Cultural difference investigation

Shorehamax

2024-07-06

Dataset

```
Spotify <- read.csv("C:/Users/chenh/Downloads/Spotify (Responses) - Form Responses 1.csv")
Spotify
```

```
##    Genz Exploring.the.outside. Understanding.of.the.local.culture  X X.1
## 1    No                   Yes                                   5 NA
## 2   Yes                   Yes                                   1 NA
## 3   Yes                   Yes                                   3 NA
## 4   Yes                   Yes                                   4 NA
## 5   Yes                   Yes                                   5 NA
## 6    No                   Yes                                   5 NA
## 7    No                   Yes                                   3 NA
## 8    No                   Yes                                   5 NA
## 9   Yes                   Yes                                   4 NA
## 10  Yes                   Yes                                   3 NA   ‘
## 11  Yes                   Yes                                   5 NA
## 12  Yes                   Yes                                   3 NA
## 13  Yes                   Yes                                   4 NA
## 14  Yes                   Yes                                   3 NA
## 15  Yes                   Yes                                   5 NA
## 16  Yes                   Yes                                   5 NA
## 17   No                   Yes                                   3 NA
## 18   No                   Yes                                   3 NA
## 19   No                   Yes                                   3 NA
## 20   No                   Yes                                   4 NA
## 21  Yes                   Yes                                   5 NA
## 22  Yes                   Yes                                   1 NA
## 23  Yes                   Yes                                   3 NA
## 24  Yes                   Yes                                   5 NA
## 25  Yes                   Yes                                   4 NA
## 26  Yes                   Yes                                   4 NA
## 27  Yes                   Yes                                   4 NA
## 28  Yes                   Yes                                   3 NA
## 29  Yes                   Yes                                   5 NA
## 30  Yes                   Yes                                   1 NA
## 31   No                    No                                   3 NA
## 32  Yes                    No                                   4 NA
## 33  Yes                    No                                   2 NA
## 34  Yes                    No                                   2 NA
```

```
## 35  Yes                    No                                   4 NA
## 36  Yes                    No                                   5 NA
## 37   No                    No                                   3 NA
## 38   No                    No                                   3 NA
## 39   No                    No                                   2 NA
## 40   No                    No                                   3 NA
## 41  Yes                    No                                   3 NA
## 42  Yes                    No                                   3 NA
## 43  Yes                    No                                   3 NA
## 44  Yes                    No                                   3 NA
## 45  Yes                    No                                   3 NA
## 46  Yes                    No                                   5 NA
## 47  Yes                    No                                   3 NA
## 48   No                    No                                   1 NA
```

```r
attach(Spotify)
```

ANOVA

```r
# Correcting the length of vectors to match
Genz <- c("No", "Yes", "Yes", "Yes", "Yes", "No", "No", "No", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes",
Exploring_the_outside <- c("Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes",

# Trimming the vectors to match in length
Genz <- Genz[1:length(Exploring_the_outside)]

# Create a dataframe from the corrected data
data <- data.frame(
  Genz = as.factor(Genz),
  Exploring_the_outside = as.factor(Exploring_the_outside),
  Understanding_of_the_local_culture = c(5, 1, 3, 4, 5, 5, 3, 5, 4, 3, 5, 3, 4, 3, 5, 5, 3, 3, 3, 4, 5,
)

# Analyze using factorial design
result <- aov(Understanding_of_the_local_culture ~ Genz * Exploring_the_outside, data = data)
summary(result)
```

```
##                            Df Sum Sq Mean Sq F value Pr(>F)
## Genz                        1   1.27   1.274   0.985 0.3259
## Exploring_the_outside       1   4.01   4.015   3.103 0.0844 .
## Genz:Exploring_the_outside  1   0.32   0.321   0.248 0.6205
## Residuals                  49  63.41   1.294
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1 observation deleted due to missingness
```

