Formative Assessment 9

APM1111 Statistical Theory

Lopez, Ian Darylle C.   Pineda, Erianne L.   Sindayen, Lorenzo Danilo V.

GitHub Link:
https://github.com/ChewyGnome/APM1111

Dataset

```r
data_table <- data.frame(
  Fertilizer = c("Blend X", "", "", "", "",
                 "Blend Y", "", "", "", "",
                 "Blend Z", "", "", "", ""),
  Wheat = c(123,156,112,100,168,
            135,130,176,120,155,
            156,180,147,146,193),
  Corn = c(128,150,174,116,109,
           175,132,120,187,184,
           186,138,178,176,190),
  Soy = c(166,178,187,153,195,
          140,145,159,131,126,
          185,206,188,165,188),
  Rice = c(151,125,117,155,158,
           167,183,142,167,168,
           175,173,154,191,169)
)

kable(data_table,
      caption = "Crop Yield by Fertilizer Blend and Crop Type")
```

Table 1: Crop Yield by Fertilizer Blend and Crop Type

Fertilizer	Wheat	Corn	Soy	Rice
Blend X	123	128	166	151
	156	150	178	125
	112	174	187	117
	100	116	153	155
	168	109	195	158

Fertilizer	Wheat	Corn	Soy	Rice
Blend Y	135	175	140	167
	130	132	145	183
	176	120	159	142
	120	187	131	167
	155	184	126	168
Blend Z	156	186	185	175
	180	138	206	173
	147	178	188	154
	146	176	165	191
	193	190	188	169

Introduction

The study aims to evaluate the effectiveness of three different fertilizer blends (Blends X, Y, and Z) across four crop types (Wheat, Corn, Soy, and Rice) and assess their impact on crop yield. The objectives are to determine the effects of the different fertilizer blends and the crop types on crop yield as well as to identify any interaction effects. In order to meet these criteria, a two-way ANOVA was utilized to analyze the given data.

Problem

A new fertilizer has been developed to increase the yield on crops, and the makers of the fertilizer want to better understand which of the three formulations (blends) of this fertilizer are most effective for wheat, corn, soybeans and rice (crops). They test each of the three blends on 5 samples of each of the four types of crops.

Assumptions

Assumption #1: Your dependent variable should be measured at the continuous level (i.e., they are interval or ratio variables).

Assumption #2: Your two independent variables should each consist of two or more categorical, independent groups.

Assumption #3: You should have independence of observations, which means that there is no relationship between the observations in each group or between the groups themselves.

Assumption #4: There should be no significant outliers.

Assumption #5: Your dependent variable should be approximately normally distributed for each combination of the groups of the two independent variables.

Assumption #6. There needs to be homogeneity of variances for each combination of the groups of the two independent variables.

Hypotheses

A. Main Effect of Fertilizer

Null Hypothesis (H_0)**:** There is no difference in mean crop yield between the three fertilizer blends.

Alternative Hypothesis (H_A)**:** At least one fertilizer blend has a different mean yield.

B. Main Effect of Crop

Null Hypothesis (H_0)**:** There is no difference in mean crop yield between the four crop types.

Alternative Hypothesis (H_A)**:** At least one crop type has a different mean yield.

C. Interaction Effect Between Fertilizer and Crop

Null Hypothesis (H_0)**:** There is no significant interaction effect on yield between fertilizer blend and crop type.

Alternative Hypothesis (H_A)**:** There is a significant interaction effect on yield between fertilizer blend and crop type.

Checking of Assumptions

Assumption #1: Your dependent variable should be measured at the continuous level (i.e., they are interval or ratio variables).

Remark: The dependent variable, crop yield, is measured at the continuous level.

Assumption #2: Your two independent variables should each consist of two or more categorical, independent groups.

Remark: The two independent variables, fertilizer blend and crop type, each consist of two or more categorical, independent groups.

Assumption #3: You should have independence of observations, which means that there is no relationship between the observations in each group or between the groups themselves.

Remark: The samples are independent of one another, where there is no relationship between the observations in each group of the independent variable or between the groups themselves.

Assumption #4: There should be no significant outliers.

