End Trimester Exam

Nzambuli Daniel

2024-08-03

Question 1

Imagine you are conducting a factorial experiment to study the effects of diet, exercise, and supplementation on health outcomes. The factors and their levels are as follows:

Factor A: Diet (3 levels: Low, Medium, High) - Fixed Effect

Factor B: Exercise (2 levels: None, Daily) - Fixed Effect

Factor C: Supplement (2 levels: With, Without) - Fixed Effect

1. Calculate the total number of possible treatment combinations for the given levels of each factor

```
library(combinat)
```

```
##
## Attaching package: 'combinat'
```

```
## The following object is masked from 'package:utils':
##
##     combn
```

```
treatment_comb = combn(c("Diet_Low", "Diet_Med", "Diet_High",
                         "Exercise_None", "Exercise_Daily",
                         "Supp_with", "Supp_WO"), 3)
```

```
check_combn = function(comb){
  if ("Diet_Low" %in% comb && "Diet_Med" %in% comb)
    return(F)
  if ("Diet_Low" %in% comb && "Diet_High" %in% comb)
    return(FALSE)
  if ("Diet_Med" %in% comb && "Diet_High" %in% comb)
    return(FALSE)
  if ("Exercise_None" %in% comb && "Exercise_Daily" %in% comb)
    return(FALSE)
  if ("Supp_with" %in% comb && "Supp_WO" %in% comb)
    return(FALSE)
  return(T)
}
```

```
treatment_comb = treatment_comb[, apply(treatment_comb, 2, check_combn)]
treatment_comb
```

```
##      [,1]            [,2]            [,3]             [,4]
## [1,] "Diet_Low"      "Diet_Low"      "Diet_Low"       "Diet_Low"
## [2,] "Exercise_None" "Exercise_None" "Exercise_Daily" "Exercise_Daily"
```

```
## [3,] "Supp_with"     "Supp_WO"       "Supp_with"      "Supp_WO"
##      [,5]            [,6]            [,7]             [,8]
## [1,] "Diet_Med"      "Diet_Med"      "Diet_Med"       "Diet_Med"
## [2,] "Exercise_None" "Exercise_None" "Exercise_Daily" "Exercise_Daily"
## [3,] "Supp_with"     "Supp_WO"       "Supp_with"      "Supp_WO"
##      [,9]            [,10]           [,11]            [,12]
## [1,] "Diet_High"     "Diet_High"     "Diet_High"      "Diet_High"
## [2,] "Exercise_None" "Exercise_None" "Exercise_Daily" "Exercise_Daily"
## [3,] "Supp_with"     "Supp_WO"       "Supp_with"      "Supp_WO"
```

Answer

> There are `12 treatment combinations` based on the information given

2. Create a full factorial design matrix that lists all possible combinations of the levels of Factors A, B, and C. Each row should represent a unique treatment combination.

```
full_design = data.frame(t(treatment_comb))

# rename columns
colnames(full_design) = c("Diet", "Exercise", "Supplement")

# remove the preceeding labels
full_design$Diet = gsub("Diet_", "", full_design$Diet)
full_design$Exercise = gsub("Exercise_", "", full_design$Exercise)
full_design$Supplement = gsub("Supp_", "", full_design$Supplement)
full_design
```

```
##    Diet Exercise Supplement
## 1   Low     None       with
## 2   Low     None         WO
## 3   Low    Daily       with
## 4   Low    Daily         WO
## 5   Med     None       with
## 6   Med     None         WO
## 7   Med    Daily       with
## 8   Med    Daily         WO
## 9  High     None       with
## 10 High     None         WO
## 11 High    Daily       with
## 12 High    Daily         WO
```

3. Assume you have collected data on a response variable (e.g., a health outcome) for each treatment combination. Perform a factorial ANOVA to assess the main effects and interaction effects of Factors A, B, and C on the response variable.

The following are the health outcomes that can be observed in a study using the specified factors above (Mayo Clinic, 2024)

1. **Blood Pressure**:
    - **Diet**: A high-quality diet (High) can lead to lower blood pressure compared to a low-quality diet (Low).
    - **Exercise**: Daily exercise can help reduce blood pressure.
    - **Supplement**: Certain supplements (e.g., omega-3 fatty acids) can also contribute to lower blood pressure.
2. **Body Weight/BMI**:
    - **Diet**: A high-quality diet can help maintain or reduce body weight.
    - **Exercise**: Daily exercise is effective in weight management.
    - **Supplement**: Some supplements (e.g., protein supplements) can aid in weight management.
3. **Cholesterol Levels**:
    - **Diet**: A high-quality diet can lower LDL (bad) cholesterol and increase HDL (good) cholesterol.
    - **Exercise**: Regular exercise can improve cholesterol levels.
    - **Supplement**: Supplements like fish oil can positively affect cholesterol levels.
4. **Blood Sugar Levels**:
    - **Diet**: A high-quality diet can help regulate blood sugar levels.
    - **Exercise**: Daily exercise can improve insulin sensitivity and lower blood sugar levels.
    - **Supplement**: Some supplements (e.g., chromium) can help manage blood sugar levels.
5. **Mental Health**:
    - **Diet**: A high-quality diet can improve mood and cognitive function.
    - **Exercise**: Regular exercise is known to reduce symptoms of depression and anxiety.
    - **Supplement**: Certain supplements (e.g., omega-3 fatty acids) can support mental health.
6. **Inflammation**:
    - **Diet**: A high-quality diet can reduce inflammation.
    - **Exercise**: Regular exercise can lower inflammatory markers.
    - **Supplement**: Anti-inflammatory supplements (e.g., turmeric) can help reduce inflammation.

The unit of interest is blood pressure (Mayo Clinic, 2024)

- the diastolic bp is normal at 120 mmHg drops by 4-12 mmHg
- the systolic bp is normal at 80 mmHg drops by 3-6 mmHg

My experiment uses the diastolic bp

Simulate the data

```
reps = 12
replicates = as.factor(rep(seq(1, 12), each = reps)) #repeat the collection of data 6 times
treatments = as.factor(rep(seq(1, 12), reps))
print("replicates")
```

```
## [1] "replicates"
```

```
replicates
```

```
##   [1] 1  1  1  1  1  1  1  1  1  1  1  1  2  2  2  2  2  2  2  2  2  2  2  2  3
##  [26] 3  3  3  3  3  3  3  3  3  3  3  4  4  4  4  4  4  4  4  4  4  4  4  5  5
##  [51] 5  5  5  5  5  5  5  5  5  5  6  6  6  6  6  6  6  6  6  6  6  6  7  7  7
##  [76] 7  7  7  7  7  7  7  7  7  8  8  8  8  8  8  8  8  8  8  8  8  9  9  9  9
## [101] 9  9  9  9  9  9  9  9  10 10 10 10 10 10 10 10 10 10 10 10 11 11 11 11 11
## [126] 11 11 11 11 11 11 11 12 12 12 12 12 12 12 12 12 12 12 12
## Levels: 1 2 3 4 5 6 7 8 9 10 11 12
```

```
print("treatments")
```

```
## [1] "treatments"
```

```
treatments
```

```
##   [1] 1  2  3  4  5  6  7  8  9  10 11 12 1  2  3  4  5  6  7  8  9  10 11 12 1
##  [26] 2  3  4  5  6  7  8  9  10 11 12 1  2  3  4  5  6  7  8  9  10 11 12 1  2
##  [51] 3  4  5  6  7  8  9  10 11 12 1  2  3  4  5  6  7  8  9  10 11 12 1  2  3
##  [76] 4  5  6  7  8  9  10 11 12 1  2  3  4  5  6  7  8  9  10 11 12 1  2  3  4
## [101] 5  6  7  8  9  10 11 12 1  2  3  4  5  6  7  8  9  10 11 12 1  2  3  4  5
## [126] 6  7  8  9  10 11 12 1  2  3  4  5  6  7  8  9  10 11 12
## Levels: 1 2 3 4 5 6 7 8 9 10 11 12
```

