Exam 1

Cody Frisby

February 8, 2016

1)

The mean of A is 7.3333333. The mean of B is 10.3333333. The critical value for our test is the mean of B minus the mean of A, 3. The p value we are calculating is

$$P(\bar{y}_B - \bar{y}_A \geq 3)$$

If this value is less than $\alpha = 0.05$ we conclude that the treatment group B does have an effect on the response when compared to treatment group A.

2.5%	-3.666667
97.5%	3.666667

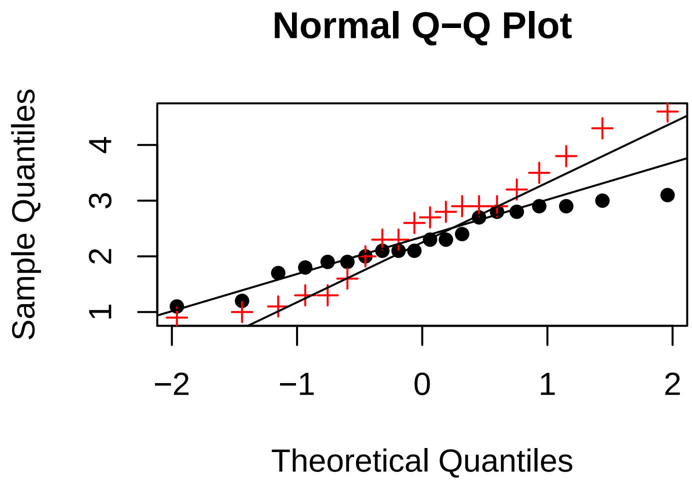
Here is the 95% quantiles of our test data. We can clearly see that 3 is contained with it. We fail to reject the null hypothesis that $\mu_B = \mu_A$.

The probability of observing a value greater than or equal to 3 is **0.15**. We conclude that there is not a significant affect from B when compared to A. Below we display 10 of the assignments of our data

A	A	A	B	B	B	meansdiff
12	10	10	9	5	7	-3.6666667
12	9	10	10	5	7	-3.0000000
12	10	7	10	5	9	-1.6666667
12	9	7	10	5	10	-1.0000000
12	10	5	10	9	7	-0.3333333
12	9	5	10	10	7	0.3333333
10	10	5	9	12	7	1.0000000
12	5	7	10	9	10	1.6666667
10	5	7	10	9	12	3.0000000
9	5	7	10	12	10	3.6666667

2)

An equality of variances test shows that the two groups variances are not equal. We can also see this by plotting the two variables qq plots on the same plot.



Not too bad, but we cannot conclude that the variances are equal. We will assume unequal variances to test whether or not the different line methods affect wait times.

```
    Welch Two Sample t-test

data:  s.line and m.line
t = -0.89446, df = 28.924, p-value = 0.3785
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.8052668  0.3152668
sample estimates:
mean of x mean of y
    2.255     2.500
```

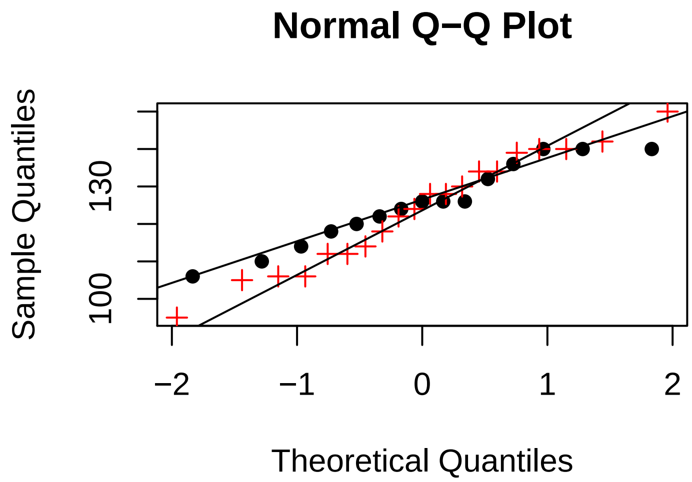
We conclude there is not a significant difference in wait times between the two line methods. The 95% confidence interval from our test contains zero, [-0.8052668, 0.3152668]. Although, we did observe a significant difference in the variation of wait times between the two line methods. This may be of interest to the executives and may warrant further experimentation.

