STAT 3480 Lab5

Shaoran Sun

March 22, 2016

1

	SS	df	MS	F
Treatment	6702.05	3	2234.02	5.26
Error	55740.83	136	424.56	
Total	64442.88	139		

2

The p-value is

```
## [1] 0.00182604
```

Since p-value is less than 0.05, we can reject the null hypothesis, and conclude that there at least one population has a different mean.

3

```
runtime.anova = lm(runtime ~ rating)
anova(runtime.anova)
```

```
## Analysis of Variance Table
##
## Response: runtime
##             Df Sum Sq Mean Sq F value   Pr(>F)
## rating       3   6701  2233.8  5.2596 0.001831 **
## Residuals  136  57759   424.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value read from the ANOVA table is the same with the one we calculated by hand.
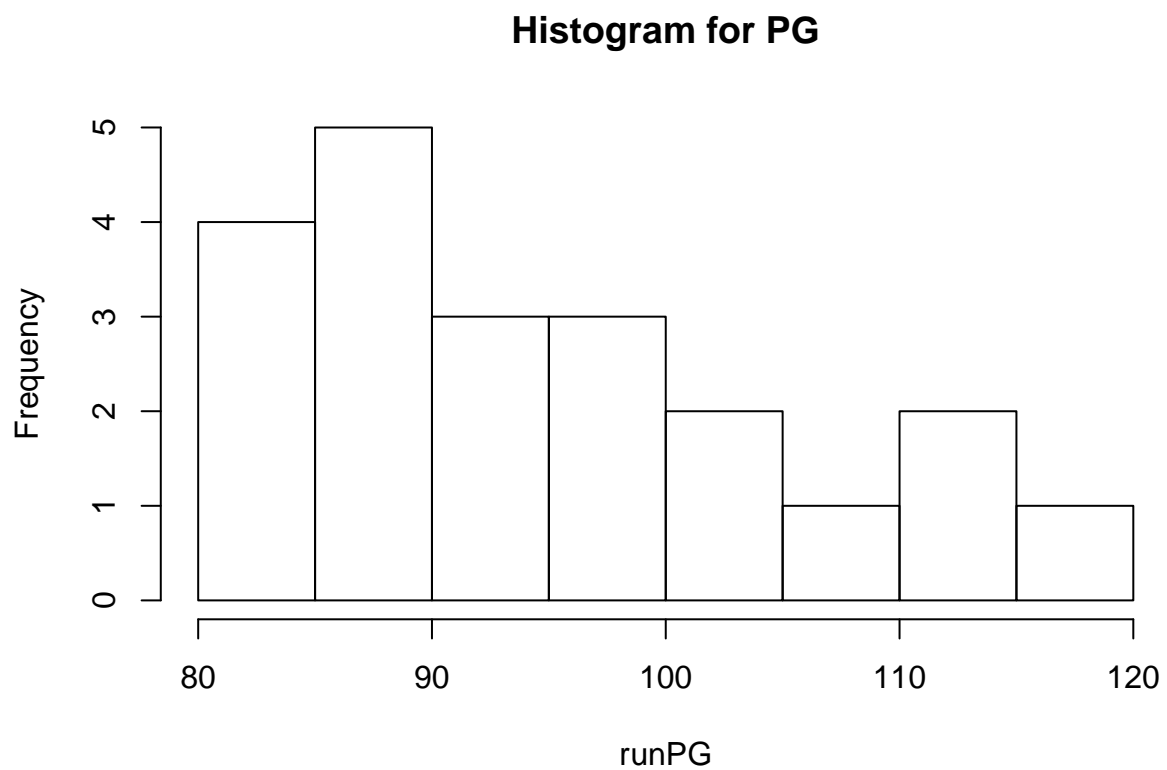
4

