Inferential Statistics Report

Minerva University

CS50: Formal Analyses

Prof. Stan

December 9, 2021

Contents

Introduction

Anxiety leads to reduction in cognitive function of the brain (Broeren & Muris, 2009). When overall physical activity of a person reduces, they face multiple mental and physical health problems like depression, cardiovascular diseases, diabetes (Warburton, 2006). We are finding an answer for whether there is significant evidence that people with higher levels of anxiety participate in different durations of physical activity then people with lower levels of anxiety. This analysis is conducted using a difference of means test and constructing confidence intervals to find insight into the theory that exercise provides distraction that can lead to reduced anxiety levels (Anderson & Shivakumar, 2013); or during high levels of anxiety a person is unable to participate in physical activity.

Data Set

Medial Temporal Lobe data has been taken from 10 men and 25 women between ages of 45 and 75. The data comes from a conducted study (Siddarth et al., 2018) and was available online (OpenIntro, n.d.). The first 10 rows of the dataset are shown (Appendix A). For this analysis we will be focusing on two variables: ham_a and sitting.

Ham_a is a Hamilton Rating Scale for anxiety. It was made to assess the level of anxiety on a 14 point scale using symptom diagnosis (University of Florida, n.d.). Ham_a acts as a qualitative ordinal variable as although it stands to have numerical

value, each numerical value denotes a degree of symptom. Additionally it is ordinal as the 14 point scale is ordered with the difference between each point unknown and depends on the person marking the value. It is being used as we need to check the anxiety levels of participants for our research question. The second variable is **sitting**, a quantitative discrete variable for the purpose of this study, as it is rounded to the nearest hour. Sitting is the self reported time by each participant and is being taken as a proxy variable for the participation in physical activity for analysis. Given that we are doing a difference of means test, hence we don't need to differentiate our variables into independent and dependent categories as we are taking sitting durations of two groups to examine higher or lower anxiety levels.[1] Additionally, we filtered out two female participants due to missing values in ham_a. Hence moving forward with 33 participants.

Analysis

Hypotheses

The data is divided into two groups for hypotheses (Appendix B):

High anxiety group (HAG) = Sitting duration of $Ham\_a \geq 4$

Low anxiety group (LAG) = Sitting duration of $Ham\_a < 4$

Null and Alternative hypotheses are formulated:

$$H_0: \mu_1 = \mu_2$$

[1] **#variables:** I have accurately identified, classified and justified the types of my variables. Given that we are doing difference of means tests, I have justified why there is no significant need of categorizing variables into independent and dependent variables.

The mean of sitting durations on participants with high-anxiety levels is equivalent to that of low-anxiety levels.

$$H_A: \mu_1 < \mu_2$$

The mean of sitting durations on participants with high-anxiety levels is less than that of low-anxiety levels.

The study is inconclusive if we fail to reject the null hypothesis. Therefore, no significant inference will be drawn about anxiety levels impact on physical activity. One-tailed test is examined as the focus is if sitting duration of high-anxiety levels is less than that of low-anxiety levels. $\alpha = 0.05$ is set as a significance level to assess the rejectability of the null hypothesis. Using scientific field's conventions, a threshold of 5% risk of committing a Type I error (rejecting the null hypothesis when it is true) is assumed to conclude that the probability of getting a sample mean more extreme than the observed $\bar{x}$ (sample mean) is not likely by random chance. Additionally, a 95% confidence interval is set to ensure that we will be 95% confident that the results capture the true population parameter. Due to computation of one p-value for a single difference of means test, bonferroni corrections are not needed.

Summary Statistics

The dataset was evaluated using Python: panda and seaborn packages for stylistic choices (Appendix A & B). Summary statistics and histograms of each group are represented (Fig 1 & Fig 2).

	High Anxiety Group (HAG)	Low Anxiety Group (LAG)
Sample Size	$n_1 = 17$	$n_2 = 16$
Mean	$\bar{x}_1 = 7.06$	$\bar{x}_2 = 7$
Standard Deviation	$s_1 = 2.93$	$s_1 = 3.78$
Median	7	6.5
Mode	10	5
Range	10	13

Fig. 1 shows summary statistics of High (HAG) and Low (LAG) anxiety levels in form of mean, standard deviation, median, mode, and range.