There are `72 values`

```
diet_levels = factor(rep(rep(c("Low", "Med", "High"), each=4), reps))
exercise_levels = factor(rep(rep(c("None", "Daily"), each = 2, times=3), reps))
supplement_levels = factor(rep(rep(c("with", "WO"), each=1, times=6), reps))



set.seed(2222)
bp_val = 120 - rnorm(12 * reps, mean = mean(seq(4, 12)), sd = sd(seq(4, 12)))

bp_data = data.frame(
  Diet = diet_levels,
  Exercise = exercise_levels,
  Supplement = supplement_levels,
  Treatment = treatments,
  Replicate = replicates,
  BP = bp_val
)

head(bp_data)
```

```
##    Diet Exercise Supplement Treatment Replicate       BP
## 1   Low     None       with         1         1 112.9258
## 2   Low     None         WO         2         1 109.4280
## 3   Low    Daily       with         3         1 107.2411
## 4   Low    Daily         WO         4         1 110.0930
## 5   Med     None       with         5         1 110.7340
## 6   Med     None         WO         6         1 112.8629
```

Confirm the data is of equal size

```
confirm_data = function(colA, colB, colRep, data){
  return(table(data[[colA]], data[[colB]], data[[colRep]]))
}
```

```
confirm_data("Diet","Exercise", "Replicate", bp_data)
```

```
## , ,  = 1
##
##
##         Daily None
##   High      2    2
##   Low       2    2
##   Med       2    2
##
## , ,  = 2
##
##
##         Daily None
##   High      2    2
##   Low       2    2
##   Med       2    2
##
## , ,  = 3
##
##
##         Daily None
##   High      2    2
##   Low       2    2
##   Med       2    2
##
## , ,  = 4
##
##
##         Daily None
##   High      2    2
##   Low       2    2
##   Med       2    2
##
## , ,  = 5
##
##
##         Daily None
##   High      2    2
##   Low       2    2
##   Med       2    2
##
## , ,  = 6
```

```
##
##
##         Daily None
##    High     2    2
##    Low      2    2
##    Med      2    2
##
## , ,  = 7
##
##
##         Daily None
##    High     2    2
##    Low      2    2
##    Med      2    2
##
## , ,  = 8
##
##
##         Daily None
##    High     2    2
##    Low      2    2
##    Med      2    2
##
## , ,  = 9
##
##
##         Daily None
##    High     2    2
##    Low      2    2
##    Med      2    2
##
## , ,  = 10
##
##
##         Daily None
##    High     2    2
##    Low      2    2
##    Med      2    2
##
## , ,  = 11
##
##
##         Daily None
##    High     2    2
##    Low      2    2
##    Med      2    2
##
## , ,  = 12
##
##
##         Daily None
##    High     2    2
##    Low      2    2
##    Med      2    2
```

```
confirm_data("Diet","Supplement", "Replicate", bp_data)
```

```
## , ,  = 1
##
```

```
##
##         with WO
##    High    2  2
##    Low     2  2
##    Med     2  2
##
## , ,  = 2
##
##
##         with WO
##    High    2  2
##    Low     2  2
##    Med     2  2
##
## , ,  = 3
##
##
##         with WO
##    High    2  2
##    Low     2  2
##    Med     2  2
##
## , ,  = 4
##
##
##         with WO
##    High    2  2
##    Low     2  2
##    Med     2  2
##
## , ,  = 5
##
##
##         with WO
##    High    2  2
##    Low     2  2
##    Med     2  2
##
## , ,  = 6
##
##
##         with WO
##    High    2  2
##    Low     2  2
##    Med     2  2
##
## , ,  = 7
##
##
##         with WO
##    High    2  2
##    Low     2  2
##    Med     2  2
##
## , ,  = 8
##
##
##         with WO
##    High    2  2
##    Low     2  2
```

```
##   Med     2  2
##
## , ,  = 9
##
##
##         with WO
##   High    2  2
##   Low     2  2
##   Med     2  2
##
## , ,  = 10
##
##
##         with WO
##   High    2  2
##   Low     2  2
##   Med     2  2
##
## , ,  = 11
##
##
##         with WO
##   High    2  2
##   Low     2  2
##   Med     2  2
##
## , ,  = 12
##
##
##         with WO
##   High    2  2
##   Low     2  2
##   Med     2  2
```

```
confirm_data("Supplement","Exercise", "Replicate", bp_data)
```

```
## , ,  = 1
##
##
##         Daily None
##   with      3    3
##   WO        3    3
##
## , ,  = 2
##
##
##         Daily None
##   with      3    3
##   WO        3    3
##
## , ,  = 3
##
##
##         Daily None
##   with      3    3
##   WO        3    3
##
## , ,  = 4
##
```

```
##
##        Daily None
##   with     3    3
##   WO       3    3
##
## , ,  = 5
##
##
##        Daily None
##   with     3    3
##   WO       3    3
##
## , ,  = 6
##
##
##        Daily None
##   with     3    3
##   WO       3    3
##
## , ,  = 7
##
##
##        Daily None
##   with     3    3
##   WO       3    3
##
## , ,  = 8
##
##
##        Daily None
##   with     3    3
##   WO       3    3
##
## , ,  = 9
##
##
##        Daily None
##   with     3    3
##   WO       3    3
##
## , ,  = 10
##
##
##        Daily None
##   with     3    3
##   WO       3    3
##
## , ,  = 11
##
##
##        Daily None
##   with     3    3
##   WO       3    3
##
## , ,  = 12
##
##
##        Daily None
```

```
##   with      3    3
##   WO        3    3
```

Perform ANOVA

```
q1_anova =  aov(BP ~ Diet * Exercise * Supplement, data = bp_data)
```

4. Provide the ANOVA table summarizing the results, including the sum of squares, degrees of freedom, mean squares, F-values, and p-values for each main effect and interaction.

```
summary(q1_anova)
```

```
##                          Df Sum Sq Mean Sq F value Pr(>F)
## Diet                      2    9.7   4.845   0.634  0.532
## Exercise                  1   12.9  12.915   1.689  0.196
## Supplement                1    3.1   3.084   0.403  0.526
## Diet:Exercise             2    5.8   2.913   0.381  0.684
## Diet:Supplement           2    3.1   1.550   0.203  0.817
## Exercise:Supplement       1    0.4   0.368   0.048  0.827
## Diet:Exercise:Supplement  2   23.1  11.527   1.508  0.225
## Residuals               132 1009.2   7.645
```

