Supplement for Lecture 22: One-Way ANOVA

Load Data

```
Exams = read.csv("Exams4.csv")
Exams
```

```
##    Student Exam Grade
## 1     Barb    1    62
## 2     Barb    2    87
## 3     Barb    3    74
## 4     Barb    4    77
## 5    Betsy    1    94
## 6    Betsy    2    95
## 7    Betsy    3    86
## 8    Betsy    4    89
## 9     Bill    1    68
## 10    Bill    2    93
## 11    Bill    3    82
## 12    Bill    4    73
## 13     Bob    1    86
## 14     Bob    2    97
## 15     Bob    3    70
## 16     Bob    4    79
## 17     Bud    1    50
## 18     Bud    2    63
## 19     Bud    3    28
## 20     Bud    4    47
```

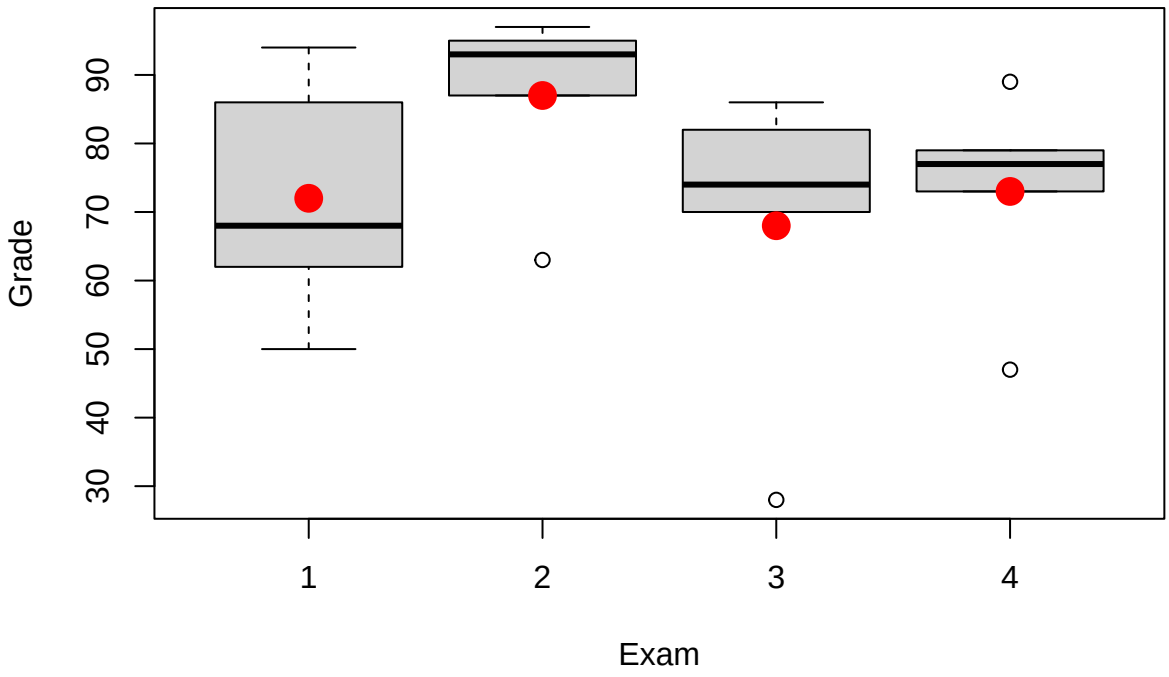
Examine Average Grades of Different Exams

```
# Overall Statistics of Grade
overall = c(length(Exams$Grade),mean(Exams$Grade),sd(Exams$Grade))

# Group Statistics of Grade
length=tapply(Exams$Grade, Exams$Exam, length)
average=tapply(Exams$Grade, Exams$Exam, mean)
st.dev=tapply(Exams$Grade, Exams$Exam, sd)

# Create Table
rbind(cbind(length,average,st.dev),overall)
```

```
##         length average   st.dev
## 1            5      72 17.88854
## 2            5      87 13.92839
## 3            5      68 23.23790
## 4            5      73 15.68439
```