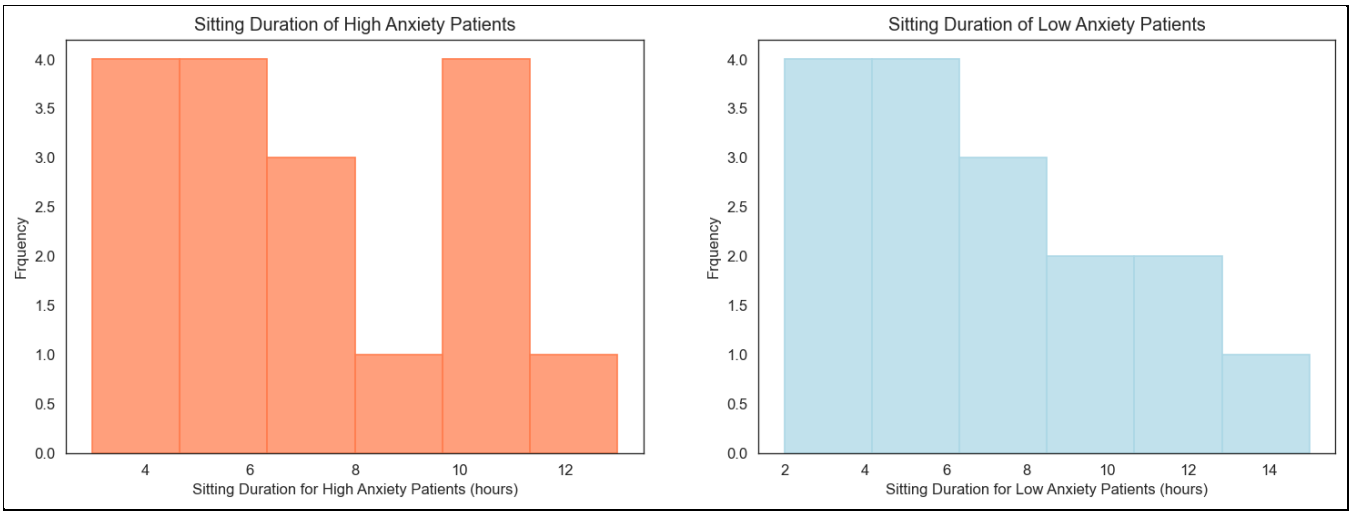


Fig. 2 shows histogram of the sitting duration of high anxiety group and low anxiety group. Both the distributions are skewed right with Sitting duration of respective anxiety patients in hours on x-axis and frequency on y-axis.

Both the distributions for anxiety groups are skewed to the right meaning that more participants with either anxiety level sit for less durations. The mean of the groups slightly differ, with the HAG having a higher mean of 7.06 stipulating that people are more active in HAG on average.[2]

Furthermore, the LAG has a range of 13 and standard deviation of 3.78, whereas, HAG has 10 and 2.93 , respectively (Appendix B). As range is the spread of the data, greater range stipulates greater variability because the data points are more dispersed. The standard deviation tells how the measurements for each subgroup are spread out from the mean, indicating the variability compared to other subgroups. As HAG's standard deviation is lower than that of LAG, LAG's data points are closer to the mean.[3]

[2] **#dataviz:** I have plotted the data effectively using appropriate means of plotting. All necessary components are included e.g. captions, x and y labels with units. I have elaborated the trends with inferences that could be made and have chosen bin sizes to reflect smooth plots.

[3] **#descriptivestats:** I have shown all required descriptive statistics with further elaboration on the ones that are most needed. I have shown clear steps of using libraries to find the statistics, found appropriate and smart interpretations from the statistics and have plotted histograms for each subgroup with multiple descriptive statistics to make a strong analysis.