> H_0 there is no statistically significant difference in the mean blood pressure because of exercising
>
> H_1 there is at-least one mean blood pressure that is statistically and significantly different from the other mean blood pressure from exercising
>
> **Observation**
>
> the p-values are
> $$\{0.532, 0.196, 0.526, 0.684, 0.817, 0.827, 0.225\} \ and$$
> $$\forall x \in p \ values$$
> $$x \geq 0.05$$
>
> **Conclusion**
>
> The p-values $\geq$ 0.05. Because of this I `fail to reject` H_0 conclude that there is no statistically significant difference in the mean blood pressure after starting exercise with diet and supplements.
>
> The interactions **Diet:Exercise**, **Diet:Supplements**, **Exercise:Supplement** and **Diet:Exercise:Supplement** cause no significant difference in the mean of the diastolic blood preasure.
>
> **Interpretation**
>
> **Diet, Exercise** and **Supplement** do not have an impact on the health effects of individuals.

5. Interpret the p-values to determine which factors and interactions have statistically significant effects on the response variable. Discuss the implications of these findings.

> The lack of statistically significant results suggests that neither the individual factors (diet, exercise, supplement) nor their interactions had a discernible impact on altering blood pressure within the context and constraints of your experimental design.
>
> There are **12 replicates** of **12 treatments**. This may not be enough samples to identify potential variations. Additionally, the treatments for supplements and exercise may be limited by just being classified into two groups. Increasing the variation in the treatments may help expose more variation in the diastolic blood pressure used as a measure for health outcomes.
>
> There is a distinction between clinical significance and statistical significance. Gay et al. (2016) suggest a sample of about 23000 individuals to observe variation in blood pressure. Jurik and Stastny (2019) support this by proposing dietary supplements like sodium, calcium, magnesium and potassium as potential features that can be observed. When it comes to exercise (VeryWellHealth, 2024) proposes observation of blood pressure from a level of hypertension. These are some features that can be observed to help make the statistical analysis similar to a clinical analysis.

Question 2

Consider a split-plot experiment designed to investigate the effects of soil type and watering frequency on plant growth. The experiment involves the following factors:

Main Plot Factor: Soil Type (2 levels: Sandy, Clay) - Fixed Effect

Subplot Factor: Watering Frequency (3 levels: Low, Medium, High) - Random Effect

1. Explain how the split-plot design is structured, including the distinction between main plots and subplots. Describe how the treatments are assigned to whole plots (main plots) and split-plots (subplots) within each whole plot.

split-plot designs assign the **main treatment** to **the larger plot** because they are easier and more economical to change

The **sub-plots** within these larger plots receive **the secondary treatment** .

from the above plots there are two large plots that receive two treatments `sandy` and `clay`

and within these two plots there are 3 treatments low, medium, high watering

Main Plot 1 (Sandy)	Main Plot 2 (Clay)
low watering	**low** watering
medium watering	**medium** watering
high watering	**high** watering

2. Create a design matrix that shows all treatment combinations, detailing how the levels of the main plot factor (Soil Type) and the subplot factor (Watering Frequency) are arranged.

```
q2_design = data.frame(soil = rep(c("Clay", "Sandy"), each = 3),
                       watering = rep(c("Low", "Medium", "High"), times = 2))
q2_design
```

```
##    soil watering
## 1  Clay      Low
## 2  Clay   Medium
## 3  Clay     High
## 4 Sandy      Low
## 5 Sandy   Medium
## 6 Sandy     High
```

> There are `6 combinations` of treatment designs.

3. Given data on a response variable (e.g., plant height) collected from this experiment, perform a split-plot ANOVA to evaluate the effects of Soil Type and Watering Frequency on the response variable.

> Wheat height ranges from `95.25 cm` to `100.25 cm` (El-Shafei & Mattar, 2022)
>
> **Process**
>
> 1. create `soil levels` ie. **clay, clay clay, sand, sand, sand**. This sequence is repeated 8 times resulting into `48 values`
> 2. create `water levels` ie **low, medium, high** . This sequence is repeated 16 times $16 * 3 = 48$
> 3. create `treatments` ie **1, 2 , 3, 4, 5, 6** this is because these are the treatment combinations they are repeated 8 times. $6 * 8 = 48$
> 4. create `repetitions` ie there are $48\ values$ if one repetition has 6 values. I need to repeat `1 6 times` , `2 six times` , `3 six times` … these first 3 values record 18values that mark 3 repetitions. Therefore my sequence becomes 6-ones, 6-twos, …, 6-eights.

```
plant_reps = 8

# the factors
soil_levels = as.factor(rep(c("Clay", "Sand"), each = 3, times = plant_reps))
water_levels = as.factor(rep(c("Low", "Medium", "High"), times = 2 * plant_reps))
treatments = as.factor(rep(seq(1, 6), times = plant_reps))
repetitions = as.factor(rep(seq(1, 8), each = 6))
```

```
plant_data = data.frame(
  Soil = soil_levels,
  Watering = water_levels,
  Treatment = treatments,
  Replicate = repetitions
)
plant_data
```

```
##    Soil Watering Treatment Replicate
## 1  Clay      Low         1         1
## 2  Clay   Medium         2         1
## 3  Clay     High         3         1
## 4  Sand      Low         4         1
## 5  Sand   Medium         5         1
## 6  Sand     High         6         1
## 7  Clay      Low         1         2
## 8  Clay   Medium         2         2
## 9  Clay     High         3         2
## 10 Sand      Low         4         2
## 11 Sand   Medium         5         2
## 12 Sand     High         6         2
## 13 Clay      Low         1         3
## 14 Clay   Medium         2         3
## 15 Clay     High         3         3
## 16 Sand      Low         4         3
## 17 Sand   Medium         5         3
## 18 Sand     High         6         3
## 19 Clay      Low         1         4
## 20 Clay   Medium         2         4
## 21 Clay     High         3         4
## 22 Sand      Low         4         4
## 23 Sand   Medium         5         4
## 24 Sand     High         6         4
## 25 Clay      Low         1         5
## 26 Clay   Medium         2         5
## 27 Clay     High         3         5
## 28 Sand      Low         4         5
## 29 Sand   Medium         5         5
## 30 Sand     High         6         5
## 31 Clay      Low         1         6
## 32 Clay   Medium         2         6
## 33 Clay     High         3         6
## 34 Sand      Low         4         6
## 35 Sand   Medium         5         6
## 36 Sand     High         6         6
## 37 Clay      Low         1         7
## 38 Clay   Medium         2         7
## 39 Clay     High         3         7
## 40 Sand      Low         4         7
## 41 Sand   Medium         5         7
## 42 Sand     High         6         7
## 43 Clay      Low         1         8
## 44 Clay   Medium         2         8
## 45 Clay     High         3         8
## 46 Sand      Low         4         8
## 47 Sand   Medium         5         8
## 48 Sand     High         6         8
```

