Homework 1

Orly Olbum

Problem 1

The National Bureau of Standards performed eight series of experiments between 1924 and 1935 to determine g, the acceleration due to gravity. The National Bureau of Standards (now called the National Institute of Standards and Technology) is the government agency that measures things. The following statement is taken from the NIST website: "Founded in 1901, NIST is a non-regulatory federal agency within the U.S. Department of Commerce. NIST's mission is to promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life." Thus, it is safe to assume that the NBS scientists were trying hard to measure the same quantity g (e.g., all experiments were done in the same location) throughout all eight series of experiments. The data are given in "Gravity.txt". The column "Series" represents the experiment group number (ordered by time), where series 1 represents the earliest set of experiments and series 8 the last. The column "Value" is the measurement taken in each experiment in deviations from 9.8 m/s2 times 105. For example, the value for the first measurement is 9.80076 m/s2.

(a) Consider testing the null hypothesis that the means for all 8 series are equal versus the alternative that they are not. What scientific conclusion would you draw if the data did not support the null hypothesis?

If we conducted this test for equal means accross all 8 series and concluded that we had evidence to suggest they were in fact different, we might conclude that one or more of the experiments may have been done with different quantities or under different circumstances than the others. Since we are testing multiple groups, we would then proceed to a pairwise test to determine specifically which group or groups were providing the significant result to reject the null in the initial test.

(b) Calculate the standard F statistic for this null hypothesis and obtain the p-value for this statistic. State explicitly the assumptions underlying this test procedure.

Since we are testing equality of means of multiple groups, we will use an ANOVA test, which generalizes a t-test to more than just 2 groups and uses the F-statistic. ANOVA assumes approximately normal data, homogeneity of variance, and independent observations.

For this test we have N = 81 and k = 8, making degrees of freedom for the F-test (73, 7) and a critical F value of 3.29.

Ho: means for all 8 series are equal
Ha: means for all 8 series are not equal
alpha = 0.05

```r
model = lm(Value ~ Series, data = gravity)
summary(aov(model))
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## Series       7   2798   399.7    3.62 0.00211 **
## Residuals   73   8061   110.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The model shows an F-stat of 3.62 and p-value of 0.002, leading us to reject the null hypothesis and conclude that not all means of the 8 different series are equal.

(c) Compute the likelihood ratio statistic for the null hypothesis that the means are all equal allowing for a different variance for each series (but still assuming the observations are independent and Gaussian) and assess its statistical significance. (You may be able to find a routine in R or SAS that will do this for you, or you may try to do this from first principles.)

Since we can hold the assumption of independent and normal observations but have to take into account unequal variances, we can adjust our model to take account of this, and then compare it to the model above that assumes equal variance.

