

# EXPERIMENT 3

## Anova Analysis

**Name:** Aryan Dali    **UID:** 2019120016

**Class:** TE EXTC    **Subject:** Data Analytics

The dataset used is the Heart dataset which is inbuilt in the SAS software. It has the patient's living status, heart rate, cholesterol level, age, sex and smoking levels.

After seeing the data set three Hypothesis were made:

### **Hypothesis 1**

Ho: If the cholesterol levels are high then the Weight of the person will be high.

Ha: If the cholesterol levels are high then the Weight of the person will not necessarily be high.

### **Hypothesis 2**

Ho: If the person smokes, he will die early.

Ha: If the person smokes the person will not necessarily die early.

### **Hypothesis 3**

Ho: Cholesterol levels depends on the sex of the person

Ha: Cholesterol levels does not depends on the sex of the person

## Code

```
1 data work.heart;
2 set sashelp.heart;
3 run;
4 proc anova data=work.heart;
5 class weight_status;
6 model cholesterol=weight_status;
7 run;
8
9 proc anova data=work.heart;
10 class status;
11 model smoking=status;
12 run;
13
14 proc anova data=work.heart;
15 class sex;
16 model cholesterol=sex;
17 run;
```

### The ANOVA Procedure

Class Level Information		
Class	Levels	Values
Weight_Status	3	Normal Overweight Underweight

Number of Observations Read	5209
Number of Observations Used	5051

### The ANOVA Procedure

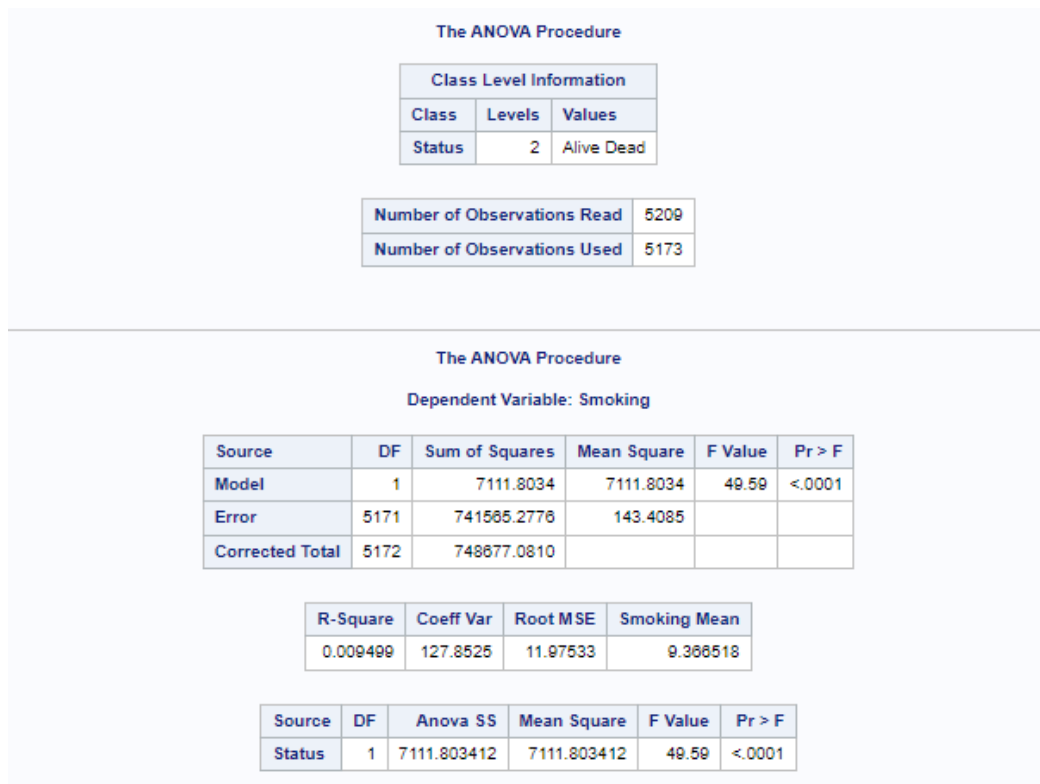
Dependent Variable: Cholesterol

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	259397.60	129698.80	65.92	<.0001
Error	5048	9932551.25	1967.62		
Corrected Total	5050	10191948.85			

