2021101113_Hypothesis_Testing_Assignment

Gowlapalli Rohit

8/2/2024

Contents

The following is the code for the data shown in the slides (Since the data on shared excel sheet was edited , the prof asked students to work on data given in the slides)

```
library(readxl)
library(ggplot2)
library(tidyr)
library(dplyr)
library(reshape2)
set.seed(0)
```

Normality Test

```
anxiety_exercise <- c(20, 23, 25, 30, 35, 29, 37, 24, 29, 31, 26, 28)
anxiety_no_exercise <- c(24, 35, 41, 21, 38, 23, 37, 44, 32, 33, 34, 42)
shapiro.test(anxiety_exercise)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  anxiety_exercise
## W = 0.97869, p-value = 0.9779
```

```
shapiro.test(anxiety_no_exercise)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  anxiety_no_exercise
## W = 0.92953, p-value = 0.3751
```

Based on the Shapiro-Wilk normality test results:

For the anxiety_exercise group:

Shapiro-Wilk statistic (W) = 0.97869 p-value = 0.9779 Since the p-value is greater than the typical significance level of 0.05, we fail to reject the null hypothesis. Therefore, we can infer that the data for anxiety levels in the exercise group follows a normal distribution.

For the anxiety_no_exercise group:

Shapiro-Wilk statistic (W) = 0.92953 p-value = 0.3751 Again, with a p-value greater than 0.05, we fail to reject the null hypothesis. Hence, we can infer that the data for anxiety levels in the no-exercise group also follows a normal distribution.

In summary, both groups' anxiety level data appear to be normally distributed based on the Shapiro-Wilk test, indicating that parametric tests such as the independent t-test can be applied reliably to compare the means of these two groups.

Choosing the right test

```r
t_test_result <- t.test(anxiety_exercise, anxiety_no_exercise,var.equal = TRUE)
print(t_test_result)
```

```
##
##  Two Sample t-test
##
## data:  anxiety_exercise and anxiety_no_exercise
## t = -2.1435, df = 22, p-value = 0.04338
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.9852541  -0.1814126
## sample estimates:
## mean of x mean of y
##  28.08333  33.66667
```

- t-Statistic:

The calculated t-statistic is -2.1435. This value represents the magnitude of the difference between the means of the two groups relative to the variability within the groups. The negative sign indicates that the mean anxiety level in the exercise group is lower than in the no-exercise group.

- Degrees of Freedom:

The degrees of freedom (df) associated with the t-test is 22. This value is determined by the sample sizes of both groups and is used to calculate the critical t-value.

- p-value:

The p-value associated with the t-test is 0.04338. This value represents the probability of observing the given t-statistic (or more extreme) under the null hypothesis that the true difference in means between the two groups is equal to zero. Since the p-value is less than the typical significance level of 0.05, we reject the null hypothesis and conclude that there is a significant difference in anxiety levels between the exercise and no-exercise groups.

- Confidence Interval:

The 95% confidence interval for the difference in means is calculated as (-10.9852541, -0.1814126). This interval provides a range of plausible values for the true difference in means between the two groups. Since the interval does not include zero, it further supports the conclusion that the mean anxiety level is significantly lower in the exercise group compared to the no-exercise group.

- Sample Estimates:

The sample estimates for the mean anxiety levels are provided as follows:

Mean anxiety level in the exercise group (mean of x): 28.08333

Mean anxiety level in the no-exercise group (mean of y): 33.66667

In summary, based on the results of the two-sample t-test, we can infer that there is a significant difference in anxiety levels between individuals who engage in regular exercise and those who do not. Specifically, individuals in the exercise group have lower mean anxiety levels compared to those in the no-exercise group.

Calculating Effect-size

```
mean_diff <- mean(anxiety_no_exercise) - mean(anxiety_exercise)
pooled_sd <- sqrt(((length(anxiety_exercise) - 1) * var(anxiety_exercise) + (length(anxiety_no_exercise)
effect_size <- mean_diff / pooled_sd
cat("Cohens-d value:",effect_size, "\n")
```

```
## Cohens-d value: 0.8750883
```

Cohen's d value of 0.8750883 indicates a large effect size according to Cohen's guidelines for interpreting effect size, where values around 0.2, 0.5, and 0.8 represent small, medium, and large effect sizes, respectively. The large effect size suggests that there is a substantial difference in anxiety levels between individuals who engage in regular exercise and those who do not. This finding underscores the potential impact of exercise on reducing anxiety levels and highlights the importance of promoting regular physical activity for mental well-being.