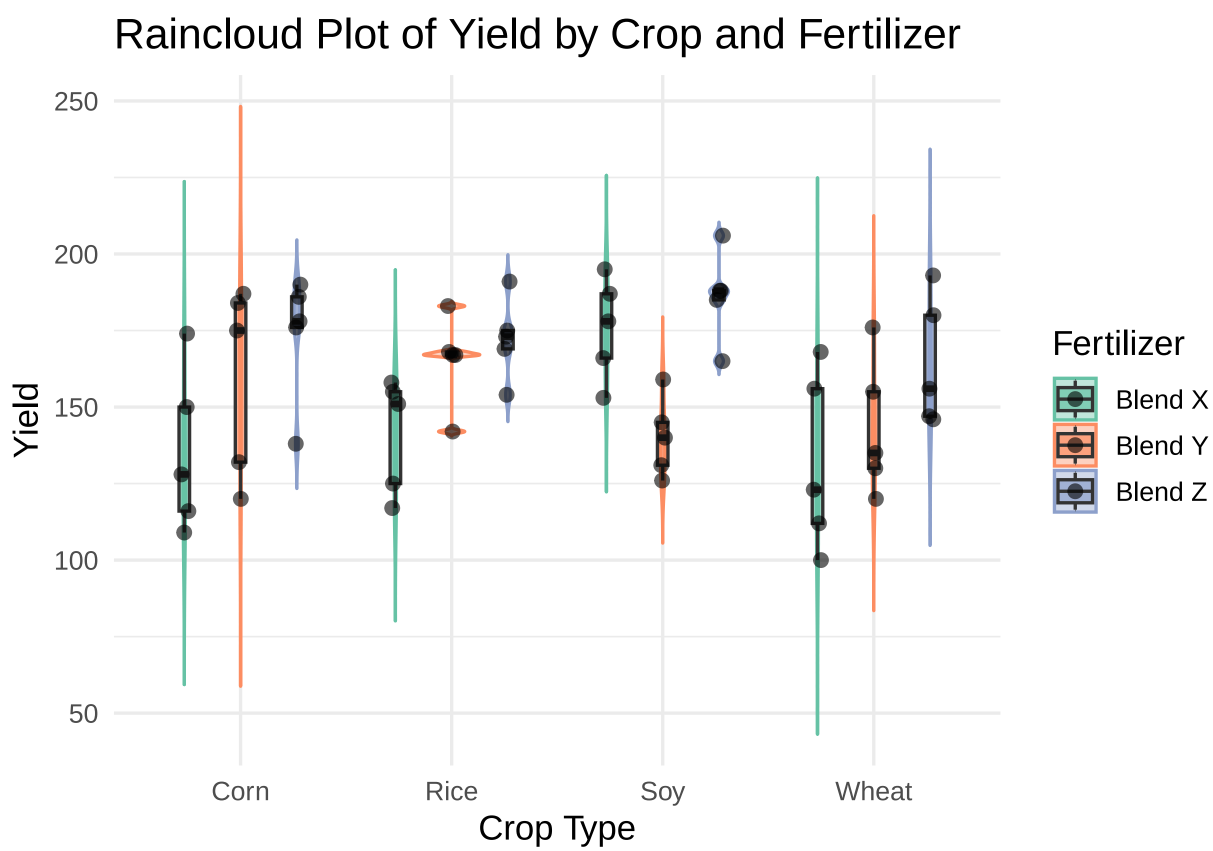
```
ggplot(df, aes(x = Crop, y = Yield, fill = Fertilizer)) +

  geom_violin(
    aes(color = Fertilizer),
    width = 0.8,
    alpha = 0.4,
    position = position_dodge(width = 0.8),
    trim = FALSE
  ) +

  geom_boxplot(
    width = 0.15,
    position = position_dodge(width = 0.8),
    alpha = 0.7,
    outlier.shape = NA
  ) +

  geom_jitter(
    position = position_jitterdodge(jitter.width = 0.1, dodge.width = 0.8),
    alpha = 0.6,
    size = 2
  ) +

  scale_fill_brewer(palette = "Set2") +
  scale_color_brewer(palette = "Set2") +
  theme_minimal(base_size = 13) +
  labs(
    title = "Raincloud Plot of Yield by Crop and Fertilizer",
    x = "Crop Type",
    y = "Yield",
    fill = "Fertilizer",
    color = "Fertilizer"
  )
```



Remark: There are no significant outliers as assessed by visual inspection of the raincloud plot.