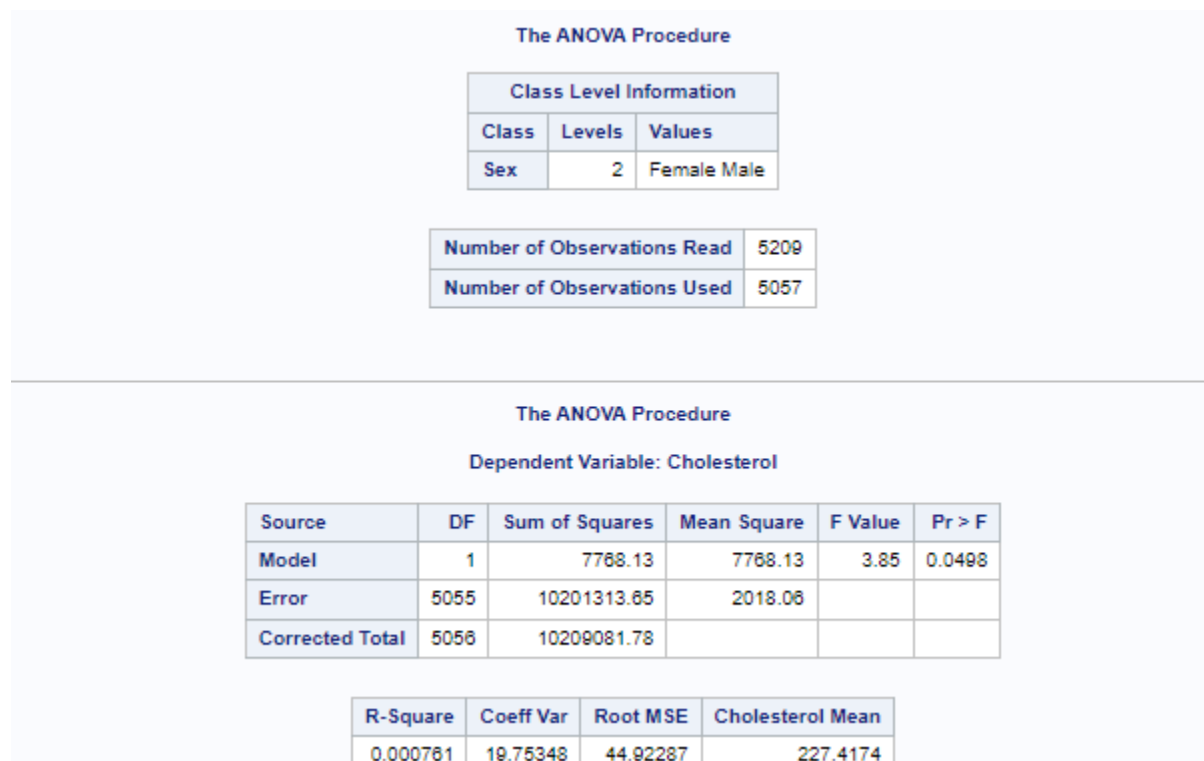
R-Square	Coeff Var	Root MSE	Cholesterol Mean
0.025451	19.50614	44.35788	227.4047

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Weight_Status	2	259397.6031	129698.8016	65.92	<.0001

Fig 1



*Fig 2*



*Fig 3*

## **Conclusion**

1. In Fig 1 for case 1 we can see that the p-value obtained is less than 0.001 therefore we accept the alternate hypothesis which is that, If the cholesterol levels are high then the Weight of the person will not necessarily be high.
2. In fig 2 for case 2 we can see that the p-value obtained is less than 0.001 therefore we accept the alternate hypothesis which is that, If the person smokes the person will not necessarily die early.
3. In fig 3 for case 3 we can see that the p-value obtained is less than 0.05 therefore we accept the null hypothesis which is that, Cholesterol levels depends on the sex of the person