Reporting Statistics

```
cat("Mean (exercise):", mean(anxiety_exercise), "\n")
```

```
## Mean (exercise): 28.08333
```

```r
cat("Mean (no exercise):", mean(anxiety_no_exercise), "\n")
```

```
## Mean (no exercise): 33.66667
```

```r
cat("Variance (exercise):", var(anxiety_exercise), "\n")
```

```
## Variance (exercise): 23.90152
```

```r
cat("Variance (no exercise):", var(anxiety_no_exercise), "\n")
```

```
## Variance (no exercise): 57.51515
```

```r
cat("Observations (exercise):", length(anxiety_exercise), "\n")
```

```
## Observations (exercise): 12
```

```r
cat("Observations (no exercise):", length(anxiety_no_exercise), "\n")
```

```
## Observations (no exercise): 12
```

```r
cat("Pooled Variance:", pooled_sd^2, "\n")
```

```
## Pooled Variance: 40.70833
```

```r
cat("Hypothesized Mean Difference:", t_test_result$null.value, "\n")
```

```
## Hypothesized Mean Difference: 0
```

```r
cat("Degrees of Freedom:", t_test_result$parameter, "\n")
```

```
## Degrees of Freedom: 22
```

```r
cat("t Stat:", t_test_result$statistic, "\n")
```

```
## t Stat: -2.14352
```

```r
cat("P(T<=t) one-tail:", t_test_result$p.value / 2, "\n")
```

```
## P(T<=t) one-tail: 0.02169075
```

```r
cat("t-Critical on one-tail:", qt(0.05, t_test_result$parameter, lower.tail = FALSE), "\n")
```

```
## t-Critical on one-tail: 1.717144
```

```r
cat("P(T<=t) two-tail:", t_test_result$p.value, "\n")
```

```
## P(T<=t) two-tail: 0.0433815
```

```r
cat("t-critical two-tail:", qt(0.025, t_test_result$parameter, lower.tail = FALSE), "\n")
```

```
## t-critical two-tail: 2.073873
```

- Mean Comparison:

The mean anxiety level for the exercise group is 28.08333, while for the no-exercise group, it is 33.66667. This indicates that, on average, individuals who engage in regular exercise have lower anxiety levels compared to those who do not.

- Variance Comparison:

The variance of anxiety levels in the exercise group (23.90152) is lower than that in the no-exercise group (57.51515). This suggests that anxiety levels within the exercise group are more consistent or less dispersed compared to the no-exercise group.

- Sample Size:

Both groups have the same sample size of 12 observations each, ensuring balanced comparison between the groups.

- Effect Size (Pooled Variance):

The pooled variance, calculated as 40.70833, is used to assess the magnitude of the difference in means between the two groups. A higher pooled variance suggests greater variability in the combined data set.

- Hypothesized Mean Difference:

The null hypothesis assumes that there is no difference in the mean anxiety levels between the exercise and no-exercise groups (Hypothesized Mean Difference = 0).

- Degrees of Freedom:

The degrees of freedom for the t-test are calculated as 22, which is determined by the sample sizes of both groups (12 each).

- t Statistic:

The calculated t statistic is -2.14352. This indicates the magnitude of the difference between the sample means relative to the variability of the data. A negative t statistic suggests that the mean anxiety level in the exercise group is lower than in the no-exercise group.

- p-values:

The p-value for a one-tailed test is 0.02169075, and for a two-tailed test, it is 0.0433815. These values represent the probability of observing the given t statistic (or more extreme) under the null hypothesis. Since both p-values are less than the typical significance level of 0.05, we reject the null hypothesis and conclude that there is a significant difference in anxiety levels between the two groups.

- t-Critical Values:

The critical t-values for both one-tailed (1.717144) and two-tailed (2.073873) tests are provided. These values represent the threshold beyond which we would reject the null hypothesis. In this case, the absolute value of the calculated t-statistic (-2.14352) exceeds both critical values, supporting the rejection of the null hypothesis.

- Conclusion:

In conclusion, based on the results of the independent t-test, we can infer that engaging in regular exercise is associated with significantly lower anxiety levels compared to not engaging in exercise.