3)

First we run an equal variances test using R and display the output. Then we display the qqnorm plot with both variables qqlines displayed.

```
    F test to compare two variances

data:  f and m
F = 0.51882, num df = 14, denom df = 19, p-value = 0.2147
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1960097 1.4842096
sample estimates:
ratio of variances
         0.5188235
```



We don't conclude that the variances are unequal. We now run a t test to compare the means of the male group (m) with the mean of the female group (f).

```
    Two Sample t-test

data:  m and f
t = -0.30237, df = 33, p-value = 0.7643
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -10.691167   7.924501
sample estimates:
mean of x mean of y
 123.9500  125.3333
```

Note: An equal variances test shows that there is not a significant difference between the groups variances. Also note that a t test using equal var assumption and a t test using unequal var assumption arrives at the same result. Also, note that zero is contained in our 95% confidence interval, $[-10.6911674, 7.9245007]$.

4)

A) This is a single factor experiemnt. Our single factor has 4 levels. Here I display the data organized by levels.

Trt1	Trt2	Trt3	Trt4
3219	2358	3407	3220
3139	2911	3511	2920
3250	2193	3022	2554
3084	2698	2980	2542

B) We fill in an ANOVA table here

Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
1.262136×10^6	3	4.20712×10^5	5.8238767	0.0107745
8.6687×10^5	12	7.2239167×10^4		
2.129006×10^6	15			

C) Confidence Interval for Treatment 1 mean:

This can be computed using the formula

$$\bar{x} \pm t_{\alpha/2:df} \frac{\hat{\sigma}}{\sqrt{n}}$$

The 95% confidence interval for treatment group 1 is [3052.7880317, 3293.2119683].

D) Treatment means comparision

We are going to use Tukey's test to compare all the means.

The computation is as follows:

$$q = \frac{\bar{y}_{max} - \bar{y}_{min}}{\sqrt{MS_E/n}}$$

where $\bar{y}_{max}$ and $\bar{y}_{min}$ are the largest and smallest sample means, respectively, out of p groups. So for our data the min was from group two while the max was from group three. So

$$5.1344357 = \frac{3230 - 2540}{\sqrt{\frac{7.2239167 \times 10^4}{4}}}$$

Tukey's test declares two means significantly different if the absolute value of their sample differences exceeds

$$T_\alpha = q_\alpha(a, f)\sqrt{\frac{MS_E}{n}}$$

The value for $q_{0.05}(4, 12) = 4.1986602$, So

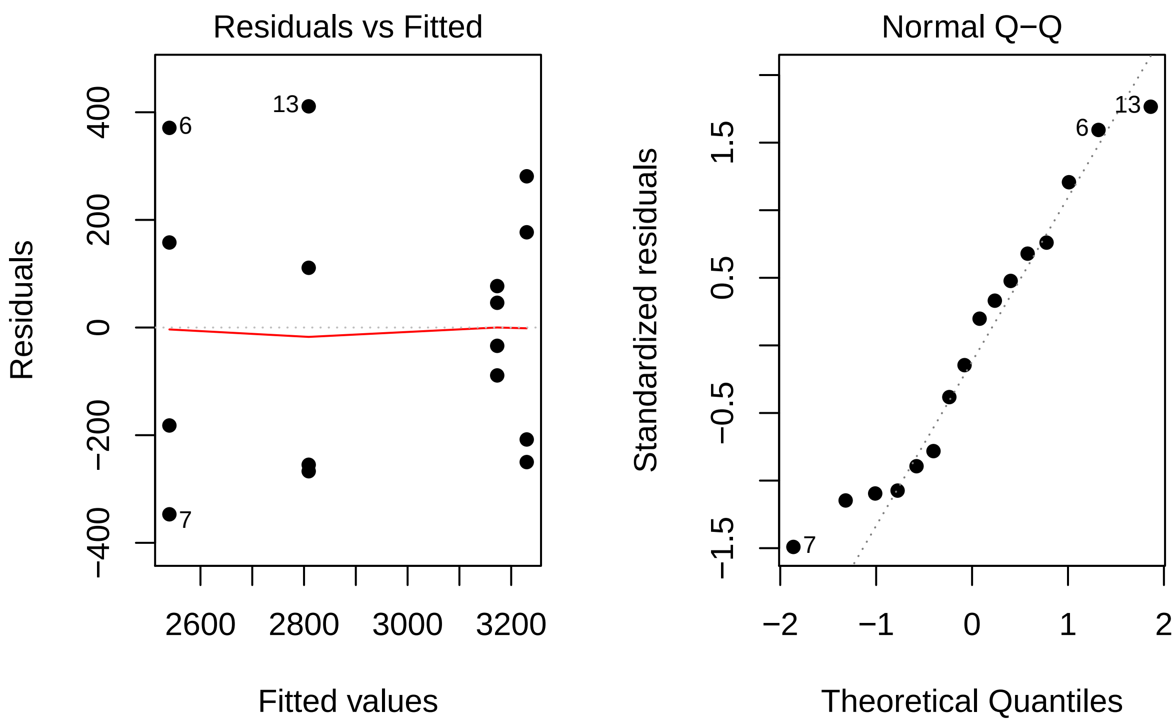
$$564.2441947 = 4.1986602\sqrt{\frac{7.2239167 \times 10^4}{4}}$$

