ADVANCED DATA ANALYSIS

HW2

Fan Yang
UNI: fy2232
01/31/2018

Problem 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 so me causal effect.

> (a) Complete the table above.
> (b) Test the null hypothesis is that all the chicken feeds have the same effect on the length of the major axis against the alternative is that the feed has some causal effect. Use $\alpha = 0.05$.

(a)

In [1]:
```
ANOVA1 = matrix(NA,nrow = 3, ncol = 4)
colnames(ANOVA1) = c("df","SS","MS","F")
rownames(ANOVA1) = c("feed","error","total")
k = 3
n = 10
ANOVA1[,1] = c(k-1,n-k,n-1)
ANOVA1[,2] = c(23.43,28.10-23.43,28.10)
ANOVA1[1:2,3] = ANOVA1[1:2,2] / ANOVA1[1:2,1]
ANOVA1[1,4] = ANOVA1[1,3] / ANOVA1[2,3]
ANOVA1
```

	df	**SS**	**MS**	**F**
feed	2	23.43	11.7150000	17.55996
error	7	4.67	0.6671429	NA
total	9	28.10	NA	NA

(b)

We reject H_0 if $F > F(\alpha, k-1, n-k)$
$F(\alpha, k-1, n-k) = F(0.05, 2, 7) = 4.73741$, which is less than $F = 17.55996$, therefore, we reject H_0 that not all the chicken feeds have the same effect on the length of the major axis

In [2]:
```
qf(0.95,2,7)
```

4.73741412777588

Problem 2

Suppose you want to compare the types of popcorn popper and the brand of pop- corn 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

> (a) Complete the table above.
> (b) Test H_0 : No interaction against H_1 : there is an interaction, use $\alpha = 0.05$.
> (c) It is decided to fit a model without an interaction and the partial results are
> (d) Complete the table above.
> (e) Test H_0 : No popper effect against H_1 : there is a popper effect. Use $\alpha = 0.05$.
> (f) Test H_0 : No corn effect against H_1 : there is a corn effect. Use $\alpha = 0.05$.

(a)

In [3]:
```
ANOVA2a = matrix(NA,nrow = 5, ncol = 4)
colnames(ANOVA2a) = c("df","SS","MS","F")
rownames(ANOVA2a) = c("Propper(A)","Corn(B)","Interaction(A*B)",
                      "Error","Total")
a = 2; b = 3; n = 3
# compute degree of freedom
ANOVA2a[,1] = c(a-1,b-1,(a-1)*(b-1),a*b*(n-1),a*b*n-1)
#compute SS
ANOVA2a[,2] = c(4.5,15.75,NA,1.67,22.00)
ANOVA2a[3,2] = ANOVA2a[5,2] - sum(ANOVA2a[1:4,2],na.rm = TRUE)
#compute MS
ANOVA2a[1:4,3] = ANOVA2a[1:4,2] / ANOVA2a[1:4,1]
#compute F
ANOVA2a[1:3,4] = ANOVA2a[1:3,3] / ANOVA2a[4,3]
ANOVA2a
```

	df	SS	MS	F
Propper(A)	1	4.50	4.5000000	32.3353293
Corn(B)	2	15.75	7.8750000	56.5868263
Interaction(A*B)	2	0.08	0.0400000	0.2874251
Error	12	1.67	0.1391667	NA
Total	17	22.00	NA	NA

(b)

We reject H_0 if $F = \frac{MSAB}{MSE} > F(1-\alpha,(a-1)(b-1),ab(n-1))$
As we computed in the ANOVA table, $F = 0.2874251$
and $F(1-\alpha,(a-1)(b-1),ab(n-1)) = F(0.95,2,12) = 3.88529$, which is greater than $F = 0.2874251$, therefore, we can not reject H_0 that no interaction exists.