Assumptions check

```r
library(car)
```

```
## Warning: package 'car' was built under R version 4.2.3
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.2.3
```

```r
# Levene's test for equal variance
leveneTest(Understanding_of_the_local_culture ~ Genz, data = data)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  0.0337 0.8551
##       51
```

```r
leveneTest(Understanding_of_the_local_culture ~ Exploring_the_outside, data = data)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  0.6332 0.4298
##       52
```

```r
# Shapiro-Wilk test for normality of residuals
shapiro.test(residuals(result))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(result)
## W = 0.93834, p-value = 0.008732
```

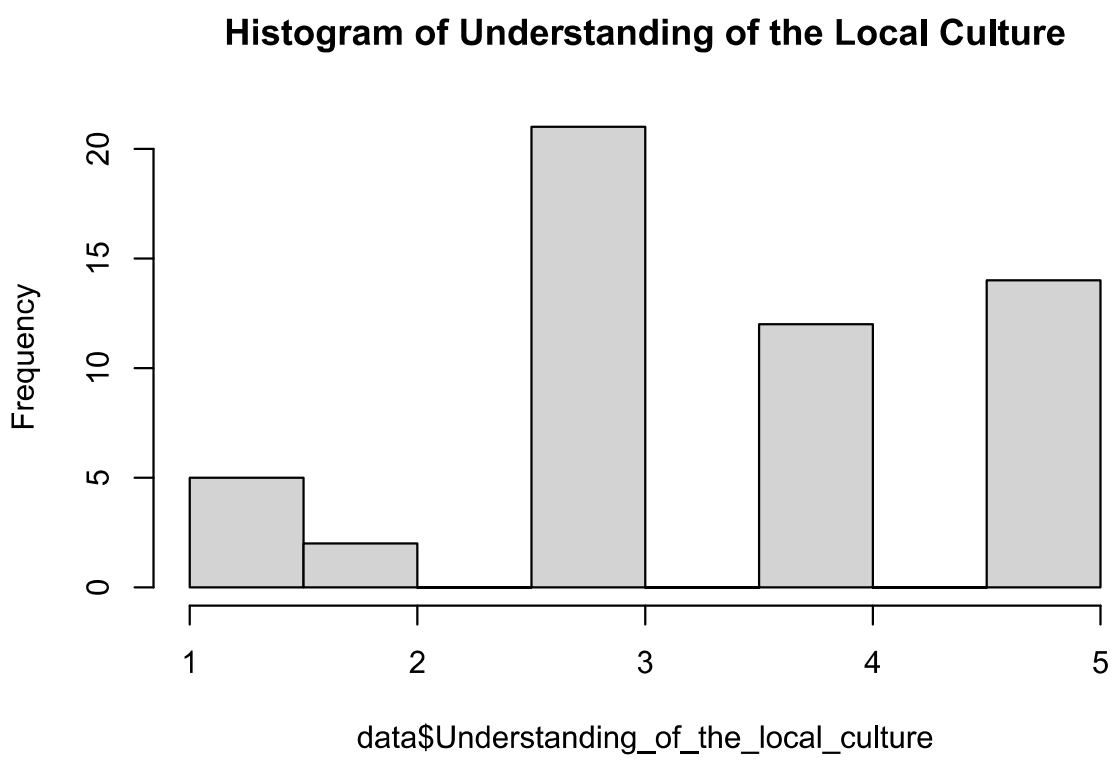
```r
# Shapiro-Wilk test for normality of the response variable
shapiro.test(data$Understanding_of_the_local_culture)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$Understanding_of_the_local_culture
## W = 0.86579, p-value = 2.278e-05
```

