Assignment 4
Part 1

1a)
dir = "..."
file1 = "rat.csv"
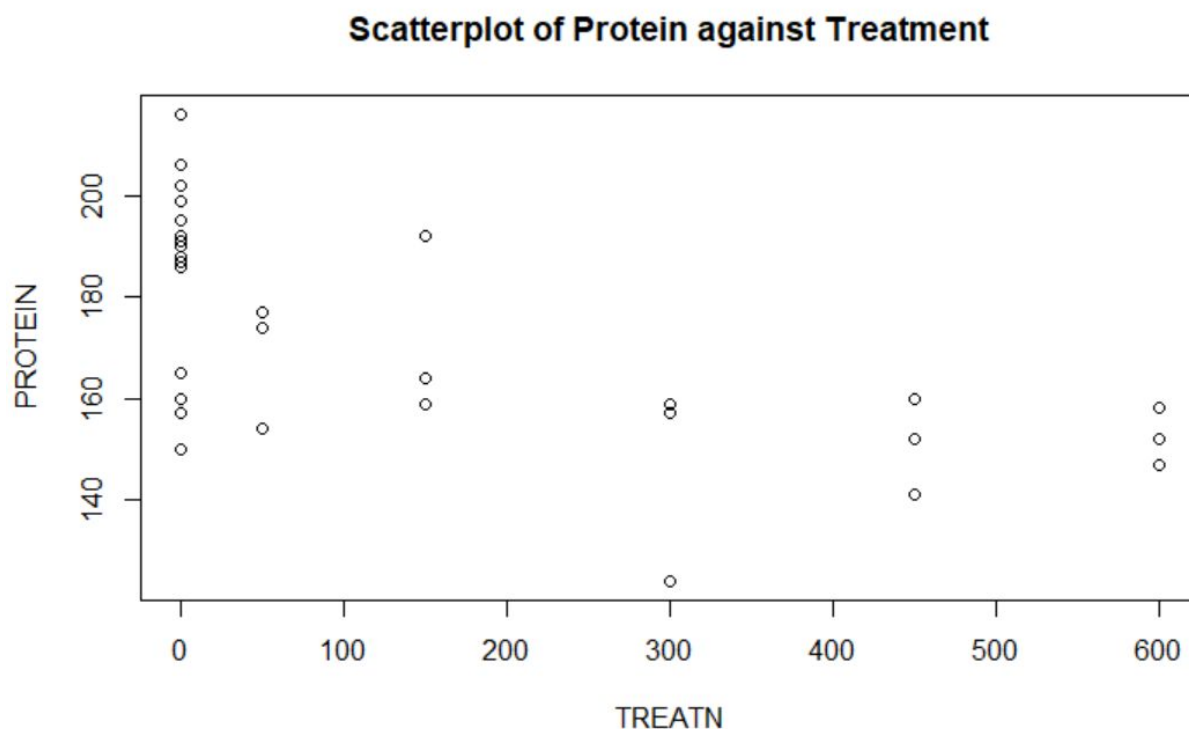dfRat = read.table(file=paste(dir,file1, sep=""), header=TRUE, sep=',')

PROTEIN = dfRat[,1]
TrtGroup = dfRat[,2]
TREATN = dfRat[,3]
plot(TREATN, PROTEIN)

