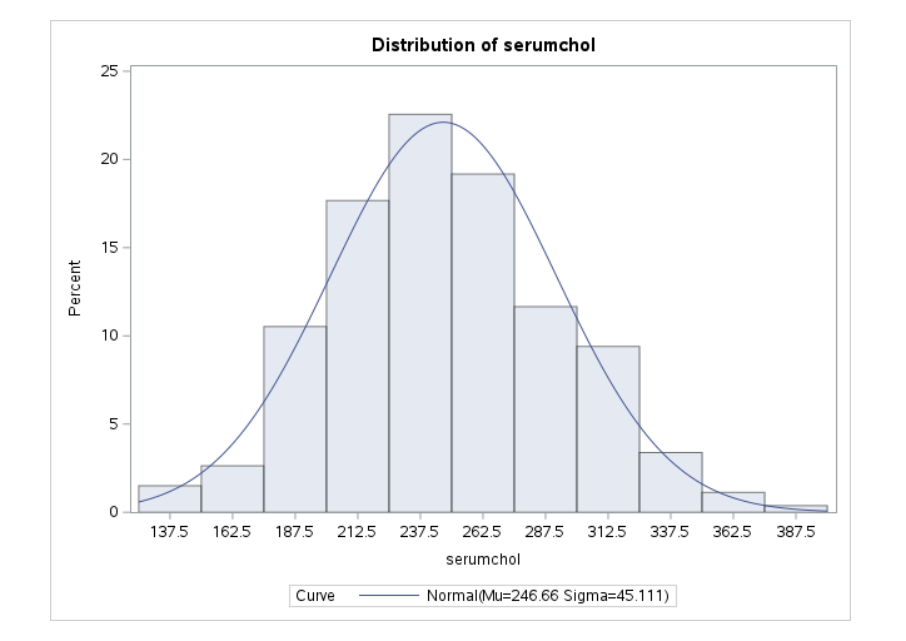
**Serum Cholesterol**  
  
  
  
We see that the mean serum cholesterol level for patient with heart disease is higher than the mean for patients without heart disease. The difference is about 13 mg/dl.

**Max Heart Rate**  
  
  
  
We see that patients with heart disease appear to have a mean max heart rate that is lower than patients without heart disease. The difference is about 19.7 bpm.

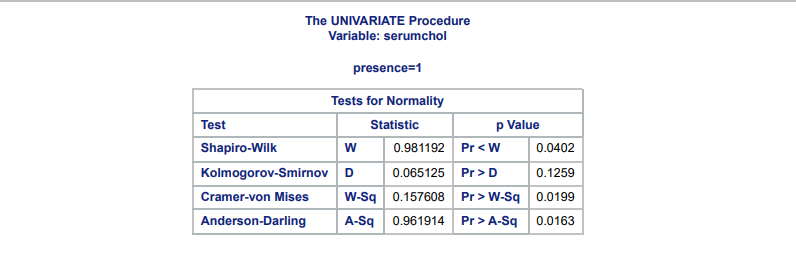
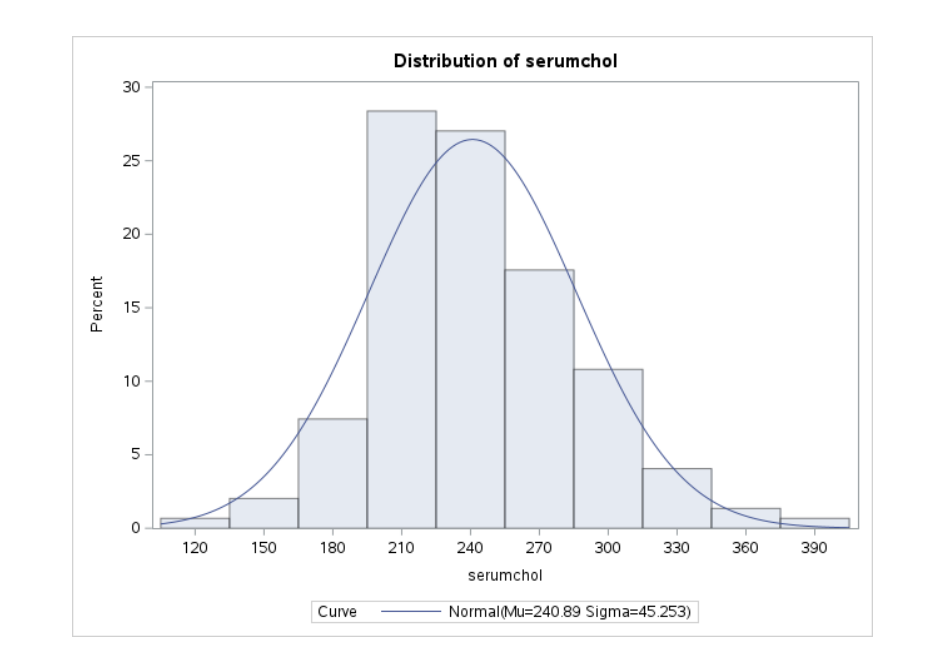
We see that for both Blood Pressure & Serum Cholesterol, that patients with heart disease have a high mean for both these measures. But for Max Heart Rate, it is the opposite and patients with heart disease have a lower mean Max Heart Rate compared to patients without heart disease.

