Statistics HW3

109550121 温柏萱

Problem 1

(a)

```
> data <- read.csv('./wastewater.csv', header=TRUE, sep=",")
> data
     AF   FS  FCC
1  34.6 38.8 26.7
2  35.1 39.0 26.7
3  35.3 40.1 27.0
4  35.8 40.9 27.1
5  36.1 41.0 27.5
6  36.5 43.2 28.1
7  36.8 44.9 28.1
8  37.2 46.9 28.7
9  37.4 51.6 30.7
10 37.7 53.6 31.2
> data_long <- stack(data)
> data_long
   values ind
1    34.6  AF
2    35.1  AF
3    35.3  AF
4    35.8  AF
5    36.1  AF
6    36.5  AF
7    36.8  AF
8    37.2  AF
9    37.4  AF
10   37.7  AF
```

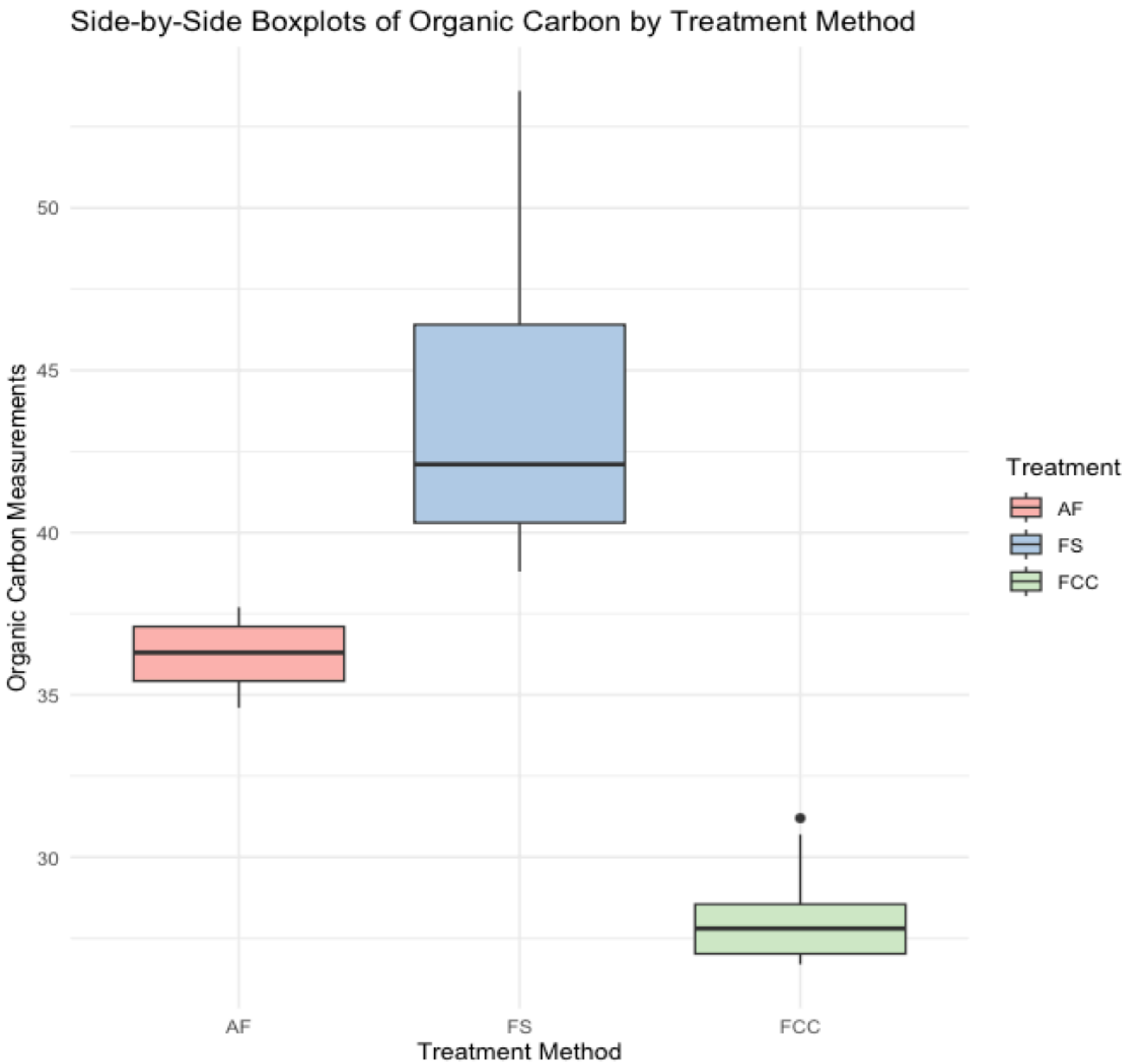
```
11    38.8   FS
12    39.0   FS
13    40.1   FS
14    40.9   FS
15    41.0   FS
16    43.2   FS
17    44.9   FS
18    46.9   FS
19    51.6   FS
20    53.6   FS
21    26.7  FCC
22    26.7  FCC
23    27.0  FCC
24    27.1  FCC
25    27.5  FCC
26    28.1  FCC
27    28.1  FCC
28    28.7  FCC
29    30.7  FCC
30    31.2  FCC
> names(data_long) <- c("OrganicCarbon", "Treatment")
> anova_result <- aov(OrganicCarbon ~ Treatment, data=data_long)
> summary(anova_result)
            Df Sum Sq Mean Sq F value   Pr(>F)    
Treatment    2 1251.5   625.8   60.63 1.03e-10 ***
Residuals   27  278.7    10.3                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

The null hypothesis $H_0 = \mu_1 = \mu_2 = \mu_3$ suggests that the influence of all the treatments AF, FS, FCC are equal. The alternative hypothesis $H_a :$ not H_0 suggests that at least one treatment differs significantly in its effectiveness.

The p-value of the ANOVA test was 1.03×10^{-10}, which is significantly smaller than 0.05, then the null hypothesis can be rejected, therefore, at least one of the treatment methods is effective.

(b)