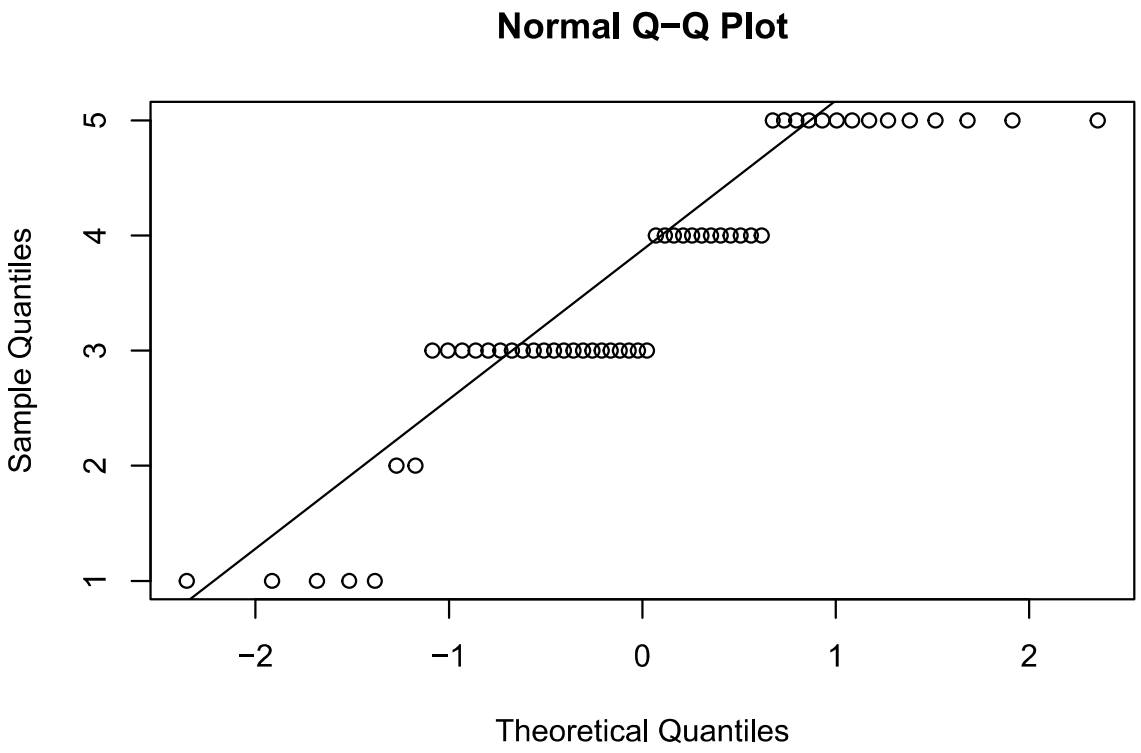
Normality assumption is violated, we need to perform transformation later.

Visualisation of distribution of the response variable

```r
# Histogram
hist(data$Understanding_of_the_local_culture, main = "Histogram of Understanding of the Local Culture")
```