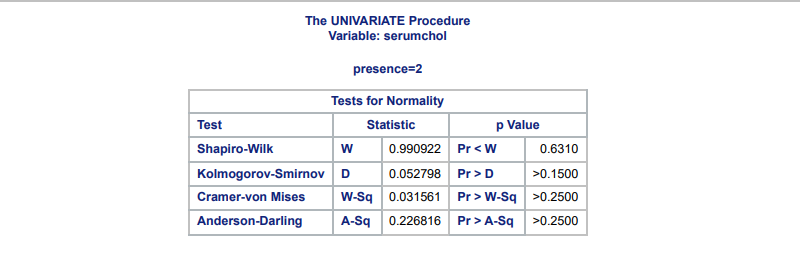
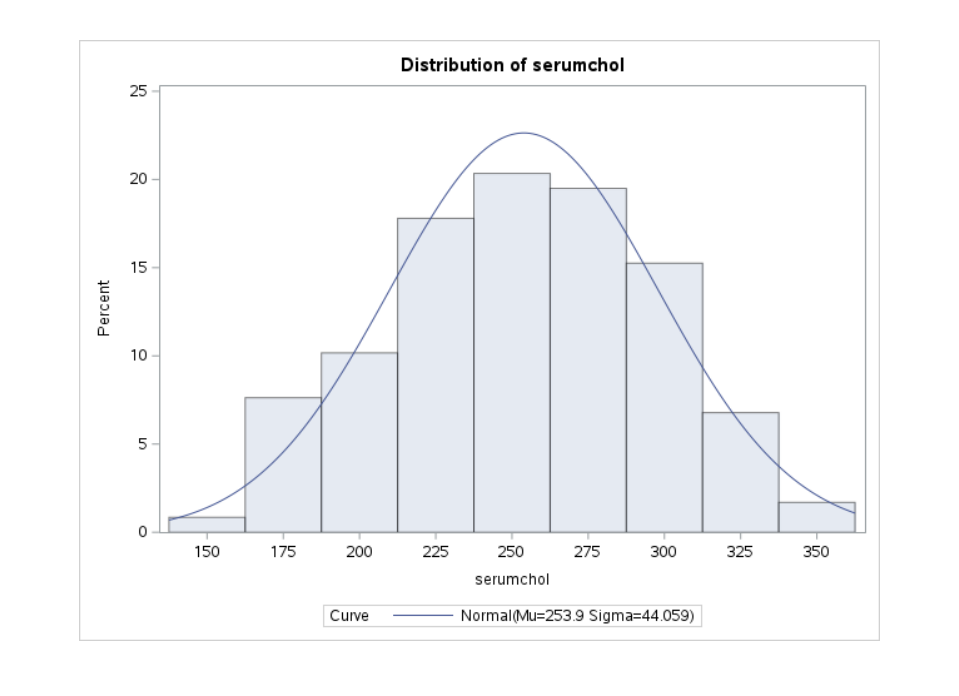
## Problem 2