Simulate plant height

```
set.seed(223)

rand_q2 = rnorm(48, mean = 0, sd = 1)

plant_data$Wheat_height = with(plant_data,
                               95.25 + (rand_q2 - min(rand_q2)) * (100.25 - 95.25)/ (max(rand_q
2) - min(rand_q2)) - as.numeric(Soil) * 10.16  + as.numeric(Watering) * 9.625) # the height inc
reases between 9.25 cm and 10 cm based on watering (El-Shafei & Mattar, 2022)
# the height reduces by between 5.08 cm  and 15.24 cm (University of Minnesota, 2018)
```

```
head(plant_data)
```

```
##   Soil Watering Treatment Replicate Wheat_height
## 1 Clay      Low         1         1    108.33467
## 2 Clay   Medium         2         1    115.46160
## 3 Clay     High         3         1     97.39783
## 4 Sand      Low         4         1     97.36965
## 5 Sand   Medium         5         1    107.63815
## 6 Sand     High         6         1     87.51968
```

Confirm the data is well written

```
confirm_data("Soil", "Watering", "Replicate", plant_data)
```

```
## , ,  = 1
##
##
##         High Low Medium
##   Clay     1   1      1
##   Sand     1   1      1
##
## , ,  = 2
##
##
##         High Low Medium
##   Clay     1   1      1
##   Sand     1   1      1
##
## , ,  = 3
##
##
##         High Low Medium
##   Clay     1   1      1
##   Sand     1   1      1
##
## , ,  = 4
##
##
##         High Low Medium
##   Clay     1   1      1
##   Sand     1   1      1
##
## , ,  = 5
##
```

```
##
##         High Low Medium
##    Clay    1   1      1
##    Sand    1   1      1
##
## , ,  = 6
##
##
##         High Low Medium
##    Clay    1   1      1
##    Sand    1   1      1
##
## , ,  = 7
##
##
##         High Low Medium
##    Clay    1   1      1
##    Sand    1   1      1
##
## , ,  = 8
##
##
##         High Low Medium
##    Clay    1   1      1
##    Sand    1   1      1
```

Perform ANOVA

```
library(nlme)
q2_anova = lme(Wheat_height ~ Soil*Watering, random = ~ 1 | Replicate/Watering, data = plant_da
ta)
q2_confirm = aov(Wheat_height ~ Soil*Watering + Watering/Replicate + Replicate, data=plant_dat
a)
q2_table = anova(q2_anova)
```

4. Provide the ANOVA table, including the sources of variation (main plot factor, subplot factor, interaction), sum of squares, degrees of freedom, mean squares, F-values, and p-values.

Using AOV to get the table

```
summary(q2_confirm)
```

```
##                     Df Sum Sq Mean Sq  F value Pr(>F)
## Soil                 1 1077.7  1077.7  815.682 <2e-16 ***
## Watering             2 2965.0  1482.5 1122.101 <2e-16 ***
## Replicate            7    7.2     1.0    0.774  0.616
## Soil:Watering        2    3.9     1.9    1.465  0.254
## Watering:Replicate 14   31.0     2.2    1.675  0.138
## Residuals           21   27.7     1.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA table generated

```
q2_table
```

```
##                  numDF denDF   F-value p-value
## (Intercept)          1    21 275546.05  <.0001
## Soil                 1    21    815.68  <.0001
## Watering             2    14    816.21  <.0001
## Soil:Watering        2    21      1.47  0.2537
```

5. Interpret the p-values to identify which factors and interactions have statistically significant effects on the response variable. Discuss any significant interactions and their practical implications

> **Based on Soil-type**
>
> H_0 There is no statistically significant difference in the mean wheat height due to the type of soil that is: clay versus sand
>
> H_1 There is at least one statistically significant different mean wheat height due to the type of soil
>
> **Based on Watering**
>
> H_0 There is no statistically significant difference in the mean wheat height due to the type of watering done that is: low, medium and high
>
> H_1 There is at least one statistically significant different mean wheat height due to the type of watering done
>
> **Based on the interaction between Soil-type and Watering**
>
> H_0 There is no statistically significant difference in the mean wheat height due to the interaction between the soil used and the type of watering done on the soil type.
>
> H_1 There is at least one statistically significant different mean wheat height due to the interaction between the soil used and the type of watering done on the soil type.
>
> **Observations**
>
> 1. there are two p-values that are lower than the significant value of `0.05` . The p-values are for **Soil** and **Watering**
> 2. The p-value of the interaction between **Soil and Watering** and the regular **Replicate** are larger than `0.05`
>
> **Interpretation**
>
> There is at least one mean that is statistically and significantly different from the other means height of the plants that are grown in different types of soil.

There is also at least one mean height in the plants that is statistically and significantly different from the height of the other plants grown under different types of irrigation

Conclusion

There needs to be a post-hoc analysis to identify which means are actually different based on the soil type and watering frequency

Post-hoc analysis

```
plant_lsd = agricolae::LSD.test(q2_confirm, c("Soil", "Watering"))$groups
plant_lsd
```