Results

Conditions for Inference

T-distribution and difference of means test is used to calculate t-value (t), p-value (p), confidence interval (CI), and effect size (d) by ensuring subgroups are independent. Hence the following conditions need to be met:

1. The data set is a simple random sample (33 middle-aged people with anxiety levels) of less than 10% of the population(U.S Census Bureau, n.d.) (National Institute of Mental Health, n.d.).
2. The sample size is sufficiently large (n > 30).
3. The skew and outliers are moderately too extreme and acceptable considering the moderate sample size (n = 16)

Accordingly, we cannot conclude with certainty that these two sub-groups are independent as the sample size of the subgroups is n < 30. However, we will assume that the samples are independent for the continuation of the study while keeping in mind the uncertainty of the results. Then, we measure the p-value to quantify the strength of the evidence against H_0 and in favor of H_A. The t-distribution is used for the sampling distribution of sample mean when H_0 is true because the population standard deviation is unknown. Given that we are assuming the sample size is approximately > 30, we can use CLT and conclude that the t-distribution approaches a normal distribution.[4]

[4] **#distributions:** I have provided appropriate context of the chosen distribution with relevant calculations needed to fulfill CLT to make a normal distribution. I have mentioned and justified the limitations to make a normal distribution due to sample size and its consequences on the study.

Differences of Means Test

Given that my sample size of each subgroup < 30 hence I will use t-distribution instead of z-distribution. To assess the statistical significance, p-value is computed using the t-value ($t = \frac{abs(\bar{x}_2 - \bar{x}_1)}{SE}$) and Standard Error $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$. A conventional estimate for degrees of freedom of the t-distribution is taken using df =15. Using Python, t-value is calculated: t = 0.0498. (Appendix D). T-value represents the number of standard deviations away from the mean in a normal distribution; t-value closer to 0 provides evidence for the null hypothesis. Hence, t-value of 0.0498 indicates that there is no significant difference between the means. Using t-value, a p-value is calculated in python using cdf function, converting t-score to the area (Appendix D). Conducting a one-tailed test, p-value = 0.48 (Appendix D. Given that p-value is the probability of observing data at least as favorable to the alternative hypothesis if the null hypothesis is true, there is a 48% chance of obtaining a statistically significant result due to random chance. As p-value > alpha, we fail to reject the null hypothesis.[5]

The effect size measures the practical significance. However, p < alpha calculation of an effect size is not needed as we concluded that the intervention does not work. To show that we understand the concept, effect size is calculated (Appendix E) using

[5] **#probability:** I successfully calculated a p-value and justified its need. I mentioned the method of calculation with detailed steps mentioned in text and shown in the appendix. I made relevant and reasonable inferences from it while comparing it to the alpha value.

Cohen's d as each subgroup has a different sample size. We result in an effect size of approximately 0.018 (Appendix E) . From the common interpretation of effect size, effect is small when the d-value is near absolute significance of 0.3. As our d-value is less than 0.3, we conclude a small practical difference in means: the size of the strength difference between HAG and LAG is small.[6]

Confidence Intervals

We calculate confidence intervals as significance tests neglects some p-values on the borderline. So the confidence level is constructed to define the distance for how close the confidence limits are to sample mean. Because we set alpha as 0.05, we use the CI 95% for the confidence interval as both of them need to coincide to give the accurate values. CI is:

CI for whole sample = [4.6244, 9.4362]

CI for Differences of Means Test = [-2.129, 2.012]

First evaluates whether the null value is within the CI; If the null value is in the range of plausible values from CI, we fail to reject H_0 as the null value represents the value of the parameter if the null hypothesis is true. CI for difference of means captures the true difference in true population mean.

[6] **#significance:** I effectively calculated statistical and practical significance tests and interpreted the results with relevant interpretations. I clarified the insignificance of calculating effect size and reason for choosing Cohen's d when we failed to reject the null hypothesis to show understanding of the practical significance concept. Also

Critical-t-value (t*) is used for 95% CI to indicate the length of distance we need to cover on both sides from the center of a t-distribution to obtain an area of 95% (Appendix F). The null value is equivalent to the LAG mean because it represents the value of the parameter if the null hypothesis is true. Hence, $\mu = 7\ hours$. The null value is within the 95% CI of the sample; therefore, we fail to reject H_0. Given that CI for Differences of Means Test is very narrow with small lower and upper bounds. This indicates that we are 95% confident that the true difference in true population mean for subgroups lies within -2.129 and 2.012 due to small range (Appendix F). However, as the number represents the differences in means, the small number indicates that we are 95% confident that we have a good chance of finding a minimum difference in sitting duration for each subgroup. By Evaluating CIs, we extract two conclusions:

1. Further evidence in failing to reject H_0 as the null value is within the sample CI.
2. Minute difference in true difference in true population for subgroups.