```
model1 = gls(Value ~ Series, data = gravity, method = "ML")
model2 = gls(Value ~ Series, data = gravity, method = "ML",
             weights = varIdent(form = ~ 1|Series))
anova(model1, model2)
```

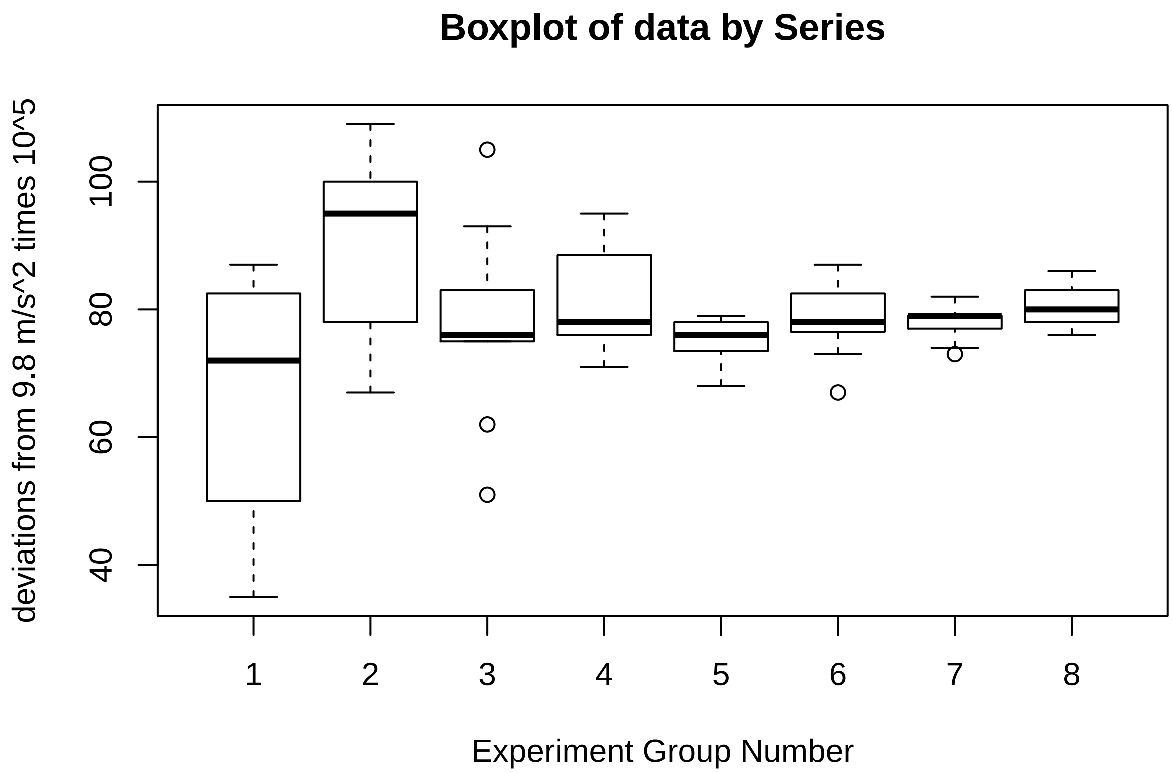
```
##        Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## model1     1  9 620.4950 642.0451 -301.2475
## model2     2 16 561.5651 599.8763 -264.7826 1 vs 2 72.92989  <.0001
```

Comparing the equal variance model with the unequal variance model, we get a Likelihood Ratio statistic of 72.93 and a p-value less than 0.0001, leading us to reject the null that means are equal.

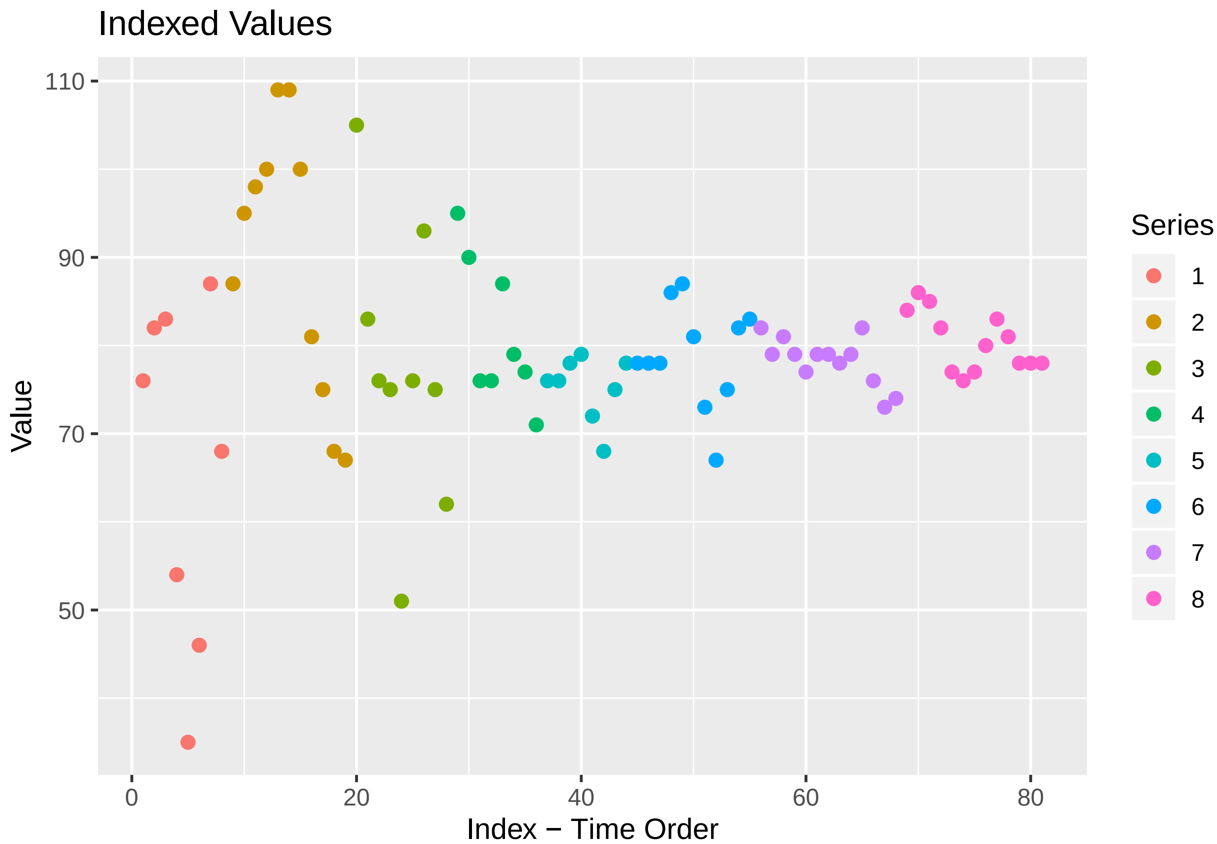