```
> library(ggplot2)
> ggplot(data_long, aes(x=Treatment, y=OrganicCarbon, fill=Treatment)) +
+   geom_boxplot() +
+   labs(title="Side-by-Side Boxplots of Organic Carbon by Treatment
Method",
+        x="Treatment Method", y="Organic Carbon Measurements") +
+   theme_minimal() +
+   scale_fill_brewer(palette="Pastel1")
```



The FCC method is the best for reducing organic carbon in wastewater. It not only achieves the lowest median levels of organic carbon but also shows the least variability among the measurements, making it the most effective and consistent treatment method. This aligns with our earlier statistical analysis from the ANOVA, which pointed out significant differences in the effectiveness of the treatment methods, with FCC standing out as the most efficient.

Problem 2

(a)

```
> data <- read.csv('./Fern.csv', header=TRUE)
> data
  wave_light Block_age Response_area
1      420nm     young        1017.0
2      460nm     young         929.0
3      600nm     young         939.8
4      720nm     young        1081.5
5      420nm       old         854.7
6      460nm       old         689.9
7      600nm       old         841.5
8      720nm       old         797.4
> data$wave_light <- as.factor(data$wave_light)
> data$Block_age <- as.factor(data$Block_age)
> anova_result <- aov(Response_area ~ wave_light + Block_age, data = data)
> summary(anova_result)
           Df Sum Sq Mean Sq F value Pr(>F)
wave_light  3  21954    7318   2.163 0.2713
Block_age   1  76793   76793  22.697 0.0176 *
Residuals   3  10150    3383
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(a)

The blocking variable is this case is age, with two levels `Young` and `Old`. This variable is used to control for the variability in plant growth due to age, thereby isolating the effect of the light treatments on fern growth.

The p-value of `block_age` is $0.0176 < 0.05$, indicating that the differences in growth responses between young and old plants are statistically significant.