These conclusions are derived with the assumption that each subgroups are simple random samples, abiding by the 10% rule for independence.[7]

Conclusion

Observations made are: The p-value is less than alpha; The null value is within the sample confidence interval; The effect size is less than 0.3. Due to the

[7] #**confidenceintervals**: I have discussed possible types of confidence intervals and have made the most significant one with analysis and detailed justification. I have successfully applied confidence intervals to estimate parameters of the population like CI for difference of means for true difference in true population mean to make reasonable conclusions.

observation of these conditions, we cannot generalize conclusions about the population from the sample. Several assumptions are made, but because they are true for this sample. Hence, the derived inductive conclusion is strong and reliable with the exception that CLT was not fully met due to a sample size < 30. However, limitations exist because several assumptions are made using CLT. To strengthen the argument, we need to adjust the significance level to 0.1, committing less Type II errors.

Hence, this study is inconclusive, and we cannot conclude any reasonable difference in effect of HAGs on sitting duration compared to LAGs. Also due to small sample size we cannot infer the population. Hence, further research is needed.[8]

Word Count: 1605[9]

[8] **#induction:** I have stated an inductive argument using the observations made whose type is identified and justified. I have further discussed the strength and reliability of the argument and the methods to improve them.

[9] **#professionalism:** I have provided a word count, I have restricted myself not to exceed the 10% word count rule. I have included relevant figures to communicate effectively and have provided captains to clarify the content of the figures. Lastly, I have provided bookmarks for the appendix to make it easy for the professor to view my appendix while reading the text; however, sometimes these links don't work on the forum but it is added for ease.

References

1. Anderson, E., & Shivakumar, G. (2013). Effects of exercise and physical activity on anxiety. Frontiers in Psychiatry, 4. https://doi.org/10.3389/fpsyt.2013.00027
2. Broeren, S., & Muris, P. (2009). The relation between cognitive development and anxiety phenomena in children. Journal of Child and Family Studies, 18(6), 702–709. https://doi.org/10.1007/s10826-009-9276-8
3. OpenIntro. (n.d.). *Medial temporal lobe (MTL) and other data for 26 participants*. OpenIntro. Retrieved December 12, 2021, from https://www.openintro.org/data/index.php?data=mtl
4. Siddarth, P., Burggren, A. C., Eyre, H. A., Small, G. W., & Merrill, D. A. (2018). Sedentary behavior associated with reduced medial temporal lobe thickness in middle-aged and older adults. PLOS ONE, 13(4). https://doi.org/10.1371/journal.pone.0195549
5. U.S. Census Bureau. (n.d.). U.S. population by sex and age. Infoplease. Retrieved December 13, 2021, from https://www.infoplease.com/us/population/us-population-sex-and-age.
6. U.S. Department of Health and Human Services. (n.d.). Any anxiety disorder. National Institute of Mental Health. Retrieved December 13, 2021, from https://www.nimh.nih.gov/health/statistics/any-anxiety-disorder.
7. University of Florida. (n.d.). Hamilton Anxiety Rating Scale (HAM-A) - university of Florida. Retrieved December 13, 2021, from https://dcf.psychiatry.ufl.edu/files/2011/05/HAMILTON-ANXIETY.pdf.

8. Warburton, D. E. R. (2006). Health benefits of physical activity: The evidence. Canadian Medical Association Journal, 174(6), 801–809. https://doi.org/10.1503/cmaj.051351

Reflection

I appreciated the professor's feedback on variables and how I can get a better score in it by providing detailed distinction between types of variables and their use. This feedback provided me effective guidance to look out for potholes while doing my current assignment. Furthermore, I carefully used professor's advice on using significance tests and effectively using CLT to prove I can use a normal distribution. The discussion in the class about the irrelevance of plotting sampling distribution really saved a lot of time here and gave me a headstart in the assignment.

Having these tools of significance tests and confidence intervals would have helped me analyze which variables I should use and whether I could make an effective conclusion from it and it could help me know whether to continue with this sample collection as the number of ambulances in a specific duration is a hard variable to collect.

