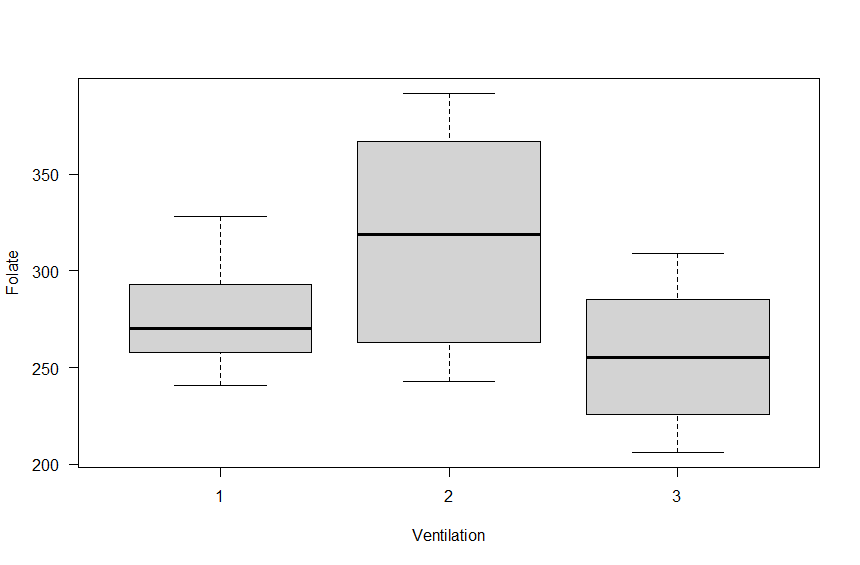
Question 1



It will be assumed that ventilation method 1 and 3 are quite similar due to the amount of overlap between them. There is a noticeable amount of overlap between the three methods, the second method has a wider variance in the folate level of patients

Estimate Std. Error 2.5 % 97.5 %

(Intercept) 278.00000 20.44662 235.20474 320.79526

ventilation2 38.62500 26.06443 -15.92847 93.17847

ventilation3 -21.55556 25.50141 -74.93063 31.81952

From the above confidence interval output for the coefficients of the predictors, it is noticed that method 1 is used as baseline; the mean red cell folate level in patients receiving method 1 ventilation during anaesthesia is 278 and the 95% interval is 235.2 to 320.8

* The mean red cell folate level if using ventilation method 2 is higher than method 1 by 38.63
* But not significantly higher as its 95% interval of -15.93 to 93.82 includes 0
* The mean red cell folate level of using ventilation method 3 is less than method 1 by 21.56 and its 95% interval of -74.93 to 31.82 includes 0; therefore, it is not significantly lower

However, we cannot explicitly compare between ventilation method 2 and 3, therefore, this estimate can be done by using Contrast method to compare the mean of using method 2 to method 3 and find out if there is a difference between both.

* Compare mean of red cell folate level of method 2 and method 3

H0: µ2 = µ3

µ2 - µ3 = 0

0\*µ1 + 1\*µ2 - 1\*µ3 = 0

Coefficients of the contrast are then (0, 1, -1)

* Compare mean of red cell folate level of method 1 to method 2 and 3

H0: µ1 = µ2 + µ3

µ1 - µ2 - µ3 = 0

1\*µ1 - 1\*µ2 - 1\*µ3 = 0

Coefficient of contrast is (1, -1, -1); therefore, the orthogonal set refers to a contrast of µ1 against µ2 and µ3

C1 = (0, 1, -1), C2 = (1, -1, -1)

**To test for orthogonality:**

* **C1 \* C2**

(0, 1, -1) \* (1, -1, -1)

**(0\*1) + (1\*1) + (-1\*1) = 0; therefore, the contrast is orthogonal**

Estimate Std. Error 2.5 % 97.5 %

(Intercept) 282.267361 11.63455 257.915964 306.61876

m2\_v\_m31 30.090278 11.10797 6.841029 53.33953

orth\_set1 -4.267361 11.63455 -28.618758 20.08404

The above output is the 95% interval output of a new model created using the 2 contrast and the output can be interpreted as:

* The intercept is the average across the 3 groups and is estimated as 282.27
* Coefficient of m2\_v\_m3 is 30.09 implies that the differences between mean red cell folate level of patients receiving method 1 to method 2.
* Also, their difference is significantly higher than the ground mean (intercept) because the 95% confidence interval (6.84 to 43.34) does not include 0