```
runPG13 <- runtime[which(rating == "PG-13")]
runPG <- runtime[which(rating == "PG")]
runG <- runtime[which(rating == "G")]
runR <- runtime[which(rating == "R")]
```
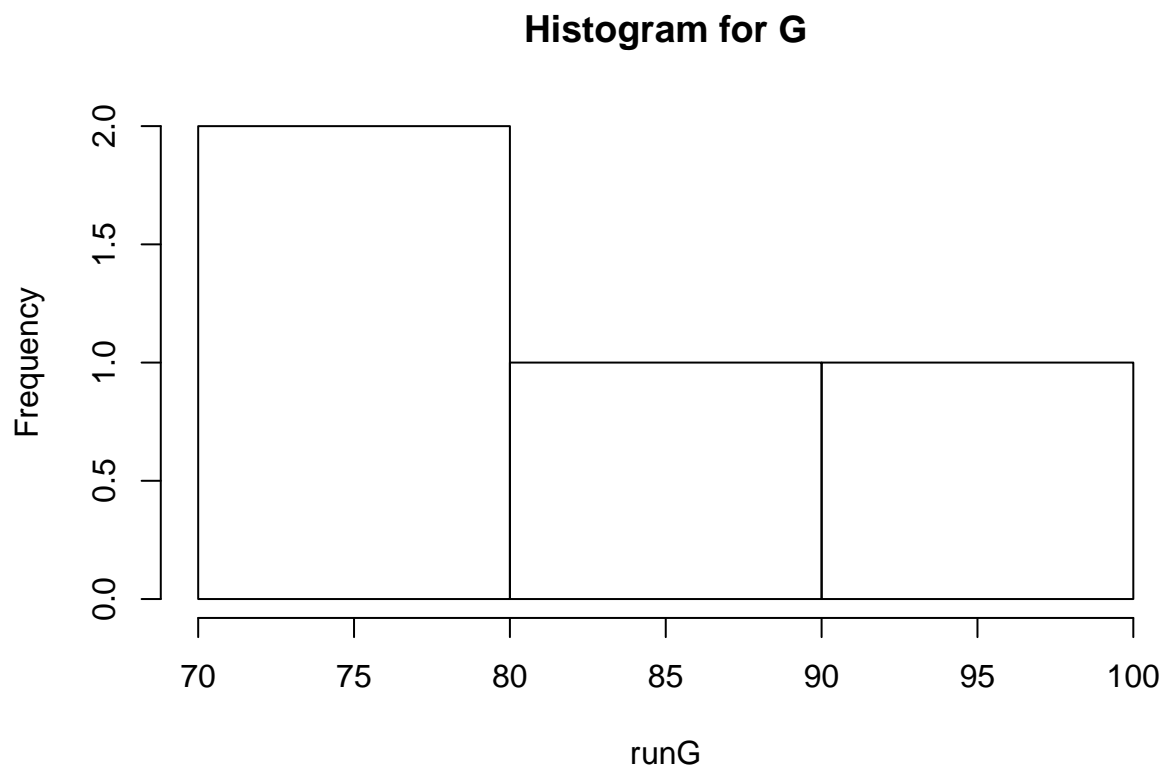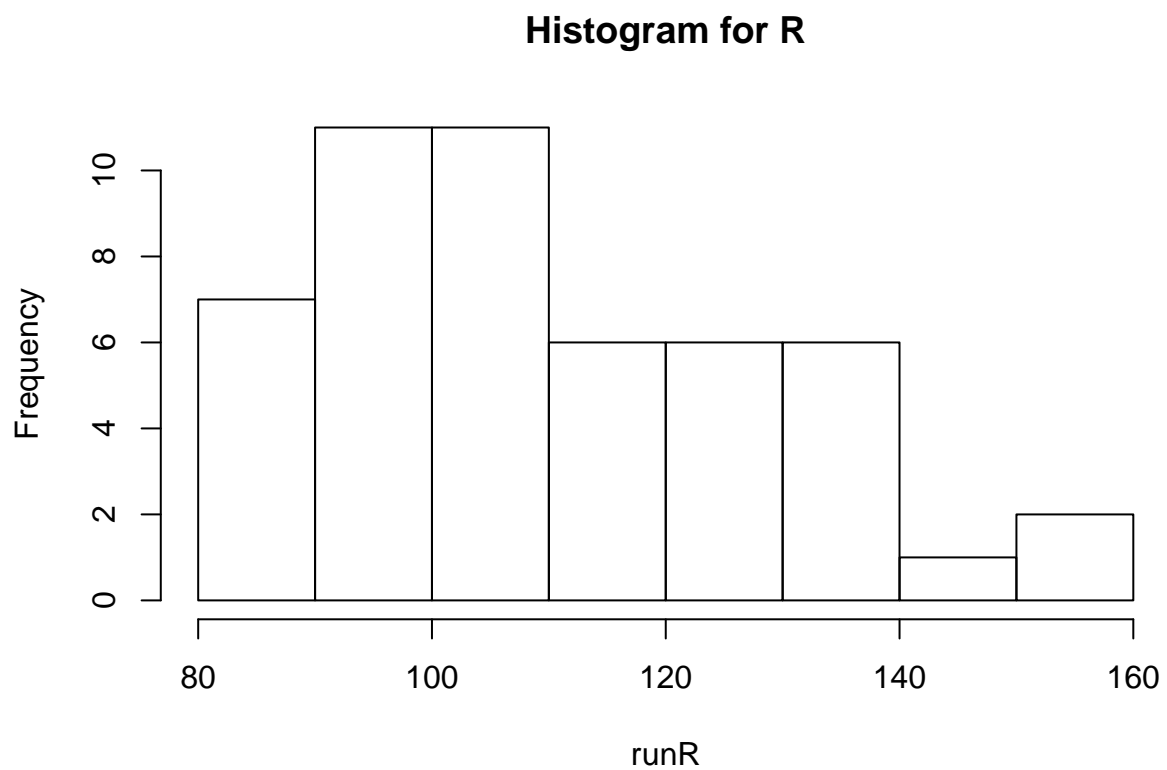
```
hist(runPG13, main="Histogram for PG13")
```

Histogram for PG13

runPG13
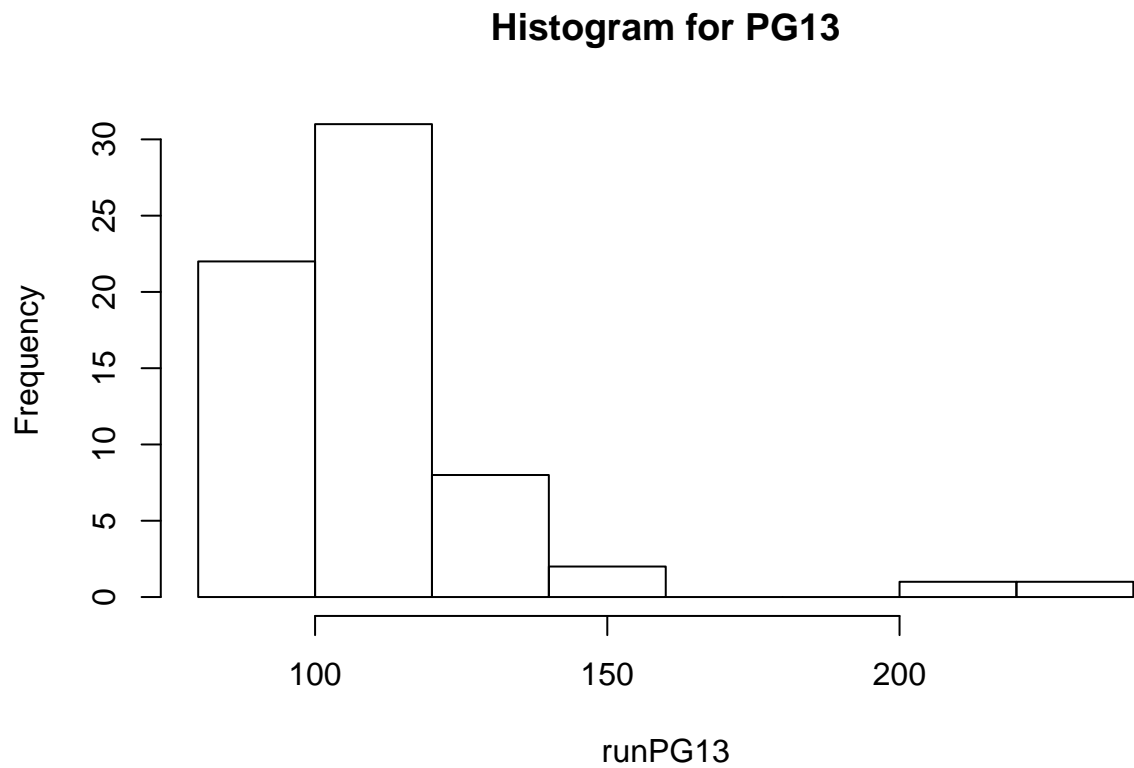
```
hist(runPG, main="Histogram for PG")
```

Histogram for PG

runPG

