02441 Applied Statistics and Statistical Software

Exercise 3B - Fertilizer

In a pilot study two different fertilizers were tested on three different fields. The yield (compared to a reference) is given in the table below

Fertilizer/Field	a	b	c
a	10.3	-4.95	-6
b	10.3	-4.95	-11

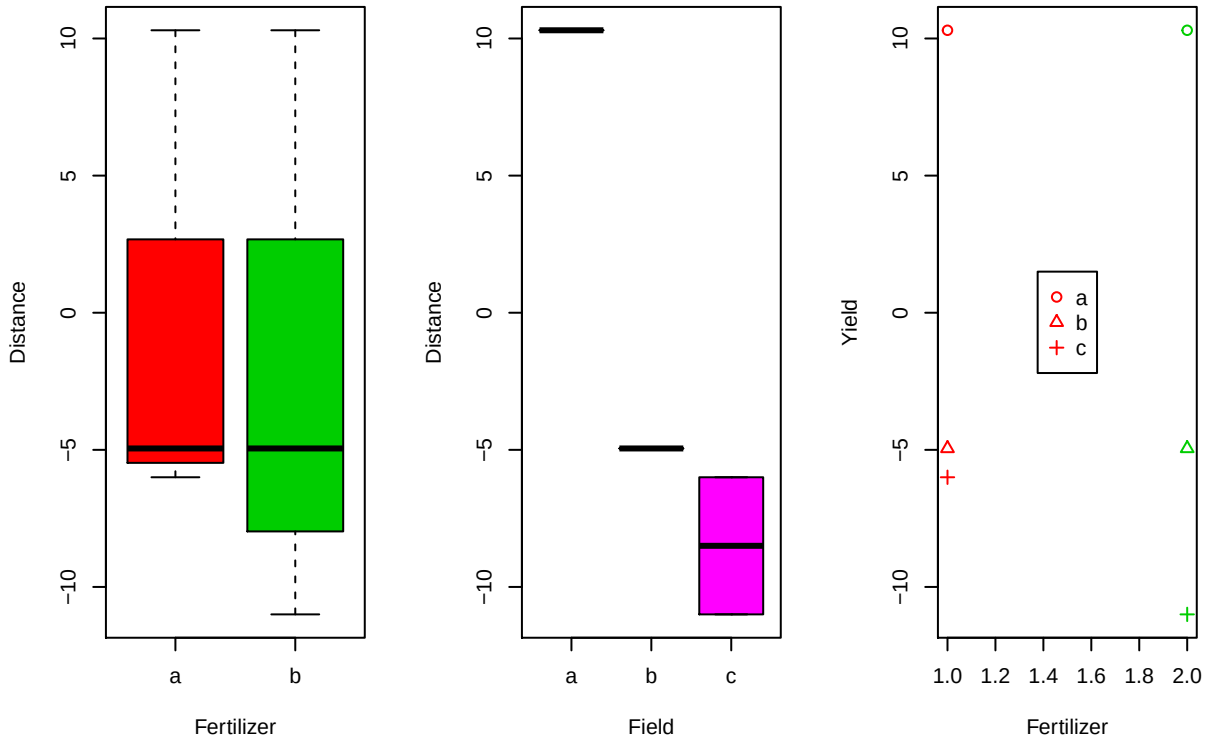
1. Type the data into R so that appropriate analysis can be carried out

```
df <- data.frame(Fertilizer=c(rep("a",3),rep("b",3)), Field=rep(c("a","b","c"),2), Yield=c(10.3,-4.95, 
fertilizer <- df$Fertilizer
field <- df$Field
yield <- df$Yield
```

2. Determine whether Fertilizer and/or Field influence on the yield

Start by making a one-way anova

```
# Plot of the data
par(mfrow = c(1,3))
boxplot(yield~fertilizer, df, xlab = "Fertilizer", ylab = "Distance", col = c(2,3))
boxplot(yield~field, df, xlab = "Field", ylab = "Distance", col=c(4,5,6))
plot(as.numeric(fertilizer), yield, col = as.numeric(fertilizer)+1,
     pch = as.numeric(field), xlab = "Fertilizer", ylab = "Yield")
legend("center", legend = c("a", "b", "c"),
       col = c(2,2,2,3,3,3), pch = c(1,2,3,1,2,3))
```