and any mean differences whose absolute value exceeds this number we would conclude the population means to be statistically different. Here is a table of all the comparisons using Tukey's test. We see the difference, confidence intervals, and p value. The two comparisons who absolute value exceeds 564.24 are significant.

	diff	lwr	upr	p adj
Trt2-Trt1	-633	-1197.2442	-68.75581	0.0266335
Trt3-Trt1	57	-507.2442	621.24419	0.9901401
Trt4-Trt1	-364	-928.2442	200.24419	0.2724905
Trt3-Trt2	690	125.7558	1254.24419	0.0157445
Trt4-Trt2	269	-295.2442	833.24419	0.5139356
Trt4-Trt3	-421	-985.2442	143.24419	0.1742495

We conclude that there is a significant difference between groups 2 and 1, and between groups 3 and 2.

E) Normal probability and Residuals Plots



The normal probability plot of the residuals looks OK. We have a few outliers but they are at the tails. The assumption of normality is OK here. The residuals by predicted plot is on the left. There doesn't appear to be any "structure" we need to be concerned about. There may be **nonconstant variance** to be concerned of. One of the treatment groups appears to have smaller variance than the others. However, the variance does not appear to increase as the predicted response increases.

5)

A) Write out a model

A model for our data could be expressed as

$$y_{ij} = \mu + \tau_i + \epsilon_{ij} \begin{cases} i = 1, 2, ..., a \\ j = 1, 2, ..., n \end{cases}$$

where $\mu_i = \mu + \tau_i,\ i = 1, 2, ..., a$. Here, μ is the overall mean, so for our purposes it would be the mean for all 20 observations. And the 5 different doses of ammonium nitrate, μ_i is the group treatment means for $i = 1, 2, 3, 4, 5$. τ_i is the effect the ammonium nitrate has on lettuce yield from the ith treatment group. The hypothesis we are testing is the effect ammonium nitrate has on lettuce yield. It can be represented as

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = 0$$

$$H_1 : \tau_i \neq 0 \quad for\ at\ least\ one\ i$$

where we assume our data to take the form

$$y_{ij} \sim N(\mu + \tau_i, \sigma^2)$$

or equivalently

$$\epsilon_{ij} \sim N(\mu + \tau_i, \sigma^2)$$

and that the observations are mutually independent and identically distributed.

B) ANOVA table

Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
4824	4	1206	5.2556653	0.0075319
3442	15	229.4666667		
8266	19			

C)

To test the hypothesis that the 5 levels of ammonium nitrate have the same affect on lettuce yield we would simply run an analysis of variance. If we find our F statistic to be greater than 4.8932096 we conclude that there is an affect on lettuce yield from ammonium nitrate. So, 5.2556653 > 4.8932096 we conclude that ammonium nitrate concentration has a statistically significant effect on the yield of lettuce.

D)

We'd like to test the control group, 0 lbs/acre ammonium nitrate, against all the other treatment groups. Our hypothesis and the contrast are

$$H_0 : \mu_1 = \mu_2 + \mu_3 + \mu_4 + \mu_5$$

$$Contrast = -\bar{y}_{1.} + \bar{y}_{2.} + \bar{y}_{3.} + \bar{y}_{4.} + \bar{y}_{5.}$$

The coefficients for the contrast will be -4, 1, 1, 1, 1 for groups 0, 50, 100, 150, and 200 respectively. Using the formula for calculating the t_0 statistic for the contrasts

$$t_0 = \frac{\sum_{i=1}^{a} c_i \bar{y}_{i.}}{\sqrt{\frac{MS_E}{n} \sum_{i=1}^{a} c_i^2}}$$

