Advanced Data Analysis HW2

Ao Liu, al3472

1.

Suppose you have three different feeds that may affect the size of eggs that chickens lay. You randomly assign 10 chickens to each one of the three feeds and record the size of the eggs (maximum length, in centimeters) that the chickens lay the following week. The null hypothesis is that all the chicken feeds have the same effect on the length of the major axis. The alternative is that the feed has some causal effect. A partial output is

Source	df	SS	MS	F
feed		23.43		
error				
total		28.10		

(a) Complete the table above

Answer:

Source	df	SS	MS	F
feed	2	23.43	11.715	67.731
error	27	4.67	0.1730	
total	29	28.10		

(b)

Test the null hypothesis that all the chicken feeds have the same effect on the length of the major axis against the alternative that the feed has some causal effect. Use $\alpha = 0.05$.

Answer:

```
>qf(.95, df1=2, df2=27)
[1] 3.354131
```

Since $F = 67.731 > F(0.95, 2, 29) = 3.354131$, we reject the null hypothesis that all the chicken feeds have the same effect.

2.

Suppose you want to compare the types of popcorn popper and the brand of popcorn with respect to their yield (in terms of cups of popped corn). Factor A is the type of popper: oil-based versus air-based. Factor B is the brand of popcorn: gourmet versus national brand versus generic. For each combination of popper type and brand, you took three separate measurements. The ANOVA table is

Source	df	SS	MS	F
Propper(A)		4.5		
Corn(B)		15.75		
Interaction (A*B)				
error		1.67		
total		22.00		

(a)

Complete the table above.

Answer:

Source	df	SS	MS	F
Propper(A)	1	4.5	4.5	32.374
Corn(B)	2	15.75	7.875	56.655
Interaction (A*B)	2	0.08	0.04	0.2878
error	12	1.67	0.139	
total	17	22.00	1.294	

(b)

Test H_0 : No interaction against H1 : there is an interaction, use $\alpha = 0.05$.

Answer:

```
>qf(.95, df1=2, df2=12)
[1] 3.885294
```

Since $F = 32.374 > F(0.95, 1, 12) = 3.885294$, we reject the null hypothesis that there is no interaciton.

(c)

It is decided to fit a model without an interaction and the partial results are

Source	df	SS	MS	F
Propper(A)		4.5		
Corn(B)		15.75		
error		1.67		
total		22.00		

Answer:

(d)

Complete the table above.

Answer:

Source	df	SS	MS	F
Propper(A)	1	4.5	4.5	37.72
Corn(B)	2	15.75	7.875	66.01
error	14	1.67	0.1193	
total	17	22.00		

(e)

Test H_0: No popper effect against H_1: there is a popper effect. Use $\alpha = 0.05$

Answer:

```
>qf(.95, df1=1, df2=14)
[1] 4.60011
```

Since $F = 37.72 > F(0.95, 1, 14) = 4.6001$, we reject the null hypothesis that there is no popper effect

(f)

Test H_0: No corn effect against H_1 : there is a corn effect. Use $\alpha = 0.05$.

Answer:

```
>qf(.95, df1=2, df2=14)
[1] 3.738892
```

Since $F = 66.01 > F(0.95, 2, 14) = 3.738892$, we reject the null hypothesis that there is no corn effect

3.

In this exercise A and B are two fertilizers types, M, N, O and P are four wheat types and yijk values are wheat yields in bushels per plot (one third of an acre) corresponding to the different combinations of fertilizer type and wheat type. Also, assume that this data was obtained by using a completely randomized experimental design. (see HW2data.csv)