```
##               Wheat_height groups
## Clay:Medium    116.31922      a
## Sand:Medium    106.96696      b
## Clay:Low       106.72105      b
## Sand:Low        97.86966      c
## Clay:High       97.50716      c
## Sand:High       87.28115      d
```

There are `4 groups`

1. **Group 1**
   - Plants grown in **Clay soil** with **Medium watering**
2. **Group 2**
   - Plants grown in **Sand soil** with **Medium watering**
   - Plants grown in **Clay soil** with **Low watering**
3. **Group 3**
   - plants grown in **Sand soil** with **Low watering**
   - plants grown in **Clay soil** with **High watering**
4. **Group 4**
   - Plants grown in **Sand soil** with **High watering**

The groups are also rankings on how the plant height is affected.

- This means that clay soil with medium watering has the highest wheat height
- It also means that sand soil with high watering has the lowest wheat height

Conclusion

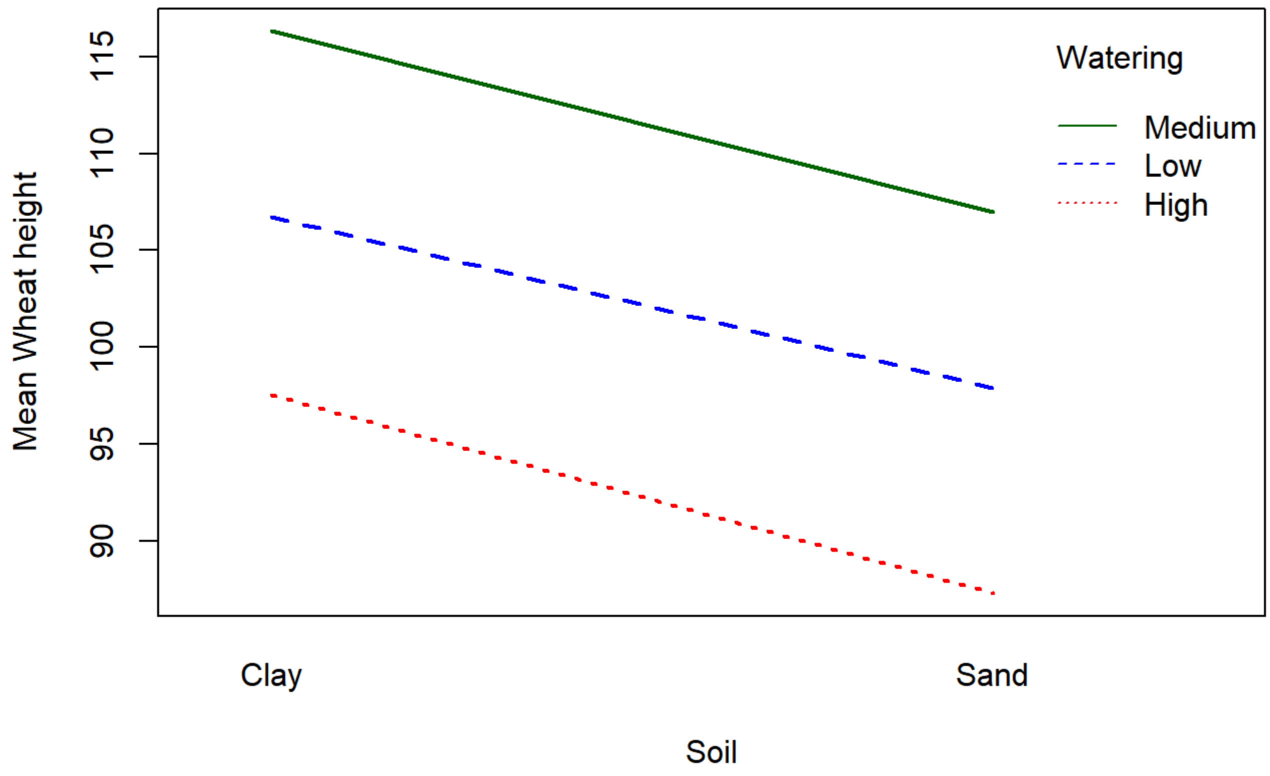
Micheal (2024) suggests that the plant height improves the yield of the plants. This is backed-up by the presence of two alleles in wheat; Rht-B1a and Rht-D1b. These alleles allow for increased wheat yield based on height in the event wheat grows in deep soil moisture (Gao et al., 2020).

From the observations clay soil with medium watering achieves the highest deep soil moisture. This allows the wheat to grow taller and may also lead to the best yields observed.

Sand soil with high watering has low deep soil moisture leading to the lowest wheat height.

Interactive plots

```
interaction.plot(
  plant_data$Soil, plant_data$Watering, plant_data$Wheat_height,
  xlab = "Soil", ylab = "Mean Wheat height", trace.label = "Watering",
  col = c("red", "blue", "darkgreen"), lwd = 2
)
```



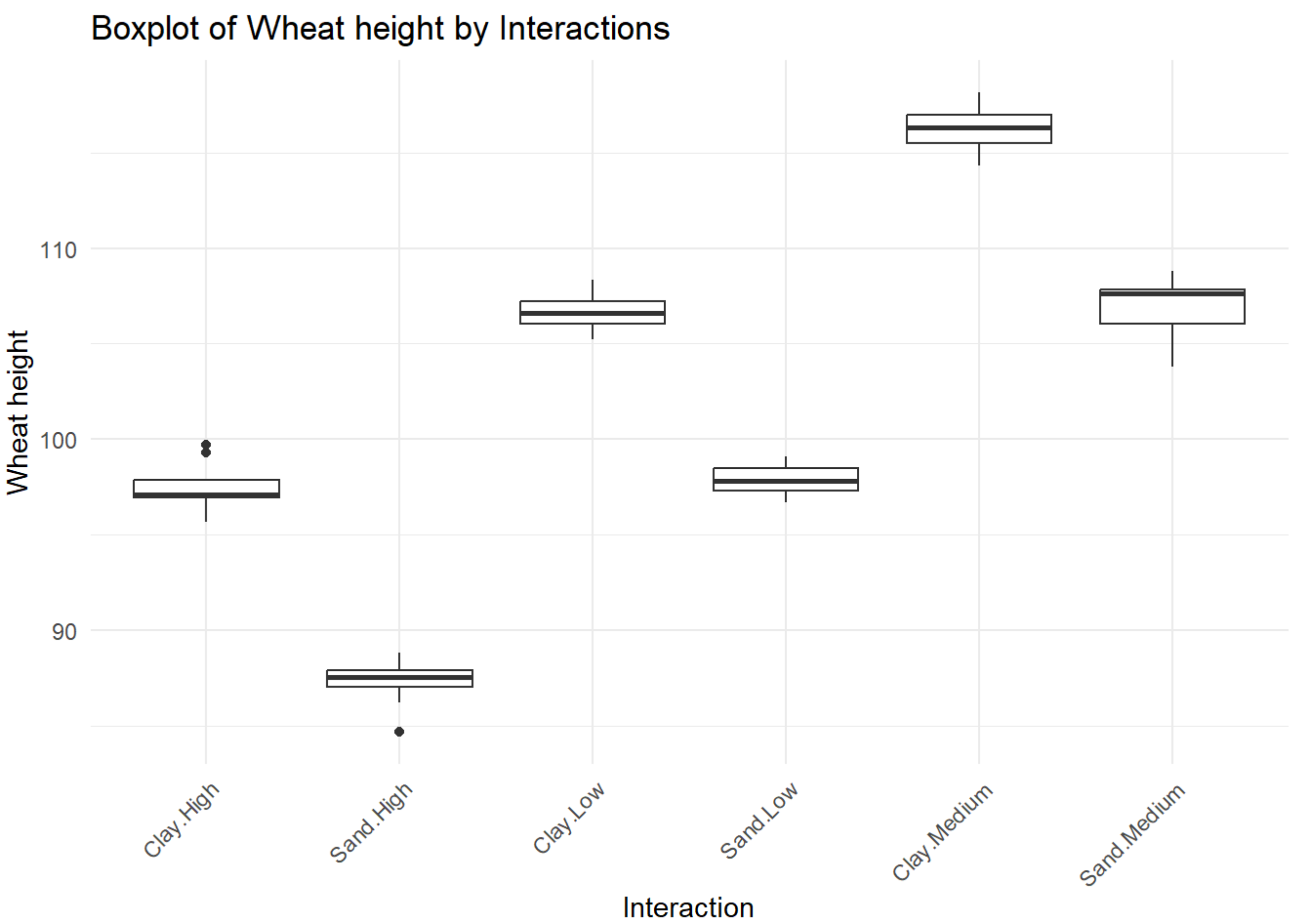
> **Observation**
>
> - The difference between the mean wheat height for clay soil is **higher** than the mean wheat height in sand soil in-spite of the watering done
> - There appears to be a slight convergence between the mean wheat height for medium and low watering when plantation moves from clay to sand soil
> - There also appears to be a slight convergence between the mean wheat height for medium and low watering when plantation moves from clay to sand soil
> - The variation between the height in clay soil is narrower than in the case of sand soil based on the type of watering done to the wheat crop.
>
> **Implication**
>
> - The impact of changing watering frequency from low to high is more noticeable in sandy soil than in clay. This could be due to the intrinsic properties of the soils affecting how water is retained and utilized by the plants

Box plot