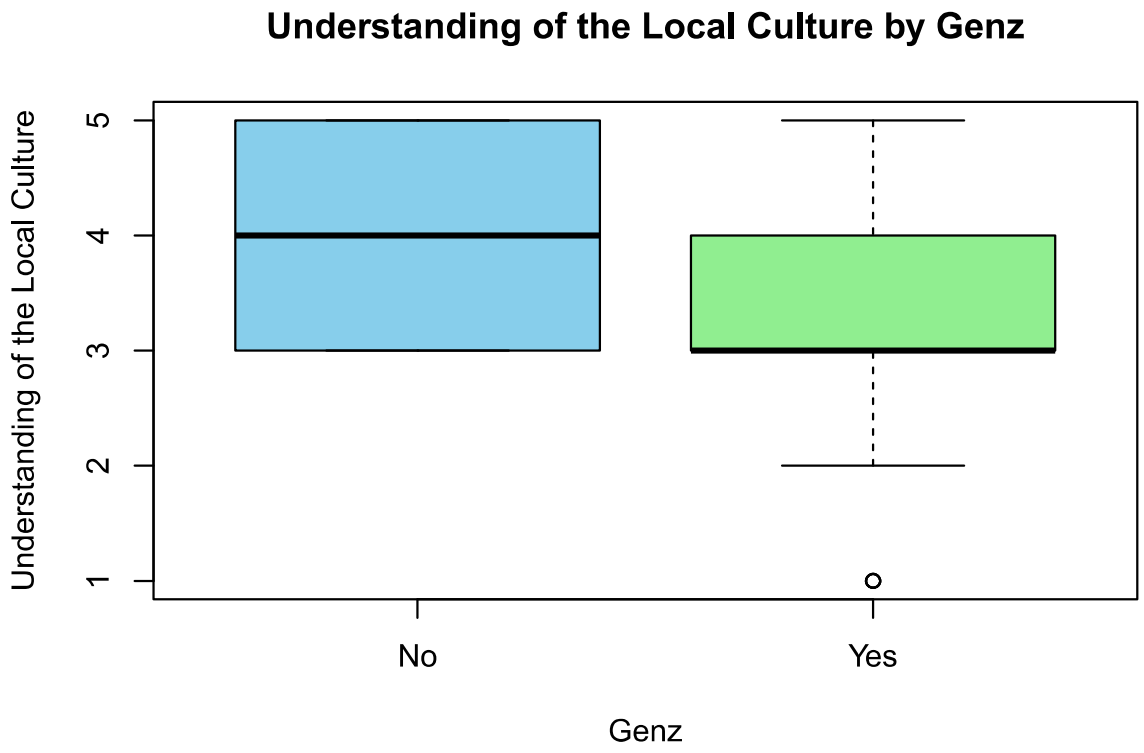



```r
# Q-Q plot
qqnorm(data$Understanding_of_the_local_culture)
qqline(data$Understanding_of_the_local_culture)
```