(d) Assuming the observations in each series are given in time order taken, plot the observations in a way that preserves both the time order of the series and the time order of the observations within each series. What does your plot suggest about the validity of the various assumptions you made in parts (b) and (c)?

Since we know that Series 1 is the earliest and Series 8 is the latest, we can plot the Values and the time order will hold.

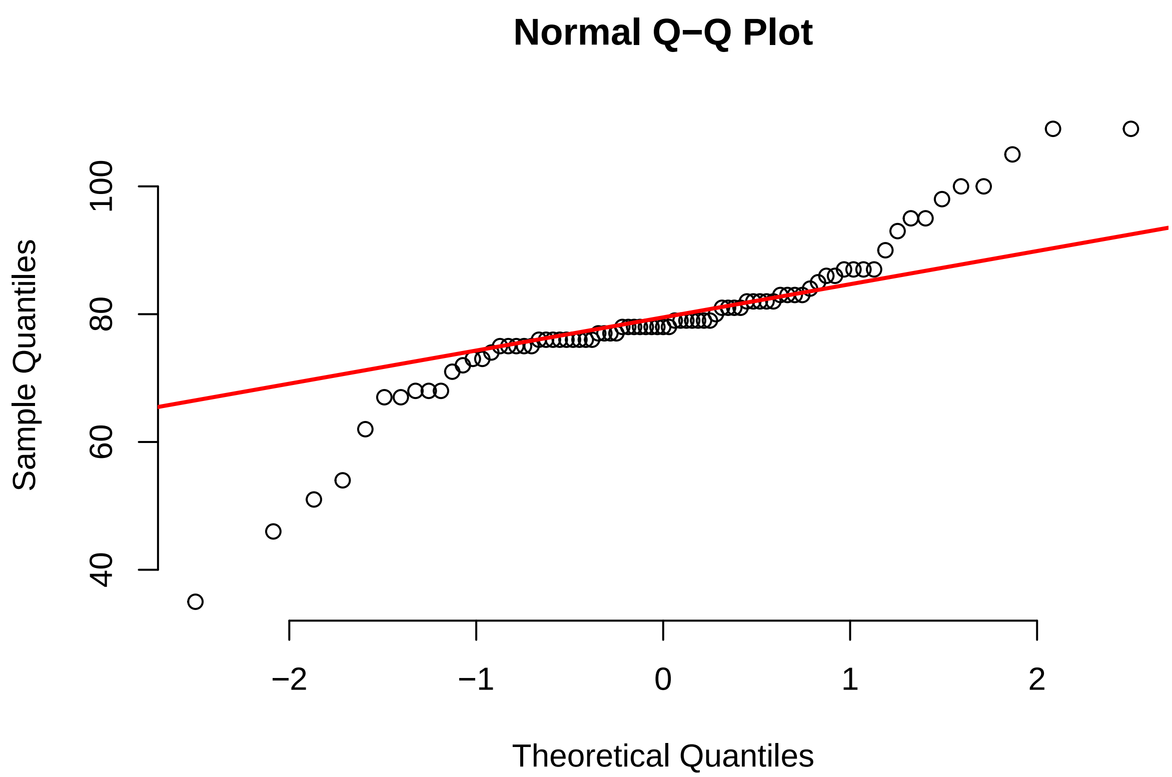
```
gravity$index = index = 1:nrow(gravity)
plot(gravity$Series, gravity$Value, main = "Boxplot of data by Series",
     xlab = "Experiment Group Number", ylab = "deviations from 9.8 m/s^2 times 10^5")
```




