HypothesisTesting.r

aravi

2024-01-15

```
#Setting Directory
getwd()
```

```
## [1] "C:/Users/aravi/Desktop/StatisticsUsingR/SocialMediaStudy"
```

```
setwd("C:\\Users\\aravi\\Desktop\\StatisticsUsingR\\SocialMediaStudy")
getwd()
```

```
## [1] "C:/Users/aravi/Desktop/StatisticsUsingR/SocialMediaStudy"
```

```
data <- read.csv("IPSMOI.csv")

#HRSPD - Hours Spend per Day
#SIEFF - Social Interaction Effects Friends and Family
#ACPA - Academic Performance Affect
#MHAF - Mental Health Affected
#QRI - Quality and Relevance of Information
#PC - Privacy Concerns in social Media
#CCF - How Often Content Creation
#SSCC - Support Community and social Cause
#UEP - Use on Education Purpose
#CASI - Creating Awareness On Social Issues
#PDDA - Purchase decisions due to advertising
#SOSP - How Often Social Media Use
#SMPF - Social Media Platform
#RPS - Review and Update Privacy Settings
#TCEM - Type of Content Engagement Most
#OLF - Online Friends
#APOC - Actively Participate in Online Community
#SMIF - Social Media Influencer FolLowing
#PFO - Provide Feedback or opinions
#PG - Personal Growth
```

```
# Remove irrelavant Columns
data <- data[-c(1, 2, 3)]

# Change Column Names to abbreviations
colnames(data) <- c("AgeGroup", "Gender", "CareerStatus", "ResidentialArea",
                    "HRSPD", "SIEFF", "ACPA", "MHAF", "QRI",
                    "PC", "CCF", "SSCC", "UEP", "CASI", "PDDA",
                    "SOSP", "SMPF", "RPS", "TCEM", "OLF", "APOC", "SMIF",
                    "PFO", "PG")


#data$MHAF
data$MHAF<-as.numeric(
factor(data$MHAF,
    levels = c("Extremely Negative",
                    "Negative",
                    "No Impact",
                    "Positive",
                    "Extremely Positive")))

threshold_hours = 3  # Adjust the threshold as needed
# Create a new variable 'UserType' based on the threshold
data$UserType <- ifelse(data$HRSPD > threshold_hours, 'Heavy User', 'Light User')

data$SIEFF<-as.numeric(
  factor(data$SIEFF,
         levels = c("Significantly Negative",
                    "Slightly Negative",
                    "No Impact",
                    "Slightly Positive",
                    "Significantly Positive")))
```

```
# Investigating the Relationship between Social Media Usage and Mental Health

# 1 Sample t-test:
# Null Hypothesis (H0): The average impact of social media on mental health
#  is equal to the expected level of impact.

# Alternative Hypothesis (H1): The average impact of social media on mental
# health is different from the expected level of impact.
result <- t.test(data$MHAF, mu = 2, alternative = "two.sided")
result
```

```
##
##  One Sample t-test
##
## data:  data$MHAF
## t = 20.281, df = 231, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 2
## 95 percent confidence interval:
##  3.058512 3.286316
## sample estimates:
## mean of x
##  3.172414
```

```
# The alternative hypothesis is that the true mean is not equal to 2,
# and the 95 percent confidence interval for the mean is between 3.058512 and 3.286316.
# The sample mean is 3.172414.

# Since the p-value is significantly less than the commonly
# used significance level of 0.05, you would reject the null
# hypothesis. This suggests that the average impact of social media on
# mental health is significantly different from the expected level of 2.




# 2 Sample Independent t-test:
# Null Hypothesis (H0): There is no significant difference in mental health scores
#  between heavy social media users and light users.

# Alternative Hypothesis (H1): There is a significant difference in mental health scores between
# heavy social media users and light users
heavy_users <- data[data$UserType == 'Heavy User', 'MHAF']
light_users <- data[data$UserType == 'Light User', 'MHAF']
# Perform the t-test
result <- t.test(heavy_users, light_users)
# Display the t-test result
result
```

```
##
##  Welch Two Sample t-test
##
## data:  heavy_users and light_users
## t = 1.4356, df = 223.47, p-value = 0.1525
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.06151738  0.39165116
## sample estimates:
## mean of x mean of y
##  3.244275  3.079208
```