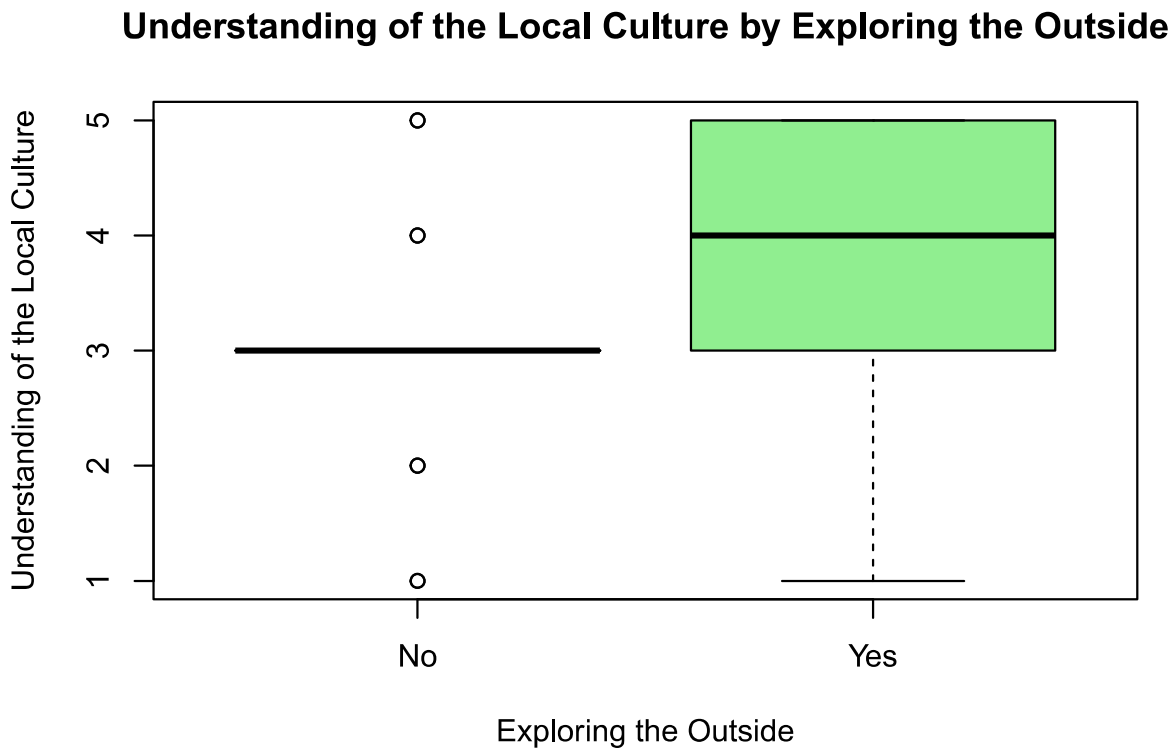


boxplot

```
# Boxplot for Understanding of the Local Culture across Genz levels
boxplot(Understanding_of_the_local_culture ~ Genz, data = data,
        main = "Understanding of the Local Culture by Genz",
        xlab = "Genz", ylab = "Understanding of the Local Culture",
        col = c("skyblue", "lightgreen"))
```




```r
# Boxplot for Understanding of the Local Culture across Exploring_the_outside levels
boxplot(Understanding_of_the_local_culture ~ Exploring_the_outside, data = data,
        main = "Understanding of the Local Culture by Exploring the Outside",
        xlab = "Exploring the Outside", ylab = "Understanding of the Local Culture",
        col = c("skyblue", "lightgreen"))
```



QQplot

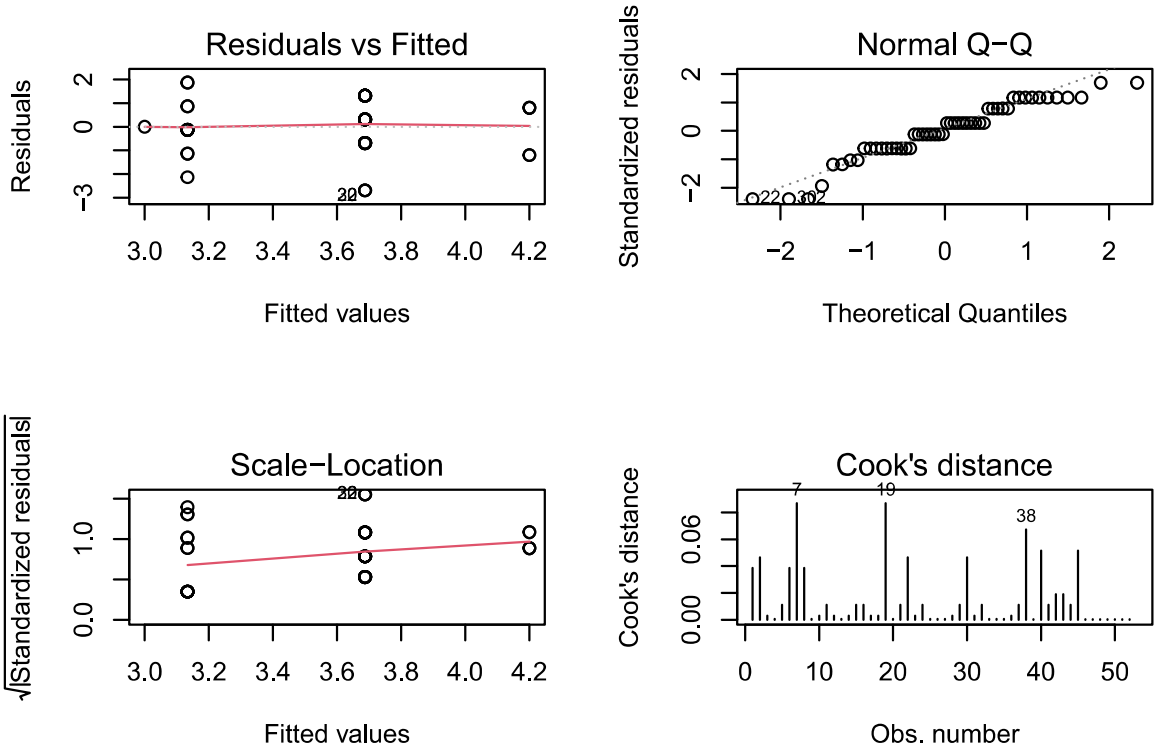
```
# Set up a 2x2 grid for multiple plots
par(mfrow = c(2, 2))

# Q-Q plot for understanding of the local culture
plot(result, 1)

# Residuals vs. Fitted values plot
plot(result, 2)
```

```
## Warning: not plotting observations with leverage one:
##   53
```

```
# Normal Q-Q plot of residuals
plot(result, 3)
```

