MTH441 DOE Lab Assignment

Kaushik Raj Nadar (208160499)

2024-11-05

Problem 1

Create the Data Table

```r
if (!requireNamespace("dplyr", quietly = TRUE)) {
  install.packages("dplyr")
}
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```r
# Create the data frame with the given data
data <- data.frame(
  Fertilizer = rep(c("F1", "F2", "F3"), each = 3),
  Crop = rep(c("Corn", "Rice", "Wheat"), times = 3),
  MonthsHealthy = c(6, 4, 6, 5, 4.2, 5, 6, 5, 5.5)
)

# Check the structure of the data
print(data)
```

```
  Fertilizer  Crop MonthsHealthy
1         F1  Corn           6.0
2         F1  Rice           4.0
3         F1 Wheat           6.0
4         F2  Corn           5.0
5         F2  Rice           4.2
6         F2 Wheat           5.0
7         F3  Corn           6.0
8         F3  Rice           5.0
9         F3 Wheat           5.5
```

Perform two-way ANOVA

```
# Perform the two-way ANOVA
anova_result <- aov(MonthsHealthy ~ Fertilizer, data = data)

summary(anova_result)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
Fertilizer   2  0.976  0.4878   0.814  0.486
Residuals    6  3.593  0.5989
```

Fertilizer: The p-value for the Fertilizer factor is 0.486, which is much greater than the common significance level of 0.05. This indicates that there is no statistically significant difference in the response variable (e.g., crop yield or health) across the three types of fertilizer tested. In other words, based on this data, none of the fertilizers appear to have a significantly different effect on the response variable.

Residuals: The residuals represent the unexplained variation in the response variable after accounting for the effects of the fertilizers. The residual sum of squares and mean square provide an estimate of within-group variability, or the variation that cannot be attributed to differences between the fertilizers.

Problem 2

```
# Create the data frame based on the provided data
data <- data.frame(
  Restaurant = rep(c("R1", "R2", "R3", "R4", "R5", "R6"), each = 3),
  Item = rep(c("Item1", "Item2", "Item3"), times = 6),
  Sales = c(31, 27, 24,
```

```r
              31, 28, 31,
              45, 29, 46,
              21, 18, 48,
              42, 36, 46,
              32, 17, 40)
)

# Check the structure of the data
print(data)
```

```
   Restaurant  Item Sales
1          R1 Item1    31
2          R1 Item2    27
3          R1 Item3    24
4          R2 Item1    31
5          R2 Item2    28
6          R2 Item3    31
7          R3 Item1    45
8          R3 Item2    29
9          R3 Item3    46
10         R4 Item1    21
11         R4 Item2    18
12         R4 Item3    48
13         R5 Item1    42
14         R5 Item2    36
15         R5 Item3    46
16         R6 Item1    32
17         R6 Item2    17
18         R6 Item3    40
```

```r
# Perform the ANOVA for Randomized Block Design
rbd_model <- aov(Sales ~ Item + Restaurant, data = data)

# Display the ANOVA table
summary(rbd_model)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
Item         2  538.8  269.39   4.959 0.0319 *
Restaurant   5  559.8  111.96   2.061 0.1547
Residuals   10  543.2   54.32
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Item: The p-value for the Item factor is 0.0319, which is less than the significance level of 0.05. This indicates that there are statistically significant differences in sales figures among the three menu items. In other words, the popularity of the items is not equal, with at least one item having a different level of popularity compared to the others.

Restaurant: The p-value for the Restaurant factor is 0.1547, which is greater than 0.05. This suggests that there are no statistically significant differences in sales figures across the six restaurants. While there may be some variability in sales figures from one restaurant to another, this variability does not significantly impact the overall analysis. Thus, restaurant location does not play a major role in the sales of these items within this dataset.

Residuals: The residuals represent the unexplained variation in sales after accounting for the effects of both Item and Restaurant. The residual mean square (54.32) provides an estimate of the within-group variability, or how much variation remains unexplained by the factors in the model.

Problem 3

```
# Load necessary package for Latin Square analysis
if (!requireNamespace("agricolae", quietly = TRUE)) {
  install.packages("agricolae")
}
library(agricolae)

# Create data frame based on the problem's data
data <- data.frame(
  Tillage = rep(c("till1", "till2", "till3", "till4", "till5"), each = 5),
  Fertilizer = rep(c("fertilizer1", "fertilizer2", "fertilizer3", "fertilizer4", "fertilizer5
  Seed = c("A", "C", "B", "D", "E",
           "E", "B", "C", "A", "D",
           "C", "A", "D", "E", "B",
           "B", "D", "E", "C", "A",
           "D", "E", "A", "B", "C"),
  Productivity = c(42, 47, 55, 51, 44,
                   45, 54, 52, 44, 50,
                   41, 46, 57, 47, 48,
                   56, 52, 49, 50, 43,
                   47, 49, 45, 54, 46)
)

# Check the structure of the data
print(data)
```

```
   Tillage  Fertilizer Seed Productivity
1    till1 fertilizer1    A           42
2    till1 fertilizer2    C           47
3    till1 fertilizer3    B           55
4    till1 fertilizer4    D           51
5    till1 fertilizer5    E           44
6    till2 fertilizer1    E           45
7    till2 fertilizer2    B           54
8    till2 fertilizer3    C           52
9    till2 fertilizer4    A           44
10   till2 fertilizer5    D           50
11   till3 fertilizer1    C           41
12   till3 fertilizer2    A           46
13   till3 fertilizer3    D           57
14   till3 fertilizer4    E           47
15   till3 fertilizer5    B           48
16   till4 fertilizer1    B           56
17   till4 fertilizer2    D           52
18   till4 fertilizer3    E           49
19   till4 fertilizer4    C           50
20   till4 fertilizer5    A           43
21   till5 fertilizer1    D           47
22   till5 fertilizer2    E           49
23   till5 fertilizer3    A           45
24   till5 fertilizer4    B           54
25   till5 fertilizer5    C           46
```

```r
# Perform ANOVA for Latin Square Design
latin_square_model <- aov(Productivity ~ Seed + Fertilizer + Tillage, data = data)

# Display the ANOVA table
summary(latin_square_model)
```

```
            Df Sum Sq Mean Sq F value   Pr(>F)
Seed         4 286.16   71.54  12.836 0.000271 ***
Fertilizer   4 109.36   27.34   4.906 0.014105 *
Tillage      4  17.76    4.44   0.797 0.549839
Residuals   12  66.88    5.57
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Seed: The very low p-value ($p < 0.001$) indicates that there are statistically significant differences in productivity across the different seed types (A, B, C, D, E). This means the type of seed used has a substantial impact on productivity. Since this factor is highly significant, follow-up tests (such as Tukey's HSD) could identify specific pairs of seeds with significant differences.

Fertilizer: The p-value for Fertilizer (p 0.014) is also significant, indicating that the type of fertilizer has a statistically significant effect on productivity. Though this is a blocking factor in our Latin square design (i.e., not the primary focus), it's meaningful that fertilizer type still influences productivity.

Tillage: The high p-value (p 0.55) for Tillage indicates that tillage type does not have a statistically significant effect on productivity in this experiment. This suggests that the variation in tillage types does not contribute much to changes in productivity under the conditions tested.

Residuals: The residuals represent unexplained variability in productivity that isn't accounted for by Seed, Fertilizer, or Tillage. The relatively low residual sum of squares indicates that the model explains a substantial portion of the variability in productivity.