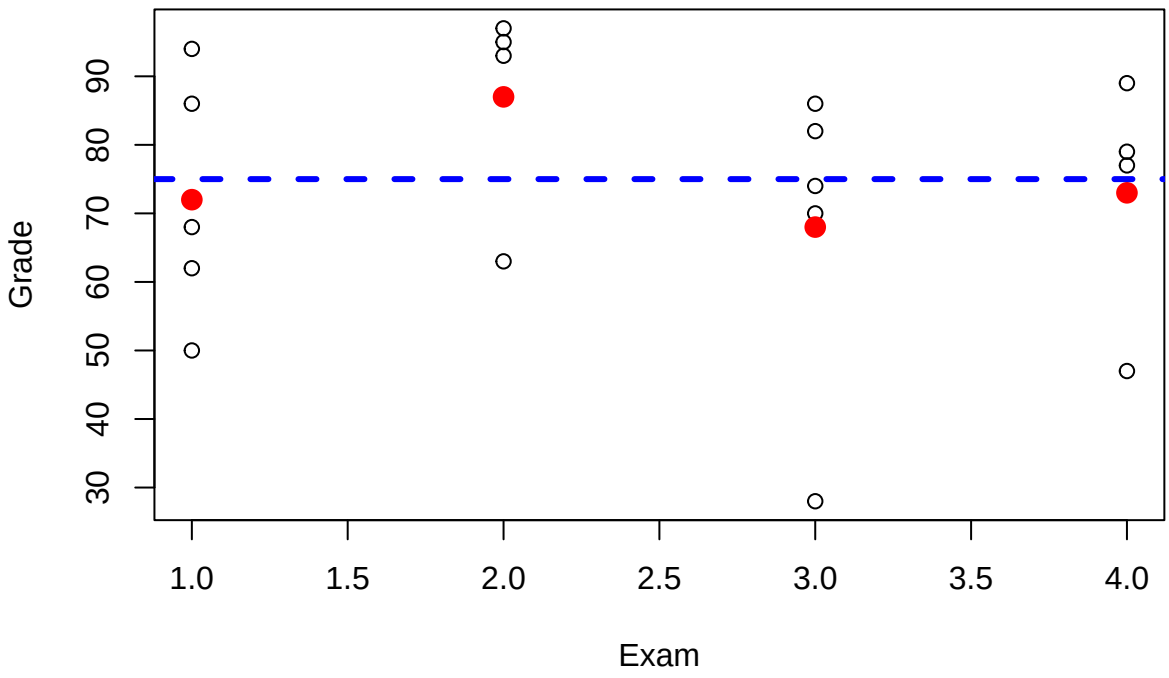
```
## overall     20      75 18.10786
```

```
# Create Visual
boxplot(Grade~Exam,data=Exams)
points(average,col="red",pch=16,cex=2)
```



ANOVA For Testing Differences in Mean Grades for the Different Exams

```
# Visual
plot(Grade ~ Exam, data = Exams)
points(average, col="red", pch=16,cex=1.5)
abline(h = mean(Exams$Grade), col = "blue",lty=2,lwd=3)
```




```
# SST vs SSE vs SSGroups
SST = sum((Exams$Grade-overall[2])^2)
SST
```

```
## [1] 6230
```

```
SSE = sum((Exams$Grade - rep(average,5))^2)
SSE
```

```
## [1] 5200
```

```
SSGroups = SST-SSE
SSGroups
```