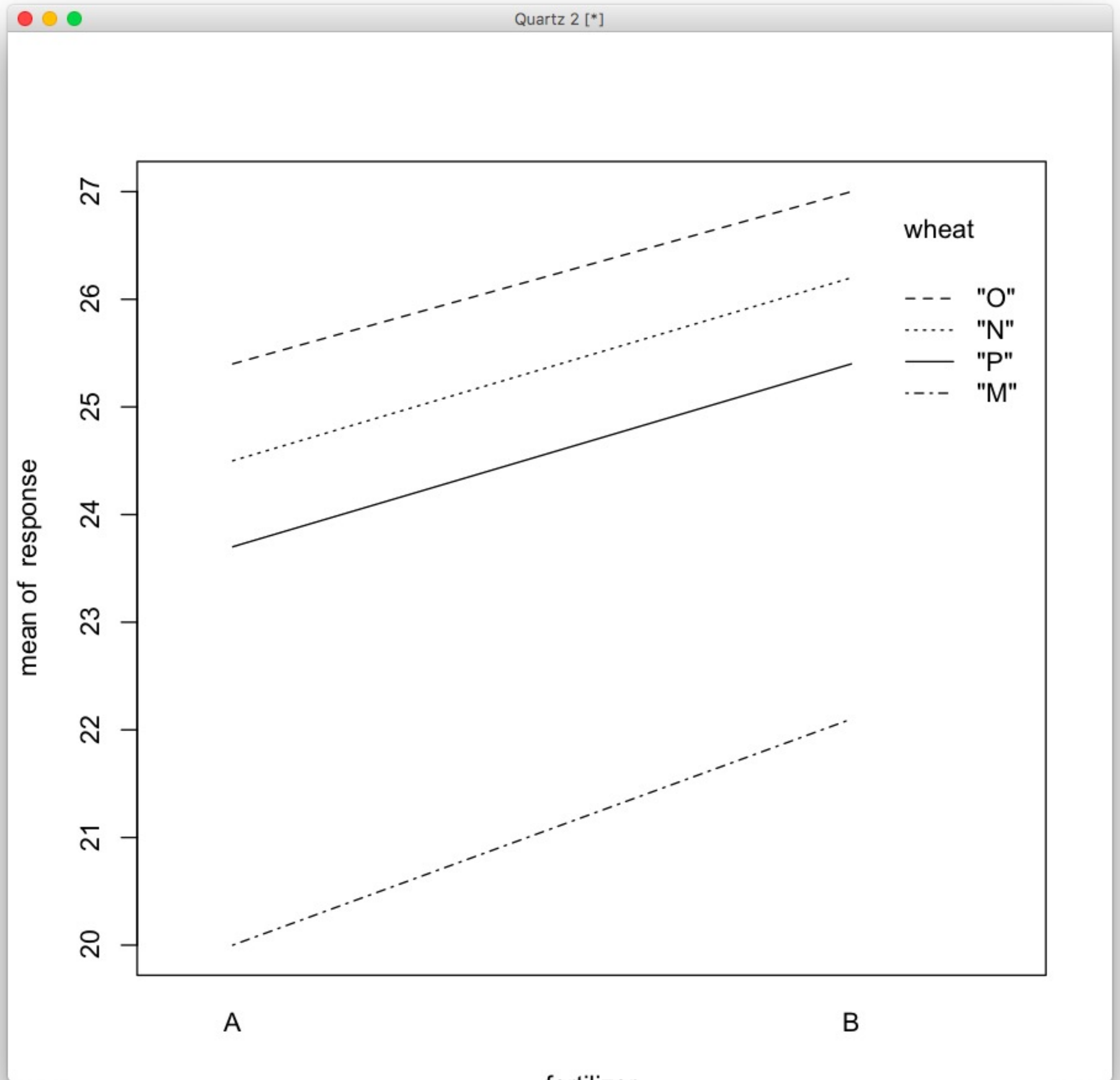
(a)

Construct an interaction plot? Does it suggest that there is an interaction between fertilizer type and wheat type?

Answer:

```
>data <- read.csv("HW2DATA.csv")
>fertilizer <- c(rep("A",12),rep("B",12))
>wheat <- data[,2]
>response <- data[,3]
>interaction.plot(fertilizer ,wheat,response)
```

we get the following interaction plot:



The four lines are parallel to each other, suggesting that the there is no interaction between fertilizer type and wheat type.

(b)

Test H_0 : No interaction against H_1 : there is an interaction, use $\alpha = 0.05$.

