PSTAT 122 - HW2

Emily Lu

May 23, 2020

3.14 A pharmaceutical manufacturer wants to investigate the bioactivity of a new drug. A completely randomized single-factor experiment was conducted with three dosage levels, and the following results were obtained.

Dosage	Observations			
20g	24	28	37	30
30g	37	44	31	35
40g	42	47	52	38

(a) Is there evidence to indicate that dosage level affects bioactivity? Use $\alpha = 0.05$.

Sol. $H_0: \mu_1 = \mu_2\mu_3$ vs $H_1: \mu_i \neq \mu_j$ for at least one pair (i, j)

```
bioactivity <- data.frame(obs = c(24, 28, 37, 30, 37, 44, 31, 35,
      42, 47, 52, 38), dosage = rep(c(20, 30, 40), each= 4))

a <- 3
n <- 4
N <- a*n

# SST
SST <- sum( (bioactivity$obs - mean(bioactivity$obs))^2 )
cat(sprintf('SST: %s\n', SST))
```

```
## SST: 738.916666666667
```

```
# SS_Treatments
df.groupmean <- bioactivity %>%
  group_by(dosage) %>%
  summarise(groupmean = mean(obs))
SS.treatment <- n*sum((df.groupmean$groupmean - mean(bioactivity$obs))^2)
cat(sprintf('SS_Treatments: %s\n', SS.treatment))
```

```
## SS_Treatments: 450.666666666667
```

```
# SSE
SS.E <- SST - SS.treatment
cat(sprintf('SSE: %s\n', SS.E))
```

```
## SSE: 288.25
```

```
# Comparing MS_Treatments and MS_E
MS_Treatments <- SS.treatment/(a-1)
MS_E <- SS.E/(N-a)
cat(sprintf('\nMS_Treatments: %s\n', MS_Treatments))
```

```
##
## MS_Treatments: 225.333333333333
```

```
cat(sprintf('MS_E: %s\n', MS_E))
```

```
## MS_E: 32.0277777777778
```

```
# F-test
F0 <- MS_Treatments/MS_E
cat(sprintf('\nF0: %s\n', F0))
```

```
##
## F0: 7.03555941023417
```

```
cat(sprintf('F_(0.05, 2, 9): %s', qf(0.05, a-1, N-a, lower.tail = F)))
```

```
## F_(0.05, 2, 9): 4.25649472909375
```

From the results above, we could see that the between-treatment mean square (225.333) is many times larger than the error mean square (32.028). This indicates that it is unlikely that the treatment means are equal. More formally, at a signficance level of $\alpha = 0.05$, our F ratio ($F_0 = 225.333/32.028 = 7.036$) is greater than $F_{0.05,2,9} = 4.256$. Therefore, we could reject H_0 and conclude that the treatment means differ, i.e. there is evidence to indicate that dosage level affects bioactivity.

(b) If it is appropriate to do so, make comparisons between the pairs of means. What conclusions can you draw?

Since there does appear to be a difference in the dosages from part (a), the comparison of means is appropriate.

1. $H_0 : C_1 : \mu_1 = \mu_2$

```
ybar1 <- df.groupmean[df.groupmean$dosage == 20, ]$groupmean
ybar2 <- df.groupmean[df.groupmean$dosage == 30, ]$groupmean

SS.C1 <- (ybar1 - ybar2)^2 / (1/n * 2)

# F-test
F0 <- (SS.C1/1) / MS_E
cat(sprintf('\nF0: %s\n', F0))
```

```
##
## F0: 3.05984388551605
```

```
cat(sprintf('F_(0.05, 1, 9): %s', qf(0.05, 1, N-a, lower.tail = F)))
```

```
## F_(0.05, 1, 9): 5.11735502919922
```

Since $F_0 < F_{0.05,1,9}$ for the dosage levels between 20g and 30g, we would not reject the null hypothesis, i.e. there is no significant difference among the average bioactivity of the 20g and the average bioactivity of the 30g.