In [4]:
```
qf(0.95,2,12)
```

3.88529383465239

(c)(d)

In [5]:
```
ANOVA2d = matrix(NA,nrow = 4, ncol = 4)
colnames(ANOVA2d) = c("df","SS","MS","F")
rownames(ANOVA2d) = c("Propper(A)","Corn(B)","Error","Total")
a = 2; b = 3; n = 3
# compute degree of freedom
ANOVA2d[,1] = c(a-1,b-1,(a*b*n-1)-(a-1)-(b-1),a*b*n-1)
#compute SS
ANOVA2d[,2] = c(4.5,15.75,NA,22.00)
ANOVA2d[3,2] = ANOVA2d[4,2] - sum(ANOVA2d[1:3,2],na.rm = TRUE)
#compute MS
ANOVA2d[1:3,3] = ANOVA2d[1:3,2] / ANOVA2d[1:3,1]
#compute F
ANOVA2d[1:2,4] = ANOVA2d[1:2,3] / ANOVA2d[3,3]
ANOVA2d
```

	df	SS	MS	F
Propper(A)	1	4.50	4.500	36
Corn(B)	2	15.75	7.875	63
Error	14	1.75	0.125	NA
Total	17	22.00	NA	NA

(e)

We reject H_0 if $F = \frac{MSA}{MSE} > F(1-\alpha,a-1,(abn-1)-(a-1)-(b-1))$
As we computed in the ANOVA table, $F = 36$
and $F(1-\alpha,a-1,(abn-1)-(a-1)-(b-1)) = F(0.95,1,14) = 4.6001099$, which is less than $F = 36$, therefore, we reject H_0 that "no popper effect exists".

In [6]:
```
qf(0.95,1,14)
```

4.60010993666942

(f)

We reject H_0 if $F = \frac{MSB}{MSE} > F(1-\alpha, b-1, (abn-1)-(a-1)-(b-1))$
As we computed in the ANOVA table, $F = 63$
and $F(1-\alpha, b-1, (abn-1)-(a-1)-(b-1)) = F(0.95, 2, 12) = 3.88529$, which is less than $F = 63$, therefore, we reject H_0 that "no corn effect exists".

