Story_Assignment_6

2024-11-20

#Noah Krebs

```
library(tidyr)
data <- read.csv("6 ReducingPain.csv")
long_data <- pivot_longer(data,
                          cols = everything(),    # Select all columns to reshape
                          names_to = "Condition", # Name for the new variable column
                          values_to = "Score")    # Name for the value column

write.csv(long_data, "6_ReducingPain_long.csv", row.names = FALSE)
```

```
data_1 <- read.csv("6 ReducingPainLong.csv")
print(data_1)
```

```
##    Treatment PainLevel
## 1  Audiobook         5
## 2  Audiobook         6
## 3  Audiobook         7
## 4  Audiobook         2
## 5  Audiobook         6
## 6  Audiobook         3
## 7  Audiobook         4
## 8  Audiobook         8
## 9  Audiobook         5
## 10 Audiobook         4
## 11     Music         5
## 12     Music         4
## 13     Music         4
## 14     Music         7
## 15     Music         6
## 16     Music         4
## 17     Music         6
## 18     Music         4
## 19     Music         3
## 20     Music         5
## 21 Earphones         4
## 22 Earphones         8
## 23 Earphones         7
## 24 Earphones         6
## 25 Earphones        10
## 26 Earphones         6
## 27 Earphones        10
## 28 Earphones         8
## 29 Earphones         5
## 30 Earphones         6
```

```r
#Question 1b
overall_mean <- mean(data_1$PainLevel)


group_means <- tapply(data_1$PainLevel, data_1$Treatment, mean)


n <- table(data_1$Treatment)


group_SS <- sum(n * (group_means - overall_mean)^2)


total_SS <- sum((data_1$PainLevel - overall_mean)^2)


error_SS <- total_SS - group_SS


df_total <- length(data_1$PainLevel) - 1
df_groups <- length(group_means) - 1
df_error <- df_total - df_groups

MS_group <- group_SS / df_groups
MS_error <- error_SS / df_error

F_value <- MS_group / MS_error

p_value <- pf(F_value, df_groups, df_error, lower.tail = FALSE)

print(F_value)
```

```
## [1] 5.020101
```

```r
print(p_value)
```

```
## [1] 0.01400622
```

```r
#Question 1 c-d
anova <- aov(data_1$PainLevel ~ data_1$Treatment, data = data)
summary(anova)
```

```
##                   Df Sum Sq Mean Sq F value Pr(>F)
## data_1$Treatment  2   29.6  14.800    5.02  0.014 *
## Residuals        27   79.6   2.948
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Conclusion: Since the p_value is less than .05, we can conclude that at least one of the three treatm
```

```r
#Question 1 e-g
group_means <- tapply(data_1$PainLevel, data_1$Treatment, mean)
group_means <- sort(group_means) # Order means from smallest to largest

print(group_means)
```

```
##     Music Audiobook Earphones
##       4.8       5.0       7.0
```

```r
# Calculate pairwise differences
pairwise_differences <- combn(names(group_means), 2, function(x) {
  diff <- abs(group_means[x[1]] - group_means[x[2]])
  data.frame(pair = paste(x[1], "-", x[2]), difference = diff)
}, simplify = FALSE)
pairwise_differences <- do.call(rbind, pairwise_differences) # Combine into a data frame
print(pairwise_differences)
```

```
##                             pair difference
## Music         Music - Audiobook        0.2
## Music1        Music - Earphones        2.2
## Audiobook Audiobook - Earphones        2.0
```

```r
# Calculate group means for individual observations
group_means_map <- with(data, ave(data_1$PainLevel, data_1$Treatment, FUN = mean))

# Calculate Error SS
error_SS <- sum((data$PainLevel - group_means_map)^2)

# Calculate Error DF
df_error <- nrow(data) - length(group_means)

# Calculate Error MS
MS_error <- error_SS / df_error

cat("Error Mean Square (MS_error):", MS_error, "\n")
```

```
## Error Mean Square (MS_error): 0
```

```r
# Calculate Standard Error (SE) for a specific pair (Audiobook vs Music)
n <- table(data$Treatment)
SE <- sqrt(MS_error * (1 / n["Audiobook"] + 1 / n["Music"]))

cat("Standard Error (SE):", SE, "\n")
```

```
## Standard Error (SE): NA
```

```r
pairwise_differences$q_stat <- pairwise_differences$difference / SE

print(pairwise_differences)
```

```
##                             pair difference q_stat
## Music         Music - Audiobook        0.2     NA
## Music1        Music - Earphones        2.2     NA
## Audiobook Audiobook - Earphones        2.0     NA
```

```r
k <- length(group_means)
critical_q <- qtukey(0.95, k, df_error)
pairwise_differences$p_value <- ptukey(pairwise_differences$q_stat, k, df_error, lower.tail = FALSE)
print(pairwise_differences)
```

```
##                             pair difference q_stat p_value
## Music         Music - Audiobook        0.2     NA      NA
## Music1        Music - Earphones        2.2     NA      NA
## Audiobook Audiobook - Earphones        2.0     NA      NA
```

```r
anova_result <- aov(data_1$PainLevel ~ data_1$Treatment, data = data)
tukey_result <- TukeyHSD(anova_result)

#After looking at the reults I observed a significant difference in the music and earphones comparison
```

```r
#Question 1
#Tukey short
data <- data.frame(
  Treatment = c(rep("Audiobook", 10), rep("Music", 10), rep("Earphones", 10)),
  PainLevel = c(5, 6, 7, 2, 6, 3, 4, 8, 5, 4, 5, 4, 4, 7, 6, 4, 6, 4, 3, 5, 4, 8, 7, 6, 10, 6, 10, 8, 5
)
aov_res <- aov(formula = PainLevel ~ Treatment, data = data)
summary(aov_res)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Treatment    2   29.6  14.800    5.02  0.014 *
## Residuals   27   79.6   2.948
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
tukey_res <- TukeyHSD(x = aov_res, ordered = TRUE, conf.level = 0.95)
print(tukey_res)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##     factor levels have been ordered
##
## Fit: aov(formula = PainLevel ~ Treatment, data = data)
##
## $Treatment
##                       diff         lwr      upr     p adj
## Audiobook-Music        0.2 -1.70387901 2.103879 0.9633444
## Earphones-Music        2.2  0.29612099 4.103879 0.0210755
## Earphones-Audiobook    2.0  0.09612099 3.903879 0.0380535
```

```r
#Question 2
#Part A
effect_difference <- 0.5  # Smallest difference to detect
error_ms <- 6.5           # Error Mean Square
alpha <- 0.05             # Significance level
power <- 0.9              # Desired power
num_groups <- 4           # Number of streams

f <- effect_difference / sqrt(error_ms)

group_means <- c(0, 0.5, 1.0, 1.5)
overall_mean <- mean(group_means)


n <- 2
achieved_power <- 0

while (achieved_power < power) {
  total_sample_size <- n * num_groups
  df1 <- num_groups - 1
```

```
  df2 <- total_sample_size - num_groups

  ncp <- n * sum((group_means - overall_mean)^2) / error_ms

  f_crit <- qf(p=alpha, df1=df1, df2=df2, lower.tail=FALSE)

  achieved_power <- pf(q=f_crit, df1=df1, df2=df2, ncp=ncp, lower.tail=FALSE)

  if (achieved_power < power) {
    n <- n + 1
  }
}
print(n)
```

```
## [1] 75
```

```
#Question 2
# Part B
#To reach a power of at least .9, one has to take 4 groups of 75 samples.
#300 total
#Each sample of tag cost $25
#300x25 = $7500 which is less than 100,000. You will be able to complete the experiment with the desire
```