2. $H_0 : C_2 : \mu_2 = \mu_3$

```
ybar3 <-  df.groupmean[df.groupmean$dosage == 40, ]$groupmean

SS.C2 <- (ybar2 - ybar3)^2 / (1/n * 2)

# F-test
F0 <- (SS.C2/1) / MS_E
cat(sprintf('\nF0: %s\n', F0))
```

```
##
## F0: 3.99653078924545
```

```r
cat(sprintf('F_(0.05, 1, 9): %s', qf(0.05, 1, N-a, lower.tail = F)))
```

```
## F_(0.05, 1, 9): 5.11735502919922
```

Since $F_0 < F_{0.05,1,9}$ for the dosage levels between 30g and 40g, we would not reject the null hypothesis, i.e. there is no significant difference among the average bioactivity of the 30g and the average bioactivity of the 40g.

3. $H_0 : C_3 : \mu_1 = \mu_3$

```r
SS.C3 <- (ybar1 - ybar3)^2 / (1/n * 2)

# F-test
F0 <- (SS.C3/1) / MS_E
cat(sprintf('\nF0: %s\n', F0))
```

```
##
## F0: 14.050303555941
```

```r
cat(sprintf('F_(0.05, 1, 9): %s', qf(0.05, 1, N-a, lower.tail = F)))
```

```
## F_(0.05, 1, 9): 5.11735502919922
```

Since $F_0 > F_{0.05,1,9}$ for the dosage levels between 20g and 40g, we would reject the null hypothesis, i.e. there is a significant difference among the average bioactivity of the 20g and the average bioactivity of the 40g.

(c) Analyze the residuals from this experiment and comment on model adequacy.

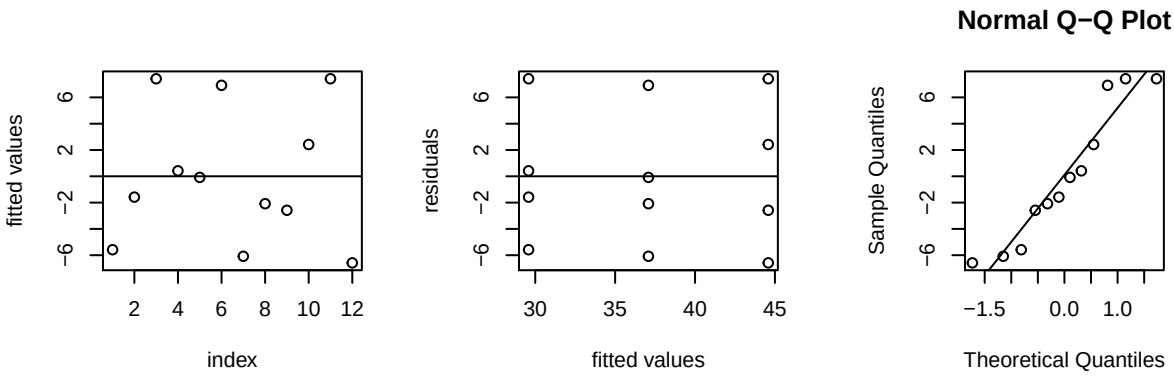
```r
fit <- aov(obs ~ dosage, data = bioactivity)

par(mfrow = c(2, 3))

#1. Independence: Plot residuals vs. index
plot(fit$residuals, ylab = 'fitted values', xlab = 'index')
abline(h = 0)

#2. Constant variance: residuals vs. fitted values
plot(fit$fitted.values, fit$residuals, xlab = "fitted values", ylab = "residuals")
abline(h = 0)

#3. Normality and Outliers: Normal Q-Q plot
par(pty="s")
qqnorm(fit$residuals)
qqline(fit$residuals)
```



From the plots above, there are nothing unusual about the residuals therefore we could assume independence, equal variance, and normality.