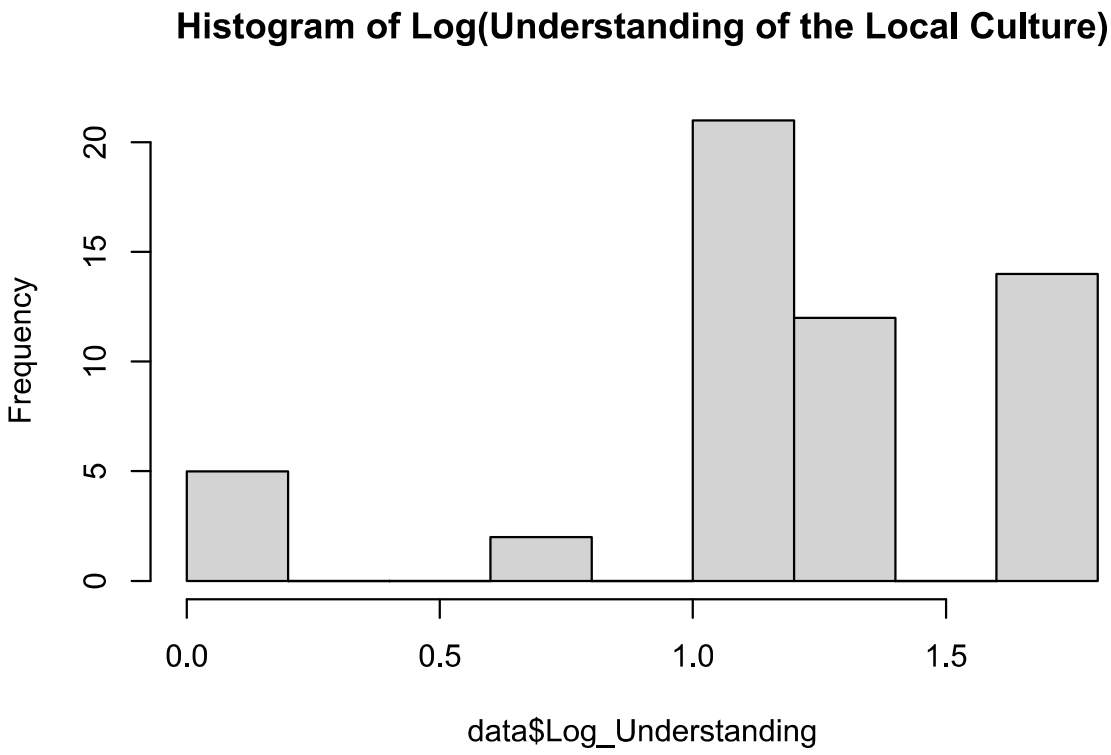
```
## Warning: not plotting observations with leverage one:
##   53
```

```
# Scale-Location plot (Square root of standardized residuals vs. fitted values)
plot(result, 4)
```