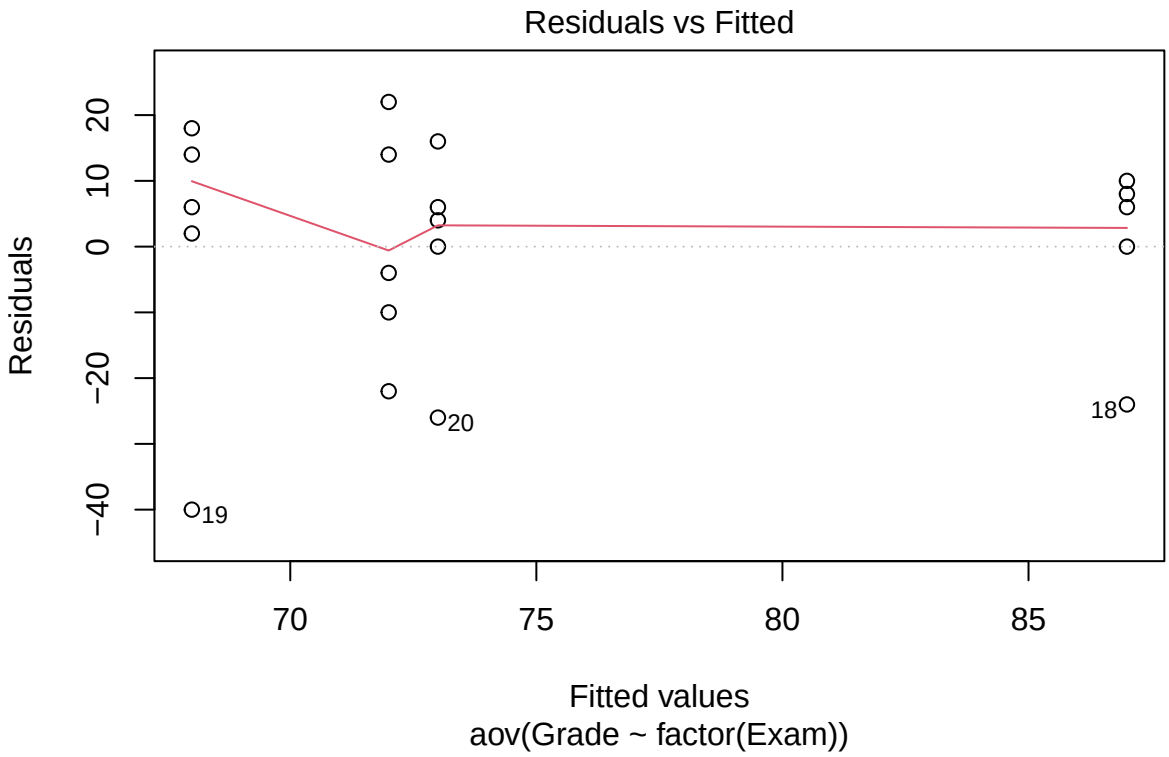
```
## [1] 1030
```

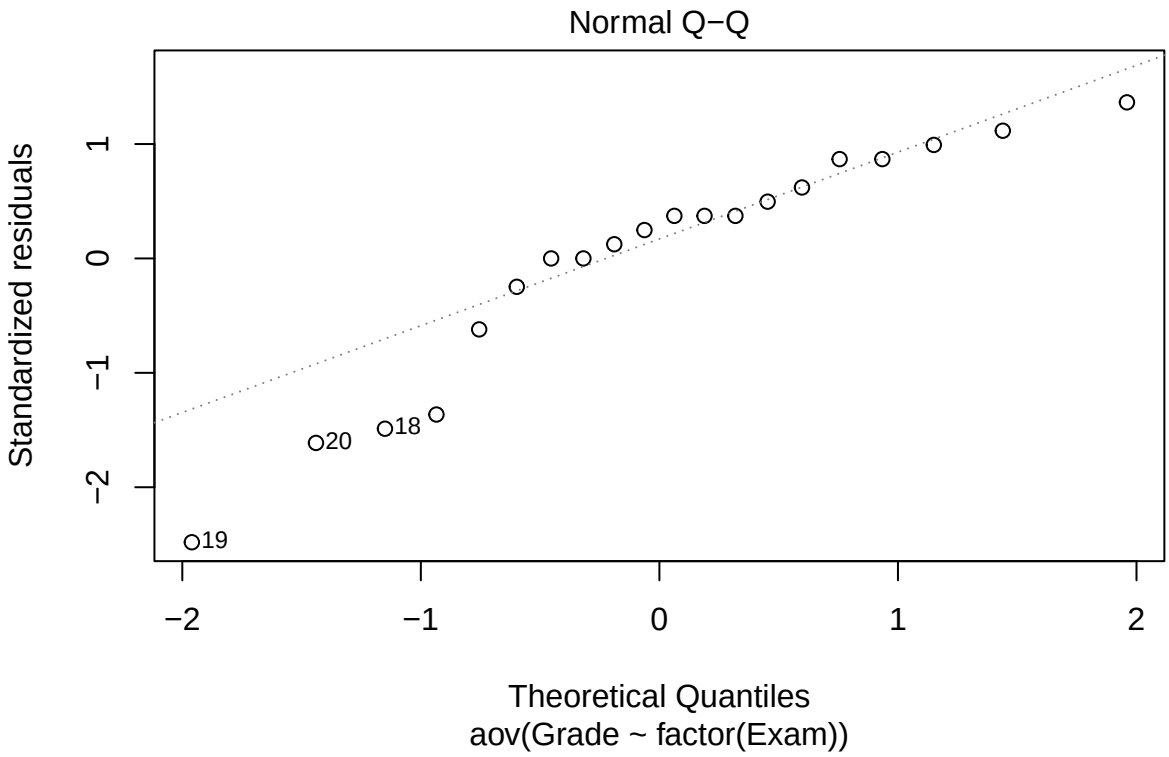
```
#ANOVA F-test
amodG = aov(Grade~factor(Exam),data=Exams)
summary(amodG)
```