3.17 A regional opera company has tried three approaches to solicit donations from 24 potential sponsors. The 24 potential sponsors were randomly divided into three groups of eight, and one approach was used for each group. The dollar amounts of the resulting contributions are shown in the following table.

Approach	Contributions (in $)						
1000	1500	1200	1800	1600	1100	1000	1250
1500	1800	2000	1200	2000	1700	1800	1900
900	1000	1200	1500	1200	1550	1000	1100

(a) Do the data indicate that there is a difference in results obtained from the three different approaches? Use $\alpha = 0.05$.

Sol. $H_0: \mu_1 = \mu_2\mu_3$ vs $H_1: \mu_i \neq \mu_j$ for at least one pair (i, j)

```
results <- data.frame(obs = c(1000, 1500, 1200, 1800, 1600,
                              1100, 1000, 1250, 1500, 1800, 2000, 1200,
                              2000, 1700, 1800, 1900, 900, 1000, 1200,
                              1500, 1200, 1550, 1000, 1100),
                      approach = rep(c(1, 2, 3), each= 8))

a <- 3
n <- 8
N <- a*n

# SST
SST <- sum( (results$obs - mean(results$obs))^2 )
cat(sprintf('SST: %s\n', SST))
```

```
## SST: 2883333.33333333
```

```
# SS_Treatments
df.groupmean <- results %>%
  group_by(approach) %>%
  summarise(groupmean = mean(obs))
SS.treatment <- n*sum((df.groupmean$groupmean - mean(results$obs))^2)
cat(sprintf('SS_Treatments: %s\n', SS.treatment))
```

```
## SS_Treatments: 1362708.33333333
```

```
# SSE
SS.E <- SST - SS.treatment
cat(sprintf('SSE: %s\n', SS.E))
```

```
## SSE: 1520625
```

```
# Comparing MS_Treatments and MS_E
MS_Treatments <- SS.treatment/(a-1)
MS_E <- SS.E/(N-a)
cat(sprintf('\nMS_Treatments: %s\n', MS_Treatments))
```

```
##
## MS_Treatments: 681354.166666667
```

```
cat(sprintf('MS_E: %s\n', MS_E))
```

```
## MS_E: 72410.7142857143
```

```r
# F-test
F0 <- MS_Treatments/MS_E
cat(sprintf('\nF0: %s\n', F0))
```

```
##
## F0: 9.40957665433621
```

```r
cat(sprintf('F_(0.05, 2, 21): %s', qf(0.05, a-1, N-a, lower.tail = F)))
```

```
## F_(0.05, 2, 21): 3.46680011154242
```

From the results above, we could see that the between-approach mean square (681354.167) is many times larger than the error mean square (72410.714). This indicates that it is unlikely that the treatment means are equal. More formally, at a signficance level of $\alpha = 0.05$, our F ratio ($F_0 = 681354.167/72410.714 = 9.410$) is greater than $F_{0.05,2,21} = 3.467$. Therefore, we could reject H_0 and conclude that the approach means differ, i.e. the data does indicate that there is a difference in results obtained from the three different approaches.

Now since it appears that there is a difference in the approaches from above, the comparison of means is appropriate.

1. $H_0 : C_1 : \mu_1 = \mu_2$

```r
ybar1 <- df.groupmean[df.groupmean$approach == 1, ]$groupmean
ybar2 <- df.groupmean[df.groupmean$approach == 2, ]$groupmean

SS.C1 <- (ybar1 - ybar2)^2 / (1/n * 2)

# F-test
F0 <- (SS.C1/1) / MS_E
cat(sprintf('\nF0: %s\n', F0))
```

```
##
## F0: 10.2734278668311
```

```r
cat(sprintf('F_(0.05, 1, 9): %s', qf(0.05, 1, N-a, lower.tail = F)))
```

```
## F_(0.05, 1, 9): 4.32479374318304
```

Since $F_0 > F_{0.05,1,21}$ for the results between approach 1 and 2, we would reject the null hypothesis, i.e. there is significant difference among the average results of approach 1 and the average results of approach 2.