Assumption #5: Your dependent variable should be approximately normally distributed for each combination of the groups of the two independent variables.

```
shapiro_results <- df %>%
  group_by(Fertilizer, Crop) %>%
  summarise(
    W = round(shapiro.test(Yield)$statistic, 4),
    p = round(shapiro.test(Yield)$p.value, 4),
    .groups = 'drop'
  )

kable(shapiro_results, caption = "Shapiro-Wilk Test of Normality")
```

Table 2: Shapiro–Wilk Test of Normality

Fertilizer	Crop	W	p
Blend X	Corn	0.9334	0.6200
Blend X	Rice	0.8397	0.1640
Blend X	Soy	0.9759	0.9114
Blend X	Wheat	0.9226	0.5467
Blend Y	Corn	0.8326	0.1456

Fertilizer	Crop	W	p
Blend Y	Rice	0.8695	0.2645
Blend Y	Soy	0.9667	0.8534
Blend Y	Wheat	0.9379	0.6510
Blend Z	Corn	0.7971	0.0767
Blend Z	Rice	0.9643	0.8373
Blend Z	Soy	0.9215	0.5398
Blend Z	Wheat	0.8667	0.2534

Remark: The dependent variable, crop yield, is approximately normally distributed for each combination of the groups of fertilizer blend and crop type, as assessed by Shapiro-Wilk test of normality, ($p > 0.05$).

Assumption #6. There needs to be homogeneity of variances for each combination of the groups of the two independent variables.

```
levene_result <- leveneTest(Yield ~ Fertilizer * Crop, data = df)

levene_table <- data.frame(
  Df = levene_result$Df,
  "F" = round(levene_result$`F value`, 4),
  "p" = round(levene_result$`Pr(>F)`, 4)
)

kable(levene_table, caption = "Levene's Test for Equality of Variance")
```

Table 3: Levene's Test for Equality of Variance

Df	F	p
11	0.6746	0.7551
48	NA	NA

Remark: The variances for each combination of the groups of fertilizer blend and crop type are homogeneous, as assessed by Levene's test of equality of variances, ($p = 0.755$).

Computation

Descriptive Statistics

```
descriptive_stats <- df %>%
group_by(Fertilizer, Crop) %>%
summarise(
Mean = round(mean(Yield), 2),
SD = round(sd(Yield), 2),
Min = min(Yield),
Max = max(Yield),
.groups = "drop"
```

```
)

kable(descriptive_stats, caption = "Descriptive Statistics for Yield by Fertilizer and Crop Type")
```

Table 4: Descriptive Statistics for Yield by Fertilizer and Crop Type

Fertilizer	Crop	Mean	SD	Min	Max
Blend X	Corn	135.4	26.60	109	174
Blend X	Rice	141.2	18.82	117	158
Blend X	Soy	175.8	16.69	153	195
Blend X	Wheat	131.8	29.06	100	168
Blend Y	Corn	159.6	31.28	120	187
Blend Y	Rice	165.4	14.74	142	183
Blend Y	Soy	140.2	12.87	126	159
Blend Y	Wheat	143.2	22.33	120	176
Blend Z	Corn	173.6	20.71	138	190
Blend Z	Rice	172.4	13.26	154	191
Blend Z	Soy	186.4	14.57	165	206
Blend Z	Wheat	164.4	21.05	146	193

ANOVA Test

```
anova_results <- df %>%
  anova_test(Yield ~ Fertilizer * Crop, effect.size = "pes")

kable(anova_results,
      caption = "Two-Way ANOVA Results",
      digits = 3)
```

Table 5: Two-Way ANOVA Results

Effect	DFn	DFd	F	p	p<.05	pes
Fertilizer	2	48	9.933	0.000	*	0.293
Crop	3	48	2.572	0.065		0.139
Fertilizer:Crop	6	48	2.347	0.046	*	0.227

Remark: There was a statistically significant interaction between fertilizer blend and crop type on yield, $F(6, 48) = 2.35, p = .046$. Thus, an analysis of simple main effects for fertilizer blend and crop type was performed with statistical significance receiving a Bonferroni adjustment and not being rejected at the p < .0125 level.

```
ggplot(df, aes(x = Crop, y = Yield, color = Fertilizer, group = Fertilizer)) +
  stat_summary(fun = mean, geom = "point", size = 3) +
  stat_summary(fun = mean, geom = "line", linewidth = 1) +
  stat_summary(fun.data = mean_se, geom = "errorbar", width = 0.2) +
  labs(title = "Interaction Effect: Fertilizer Blend by Crop Type",
       y = "Mean Yield (+/- SE)") +
```

```r
  theme_minimal() +
  scale_color_brewer(palette = "Set2")
```