```r
box_q2 = plant_data[, c(1, 2, 5)]
box_q2$interactions = interaction(box_q2$Soil, box_q2$Watering)
head(box_q2)
```

```
##   Soil Watering Wheat_height interactions
## 1 Clay      Low    108.33467     Clay.Low
## 2 Clay   Medium    115.46160  Clay.Medium
## 3 Clay     High     97.39783    Clay.High
## 4 Sand      Low     97.36965     Sand.Low
## 5 Sand   Medium    107.63815  Sand.Medium
## 6 Sand     High     87.51968    Sand.High
```

```r
library(ggplot2)
ggplot(box_q2, aes(x=interactions, y=Wheat_height)) +
  geom_boxplot() +
  theme_minimal() +
  labs(title="Boxplot of Wheat height by Interactions", x="Interaction", y="Wheat height")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Observation

Yield will be used to refer to wheat height

overall yield is used as a measure of the interquartile range represented by the bounds of the box plots with

- clay soil with medium watering has the highest overall yield
- clay soil with low watering and sand soil with medium watering have the second highest overall yield

- clay soil with high watering and sand soil with low watering have the third highest overall yield
- sand soil with high watering has the lowest yield

Implications

It is recommended that if a farmer is to plant wheat they can choose ground with clay soil and water the plants with medium watering frequency. For this simulation medium watering is taken as

$$\frac{80+100+120}{3} = 100\, plant\ evapotranspiration\ (El - Shafei\ \&\ Mattar, 2022)$$

this ensures that the plants acquire nutrients adequately because of deep soil water retention.

In the event the farmer is in a sandy soil area, medium watering achieves the best wheat height. with higher median yield than plants grown in clay soil with low watering.

Farmers are advised to avoid high watering in sandy soil regions as this leads to the lowest overall crop growth.

Question 3

You have yield data (in bushels per acre) from an experiment designed to investigate the effects of different fertilization treatments on corn growth. The experiment follows a Latin square design with the following treatments:

A: High Nitrogen

B: Low Nitrogen

C: High Phosphorus

D: Low Phosphorus

E: Control

The yields for each treatment are recorded as follows:

Treatment	Row 1	Row 2	Row 3	Row 4	Row 5
A	10.2	4.5	7.0	5.3	10.0
B	5.3	6.8	6.5	8.2	8.2
C	12.4	10.0	16.0	13.8	16.5
D	6.8	5.9	8.2	9.2	12.4
E	3.9	5.7	5.0	6.9	6.9

1. Evaluate the effects of the different fertilization treatments on corn yield, taking into account potential random effects due to environmental variability. Perform an appropriate statistical analysis to determine the significance of the treatment effects.

The data has no repeating treatment across the rows and columns.

This does not fit an LSD. It does however fit an RCBD. because the individual rows get a variety of treatments and the yield varies based on row and treatment not including the column effect and row effect expected in an LSD

```
yield_data = data.frame(Row = factor(rep(1:5, times = 5)),
                        Treatment =factor(rep(c("High N", "Low N", "High P", "Low P", "Contro
l"), each = 5), levels = c("High N", "Low N", "High P", "Low P", "Control")),
                        Yield = c(10.2, 4.5, 7, 5.3, 10.0,
```

```r
                                  5.3, 6.8, 6.5, 8.2, 8.2,
                                  12.4, 10.0, 16.0, 13.8, 16.5,
                                  6.8, 5.9, 8.2, 9.2, 12.4,
                                  3.9, 5.7, 5.0, 6.9, 6.9))
yield_data
```

```
##    Row Treatment Yield
## 1    1    High N  10.2
## 2    2    High N   4.5
## 3    3    High N   7.0
## 4    4    High N   5.3
## 5    5    High N  10.0
## 6    1     Low N   5.3
## 7    2     Low N   6.8
## 8    3     Low N   6.5
## 9    4     Low N   8.2
## 10   5     Low N   8.2
## 11   1    High P  12.4
## 12   2    High P  10.0
## 13   3    High P  16.0
## 14   4    High P  13.8
## 15   5    High P  16.5
## 16   1     Low P   6.8
## 17   2     Low P   5.9
## 18   3     Low P   8.2
## 19   4     Low P   9.2
## 20   5     Low P  12.4
## 21   1   Control   3.9
## 22   2   Control   5.7
## 23   3   Control   5.0
## 24   4   Control   6.9
## 25   5   Control   6.9
```

Creating the anova function

$$Overall\ mean = \frac{Grand\ Total}{N}$$
$$total\ sum\ squares = \sum(Yield - \mu)^2$$
$$tss = 5(\sum(treatment\ total - \mu)^2)$$
$$ess = total\ ss - tss$$
$$DF = nrow(x) - 1$$
$$DF_{error} = df_{total} - \sum df_{others}$$
$$MS = \frac{\sum squares}{DF}$$
$$F = \frac{ms_{treatment}}{ms_{error}}$$

```r
library(dplyr)
```

```
## 
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:nlme':
## 
```

```
##     collapse
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Perform ANOVA

```
q3_aov = aov(Yield~Row + Treatment, data = yield_data)
summary(q3_aov)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## Row          4  48.06   12.02   4.167   0.0168 *  
## Treatment    4 194.32   48.58  16.846 1.37e-05 ***
## Residuals   16  46.14    2.88                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

> **Observation**
>
> Treatment and Rows have p-values that are greater than `0.05`
>
> **Interpretation**
>
> H_0 There is no statistically and significantly different from the mean Brussels yield because of row and treatment
>
> H_1 There is at least one mean that is statistically and significantly different from the mean Brussels yield of the other wheat produced in this experiment
>
> because of the p-value being less than 0.05 we reject H_0 and conclude that there is at least one mean yield that is statistically and significantly different from the other mean yields

2. If necessary, conduct post-hoc tests to determine which specific treatments significantly differ from each other in terms of yield. Include a visual representation (e.g., a bar plot or box plot) comparing the treatments.

```
yield_lsd = agricolae::LSD.test(q3_aov, c("Row", "Treatment"))$groups
yield_lsd
```

```
##            Yield groups
## 5:High P    16.5      a
## 3:High P    16.0      a
## 4:High P    13.8     ab
## 1:High P    12.4    abc
## 5:Low P     12.4    abc
```