Word Count: 150

Appendix

The full Jupyter notebook file and the data can be accessed in the zipped folder submitted as a secondary file.

Appendix A: Import, Analyze, and Visualize Data

```
#import relevant packages
import pandas as pd
import numpy as np
import matplotlib
from matplotlib import pyplot as plt
import statistics
import scipy.stats
from scipy.stats import norm
import seaborn as sns
```

```
col_list = ["ham_a", "sitting"]
df = pd.read_csv("mtl.csv", usecols=col_list)#importing two specfic columns from dataset
df.head(10) #display the first 10 rows of the dataset
```

	ham_a	sitting
0	9	10
1	4	11
2	0	5
3	9	7
4	1	3
5	0	5
6	12	4
7	7	6
8	4	7
9	4	10

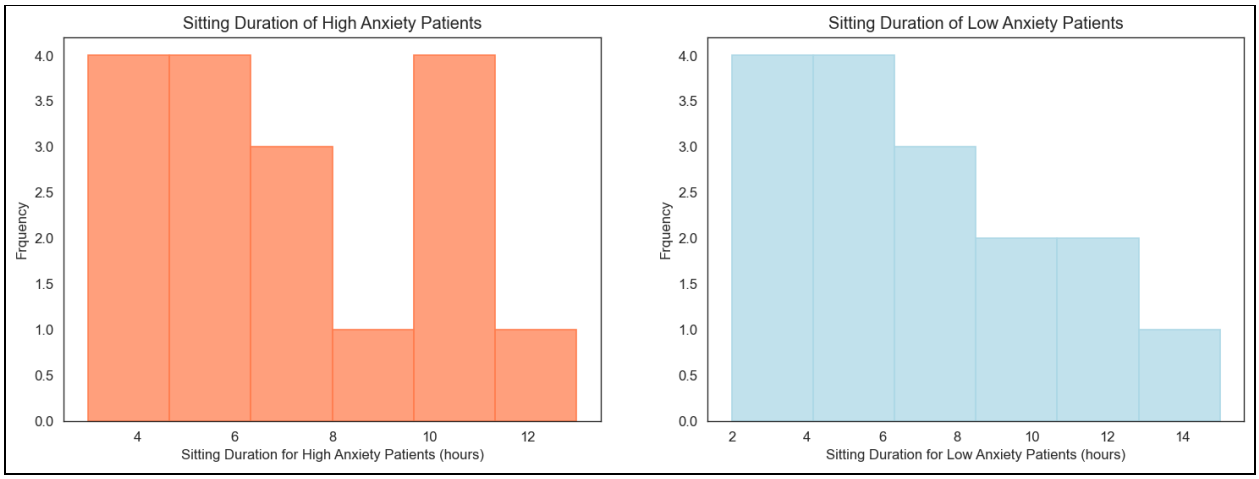
Appendix B: Examine the Subgroups

```
# the function shows the statistical summary of the high_anxiety and low_anxiety groups
def descritpive_stats(data):

    # retrieve dataset for high ham_a and low ham_a sitting duration
    sitting_high_anxiety = data[data["ham_a"] >= 4]["sitting"] # high_anxiety group
    sitting_low_anxiety = data[data["ham_a"] < 4]["sitting"]  # low_anxiety group

    # calculates high_anxiety group and low_anxiety group means
    sitting_high_anxiety_mean = statistics.mean(sitting_high_anxiety)
    sitting_low_anxiety_mean = statistics.mean(sitting_low_anxiety)

    # calculates high_anxiety group and low_anxiety group stdev
    sitting_high_anxiety_stdev = statistics.stdev(sitting_high_anxiety)
    sitting_low_anxiety_stdev = statistics.stdev(sitting_low_anxiety)

    # calculates high_anxiety group and low_anxiety group median
    sitting_high_anxiety_median = statistics.median(sitting_high_anxiety)
    sitting_low_anxiety_median = statistics.median(sitting_low_anxiety)

    # calculates high_anxiety group and low_anxiety group mode
    sitting_high_anxiety_mode = statistics.mode(sitting_high_anxiety)
    sitting_low_anxiety_mode = statistics.mode(sitting_low_anxiety)

    # calculates high_anxiety group and low_anxiety group range
    sitting_high_anxiety_range = max(sitting_high_anxiety) - min(sitting_high_anxiety)
    sitting_low_anxiety_range = max(sitting_low_anxiety) - min(sitting_low_anxiety)

    print("\t High Anxiety \t Low Anxiety")
    print("sample: ", (len(sitting_high_anxiety)), "\t\t", (len(sitting_low_anxiety)))
    print("mean:   ", round(sitting_high_anxiety_mean,2), "\t\t", round(sitting_low_anxiety_mean,2))
    print("stdev:  ", round(sitting_high_anxiety_stdev,2), "\t\t", round(sitting_low_anxiety_stdev,2))
    print("median: ", round(sitting_high_anxiety_median,2), "\t\t", round(sitting_low_anxiety_median,2))
    print("mode:   ", sitting_high_anxiety_mode, "\t\t", sitting_low_anxiety_mode)
    print("range:  ", sitting_high_anxiety_range, "\t\t", sitting_low_anxiety_range)

descritpive_stats(df)
```