Transformation

```
# Logarithmic transformation of the response variable
data$Log_Understanding <- log(data$Understanding_of_the_local_culture)

# Check the distribution of the transformed variable
hist(data$Log_Understanding, main = "Histogram of Log(Understanding of the Local Culture)")
```




```r
# Perform ANOVA on the transformed variable
result_transformed <- aov(Log_Understanding ~ Genz * Exploring_the_outside, data = data)
summary(result_transformed)
```

```
##                             Df Sum Sq Mean Sq F value Pr(>F)
## Genz                         1  0.160 0.15954   0.890  0.350
## Exploring_the_outside        1  0.294 0.29388   1.640  0.206
## Genz:Exploring_the_outside   1  0.019 0.01853   0.103  0.749
## Residuals                   49  8.781 0.17921
## 1 observation deleted due to missingness
```

Transformation failed, we have to use a non-parametric test.

##Non-parametric test

```r
# Perform Kruskal-Wallis test
kruskal.test(Understanding_of_the_local_culture ~ Genz, data = data)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Understanding_of_the_local_culture by Genz
## Kruskal-Wallis chi-squared = 0.70524, df = 1, p-value = 0.401
```

```r
# Perform Kruskal-Wallis test for the second factor
kruskal.test(Understanding_of_the_local_culture ~ Exploring_the_outside, data = data)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Understanding_of_the_local_culture by Exploring_the_outside
## Kruskal-Wallis chi-squared = 5.7059, df = 1, p-value = 0.01691
```

The p-value (0.401) is greater than the significance level of 0.05. Therefore, we do not have sufficient evidence to reject the null hypothesis. This suggests that there is no significant difference in the median understanding of the local culture between individuals belonging to Generation Z and those who do not.

The p-value (0.01691) is less than the significance level of 0.05. Thus, we have sufficient evidence to reject the null hypothesis. This indicates that there is a significant difference in the median understanding of the local culture between individuals who frequently explore the outside and those who do not.

Based on these results, while there is no significant difference in understanding of the local culture between Generation Z and non-Generation Z individuals, there is a significant difference based on the frequency of exploring the outside.

Post Hoc

```r
pairwise.wilcox.test(data$Understanding_of_the_local_culture, data$Exploring_the_outside, p.adjust.meth
```

```
## Warning in wilcox.test.default(xi, xj, paired = paired, ...): cannot compute
## exact p-value with ties
```

```
##
##  Pairwise comparisons using Wilcoxon rank sum test with continuity correction
##
## data:  data$Understanding_of_the_local_culture and data$Exploring_the_outside
##
##     No
## Yes 0.017
##
## P value adjustment method: bonferroni
```

The pairwise comparison results suggest that there is a statistically significant difference in understanding of the local culture between individuals who frequently explore the outside and those who do not, with those who explore the outside having a higher median understanding.

Median difference

```r
# Calculate median difference between groups
median_diff <- median(data$Understanding_of_the_local_culture[data$Exploring_the_outside == "Yes"]) -
               median(data$Understanding_of_the_local_culture[data$Exploring_the_outside == "No"])

# Create a subset of data for individuals who explore the outside and those who do not
```

```r
exploring_yes <- data$Understanding_of_the_local_culture[data$Exploring_the_outside == "Yes"]
exploring_no <- data$Understanding_of_the_local_culture[data$Exploring_the_outside == "No"]

# Calculate the confidence interval for the difference in medians
ci <- t.test(exploring_yes, exploring_no, alternative = "two.sided", conf.level = 0.95)$conf.int

# Print the results
cat("Difference in medians:", median_diff, "\n")
```

```
## Difference in medians: 1
```

```r
cat("95% Confidence Interval for the difference in medians:", ci[1], "to", ci[2])
```

```
## 95% Confidence Interval for the difference in medians: 0.08093507 to 1.432578
```

The results indicate:

Difference in medians: 1 95% Confidence Interval for the difference in medians: 0.0809 to 1.4326 This means that individuals who frequently explore the outside have a median understanding of the local culture that is 1 unit higher than individuals who do not explore the outside. Additionally, we are 95% confident that the true difference in medians lies between 0.0809 and 1.4326 units.