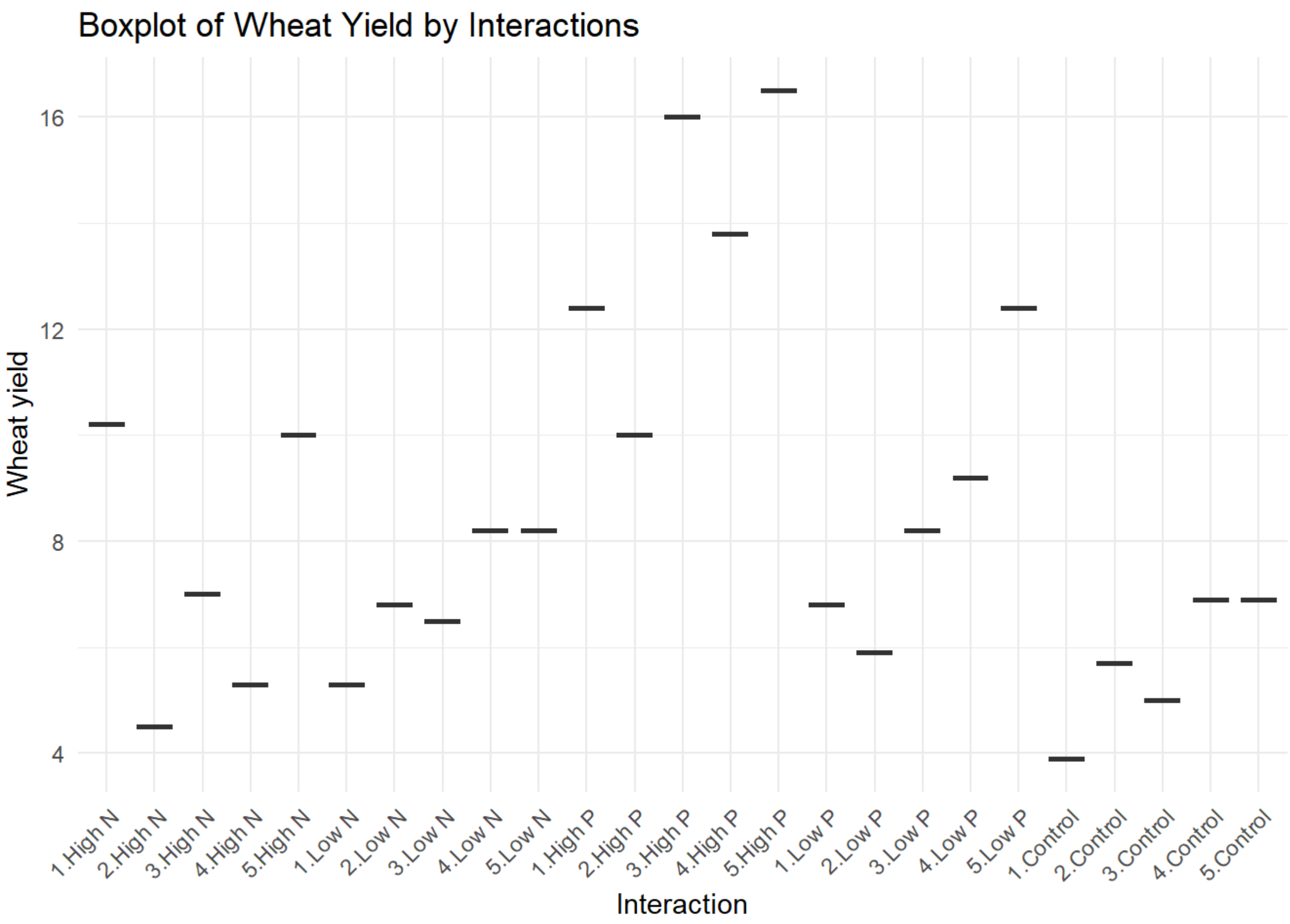
```
## 1:High N   10.2    bcd
## 2:High P   10.0   bcde
## 5:High N   10.0   bcde
## 4:Low P     9.2  bcdef
## 3:Low P     8.2  cdefg
## 4:Low N     8.2  cdefg
## 5:Low N     8.2  cdefg
## 3:High N    7.0   defg
## 4:Control   6.9   defg
## 5:Control   6.9   defg
## 1:Low P     6.8   defg
## 2:Low N     6.8   defg
## 3:Low N     6.5   defg
## 2:Low P     5.9   defg
## 2:Control   5.7   defg
## 1:Low N     5.3   defg
## 4:High N    5.3   defg
## 3:Control   5.0    efg
## 2:High N    4.5     fg
## 1:Control   3.9      g
```

> **Observation**
>
> The highest yield is observed in row 5 when there is high phosphorous
>
> The lowest yield is observed in row 1 with the control condition
>
> **Yields similar to row 1 with the control**
>
> 1. Row 3 with low phosphorous
> 2. Row 4 with low Nitrogen
> 3. Row 3 with high Nitrogen
> 4. Row 5 with the control
> 5. Row 1 with low phosphorous
> 6. Row 2 with low Nitrogen
> 7. Row 3 with low Nitrogen
> 8. Row 2 with low phosphorous
> 9. Row 2 with Control
> 10. Row 1 with low Nitrogen
> 11. Row 4 with high Nitrogen
> 12. Row 3 with Control
> 13. Row 2 with High Nitrogen
>
> It appears that when the rows are exposed to high phosphorous the yields from the plants fall in the highest two groups of plant yield
>
> Row 2 seems to need both phosphorous and nitrogen to have high yield as it appears among the group with similar features to the lowest mean yield.

Visualization

```
box_q3 = yield_data
box_q3$interactions = interaction(box_q3$Row, box_q3$Treatment)
```

```r
library(ggplot2)
ggplot(box_q3, aes(x=interactions, y=Yield)) +
  geom_boxplot() +
  theme_minimal() +
  labs(title="Boxplot of Wheat Yield by Interactions", x="Interaction", y="Wheat yield")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



> **Observation**
>
> The data in the graph matches the LSD analysis
>
> showing that row 5 with high phosphorous is still the highest yield and,
>
> row 1 with control is the lowest
>
> Phosphorous also appears to be the essential supplement needed for high yields in wheat

3. Discuss the results in the context of agricultural practices for enhancing corn growth. Consider both the statistical significance and the practical implications of the findings, offering recommendations for optimal fertilization strategies based on the results.

Implications

Yields are generally higher in treatments with high phosphorus (High P), especially when compared to the low phosphorus and some nitrogen treatments. from this we can infer that high phosphorus consistently shows yield improvements over controls and low nutrient applications, indicating its effectiveness in enhancing growth

The effect of nitrogen is less consistent compared to phosphorus. an example of such a high N treatmetment is like in the case of Row 1: High N. However there is varied results, sometimes matching yields in high phosphorus treatments but generally lower. Yields in treatments with High N are generally above the control but do not reach the levels of High P treatments, suggesting that phosphorus may be more limiting to wheat yields.

Both Low P and Low N treatments result in yields that are generally lower than the higher fertilizer treatments but are still above the control. This highlights the significant benefit of using fertilizer.

Suggestion

I suggest that the farmer conducts soil testing across different field sections to help tailor fertilizer applications to specific needs of each area. Additionally, regularly monitoring soil nutrient levels and crop health can help the farmer adjust fertilization strategies to changing conditions, maximizing yield while avoiding the overuse of fertilizers

Question 4

In a feeding trial involving 5 dairy cows, each cow receives one of 5 diets in 5 successive periods, following a Latin square design. The milk production (in liters) for each cow and period is as follows:

Cow	Period 1	Period 2	Period 3	Period 4	Period 5
1	24.1 (B)	26.3 (D)	25.6 (C)	24.9 (E)	22.8 (A)
2	22.9 (A)	23.1 (E)	25.0 (D)	24.0 (C)	26.7 (B)
3	26.5 (C)	24.2 (A)	23.1 (B)	26.2 (E)	24.0 (D)
4	26.7 (E)	27.4 (B)	25.1 (D)	25.1 (C)	29.6 (A)
5	27.8 (D)	24.7 (C)	24.6 (E)	22.2 (A)	24.7 (B)

The diets are:

A: Grass alone

B: Grass with supplement 1

C: Grass with supplement 2

D: Grass with supplement 3

E: Grass with supplement 4

1. Evaluate the effects of the different diets on milk production by performing an ANOVA to test the significance of diet effects, period effects, and cow effects. Provide the ANOVA table, including sources of variation, sum of squares, degrees of freedom, mean squares, F-values, and p-values.