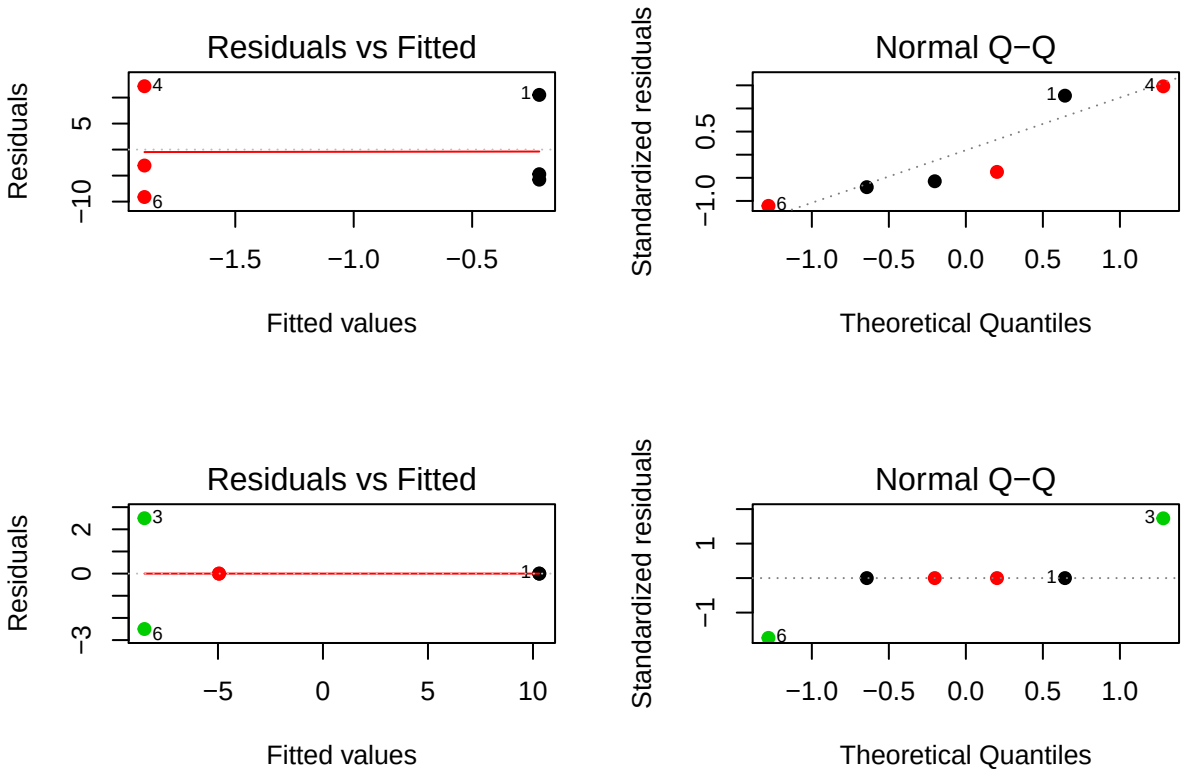

```
# Calculate one-way ANOVA for fertilizer
lm1a <- lm(yield~fertilizer, df)
anova(lm1a)
```

```
## Analysis of Variance Table
##
## Response: yield
##            Df Sum Sq Mean Sq F value Pr(>F)
## fertilizer  1   4.17   4.167  0.0409 0.8496
## Residuals   4 407.40 101.851
```

```
# Calculate one-way ANOVA for field
lm1b <- lm(yield~field, df)
anova(lm1b)
```

```
## Analysis of Variance Table
##
## Response: yield
##           Df Sum Sq Mean Sq F value   Pr(>F)
## field      2 399.07 199.535  47.888 0.005293 **
## Residuals  3  12.50   4.167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Checking model assumptions
par(mfrow = c(2,2))
plot(lm1a, col = fertilizer, which = 1:2, pch=19)
plot(lm1b, col = field, which = 1:2, pch=19)
```




```r
# Two-way Anova
# lm2 <- lm(yield~fertilizer+field, df)
# anova(lm2)
# par(mfrow = c(1,2))
# plot(lm2, col = df$field, which = 1:2, pch=19)
```

From the One-Way ANOVA, fertilizer isn't statistically significant since the p-value is $> \alpha = 0.05$, i.e. we reject the null hypothesis H_0 that the means of the groups are not the same. Field, on the other hand, is statistically significant, i.e. we fail to reject the null hypothesis H_0, so the means of the groups are the same.