In [7]:
```
qf(0.95,2,12)
```

3.88529383465239

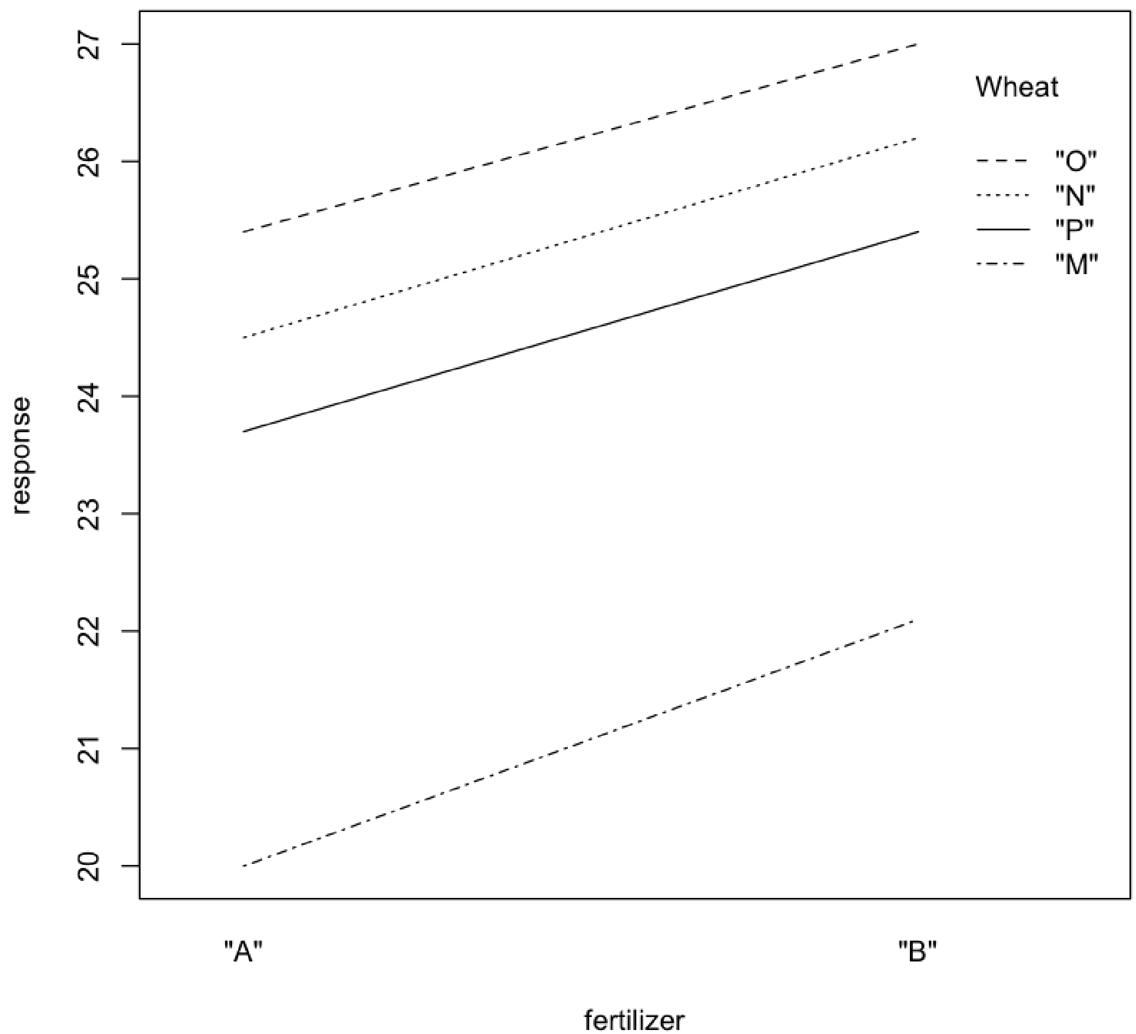
Problem 3

In this exercise A and B are two fertilizers types, M, N, O and P are four wheat types and y_{ijk} values are wheat yields in bushels per plot (one third of an acre) corresponding to the different combinations of fertilizer type 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 is suggest that there is an interaction between fertilizer type and wheat type?
> (b) Test H_0 : No interaction against H_1 : there is an interaction, use $\alpha = 0.05$.
> (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 Tukey's method to do pairwise comparisons of the different fertilizer types.
> (d) Test H_0 : No wheat effect against H_1 : there is a effect effect. Use $\alpha = 0.05$ if you reject H_0, use Tukey's method to do pairwise comparisons of the different wheat types.

(a)

In [8]:
```
data3 = read.csv("HW2data.csv", header = TRUE, stringsAsFactors=FALSE)
interaction.plot(data3$fertilizer,data3$Wheat,data3$response,
                 trace.label="Wheat",xlab="fertilizer", ylab = "response"
)
```



It suggests that there is no interaction between fertilizer type and wheat type.

(b)

In [9]:
```
summary(aov(response~fertilizer*Wheat, data=data3))
```

```
                 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
```

The p-value of the test for interaction is 0.902. Under $\alpha = 0.05$, we fail to reject H0 and conclude that there is no interaction.