### 2a.)

We will be testing the hypothesis:  
 : Distribution of Serum Cholesterol is Normal  
 : Distribution of Serum Cholesterol is not Normal  
  
We see by the p-value of the above tests that we fail to reject the null hypothesis, we don't have evidence to suggest that the distribution of Serum Cholesterol differs from the Normal Distribution. So an assumption of Normality is reasonable in this scenario for this variable. Visually we can also see the distribution of the variable is close to the Normal Distribution:  


### 2b.)

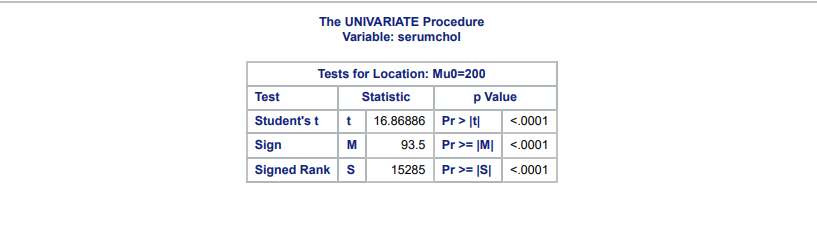
**Serum Cholesterol Normality Test for Patients without Heart Disease**  
  
  
If we use an = 0.05, We would reject the Null Hypothesis for three out of the four tests. When Patients don't have heart disease, the Normality assumption is not reasonable and we confirm this visually as we see that the peak of the data does not match the peak of the normal distribution curve.

**Serum Cholesterol Normality Test for Patients with Heart Disease**  
  
  
For patients with heart disease, we see that the Normality assumption is reasonable as based on the tests we don't have any significant p-values. Additionally we see that visually the Distribution of data matches pretty closely to a Normal Distribution.

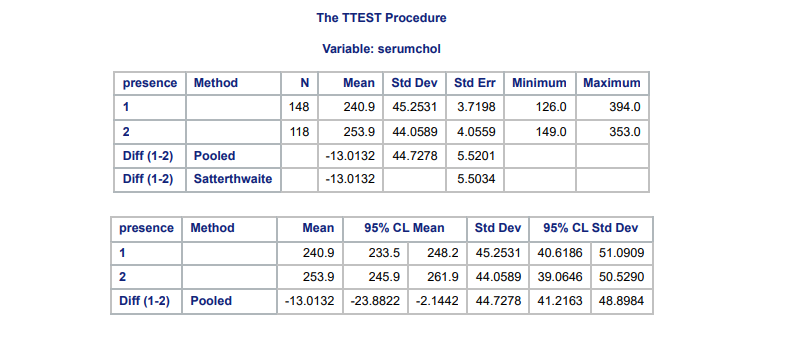
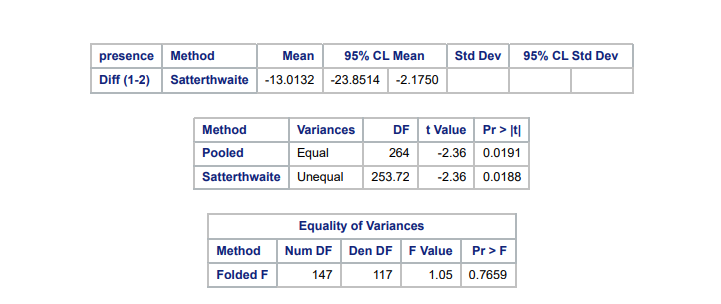
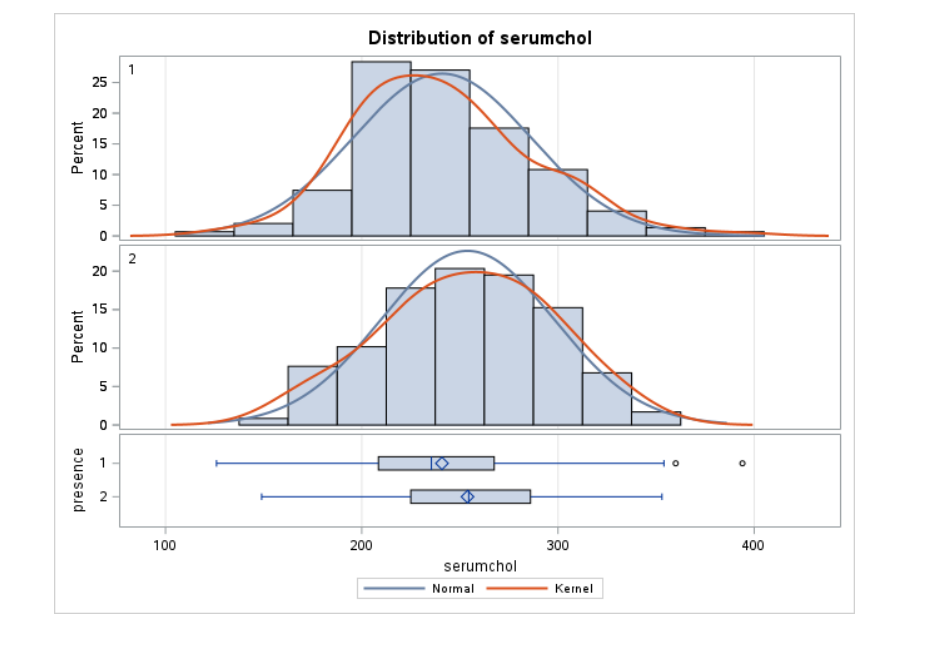
So patients without heart disease have serum cholesterol levels that differ significantly from Normality.