Analysis of Variance Table

Response: folate

Df Sum Sq Mean Sq F value Pr(>F)

m2\_v\_m3 1 15235 15234.6 7.2881 0.0142 \*

orth\_set 1 281 281.2 0.1345 0.7178

Residuals 19 39716 2090.3

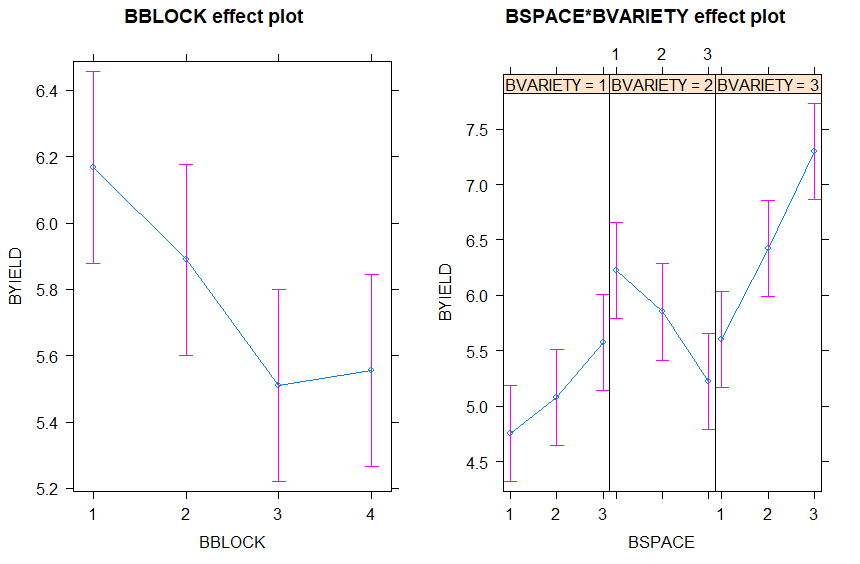
---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

From the ANOVA output, it will be seen that with a p value of 0.0142 which is less than 0.05, the difference between method 2 and 3 is significant (m2\_v\_m3)

The orth\_set is the orthogonal set and with a p value of 0.7178, the set is not significant. No significant difference between the mean of method 1 and the average of the remaining 2 methods (2 and 3)

Question 2



From the above exploratory plot,

* It is seen from the BBLOCK effect plot on the left that the mean BYIELD for **block 1** is more when compared to the other 3 blocks and the lowest BBYIELD is given by **block 2**.
* The BSPACE\*BVARIETY effect plot on the right shows the effect of the 3 variety on the yield when different row spacing is applied.
  + For Variety 1, it would be seen that the BYIELD increases as the BSPACE increases. There is some form of significant overlap in the amount of yield gotten from different row spaces.
  + For Variety 2, the yield reduces as the as BSPACE increases; also, a significant amount of overlap between row spaced can be assumed to give similar yield
  + Variety 3 gives the highest BYIELD in all row spaces compared to other Variety types; it is seen that BYIELD in the 3 different row spaces give high yield that can be assumed as significantly different from each other because there is little or no overlap in the yield for different spaces.

Call:

lm(formula = BYIELD ~ BBLOCK + BSPACE \* BVARIETY, data = barley)

Residuals:

Min 1Q Median 3Q Max

-0.6111 -0.2535 -0.0625 0.2938 0.6694

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.1361 0.2427 21.162 < 2e-16 \*\*\*

BBLOCK2 -0.2778 0.1982 -1.402 0.173792

BBLOCK3 -0.6556 0.1982 -3.308 0.002953 \*\*

BBLOCK4 -0.6111 0.1982 -3.084 0.005081 \*\*

BSPACE2 0.3250 0.2972 1.093 0.285088

BSPACE3 0.8250 0.2972 2.775 0.010511 \*

BVARIETY2 1.4750 0.2972 4.962 4.58e-05 \*\*\*

BVARIETY3 0.8500 0.2972 2.860 0.008642 \*\*

BSPACE2:BVARIETY2 -0.7000 0.4204 -1.665 0.108877

BSPACE3:BVARIETY2 -1.8250 0.4204 -4.341 0.000222 \*\*\*

BSPACE2:BVARIETY3 0.5000 0.4204 1.189 0.245909

BSPACE3:BVARIETY3 0.8750 0.4204 2.081 0.048227 \*

---

Signif. codes: 0 â€˜\*\*\*â€™ 0.001 â€˜\*\*â€™ 0.01 â€˜\*â€™ 0.05 â€˜.â€™ 0.1 â€˜ â€™ 1