2. $H_0 : C_2 : \mu_2 = \mu_3$

```r
ybar3 <-  df.groupmean[df.groupmean$approach == 3, ]$groupmean

SS.C2 <- (ybar2 - ybar3)^2 / (1/n * 2)

# F-test
F0 <- (SS.C2/1) / MS_E
cat(sprintf('\nF0: %s\n', F0))
```

```
##
## F0: 17.0921701602959
```

```r
cat(sprintf('F_(0.05, 1, 9): %s', qf(0.05, 1, N-a, lower.tail = F)))
```

```
## F_(0.05, 1, 9): 4.32479374318304
```

Since $F_0 > F_{0.05,1,21}$ for the results between approach 2 and 3, we would reject the null hypothesis, i.e. there is significant difference among the average results of approach 2 and the average results of approach 3.

3. $H_0 : C_3 : \mu_1 = \mu_3$

```
SS.C3 <- (ybar1 - ybar3)^2 / (1/n * 2)

# F-test
F0 <- (SS.C3/1) / MS_E
cat(sprintf('\nF0: %s\n', F0))
```

```
##
## F0: 0.863131935881628
```

```
cat(sprintf('F_(0.05, 1, 9): %s', qf(0.05, 1, N-a, lower.tail = F)))
```

```
## F_(0.05, 1, 9): 4.32479374318304
```

Since $F_0 < F_{0.05,1,21}$ for the results between approach 1 and 3, we would not reject the null hypothesis, i.e. there is no significant difference among the average results of approach 1 and the average results of approach 3. So to summarize, approach 2 or approach 2 is different than approach 1 and approach 3.

(b) Analyze the residuals from this experiment and comment on model adequacy.

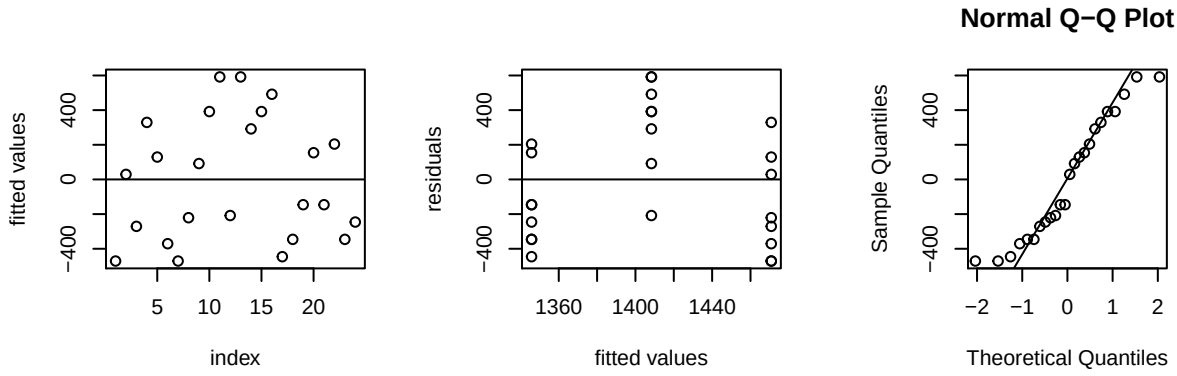
```
fit <- aov(obs ~ approach, data = results)

par(mfrow = c(2, 3))

#1. Independence: Plot residuals vs. index
plot(fit$residuals, ylab = 'fitted values', xlab = 'index')
abline(h = 0)

#2. Constant variance: residuals vs. fitted values
plot(fit$fitted.values, fit$residuals, xlab = "fitted values", ylab = "residuals")
abline(h = 0)

#3. Normality and Outliers: Normal Q-Q plot
par(pty="s")
qqnorm(fit$residuals)
qqline(fit$residuals)
```



From the plots above, there are nothing unusual about the residuals there we could assume model adequacy.