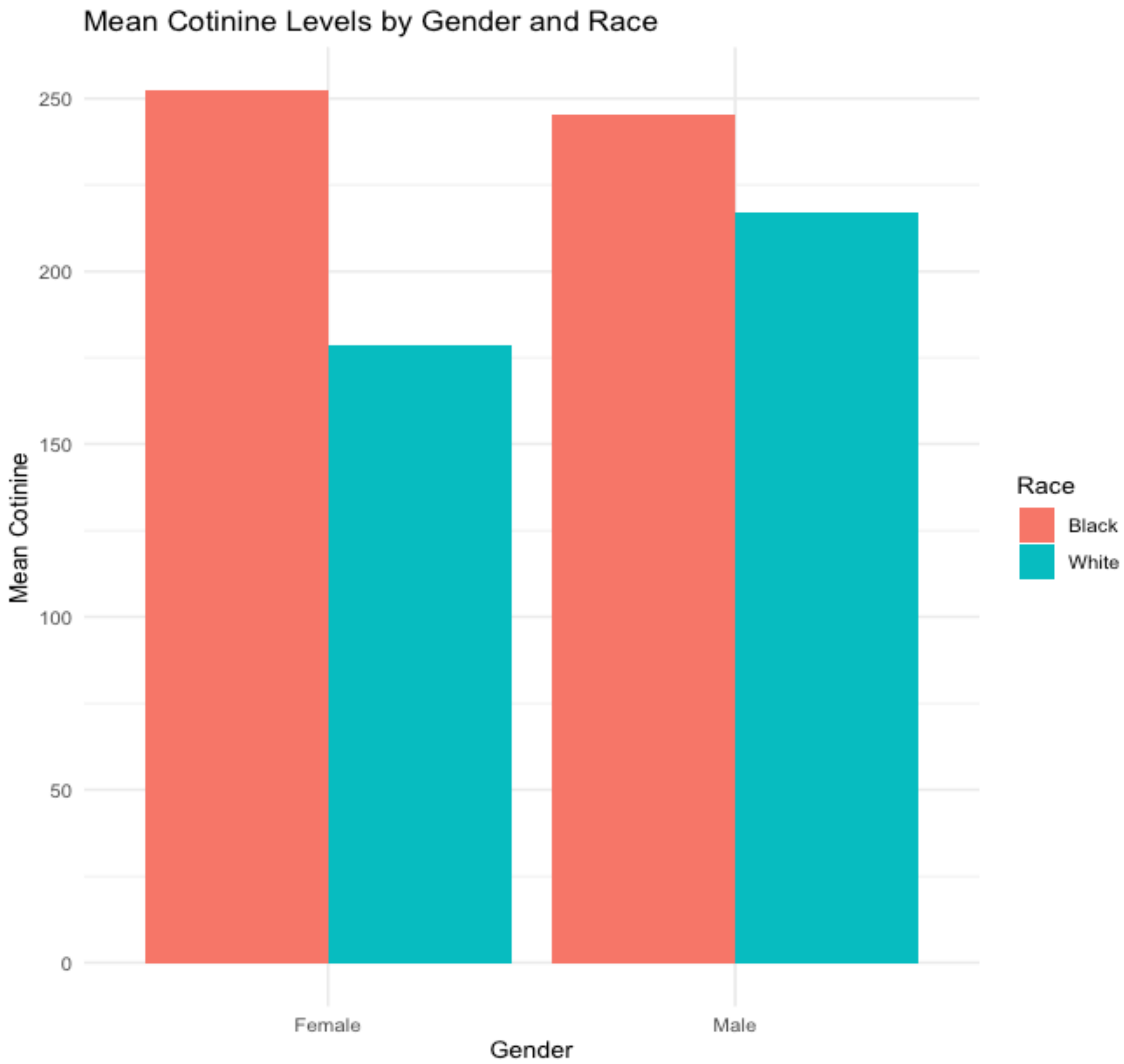
(b)

The p-value of `wave_light` is $0.2713 > 0.05$, indicating the variable does not have a significant influence on the growth response.