```r
hist(runG, main="Histogram for G")
```
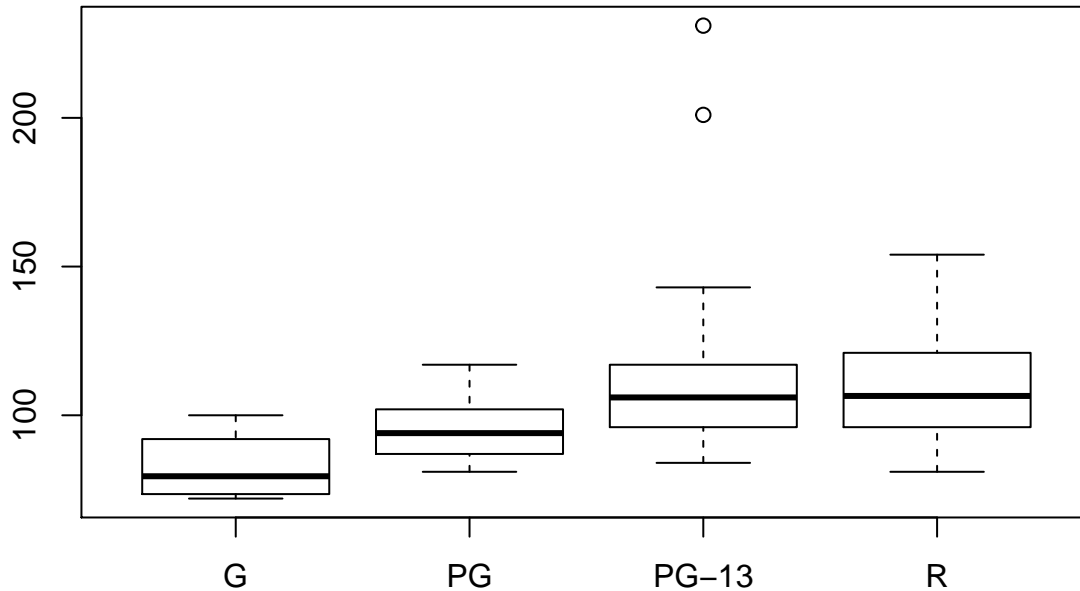
Histogram for G