```
dairy_cows = data.frame(
  Cow = as.factor(rep(1:5, each= 5)),
  Period = as.factor(rep(1:5, times = 5)),
  Treatment = as.factor(c(
    "B", "D", "C", "E", "A",
    "A", "E", "D", "C", "B",
    "C", "A", "B", "E", "D",
    "E", "B", "D", "C", "A",
    "D", "C", "E", "A", "B"
  )),
  Yield = c(
    24.1, 26.3, 25.6, 24.9, 22.8,
    22.9, 23.1, 25, 24, 26.7,
```

```
    26.5, 24.2, 23.1, 26.2, 24,
    26.7, 27.4, 25.1, 25.1, 29.6,
    27.8, 24.7, 24.6, 22.2, 24.7
  )
)

dairy_cows
```

```
##    Cow Period Treatment Yield
## 1    1      1         B  24.1
## 2    1      2         D  26.3
## 3    1      3         C  25.6
## 4    1      4         E  24.9
## 5    1      5         A  22.8
## 6    2      1         A  22.9
## 7    2      2         E  23.1
## 8    2      3         D  25.0
## 9    2      4         C  24.0
## 10   2      5         B  26.7
## 11   3      1         C  26.5
## 12   3      2         A  24.2
## 13   3      3         B  23.1
## 14   3      4         E  26.2
## 15   3      5         D  24.0
## 16   4      1         E  26.7
## 17   4      2         B  27.4
## 18   4      3         D  25.1
## 19   4      4         C  25.1
## 20   4      5         A  29.6
## 21   5      1         D  27.8
## 22   5      2         C  24.7
## 23   5      3         E  24.6
## 24   5      4         A  22.2
## 25   5      5         B  24.7
```

ANOVA table

```
dairy_anova = aov(Yield~Cow + Period + Treatment, data = dairy_cows)
summary(dairy_anova)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Cow          4  18.55   4.637   1.292  0.327
## Period       4   5.12   1.280   0.357  0.835
## Treatment    4   7.47   1.867   0.520  0.723
## Residuals   12  43.07   3.589
```

> **Observation**
>
> All the p-values are ≥ 0.05
>
> **Interpretation**
>
> H_0 there is no statistically significant difference in the mean of the milk production for cows based on the feeding period; the type of cow; and the food the cow consumes

> H_1 there is atleast one mean mil production that is statistically and significantly different from the other means of the milk production based on the feeding period; the type of cow; and the food the cow consumes
>
> because the p-value for the **type of Cows**, **the preriod of feeding**, and **the type of food the cow consumes** is greater than `0.05` we reject the null hypothesis
>
> **Conclusion**
>
> - There is no need to perform post-hoc as the means of milk production are not statistically or significantly different from each other.

2. If necessary, conduct post-hoc tests to identify which specific diets significantly differ in terms of milk production. Include a visual representation (e.g., a bar plot or box plot) comparing the diets.

3. Interpret the results in the context of nutritional strategies for dairy farming. Highlight any potential benefits or drawbacks of specific diets based on the milk production results, offering recommendations for optimal feeding practices

```
cow_lsd = agricolae::LSD.test(dairy_anova, c("Cow", "Period", "Treatment"))$groups
cow_lsd
```

```
##       Yield groups
## 4:5:A  29.6      a
## 5:1:D  27.8     ab
## 4:2:B  27.4     ab
## 2:5:B  26.7     ab
## 4:1:E  26.7     ab
## 3:1:C  26.5     ab
## 1:2:D  26.3     ab
## 3:4:E  26.2     ab
## 1:3:C  25.6     ab
## 4:3:D  25.1     ab
## 4:4:C  25.1     ab
## 2:3:D  25.0     ab
## 1:4:E  24.9     ab
## 5:2:C  24.7     ab
## 5:5:B  24.7     ab
## 5:3:E  24.6     ab
## 3:2:A  24.2     ab
## 1:1:B  24.1     ab
## 2:4:C  24.0     ab
## 3:5:D  24.0     ab
## 2:2:E  23.1      b
## 3:3:B  23.1      b
## 2:1:A  22.9      b
```

```
## 1:5:A  22.8      b
## 5:4:A  22.2      b
```

> **Recommendations**
>
> Despite the means not being significantly different from each other; it is observed that selecting cow `4` being fed at period `5` with Grass with grass alone offers the best milk yield
>
> The individual supplements have no significant effect on the ability of the cow to produce milk.
>
> To save the farmer money feeding their cows on the first and fifth period with grass only can maintain high yield with the least economical strain.
>
> The farmer could also consider acquiring cow 4 or cow 5.

References

El-Shafei, A. A., & Mattar, M. A. (2022). Irrigation scheduling and production of wheat with different water quantities in surface and drip irrigation: Field experiments and modelling using CROPWAT and SALTMED. *Agronomy*, *12*(7), 1488–1488. https://doi.org/10.3390/agronomy12071488

Gao, Z., Wang, Y., Tian, G., Zhao, Y., Li, C., Cao, Q., Han, R., Shi, Z., & He, M. (2020). Plant height and its relationship with yield in wheat under different irrigation regime. *Irrigation Science*, *38*(4), 365–371. https://doi.org/10.1007/s00271-020-00678-z

Gay, H. C., Rao, S. G., Vaccarino, V., & Ali, M. K. (2016). Effects of Different Dietary Interventions on Blood Pressure. *Hypertension*, *67*(4), 733–739. https://doi.org/10.1161/hypertensionaha.115.06853

Jurik, R., & Stastny, P. (2019). Role of nutrition and exercise programs in reducing blood pressure: a systematic review. *Journal of Clinical Medicine*, *8*(9), 1393. https://doi.org/10.3390/jcm8091393

Mayo Clinic. (2024, February 20). *Exercise: A drug-free approach to lowering high blood pressure*. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/in-depth/high-blood-pressure/art-20045206

Michael. (2024, February 17). *Uncovering the secrets: How tall does wheat grow and why it matters*. Sipsscene.com; SipsScene. https://sipsscene.com/how-tall-does-wheat-grow/

University of Minnesota. (2018, November 16). *Soil compaction*. Extension.umn.edu. https://extension.umn.edu/soil-management-and-health/soil-compaction#crop-growth-and-root-development-1147263

VeryWellHealth. (2024, April 20). *How to lower diastolic blood pressure (the bottom number)*. Verywell Health. https://www.verywellhealth.com/how-to-lower-diastolic-blood-pressure-7100411