Residual standard error: 0.4204 on 24 degrees of freedom

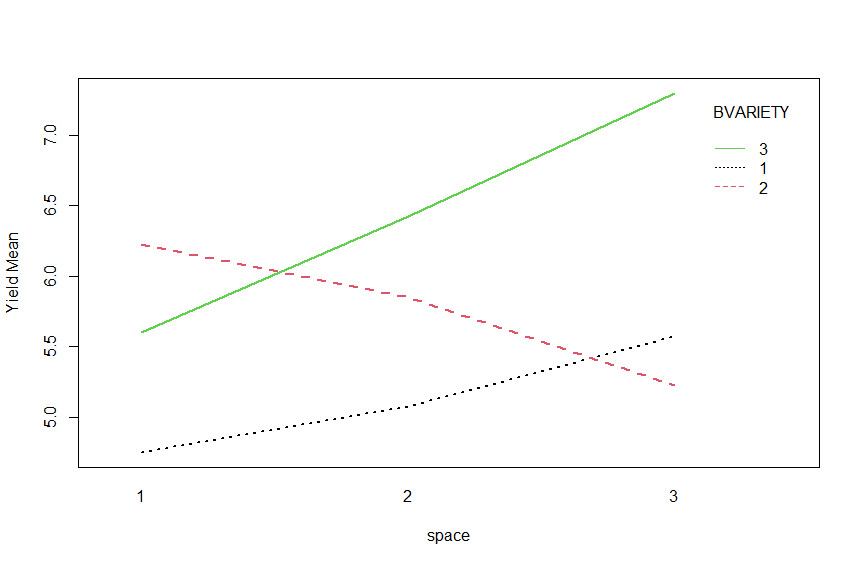
Multiple R-squared: 0.8386, Adjusted R-squared: 0.7646

F-statistic: 11.34 on 11 and 24 DF, p-value: 5.173e-07

The above output is the fitted model taking into consideration the block, space, variety types and the interaction between the space and variety types. The output shows that:

* BVariety 2 and 3 have an estimated yield that is significantly higher than variety 1 as their p-values are both greater than 0.05
* Both Bspace 2 and 3 have highere estimated yield than Bspace 1 but only Bspace 3 produces a significantly yield returns which is higher than Bspace 1.
* Block 1 has a higher estimated yield when compared to the remaining block. Block 3 and 4 are significantly lower than block 1 in estimated yield. Although, the estimated yield in block 2 is lower than block 1, but the difference is not significant
* About 76.46% of the variability is explained by the model.

The model equation is given by:



The above diagram shows the mean yield of different variety types at different row spaces:

For row space 1, it would be seen that Variety 2 produces the highest mean yield and Variety 1 had the lowest yield. While in row space 2, the mean yield for variety 2 decreases and variety produces more yield, while variety 1 slowly increases. Lastly, in row space 3, variety 2 has the lowest yield as variety 1 produces more yield than 2; but the most yield is gotten from variety 3 which continues to increase rapidly.

Therefore, it can be concluded that variety 2 is best suited for row space of 1 when trying to get the maximum yield, but variety 3 is most preferable in row space 2 and 3 as it produces the most yield

Analysis of Variance Table

Response: BYIELD

Df Sum Sq Mean Sq F value Pr(>F)

BBLOCK 3 2.5564 0.8521 4.8221 0.00912 \*\*

BSPACE 2 1.5506 0.7753 4.3872 0.02377 \*

BVARIETY 2 10.2739 5.1369 29.0694 3.872e-07 \*\*\*

BSPACE:BVARIETY 4 7.6544 1.9136 10.8289 3.679e-05 \*\*\*

Residuals 24 4.2411 0.1767

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

From the ANOVA output above, it will be seen that BBLOCK is significant to the model, also, row space (BSPACE), variety (BVARIETY) and the interaction of the space and variety is significant to the model as it explains a significant amount of variance with p-values less than 0.05.