$$4.428396 = \frac{150}{33.8723092}$$

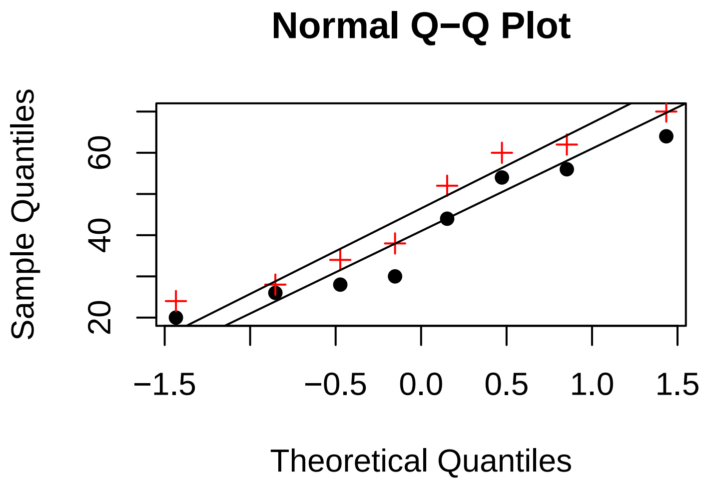
E)

Testing the above contrast at the $\alpha = 0.01$, $t_{\alpha/2,N-a} = 2.9467129$. Our computed contrast statistic was 4.428396. So, we would reject the null hypothesis. We conclude that there is a significant difference between the control group and the other groups.

6)

A) Test treatment 1 vs treatment 2

At first glance of the data we may suspect that the variance are not equal. We should check this using a qq plot of our variables.



We could also run an equal variances test using R's function var.test.

```
var.test(Trt1, Trt2)
```

```

	F test to compare two variances

data:  Trt1 and Trt2
F = 0.90553, num df = 7, denom df = 7, p-value = 0.8992
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1812904 4.5230344
sample estimates:
ratio of variances
         0.9055288
```

We conclude the variances are not unequal. We can proceed with a t test but we will run a comparison test using contrasts and ANOVA since there was a third treatment group in this experiment.
Our hypothesis and the contrast are

$$H_0 : \mu_1 = \mu_2$$

$$Contrast = 1\bar{y}_{1.} - \bar{y}_{2.} + 0\bar{y}_{3.}$$

The coefficients are 1, -1, 0 for groups 1, 2, and 3 respectively. Using the formula for calculating the t_0 statistic for the contrasts

$$t_0 = \frac{\sum_{i=1}^{a} c_i \bar{y}_{i.}}{\sqrt{\frac{MS_E}{n} \sum_{i=1}^{a} c_i^2}}$$

$$-0.8293884 = \frac{-5.75}{6.9328196}$$

Testing the above contrast at the $\alpha = 0.05$, $t_{\alpha/2,N-a} = 2.0796138$. Our computed contrast statistic was -0.8293884. We ask if the absolute value of this is greater than 2.0796138. The answer is no, so we fail to reject the null hypothesis.

B) Comparing the control to treatments 1 & 2

We'd like to test the control group against the other two treatment groups. Our hypothesis and the contrast are

$$H_0 : \mu_1 + \mu_2 = \mu_3$$

$$Contrast = \bar{y}_{1.} + \bar{y}_{2.} - \bar{y}_{3.}$$

The coefficients are given in the problem, 1, 1, 2 for groups 1, 2, and the control respectively. Using the formula for calculating the t_0 statistic for the contrasts

$$t_0 = \frac{\sum_{i=1}^{a} c_i \bar{y}_{i.}}{\sqrt{\frac{MS_E}{n} \sum_{i=1}^{a} c_i^2}}$$

$$5.7878101 = \frac{69.5}{12.0079958}$$

Testing the above contrast at the $\alpha = 0.05$, $t_{\alpha/2,N-a} = 2.0796138$. Our computed contrast statistic was 5.7878101. So, we would reject the null hypothesis. We conclude that there is a significant difference between the control group and groups 1 & 2.

C) ANOVA Table

To compute the sum of square for a contrast we take the numerator for our t_0 statistic for the contrast and square it. We then divide it by the sum of the contrast coefficients squared divided by n_i.

$$SS_C = \frac{(\sum_{i=1}^{a} c_i \bar{y}_{i.})^2}{\frac{1}{n} \sum_{i=1}^{a} c_i^2}$$

Source	Sum of Squares	D.F.	Mean Square	F_0	P-Value
Treatments	6572.5833333	2	3286.2916667	17.0933156	3.9268726×10^{-5}
$C_1 : \mu_1 = \mu_2$	(132.25)	1	132.25	0.6878851	0.4162132
$C_2 : \mu_1 + \mu_2 = \mu_3$	(6440.3333333)	1	6440.3333333	33.4987461	9.5808925×10^{-6}
Error	4037.375	21	192.2559524		
Total	1.0609958×10^4	23			