Answer:

```
>summary(aov(response~fertilizer*wheat))
```

we get the following result:

```
                 Df Sum Sq Mean Sq F value    Pr(>F)
fertilizer        1  18.90  18.904   48.63 3.14e-06 ***
wheat             3  92.02  30.674   78.90 8.37e-10 ***
fertilizer:wheat  3   0.22   0.074    0.19    0.902
```

```
Residuals          16    6.22    0.389
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value for "fertilizer:wheat" is $0.902 > \alpha = 0.05$, then we cannot reject H_0.

(c)

Fit a model without an interaction and test H_0: No fertilizer effect against H_1 : there is a fertilizer effect. Use $\alpha = 0.05$ if you reject H_0, use Tukeys method to do pairwise comparisons of the different fertilizer types.

Answer:

We fit a model without interaction:

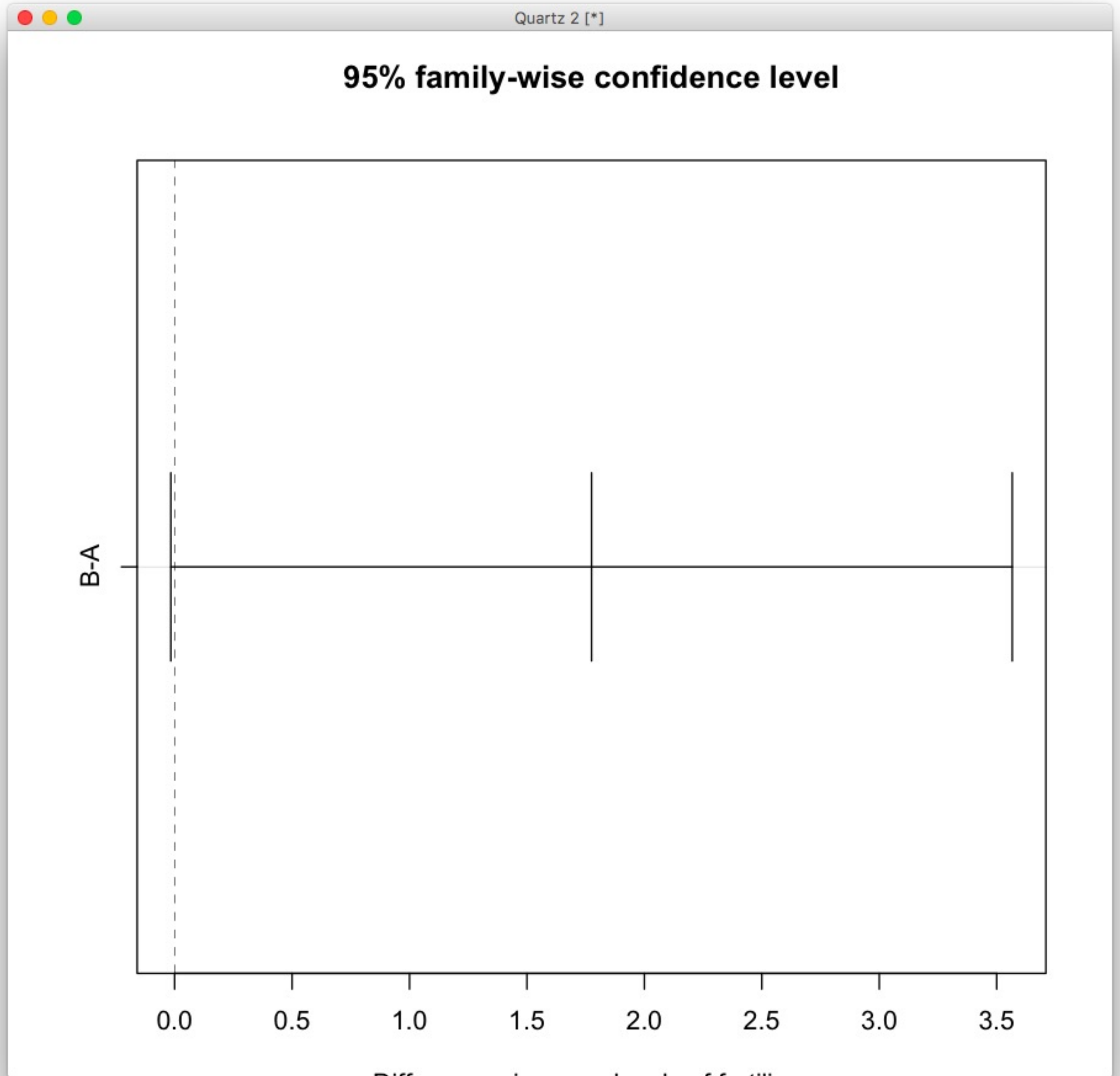
```
>summary(aov(response~fertilizer+wheat))
```

```
          Df Sum Sq Mean Sq F value    Pr(>F)
fertilizer   1  18.90  18.904   55.76 4.59e-07 ***
wheat        3  92.02  30.674   90.48 1.97e-11 ***
Residuals   19   6.44   0.339
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since p-value for fertilizer if $4.59e-07 < \alpha = 0.05$, we reject H_0.

Now we use Tukey's method to do pairwise comparisons of the different fertilizer types:

```
fit <- aov(response~fertilizer)
tk<-TukeyHSD(fit,"fertilizer")
tk
plot(tk)
```




```
            Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = response ~ fertilizer)

fertilizer
      diff          lwr        upr     p adj
B-A  1.775  -0.01614443  3.566144  0.0519242

```

Since the p-value is $0.0519242 > \alpha = 0.05$, we cannot reject that there's no difference between the two fertilizer types.

(d)

Test H_0 : No wheat effect against H_1 : there is a effect. Use $\alpha = 0.05$ if you reject H_0, use Tukeys method to do pairwise comparisons of the different wheat types.