## Problem 3

### 3a.)

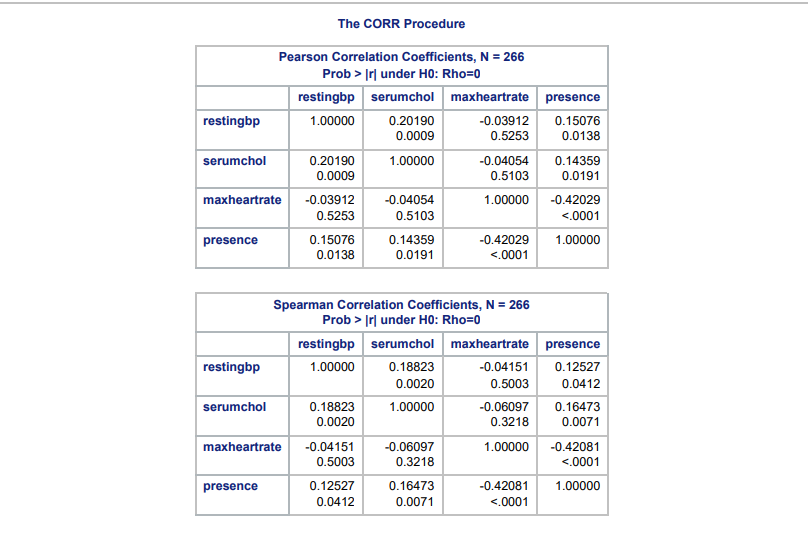
We will test for whether the mean of the Serum Cholesterol levels for our data set is greater than 200.  
 :   
 :   
  
We see that we have a really small p-value for every test, so we reject the Null Hypothesis as we have evidence to suggest that the mean Serum Cholesterol levels for patients in our dataset is significantly greater than 200.