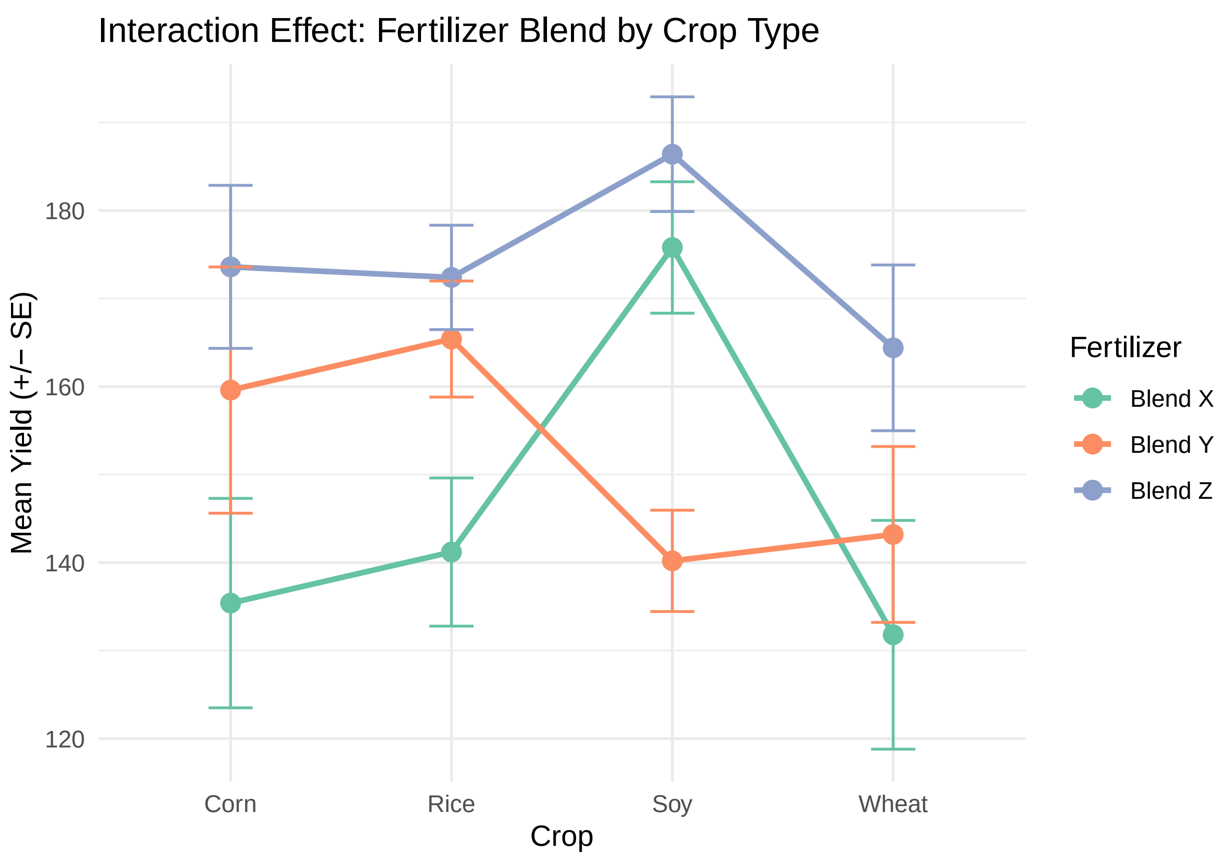


Figure 1: Interaction Plot of Fertilizer and Crop on Yield

Simple Main Effects Analysis

```r
simple_main_effects_fert <- df %>%
  group_by(Crop) %>%
  anova_test(Yield ~ Fertilizer, effect.size = "pes")

kable(simple_main_effects_fert,
      caption = "Simple Main Effects of Fertilizer at each Crop Type")
```

Table 6: Simple Main Effects of Fertilizer at each Crop Type

Crop	Effect	DFn	DFd	F	p	p<.05	pes
Corn	Fertilizer	2	12	2.649	0.111000		0.306
Rice	Fertilizer	2	12	5.380	0.021000	*	0.473
Soy	Fertilizer	2	12	13.378	0.000881	*	0.690
Wheat	Fertilizer	2	12	2.298	0.143000		0.277