	High Anxiety	Low Anxiety
sample:	17	16
mean:	7.06	7
stdev:	2.93	3.78
median:	7	6.5
mode:	10	5
range:	10	13

Appendix C: Histograms of the Subgroups

```
# this plots the histograms to see distributions of the samples of high_anxiety and low_anxiety groups
def histogram_plot(data):

    # retrieve dataset for high ham_a and low ham_a sitting duration
    sitting_high_anxiety = data[data["ham_a"] >= 4]["sitting"] # high_anxiety group
    sitting_low_anxiety = data[data["ham_a"] < 4]["sitting"]  # low_anxiety group

    # This plots the histograms parallel to each other
    fig, axs, = plt.subplots(1,2, figsize = (15, 5), dpi = 120)

    # This plots histogram for high_anxiety and low_anxiety groups
    plt1 = sns.histplot(sitting_high_anxiety, bins = 6,
                        edgecolor = "coral", color = "coral", ax=axs[0])
    plt2 = sns.histplot(sitting_low_anxiety, bins = 6,
                        edgecolor = "lightblue", color = "lightblue", ax=axs[1])

    # titles of axes and plots
    plt1.set(xlabel = "Sitting Duration for High Anxiety Patients (hours)", ylabel = "Frquency")
    plt2.set(xlabel = "Sitting Duration for Low Anxiety Patients (hours)", ylabel = "Frquency")

    plt1.title.set_text('Sitting Duration of High Anxiety Patients')
    plt.title('Sitting Duration of Low Anxiety Patients')

    plt.show()

histogram_plot(df)
```



Appendix D: Statistical Significance Difference of Means Test

```
# the function calculates p-vaue using a t-distribution and a difference of means test
def p_value(data):

    # retrieve dataset for high ham_a and low ham_a sitting duration
    sitting_high_anxiety = data[data["ham_a"] >= 4]["sitting"] # high_anxiety group
    sitting_low_anxiety = data[data["ham_a"] < 4]["sitting"]  # low_anxiety group

    # calculates high_anxiety group and low_anxiety group means
    high_anxiety_mean = statistics.mean(sitting_high_anxiety)
    low_anxiety_mean = statistics.mean(sitting_low_anxiety)

# assign values

    alpha = 0.05

    # sample size of subgroups
    n1 = len(sitting_high_anxiety)
    n2 = len(sitting_low_anxiety)

    # sample mean of each subgroup
    x1 = high_anxiety_mean
    x2 = low_anxiety_mean

    # sample stdev for each subgroup
    s1 = np.std(sitting_high_anxiety, ddof = 1) #ddof = 1 to account for sample
    s2 = np.std(sitting_low_anxiety, ddof = 1)

    # calulate SE for difference of means test
    SE = np.sqrt((s1**2)/n1 + (s2**2)/n2)

    # calculate t-value
    t = abs(x2 - x1)/SE #abs is used to eliminate negative sign
    dof = min(n1,n2) - 1 # degrees of freedom for difference of means test

    # calculate p-value
    p = scipy.stats.t.cdf(-t, dof)
    print("t-value:", round(t,4))
    print("p-value:", round(p,2))

    # reject null hypothesis
    if p < alpha:
        print("p-value = P (X > 0.05) which is approximately {}.".format(round(p,2)))
        print("We reject the Null Hypothesis.")

   # fail to reject null hypothesis
   else:
       print("p-value = P (X < 0.05) which is approximately {}.".format(round(p,2)))
       print("We fail to reject the Null Hypothesis.")

p_value(df)
```

```
t-value: 0.0498
p-value: 0.48
p-value = P (X < 0.05) which is approximately 0.48.
We fail to reject the Null Hypothesis.
```