Problem 3

(a)

```
> means <- aggregate(cotinine ~ Gender + Race, data, mean)
> print(means)
  Gender  Race cotinine
1 Female Black    252.2
2   Male Black    245.4
3 Female White    178.6
4   Male White    217.0
> ggplot(data = means, aes(x = Gender, y = cotinine, fill = Race)) +
+   geom_bar(stat = "identity", position = "dodge") +
+   labs(title = "Mean Cotinine Levels by Gender and Race",
+        x = "Gender",
+        y = "Mean Cotinine") +
+   theme_minimal()
```



Gender	Race	Cotinine
Female	Black	252.2
Female	Black	245.4
Male	White	178.6
Male	White	217.0

The gap between Black and White seems larger for females than for males. This pattern suggests a potential interaction effect where the race impact on cotinine levels might be different for males and females.

(b)

```
> data <- read.csv("Cotinine.csv")
> data$Gender <- as.factor(data$Gender)
> data$Race <- as.factor(data$Race)
> model <- aov(cotinine ~ Gender + Race + Gender:Race, data = data)
> summary(model)
            Df Sum Sq Mean Sq F value Pr(>F)
Gender       1   1248    1248   0.204  0.657
Race         1  13005   13005   2.129  0.164
Gender:Race  1   2554    2554   0.418  0.527
Residuals   16  97731    6108
```

- Gender: The p-value for gender is $0.657 > 0.05$, which is not statistically significant. This suggests that there is no significant effect of gender on cotinine levels when not considering race.
- Race: The p-value for race is $0.164 > 0.05$, which is also not statistically significant. This indicates that the race alone does not significantly affect cotinine levels.

The p-value for the interaction between gender and race is $0.527 > 0.05$, indicating that there is no significant interaction effect. This means the effect of gender on cotinine levels does not significantly differ across different races, and vice versa.

(c)

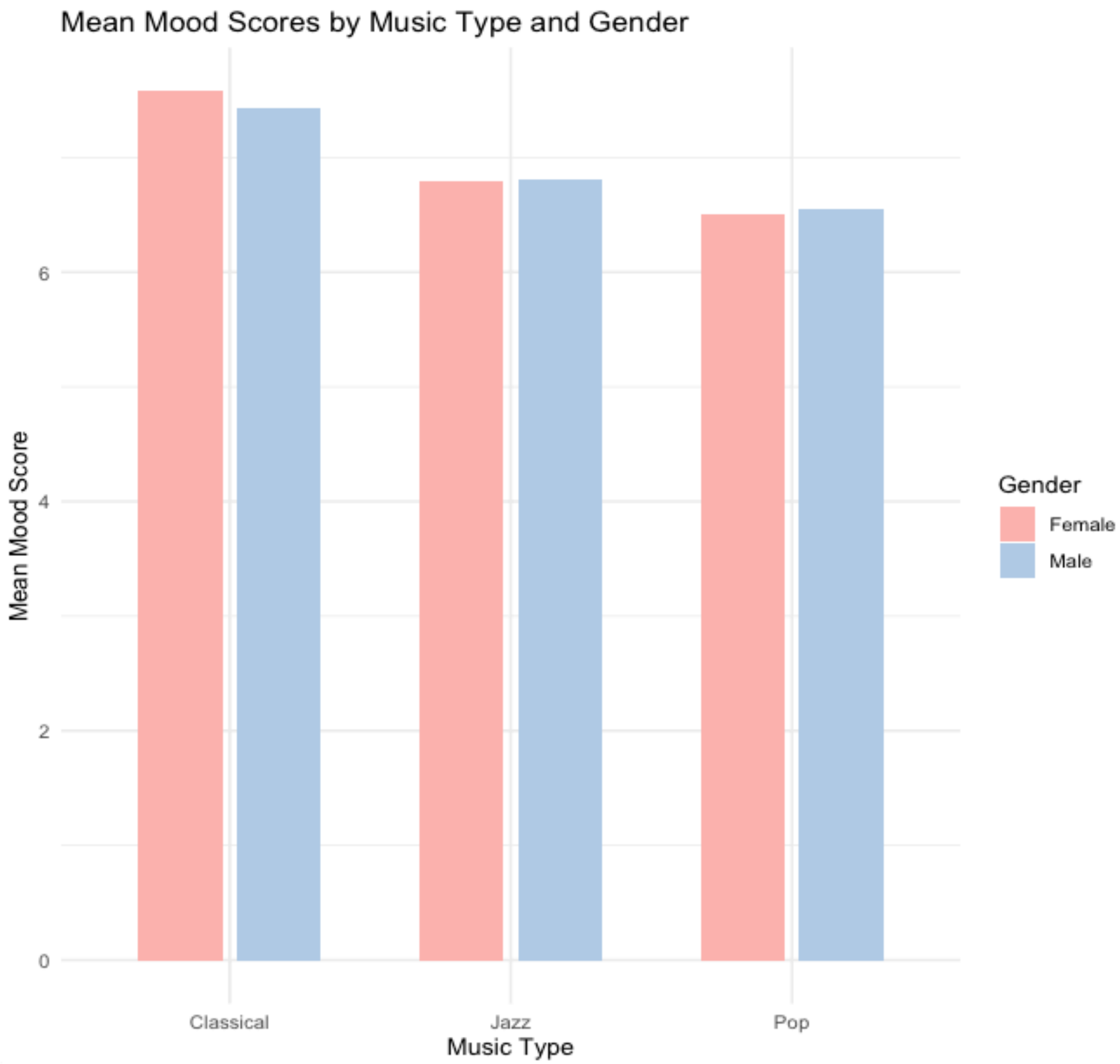
```
> model <- aov(cotinine ~ Gender + Race, data = data)
> summary(model)
            Df Sum Sq Mean Sq F value Pr(>F)
Gender       1   1248    1248   0.212  0.651
Race         1  13005   13005   2.205  0.156
Residuals   17 100285    5899
```