Crop	Effect	DFn	DFd	F	p	p<.05	pes

Remark: Based on results, Soy was the only crop with a significant difference (p = .0009). As such, all pairwise comparisons were run for the significant simple main effect, Soy.

```
simple_main_effects_crop <- df %>%
  group_by(Fertilizer) %>%
  anova_test(Yield ~ Crop, effect.size = "pes")

kable(simple_main_effects_crop,
      caption = "Simple Main Effects of Crop at each Fertilizer Blend")
```

Table 7: Simple Main Effects of Crop at each Fertilizer Blend

Fertilizer	Effect	DFn	DFd	F	p	p<.05	pes
Blend X	Crop	3	16	3.738	0.033	*	0.412
Blend Y	Crop	3	16	1.627	0.223		0.234
Blend Z	Crop	3	16	1.315	0.304		0.198

Remark: Based on results, no p-values were less than 0.0125, indicating that for any given fertilizer, the crop type did not significantly change the yield. As such, pairwise comparisons were not run for this direction.

Post Hoc Analysis

```
pairwise_comparisons <- df %>%
  group_by(Crop) %>%
  tukey_hsd(Yield ~ Fertilizer)

soy_comparisons <- pairwise_comparisons %>%
  filter(Crop == "Soy")

kable(soy_comparisons,
      caption = "Post Hoc Comparisons (Tukey) for Soy")
```

Table 8: Post Hoc Comparisons (Tukey) for Soy

Crop	term	group1	group2	null.value	estimate	conf.low	conf.high	p.adj	p.adj.signif
Soy	Fertilizer	Blend X	Blend Y	0	-35.6	-60.56413	-10.63587	0.006550	**
Soy	Fertilizer	Blend X	Blend Z	0	10.6	-14.36413	35.56413	0.513000	ns
Soy	Fertilizer	Blend Y	Blend Z	0	46.2	21.23587	71.16413	0.000926	***

Reporting

The study employed a two-way ANOVA test to evaluate the impact of fertilizer blend (Blend X, Y, and Z) and crop type (Wheat, Corn, Soy, and Rice) on crop yield. Assumptions were initially verified before performing analysis. Outliers were assessed by visual inspection of a raincloud plot; normality was assessed by Shapiro-Wilk's normality test; homogeneity of variances was assessed by Levene's test. Results showed that there were no outliers, residuals were normally distributed ($p > .05$), and homogeneity of variances was met ($p = .755$).

Based on findings, the effects of fertilizer blend had a statistically significant effect on crop yield, $F(2, 48) = 9.93, p < .001$, partial $\eta^2 = .293$, indicating that the mean yields between the Blends X, Y, and Z differed. The main effect of crop type on yield, however, was not statistically significant, $F(3, 48) = 2.57, p = .065$, partial $\eta^2 = .139$. It is important to note that there was a statistically significant interaction between fertilizer blend and crop type, $F(6, 48) = 2.35, p = .046$, partial $\eta^2 = .227$, suggesting that each fertilizer blend's effectiveness on yield depended on the crop type used. Thus, an analysis of simple main effects for fertilizer blend was performed for each crop type, with statistical significance receiving a Bonferroni adjustment and not being rejected at the $p < .0125$ level.

The simple main effects analysis indicated that the effect of fertilizer blend was statistically significant for Soy, $F(2, 12) = 13.38, p < .001$, partial $\eta^2 = .690$. This shows a significant difference in yield across different fertilizer blends for Soy. On the other hand, the simple main effects of fertilizer were not statistically significant for the other crop types: Wheat ($p = .143$), Corn ($p = .603$), or Rice ($p = .844$). All pairwise comparisons were run for the significant simple main effect (Soy) with reported 95% confidence intervals and p-values Tukey-adjusted.

The mean yields for Soy using Blends X, Y, and Z were 175.80 ($SD = 16.69$), 140.20 ($SD = 12.87$) and 186.40 ($SD = 14.57$), respectively. The post hoc Tukey comparisons revealed fertilizer–crop combinations that yielded significantly different results. For instance, Blend Z produced noticeably higher yields when applied to Soy ($M = 186.4$) compared with Blend Y ($M = 140.2$) when applied to Soy. These findings suggest that depending on the crop type, there are specific fertilizer blends that are more optimal to utilize to maximize yield.