H0: Residuals are normally distributed

Shapiro-Wilk normality test

data: lmod$residuals

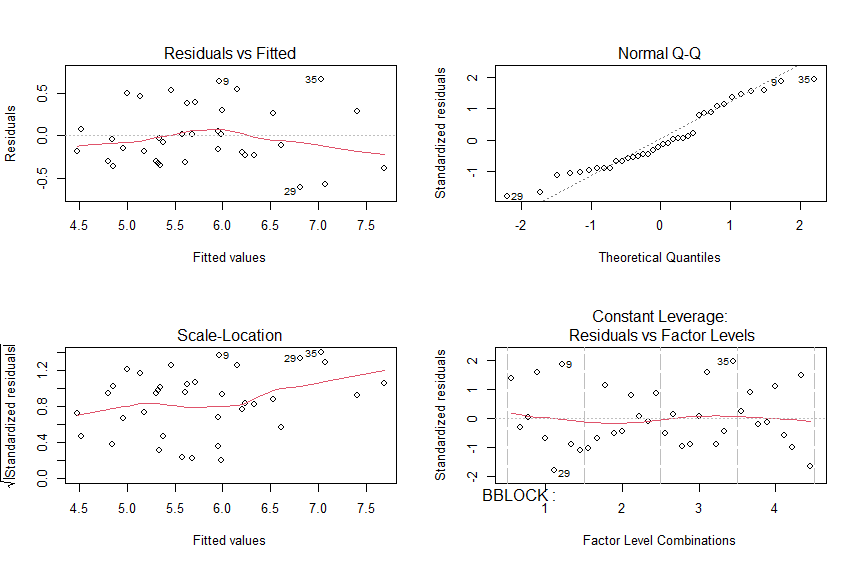
W = 0.94702, p-value = 0.08429

The above output tests the residuals to confirm that it is normally distributed; with a p value > 0.05, the null hypothesis is accepted, therefore, the residuals are normally distributed.

mean(lmod$residuals)

[1] -5.393906e-18

The above result confirms the assumption that the mean of the residuals is 0



From the assumption plot above, it can be seen that the Residuals plot shows that the residuals have a constant variance around 0. The residual is also assumed to be normally distributed as seen in Normal Q-Q plot. The residual vs Factor level shows a random pattern of residuals across the 4 blocks.

**R CODE**

setwd("C:/Users/olley/Downloads/Documents/statistics/410 ANOVA Factorial")

#install.packages("effects")

library(dae)

library(effects)

barley = read.table("Barley.txt", header=T)

str(barley)

barley$BBLOCK = as.factor(barley$BBLOCK)

barley$BSPACE = as.factor(barley$BSPACE)

barley$BVARIETY = as.factor(barley$BVARIETY)

str(barley)

# Linear model

lmod <- lm(BYIELD~BBLOCK+BSPACE\*BVARIETY, data = barley)

plot(allEffects(lmod))

summary(lmod)

anova(lmod)

interaction.plot(barley$BSPACE,barley$BVARIETY,barley$BYIELD,

col= c(1,2,3),lwd=2,xlab="space",ylab="Yield Mean",

trace.label = "BVARIETY")

?interaction.plot

print(allEffects(lmod))

par(mfrow=c(2,2))

plot(lmod)

par(mfrow=c(1,1))

boxplot(lmod$residuals)

shapiro.test(lmod$residuals)

mean(lmod$residuals)

setwd("C:/Users/olley/Downloads/Documents/statistics")

folate = read.table("folate.txt", header=T)

str(folate)

folate$ventilation <- as.factor(folate$ventilation)

plot(folate$folate~folate$ventilation,las=1,xlab="Ventilation",ylab="Folate")

df <- as.data.frame(cbind(method=c(1,2,3), mean = round(tapply(folate$folate,folate$ventilation,mean),2)))

df

model <- lm(folate~ventilation,data = folate)

summary(model)

source("Rfunctions.R")

# Interpret Coefficient

betaCI(model)

m2\_v\_m3 <- C(folate$ventilation,c(0,1,-1),1)

orth\_set <- C(folate$ventilation,c(1,-1,-1),1)

new\_model <- lm(folate~m2\_v\_m3+orth\_set,data = folate)

betaCI(new\_model)

anova(new\_model)