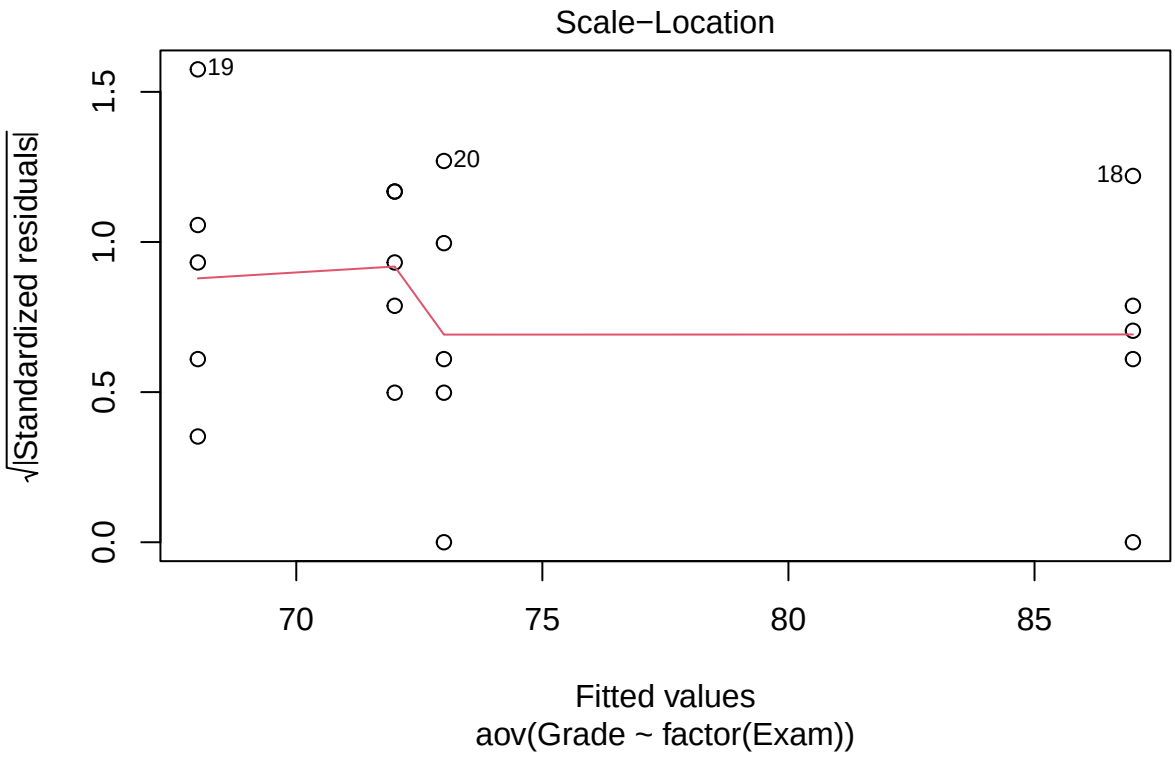
```
##              Df Sum Sq Mean Sq F value Pr(>F)
## factor(Exam)  3   1030   343.3   1.056  0.395
## Residuals    16   5200   325.0
```