Interpret the linear models

```r
# Fertilizer Linear Model
summary(lm1a)
```

```
##
## Call:
## lm(formula = yield ~ fertilizer, data = df)
##
## Residuals:
```

```
##      1      2      3      4      5      6
## 10.517 -4.733 -5.783 12.183 -3.067 -9.117
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.2167     5.8267  -0.037    0.972
## fertilizerb  -1.6667     8.2402  -0.202    0.850
##
## Residual standard error: 10.09 on 4 degrees of freedom
## Multiple R-squared:  0.01012,    Adjusted R-squared:  -0.2373
## F-statistic: 0.04091 on 1 and 4 DF,  p-value: 0.8496
```

```
# Field Linear Model
summary(lm1b)
```

```
##
## Call:
## lm(formula = yield ~ field, data = df)
##
## Residuals:
##          1          2          3          4          5          6
##  2.442e-15  4.194e-17  2.500e+00 -2.318e-15 -8.438e-17 -2.500e+00
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   10.300      1.443   7.136  0.00567 **
## fieldb       -15.250      2.041  -7.471  0.00497 **
## fieldc       -18.800      2.041  -9.210  0.00271 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.041 on 3 degrees of freedom
## Multiple R-squared:  0.9696, Adjusted R-squared:  0.9494
## F-statistic: 47.89 on 2 and 3 DF,  p-value: 0.005293
```

From the linear models, we can conclude that both fertilizers aren't statistically significant, i.e. we reject the null hypothesis. However, for the field data, all three fields are statistically significant, being field a the one with the highest yield and field c the lower.

3. Is it possible to test for interaction effects between Fertilizer and Field?

```
lm2 <- lm(yield~fertilizer*field, df)
# One-way ANOVA with interactions between fertilizer and field
anova(lm2)
```

```
## Warning in anova.lm(lm2): ANOVA F-tests on an essentially perfect fit are
## unreliable
```

```
## Analysis of Variance Table
##
```

```
## Response: yield
##                   Df Sum Sq Mean Sq F value Pr(>F)
## fertilizer         1   4.17   4.167
## field              2 399.07 199.535
## fertilizer:field   2   8.33   4.167
## Residuals          0   0.00
```

```
# Fertilizer and Field interactions linear model
summary(lm2)
```

```
##
## Call:
## lm(formula = yield ~ fertilizer * field, data = df)
##
## Residuals:
## ALL 6 residuals are 0: no residual degrees of freedom!
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          1.030e+01         NA      NA       NA
## fertilizerb         -1.737e-15         NA      NA       NA
## fieldb              -1.525e+01         NA      NA       NA
## fieldc              -1.630e+01         NA      NA       NA
## fertilizerb:fieldb  -1.154e-15         NA      NA       NA
## fertilizerb:fieldc  -5.000e+00         NA      NA       NA
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:    NaN
## F-statistic:   NaN on 5 and 0 DF,  p-value: NA
```

No, since we have one observation by factor combination. However, we can test for combinations of fields.

```
# Combining fields into "ab" and "c"
df$Field2 <- df$Field
levels(df$Field2) <- rep(c(rep("ab",2), "c"),2)
field2 <- df$Field2
lm3 <- lm(yield~fertilizer*field2, df)
# One-way ANOVA of fertilizer and field interactions
anova(lm3)
```

```
## Analysis of Variance Table
##
## Response: yield
##                    Df  Sum Sq Mean Sq F value Pr(>F)
## fertilizer          1   4.167   4.167  0.0358 0.8673
## field2              1 166.508 166.508  1.4319 0.3541
## fertilizer:field2   1   8.333   8.333  0.0717 0.8140
## Residuals           2 232.563 116.281
```

```
# Fertilizer and Field interactions linear model
summary(lm3)
```

```
##
## Call:
## lm(formula = yield ~ fertilizer * field2, data = df)
##
## Residuals:
##          1          2          3          4          5          6
##  7.625e+00 -7.625e+00 -4.441e-16  7.625e+00 -7.625e+00  8.882e-16
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.675e+00  7.625e+00   0.351    0.759
## fertilizerb          -3.932e-16  1.078e+01   0.000    1.000
## field2c              -8.675e+00  1.321e+01  -0.657    0.579
## fertilizerb:field2c -5.000e+00  1.868e+01  -0.268    0.814
##
## Residual standard error: 10.78 on 2 degrees of freedom
## Multiple R-squared:  0.4349, Adjusted R-squared:  -0.4127
## F-statistic: 0.5131 on 3 and 2 DF,  p-value: 0.7132
```

Interactions aren't statistically significant.