```
# Since the p-value (0.1525) is greater than the common significance level of 0.05,
#  we fail to reject the null hypothesis. There is insufficient evidence to suggest
#  a significant difference in mental health scores between heavy and light social media users.

# Based on the data, we do not have enough evidence to conclude that there is a significant
# difference in mental health scores between heavy and light social media users.



# One-way ANOVA:
# Null Hypothesis (H0): There is no significant difference in mental health scores
#  across multiple groups based on different social media usage levels.

# Alternative Hypothesis (H1): There is a significant difference in mental health
# scores across multiple groups based on different social media usage levels.
data$SOSP<-as.numeric(
factor(data$SOSP,
    levels = c("Rarely",
               "A few times a month",
               "A few times a week",
               "Once a day",
               "Multiple times a day")))
result_1way_anova <- aov(MHAF ~ SOSP, data = data)
summary(result_1way_anova)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## SOSP          1   0.01  0.0055   0.007  0.933
## Residuals   230 179.10  0.7787
```

```
# The one-way ANOVA results indicate that there is no significant difference
# in mental health scores across different levels of social media usage (SOSP)
# based on the provided data. The p-value (Pr(>F)) is 0.933, which is greater
# than the commonly used significance level of 0.05. Therefore, we do not reject
# the null hypothesis, suggesting no significant variation in mental health scores
#  among the groups defined by social media usage levels.


# Two-way ANOVA:
# H01:There is no significant correlation between social media usage and mental health.
# H11:There is a significant correlation between social media usage and mental health.
# H02:There is no significant difference in mental health scores between individuals residing in different areas.
# H12:There is a significant difference in mental health scores between individuals residing in different areas.
result_2way_anova <- aov(MHAF ~ SOSP * ResidentialArea, data = data)
summary(result_2way_anova )
```

```
##                       Df Sum Sq Mean Sq F value Pr(>F)
## SOSP                   1   0.01  0.0055   0.007  0.933
## ResidentialArea        2   1.07  0.5328   0.679  0.508
## SOSP:ResidentialArea   2   0.67  0.3340   0.426  0.654
## Residuals            226 177.36  0.7848
```

```
# ANOVA results suggest that none of the factors (SOSP, ResidentialArea, and their interaction)
#  have a significant impact on the dependent variable MHAF (mental health scores).
#   This is based on the p-values obtained from the F-tests.

# For 'SOSP' (social media usage), the p-value is 0.933 (> 0.05),indicating that there is no
# significant difference in mental health scores betweendifferent levels of social media usage.

# For 'ResidentialArea' (areas of residence), the p-value is 0.508 (> 0.05),
#  suggesting that there is no significant difference in mental health scores
#  among individuals residing in different areas.

# The interaction term 'SOSP:ResidentialArea' also has a p-value of 0.654 (> 0.05),
#  indicating that the combined effect of social media usage and residential area is
#   not significantly associated with mental health scores.

# Overall, based on this analysis, there is no significant evidence
#  to reject the null hypothesis for any of the factors or their interaction.
```

```r
# Exploring the Association between Social Media Interaction and Real-life Social Relationships:

# 1 Sample t-test:
# Null Hypothesis (H0): The average impact of social interaction on social media
#  on real-life relationships is equal to the expected level of impact.

# Alternative Hypothesis (H1): The average impact of social interaction on
# social media on real-life relationships is different from the expected level of impact.

# Let the expected value be 4 assuming it is HIGH
expected_level_of_impact <- 4
# One-sample t-test
t_test_result <- t.test(data$SIEFF, mu = expected_level_of_impact, alternative = "two.sided")
# Print the result
print(t_test_result)
```

```
##
##  One Sample t-test
##
## data:  data$SIEFF
## t = -7.6103, df = 231, p-value = 6.911e-13
## alternative hypothesis: true mean is not equal to 4
## 95 percent confidence interval:
##  3.289157 3.581533
## sample estimates:
## mean of x
##  3.435345
```