Appendix E: Practical Significance Test (Effect Size)

```
# this function calculates the effect size of the given sample
def effect_size(data):

    # retrieve dataset for high ham_a and low ham_a sitting duration
    sitting_high_anxiety = data[data["ham_a"] >= 4]["sitting"] # high_anxiety group
    sitting_low_anxiety = data[data["ham_a"] < 4]["sitting"]  # low_anxiety group

    # sample size of subgroups
    n1 = len(sitting_high_anxiety)
    n2 = len(sitting_low_anxiety)

    # calculates high_anxiety group and low_anxiety group means
    high_anxiety_mean = statistics.mean(sitting_high_anxiety)
    low_anxiety_mean = statistics.mean(sitting_low_anxiety)

    # sample stdev for each subgroup
    s1 = np.std(sitting_high_anxiety, ddof = 1) #ddof = 1 to account for sample
    s2 = np.std(sitting_low_anxiety, ddof = 1)

    # calculates pooled stdev
    sd_pooled = (s1 + s2)/2

    # effect size: Cohen's d (practical significance)
    d = (high_anxiety_mean - low_anxiety_mean)/sd_pooled

    print("Effect size: {}".format(round(d,3)))
effect_size(df)
```

```
Effect size: 0.018
```

Appendix F: Confidence Interval

```
# this function finds the Confidence interval for the sample and difference of means test
def confidence_interval(data):

     # retrieve dataset for high ham_a and low ham_a sitting duration
    sitting_high_anxiety = data[data["ham_a"] >= 4]["sitting"] # high_anxiety group
    sitting_low_anxiety = data[data["ham_a"] < 4]["sitting"]  # low_anxiety group
    sample = data["sitting"]

    # calculates high_anxiety group and low_anxiety group means
    high_anxiety_mean = statistics.mean(sitting_high_anxiety)
    low_anxiety_mean = statistics.mean(sitting_low_anxiety)

# assign values

    alpha = 0.05

    x̄ = statistics.mean(sample)

    n = len(sample)
    n1 = len(sitting_high_anxiety)
    n2 = len(sitting_low_anxiety)

    x1 = high_anxiety_mean
    x2 = low_anxiety_mean
    null_value = low_anxiety_mean

    s1 = np.std(sitting_high_anxiety, ddof = 1) #ddof = 1 to account for sample
    s2 = np.std(sitting_low_anxiety, ddof = 1)

    stdev = np.std(sample, ddof = 1)
    # SE for difference of means
    SE = np.sqrt((s1**2)/n1 + (s2**2)/n2)

    dof_difference_of_means = min(n1, n2) - 1

    # calculates critical t-value
    critical_tvalue = scipy.stats.t.ppf(0.95, dof_difference_of_means)
    t = abs(x2 - x1)/SE # t-value
    dof = n - 1 # degrees of freedom for sample

    #95% confidence interval using scipy.stats for the whole sample
    confidence_interval = scipy.stats.t.interval(alpha = 0.95, df = dof, loc = x̄, scale = SE)
    print("95% CI for the Sample:", confidence_interval)

    # 95% confidence interval for differences of means test
    CI_lowerbound = (x2 - x1) - critical_tvalue * SE
    CI_upperbound = (x2 - x1) + critical_tvalue * SE

    # return the CI for true difference of true means
    print("95% CI for Differences of Means test: (", round(CI_lowerbound,3), round(CI_upperbound,3), ")")

confidence_interval(df)
```

```
95% CI for the Sample: (4.624369250977776, 9.436236809628284)
95% CI for Differences of Means test: ( -2.129 2.012 )
```