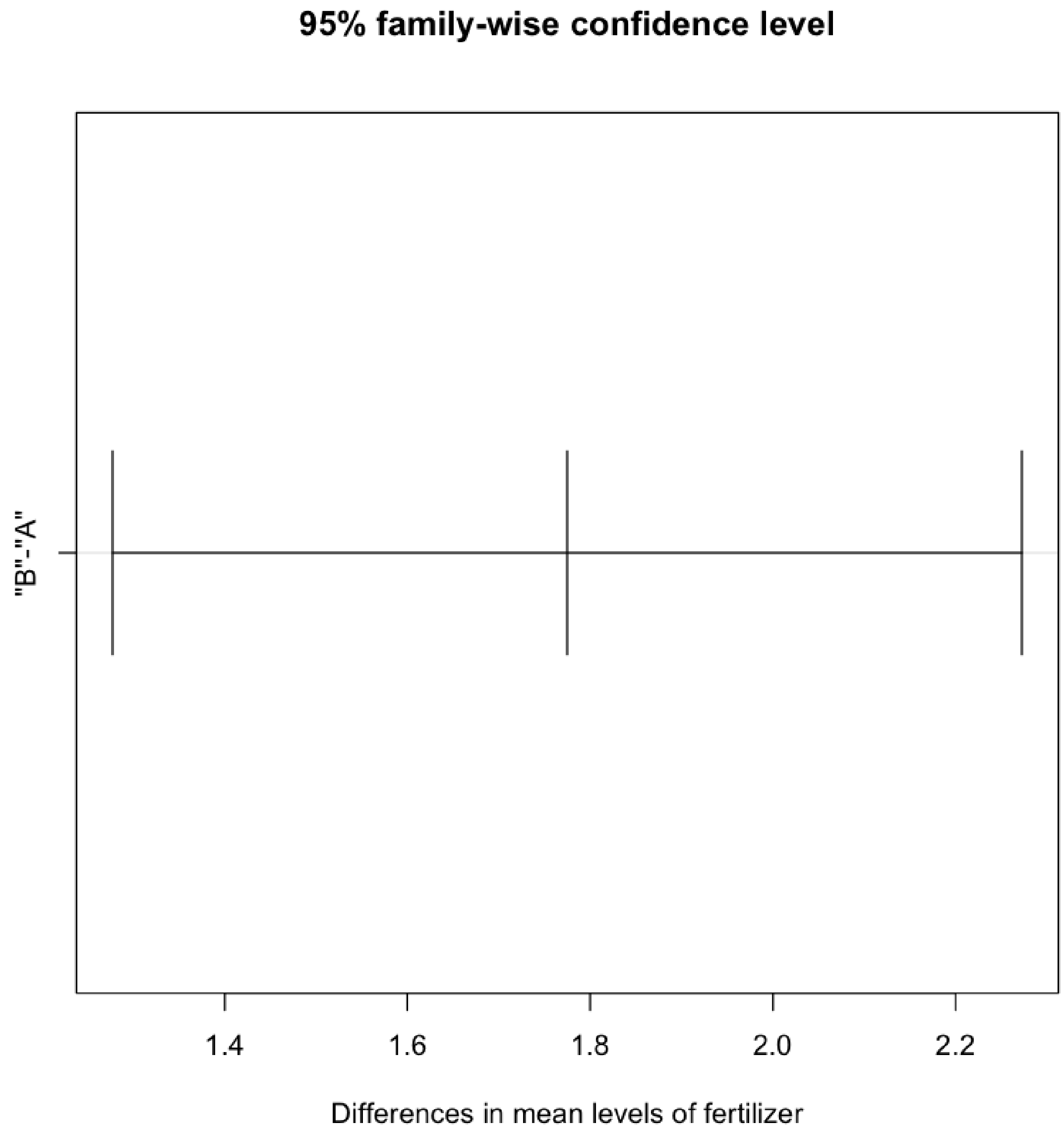
(c)

In [10]:
```
summary(aov(response~fertilizer+Wheat, data=data3))
```

```
            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
```

From this output we can conclude that fertilizer level means (p-value<0.05) are different. So we conclude there is a fertilizer effect.

In [11]:
```
fit.c <- aov(response~fertilizer+Wheat, data = data3)
tk.c <- TukeyHSD(fit.c, which = "fertilizer")
plot(tk.c)
```



In [12]:
```
tk.c
```

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

Fit: aov(formula = response ~ fertilizer + Wheat, data = data3)