```r
hist(runR, main="Histogram for R")
```

Histogram for R

Based on the 4 histograms, I feel this assumption of normality is violated. All graphs are skewed right, due to small sample size.

5

```r
boxplot(runtime ~ rating)
```


Based on the boxplot, I feel this assumption of equal variance is violated. Each variance is different with another. This is due to the /textbf{different and small sample size}, for instance, G rating only has 4 samples.

6

```r
factorial(140)/factorial(4)/factorial(21)/factorial(65)/factorial(50)
```

```
## [1] 4.37673e+64
```

There are 4.37673e64 assigments we need to consider to perform an exact permutation test. This is a very large number. Hence we are using the following code:

7

```r
runtime.anova = lm(runtime ~ rating)
summary(runtime.anova)$fstatistic[1]
```

```
##    value
## 5.259613
```

This F-statistic is taking the F-statistc of the summary of the ANOVA table. The value is 5.2596, which is the same with #1 (5.26) and #3 (5.2596).

8

Since there are 4.37673e64 assigments we need to consider to perform an exact permutation test, we are taking the number of when the *teststat* is "as or more extreme" than the *teststat.obs* in 1000 shuffled runs, and divide it by 1000 to get the *p*-value. We are always checking the upper-tail because we are checking to see if the permuted test statatistic is "as or more extreme" than the observed test statistic. The *F*-statistic is the ratio of explained variance and the unexplained variance. The larger the F is, more variance are explained, in another words, the more significant it is. This might change under the circumstances if "more extreme" is defined as smaller value. For example, to test how rare something happens, we might use less or equal to.

9

```
runtime.anova = lm(runtime ~ rating)
teststat.obs = summary(runtime.anova)$fstatistic[1]

teststat= rep(NA,1000)
for(i in 1:1000){
ratingSHUFFLE = sample(rating)
SHUFFLE.anova = lm(runtime~ratingSHUFFLE)
teststat[i]=summary(SHUFFLE.anova)$fstatistic[1]
}
sum(teststat >= teststat.obs)/1000
```

```
## [1] 0.006
```

The *p*-value is 0.008, which is less than 0.05. Hence, we reject null hypothesis, and conclude that at least one pair (i, j), with strict inequality for at least one x.