Answer:

According to the model we fit in (c), the p-value for wheat type is $4.59e-07 < \alpha = 0.05$, so we reject the hypothesis that there's no wheat efffect.

then we use the Tukey's method to do the pairwise comparisons of different wheat types:

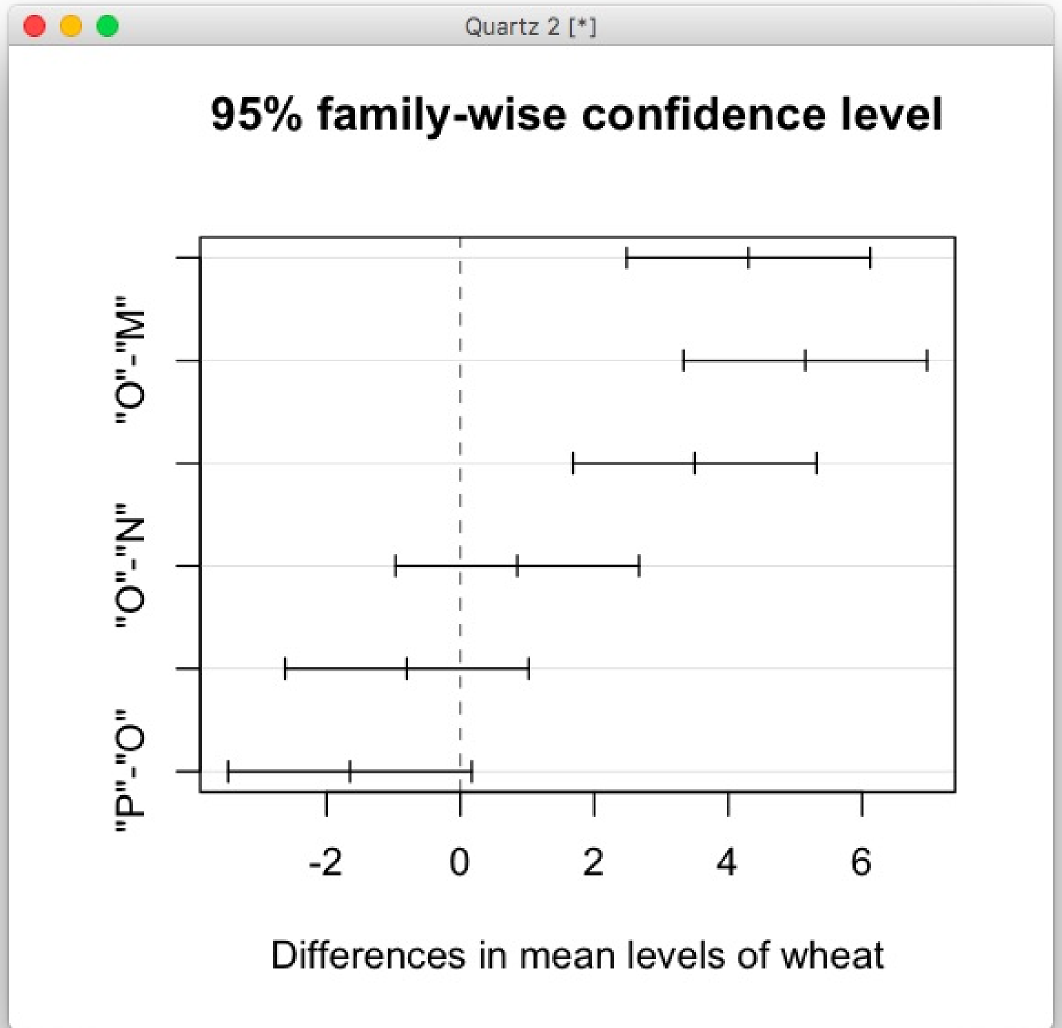
```
>fit <- aov(response~wheat)
>tk<-TukeyHSD(fit,"wheat")
>tk
>plot(tk)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = response ~ wheat)

wheat
          diff        lwr       upr     p adj
"N"-"M"   4.30  2.4808709 6.1191291 0.0000107
"O"-"M"   5.15  3.3308709 6.9691291 0.0000008
"P"-"M"   3.50  1.6808709 5.3191291 0.0001557
"O"-"N"   0.85 -0.9691291 2.6691291 0.5687888
"P"-"N"  -0.80 -2.6191291 1.0191291 0.6152451
"P"-"O"  -1.65 -3.4691291 0.1691291 0.0839841
```

we get the following plot:



According to the plot and the p-value of the test, we arrive at a conclusion that there is no significant difference between wheat type P, N and O, but there is a significant difference between wheat type M and these three wheat types.