$fertilizer
         diff      lwr      upr p adj
"B"-"A" 1.775 1.277484 2.272516 5e-07
```

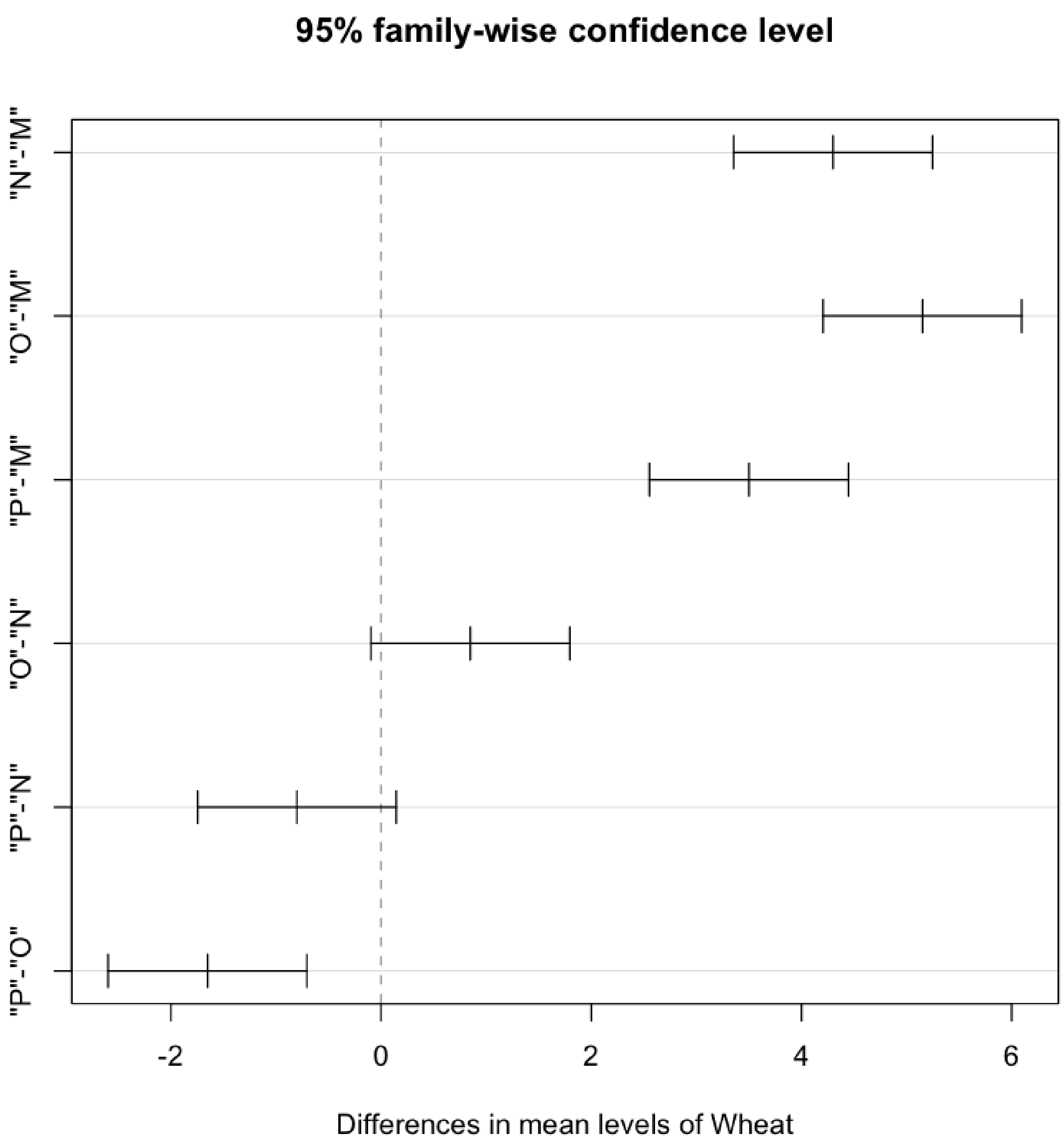
The p-value of the test is 0.0519242. Under $\alpha = 0.05$, we fail to reject H0 and conclude that there is no fertilizer effect.

(d)

response ~ fertilizer + Wheat

As discussed in (c), we can conclude that Wheat level means (p-value<0.05) are different. So we conclude there is a Wheat effect.

In [13]:
```
fit.d <- aov(response~fertilizer+Wheat , data = data3)
tk.d <- TukeyHSD(fit.d, "Wheat")
plot(tk.d)
```



In [14]:
```
tk.d
```

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

Fit: aov(formula = response ~ fertilizer + Wheat, data = data3)

$Wheat
         diff         lwr        upr     p adj
"N"-"M"  4.30  3.35476633  5.2452337 0.0000000
"O"-"M"  5.15  4.20476633  6.0952337 0.0000000
"P"-"M"  3.50  2.55476633  4.4452337 0.0000000
"O"-"N"  0.85 -0.09523367  1.7952337 0.0872269
"P"-"N" -0.80 -1.74523367  0.1452337 0.1152696
"P"-"O" -1.65 -2.59523367 -0.7047663 0.0005208
```

Under $\alpha = 0.05$, we have $\{"\ P\ ", "\ N\ "\}$ is similar and $\{"\ O\ ", "\ N\ "\}$ is similar while $\{"\ M\ "\}$ different from others.

response ~ Wheat

In [15]:
```
fit.d2 <- aov(response~Wheat , data = data3)
tk.d2 <- TukeyHSD(fit.d2, "Wheat")
tk.d2
```

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

Fit: aov(formula = response ~ Wheat, data = data3)

$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
```

Under $\alpha = 0.05$, we have two groups here$\{"\ P\ ", "\ N\ ", "\ O\ "\}$ and $\{"\ M\ "\}$.