The p-value of gender is $0.651 > 0.05$, and the p-value for race is 0.156, both of which are greater than the 5% significant level, suggesting that variations in cotinine levels across different groups are not statistically significant.

Problem 4

(a)

```
> grouped_means <- data %>%
+   group_by(MusicType, Gender) %>%
+   summarise(MeanMoodScore = mean(MoodScore), .groups = 'drop')
> ggplot(grouped_means, aes(x = MusicType, y = MeanMoodScore, fill =
Gender)) +
+   geom_bar(stat = "identity", position = position_dodge(width = 0.7),
width = 0.6) +
+   labs(title = "Mean Mood Scores by Music Type and Gender", x = "Music
Type", y = "Mean Mood Score") +
+   scale_fill_brewer(palette = "Pastel1") +
+   theme_minimal()
```



The mean score across different groups are of small difference, indicating that there might not be a strong interaction effect between music type and gender, both gender in the same music type showed similar means, indicating that the Music Type may be a more important factor.

(b)

```
> anova_result <- aov(MoodScore ~ MusicType * Gender, data = data)
> summary(anova_result)
                 Df Sum Sq Mean Sq F value  Pr(>F)
MusicType         2  10.60   5.300   7.494 0.00134 **
Gender            1   0.01   0.013   0.018 0.89295
MusicType:Gender  2   0.12   0.060   0.085 0.91868
Residuals        54  38.19   0.707
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F-value and p-value of the interaction is 0.085 and $0.91868 > 0.05$, indicating that the interaction between music type and gender does not significantly affect mood scores.

Overall, there is no interaction effect at the 5% significance level since the p-value for the interaction is much greater than 0.05. This implies that the effect of music type on mood scores is consistent across genders.

(c)

- MusicType: The F-value and p-value for Music Type is 7.494 and $0.00134 << 0.05$, which is highly significant, suggesting that different types of music elicit different responses in terms of mood scores.
- Gender: The F-value and p-value for Gender is 0.018 and $0.89295 > 0.05$, indicating no significant difference in mood scores between genders, suggesting that mood response to music is generally uniform across male and female participants, which implies that gender does not need to be a primary consideration when selecting music for influencing mood in a mixed-gender context.

To sum up, MusicType shows a significant effect on the mood scores, meaning the type of music affects the mood scores significantly. Gender does not show a significant effect, indicating that the mood response is similar across genders irrespective of the type of music played.