### 3b.)

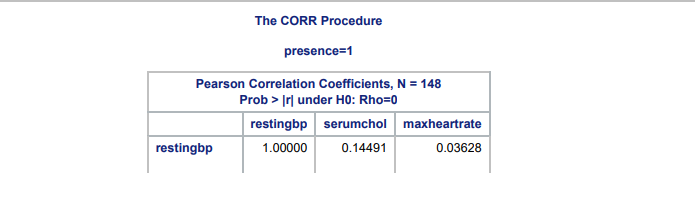
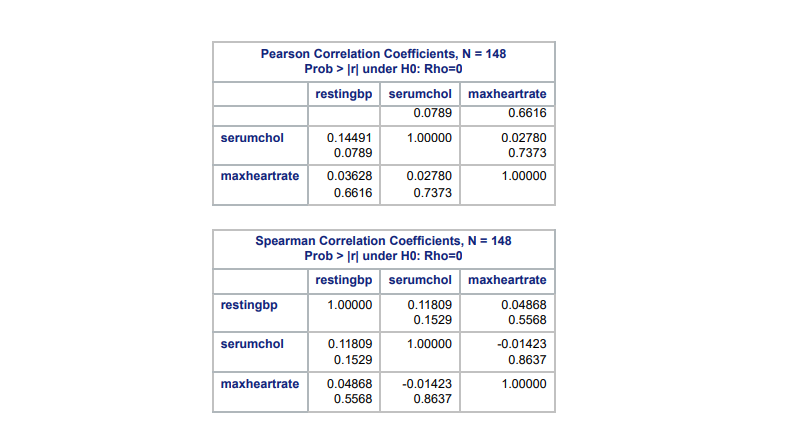
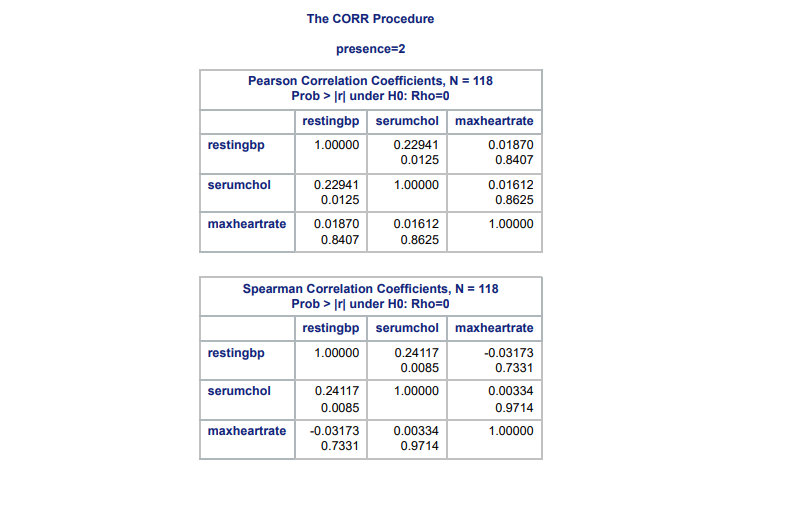
We will test whether the mean Serum Cholesterol level between patients with and without heart disease is the same.  
:   
 :   
  
  
  
Using an = 0.05 we reject the null hypothesis, regardless of whether we assume equal or unequal variances, we have evidence to suggest that the mean cholesterol level of patients with heart disease is significantly different from patients without heart disease in this dataset. In this case we see that the patients with heart disease have a higher mean serum cholesterol level. The visual plots also help us see how the distributions differ between the two patient groups.

## Problem 4

### 4a.)

  
The correlations between the variables are all either weak or likely non-existent. All the correlation values are closer to 0 than they are to one, so despite there being some correlations between the variables, none of them are very strong. The most significant correlation out of all the variables is the one between max heart rate & presence, which has a values of -0.42, which is still a weak correlation. The Spearman correlations give us similar conclusions.

### 4b.)

**Patients without Heart Disease**  
  
  
For Patients without heart disease, the correlations all still are very weak with most being not significantly different from zero. The only one that is significantly different from zero is Serum Cholesterol and Resting Blood Pressure, but the correlation values is still very small at about 0.14 which indicates the linear relationship is very weak based on Pearson correlation values, this is smaller than the value from the combined data. For Spearman the conclusions are the same expect all the correlations are not significantly different from zero based on the p-values, this includes Serum Cholesterol and Resting Blood Pressure.  
**Patients with Heart Disease**  
  
For Patients with heart disease we see that the only pair of variables with a correlation significantly different from zero is Serum Cholesterol and Resting Blood Pressure again. The relationship is still weak because the correlation value is still about 0.23 but this is higher than the combined data. This conclusion is the same with the Spearman correlation values as well.