```
ggplot(gravity, aes(x = index, y = Value, color = Series)) +
  geom_point(size = 2) +
  ggtitle("Indexed Values") +
  xlab("Index - Time Order") +
  ylab("Value")
```




```r
qqnorm(gravity$Value, pch = 1, frame = FALSE)
qqline(gravity$Value, col = "red", lwd = 2)
```



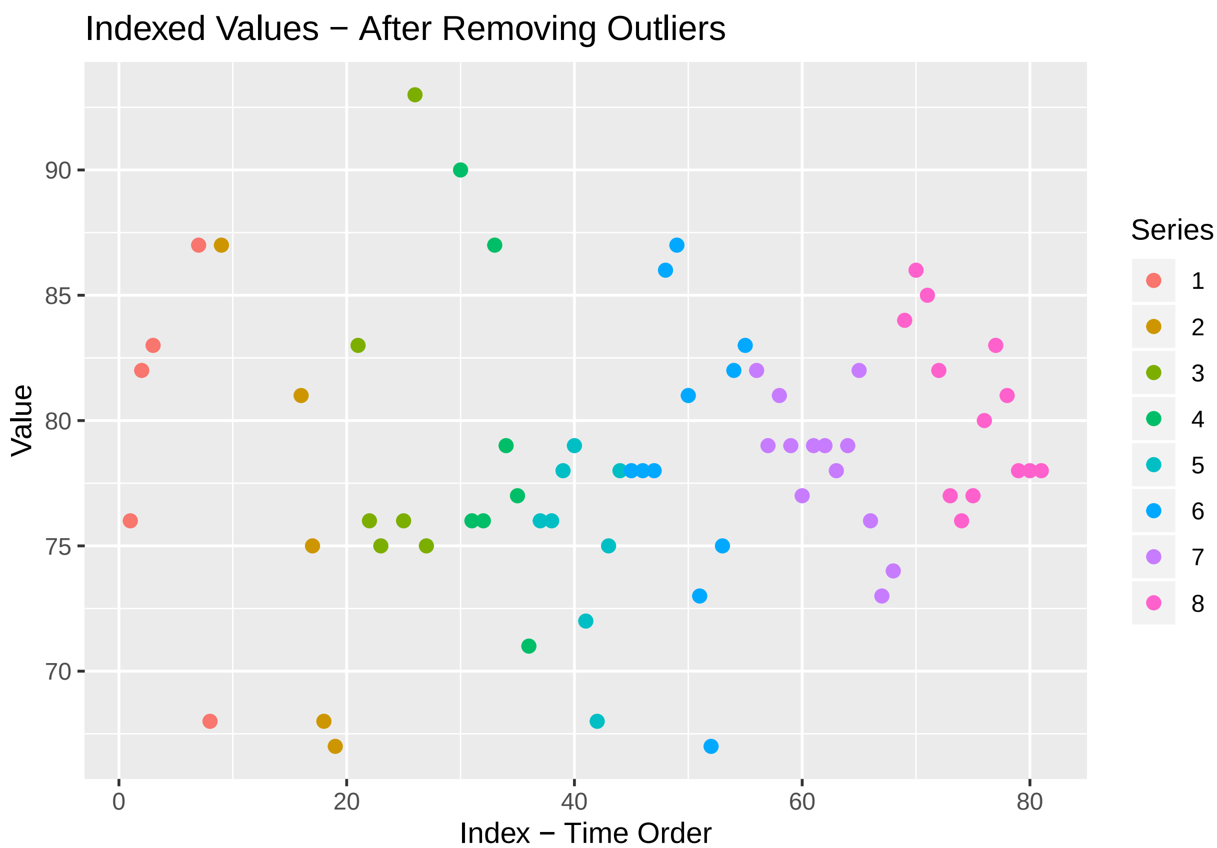
The boxplots suggests we may not have homogeneous variances, specifically from Series 1 and 2, which takes away an important assumption of the ANOVA test from above. Although the qqplot shows not-so-great normality, ANOVA is robust to not-so-normal data and our (c) assumptionsh hold. We can also see from the scatterplot that the series of experiments done later in time are closer together, which leads me to believe that the more these experiments go on, the closer to the true value of gravity we get, which would indicate more time series-behavior type data. We may need to take into account that data points are autocorrelated or that they need to be smoothed in order to truly evaluate.

(e) Doing whatever analyses you deem appropriate, provide your best estimate of g at the location of these experiments and a standard error of your estimate using only the data in "Gravity.txt".

Remove outliers and re-assess plot, and re-run the model from (b) to get a better estimate of g.

```
iqr = IQR(gravity$Value)
lo = quantile(gravity$Value, .25)[[1]] - 1.5*iqr
hi = quantile(gravity$Value, .75)[[1]] + 1.5*iqr
grav2 = subset(gravity, gravity$Value >=lo & Value <=hi)

ggplot(grav2, aes(x = index, y = Value, color = Series)) +
  geom_point(size = 2) +
  ggtitle("Indexed Values - After Removing Outliers") +
  xlab("Index - Time Order") +
  ylab("Value")
```




```r
model.update = lm(Value ~ Series, data = grav2)
est = summary((model.update))$coef[,1] # estimates of g for each Series/location
se = summary((model.update))$coef[,2] # standard errors for each Series/location
cbind(est, se)
```

```
##                    est       se
## (Intercept) 79.2000000 2.379988
## Series2     -3.6000000 3.365811
## Series3      0.4666667 3.222519
## Series4      0.2285714 3.116136
## Series5     -3.9500000 3.033901
## Series6     -0.2909091 2.870374
## Series7     -0.8923077 2.800524
## Series8      1.1846154 2.800524
```

The last table shows the estimates of g for each Series (with the Intercept defaulting to Series 1), followed by the standard errors. Further analysis of this data may include a time series analysis since we are analyzing observations within each series on a time scale. We might want to difference the data or perform moving-average analysis.