```r
# The alternative hypothesis is that the true mean is not equal to 4 (the expected level of impact).
# The 95% confidence interval for the mean of social interaction scores is (3.289157,3.581533)

# With such a low p-value, you would reject the null hypothesis,
# suggesting that the average impact of social interaction on social
# media on real-life relationships is significantly different from the

# expected level of impact (4). The negative t-value indicates
#  that the mean is significantly less than 4.



# 2 Sample Independent t-test:
# Null Hypothesis (H0): There is no significant difference in real-life relationship scores
#  between individuals with positive and negative perceptions of social media impact.

# Alternative Hypothesis (H1): There is a significant difference in real-life relationship
# scores between individuals with positive and negative perceptions of social media impact.
# Create a new column 'PerceptionCategory' based on MHAF values
data$PerceptionCategory <- ifelse(data$MHAF %in% c(1, 2), 'Negative', 'Positive')
# Filter the data based on PerceptionCategory
negative_data <- subset(data, PerceptionCategory == 'Negative')
positive_data <- subset(data, PerceptionCategory == 'Positive')
# Perform 2 Sample Independent t-test
result <- t.test(negative_data$SIEFF, positive_data$SIEFF)
# Print the result
print(result)
```

```
##
##  Welch Two Sample t-test
##
## data:  negative_data$SIEFF and positive_data$SIEFF
## t = -3.6446, df = 89.469, p-value = 0.0004489
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.9513258 -0.2800516
## sample estimates:
## mean of x mean of y
##  2.962963  3.578652
```

```r
# The negative p-value suggests strong evidence against the null hypothesis.
# Therefore, We can reject the null hypothesis and conclude that there is
#  a significant difference in the 'SIEFF' variable between individuals with
#  negative and positive perceptions of social media impact.




# One-way ANOVA:
# Null Hypothesis (H0): There is no significant difference in real-life relationship scores across multiple
#  groups based on different levels of social interaction on social media.
# Alternative Hypothesis (H1): There is a significant difference in real-life relationship scores across multiple
#  groups based on different levels of social interaction on social media.

# Create groups based on different levels of social interaction on social media
data$SIEFF_Groups <- cut(data$SIEFF, breaks = c(-Inf, 2, 4, Inf), labels = c("Low", "Medium", "High"))

# Perform one-way ANOVA
result_1way_anova <- aov(MHAF ~ SIEFF_Groups, data = data)

# Print the ANOVA table
summary(result_1way_anova)
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## SIEFF_Groups   2   8.25   4.123   5.526 0.00453 **
## Residuals    229 170.86   0.746
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# The ANOVA table shows that the factor "SIEFF_Groups" is significant (p-value = 0.00453),
# indicating that there is a significant difference in real-life relationship scores across
# different levels of social interaction on social media. The asterisks indicate the level of
#  significance, with ** meaning p-value < 0.01.



# Two-way ANOVA:
# H01:Social interaction on social media has no significant impact on real-life social relationships.
# H11:Social interaction on social media positively influences real-life social relationships.
# H02:There is no significant difference in real-life relationship scores between individuals of different age groups.
# H12:There is a significant difference in real-life relationship scores between individuals of different age groups.

# result_2way_anova <- aov(SIEFF ~ AgeGroup * positive_data, data = data)
result_2way_anova <- aov(SIEFF ~ SOSP * AgeGroup, data = data)

# Print the ANOVA table
summary(result_2way_anova)
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## SOSP           1   0.02   0.024   0.020 0.88867
## AgeGroup       4  17.51   4.377   3.602 0.00722 **
## SOSP:AgeGroup  4   7.71   1.928   1.586 0.17886
## Residuals    222 269.79   1.215
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# The Analysis suggests that both "AgeGroup" and "PositivePerception" have a significant impact on the "SIEFF" variable,
# as indicated by the low p-values (Pr(>F)).

# AgeGroup: The p-value is less than 0.05 (***), indicating that there is a significant difference in social
# interaction effects among different age groups.

# PositivePerception: The p-value is extremely low (< 2e-16) (***), suggesting a highly significant difference
# in social interaction effects between individuals with positive and negative perceptions.

# Interaction (AgeGroup:PositivePerception): The p-value is 0.0274 (*), indicating a significant interaction
# effect between AgeGroup and PositivePerception on social interaction.

# This implies that both age group and perception of social media impact significantly contribute to the variations
#  observed in the "SIEFF" variable, and there is also an interaction effect between age group and perception.
```