Scatterplot of Protein against Treatment

1b)
Simple Linear Regression Model = Yhat = B0 + B1X + E

SLR Assumptions:
1. Normality: For each value of the explanatory variable (X) there is a subpopulation of response variables (Y) that is normally distributed
2. Linearity: The means of the subpopulation for each value of X fall on the straight line B0 + B1X
3. Constant Variance: All subpopulations have the same standard deviation
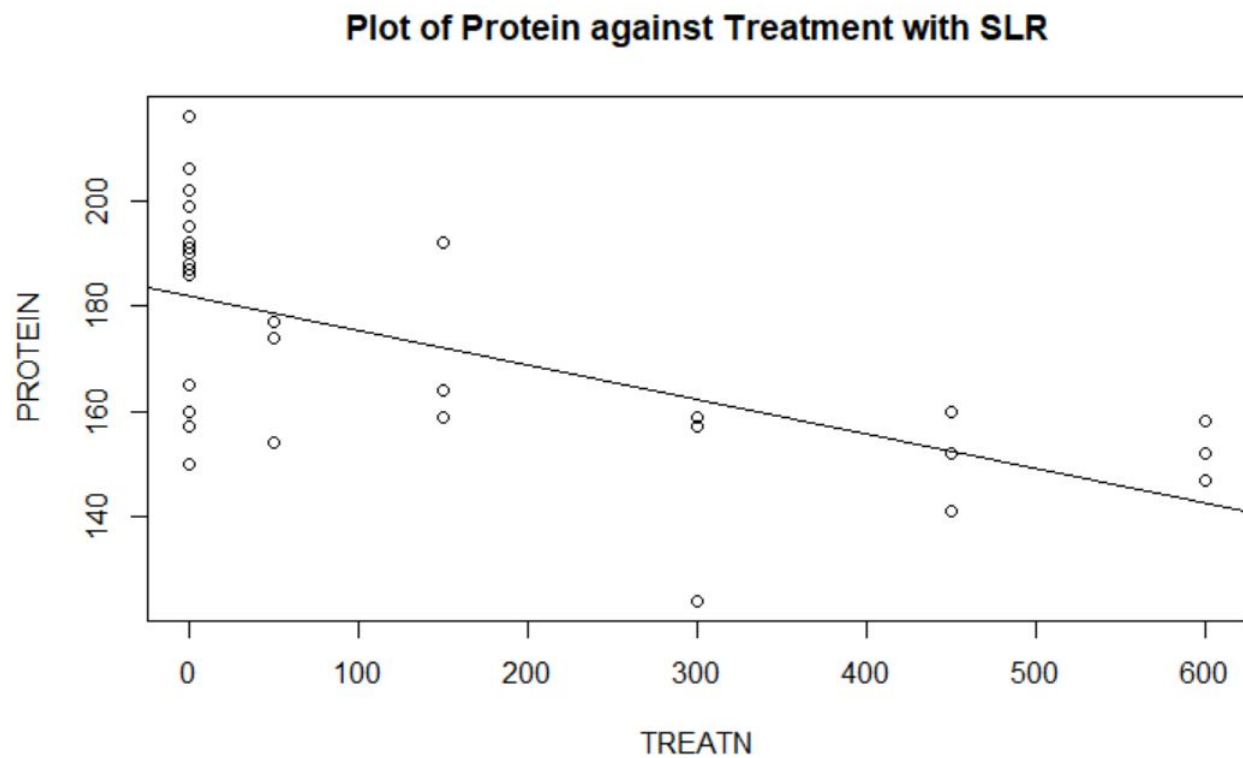4. Independence: All observations are independent

1c)

```
M1 = lm(PROTEIN~TREATN) #SLR
plot(TREATN, PROTEIN, main="Plot of Protein against Treatment with SLR")
abline(M1)
```

Intercept = 181.9998
Slope = -0.0658

1d)

```
M2 = lm(PROTEIN~TrtGroup) #ANOVA
AnovaM2 = anova(M2)
```

```
Analysis of Variance Table

Response: PROTEIN
          Df Sum Sq Mean Sq F value   Pr(>F)
TrtGroup   5 7222.5 1444.51  4.8827 0.003196 **
Residuals 24 7100.3  295.84
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since we now fit an ANOVA model, we use the ANOVA assumptions

1. Observations are independent
   - within groups
   - between groups
   - determined by the nature of the data
2. Observations are normally distributed
3. There is equal variance among the groups

1e)
We perform a Lack-of-Fit/Goodness-of-Fit/Extra Sum of Square F-test by using an ANOVA to test the fit of the reduced model (TREATN) against the data.

anova(M1, M2)
Fobs = 0.297
qf(0.95, df1=4, df2=24) = FCritical = 2.77

Table 1

```
> anova(M1, M2)
Analysis of Variance Table

Model 1: PROTEIN ~ TREATN
Model 2: PROTEIN ~ TrtGroup
  Res.Df    RSS Df Sum of Sq        F Pr(>F)
1     28 8642.5
2     24 7100.3  4    1542.2 1.3032  0.297
```

5 Step Hypothesis test:

1. H0: The reduced model adequately explains the data
   H1: The full model is necessary to explain the data

2. Significance level, a = 0.05

3. From Table 1 we can see that Fobs = 0.297

4. We will reject the null hypothesis is Fobs > Fcritical. We can see that the Fobs = 0.297 is < Fcritical = 2.77, so we do not have evidence to reject the null.

5. There is a lack of evidence to suggest the multiple treatment group ANOVA model fits the data better than the SLR model.