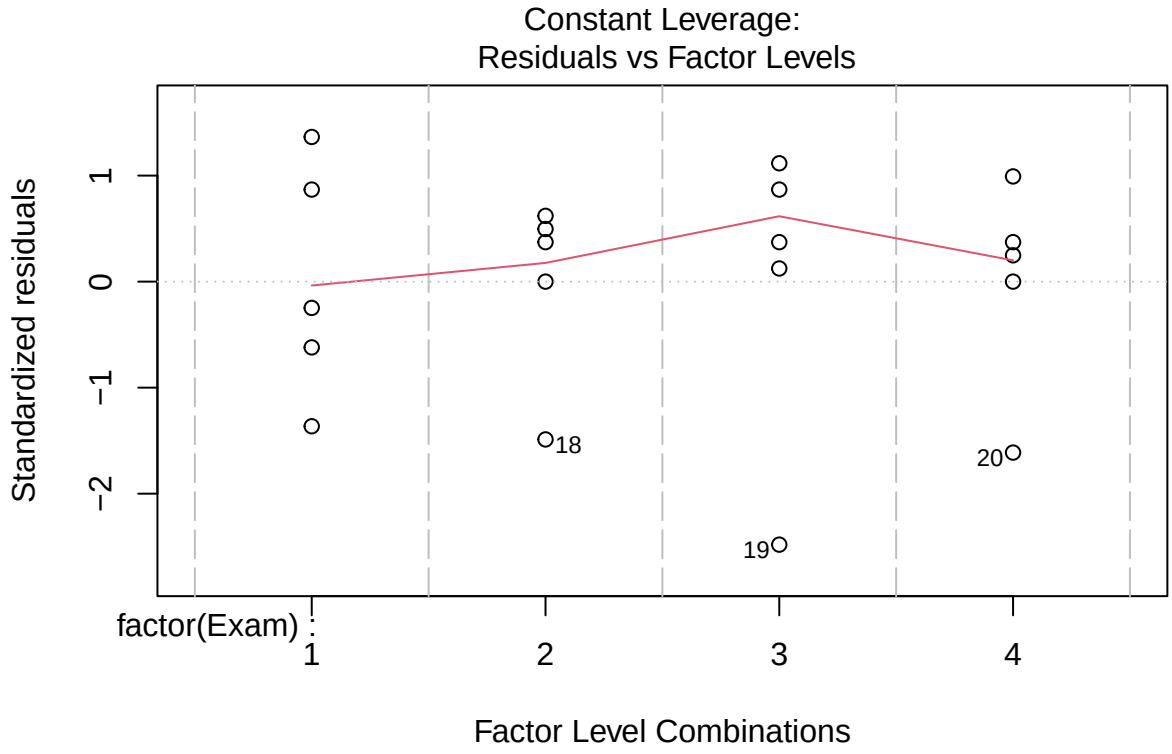
Checking Assumptions

```
plot(amodG) #Check Plots of Residuals
```


Residuals vs Fitted
Residuals
20
10
0
−20
−40
70
75
80
85
Fitted values
aov(Grade ~ factor(Exam))
19
20
18


Normal Q−Q
Standardized residuals
−2
−1
0
1
19
20
18
−2
−1
0
1
2
Theoretical Quantiles
aov(Grade ~ factor(Exam))


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
19
20
18
70
75
80
85
Fitted values
aov(Grade ~ factor(Exam))




```
tapply(Exams$Grade, Exams$Exam, sd) #Is Largest More than Double Smallest
```

```
##        1        2        3        4
## 17.88854 13.92839 23.23790 15.68439
```

```
23.24>2*13.93
```

```
## [1] FALSE
```