Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen International Center for Information Technology

Human-Centered Interaction in Robotics

HCIR Assignment-5

Shrikar Nakhye

Trushar Ghanekar

Keerthan Shekarappa

Mohammed Ibrahim Memon

July 2024

Contents

1

Task 1. Designing a Human-Robot Interaction Experiment (45%)

1.1 Suppose that you are developing a scenario where the Atlas robot should assist a plumber by bringing necessary tools and objects when requested. Let's say, you have developed several ways in which the robot can handover an object to the plumber. Now, you would like to evaluate these by conducting an empirical study with experienced plumbers. (You can watch this video for inspiration https://www.youtube.com/watch?v=-e1_QhJ1EhQ)

1.1.1 Define at least one research question that you would like to investigate in the context of human-robot interaction in the above scenario.

How do different object handover methods of the Atlas robot impact the efficiency, accuracy, and user satisfaction of experienced plumbers in a simulated worksite environment?

Sub-questions

Efficiency:

- Which handover method minimizes the time taken for plumbers to receive the requested tools?

Accuracy:

- How do different handover methods affect the accuracy of the Atlas robot in delivering the correct tools without errors?

User Satisfaction:

- Which handover method is perceived as the most comfortable and satisfying by the plumbers?

1.1.2 Identify the constructs involved in this research question(s).

To evaluate the Atlas robot's object handover methods in assisting plumbers, the following constructs are involved:

1. Efficiency

Time Taken: The duration from the plumber's request for a tool to the moment they receive it.
Task Completion Time: The total time taken to complete a plumbing task with the assistance of the robot.

2. Accuracy

Error Rate: The frequency of errors in tool delivery, such as wrong tools handed over or tools dropped.
Correct Tool Delivery: The accuracy with which the robot delivers the correct tool requested by the plumber.

3. User Satisfaction

Comfort Level: The plumber's subjective rating of their comfort during the handover process.
Satisfaction Rating: The plumber's overall satisfaction with the interaction and assistance provided by the robot.
Ease of Use: How easy and intuitive the plumbers find the interaction with the robot's handover methods.
Safety Perception: The plumber's perception of safety while receiving tools from the robot.

Additional Constructs (Contextual Factors)

Task Complexity: The complexity of the plumbing tasks being performed, which may affect the perceived usefulness and effectiveness of each handover method.
Experience Level: The level of experience of the plumbers participating in the study, which could influence their interaction with the robot.

Conclusion: These constructs provide a comprehensive framework to evaluate the performance and user experience of the Atlas robot's handover methods in a realistic work environment.

1.1.3 Formulate at least one hypothesis based on the research question(s) and the identified constructs.

H1: The direct handover method by the Atlas robot will result in the highest efficiency, accuracy, and user satisfaction among experienced plumbers compared to other handover methods (placing on a tray, placing on the ground, using a conveyor belt).

Supporting Hypotheses Based on Constructs

Efficiency

H1a

The direct handover method will minimize the time taken for plumbers to receive requested tools compared to other handover methods.

H1b

The total task completion time will be shorter when using the direct handover method compared to other handover methods.

Accuracy

H1c

The direct handover method will have the lowest error rate in tool delivery compared to other handover methods.

H1d

The direct handover method will have the highest rate of correct tool delivery compared to other handover methods.

User Satisfaction

H1e

Plumbers will report higher comfort levels with the direct handover method compared to other handover methods.

H1f

Plumbers will report higher overall satisfaction with the direct handover method compared to other handover methods.

H1g

Plumbers will find the direct handover method easier to use compared to other handover methods.

H1h

Plumbers will perceive the direct handover method as safer compared to other handover methods.

Conclusion

These hypotheses are designed to test whether the direct handover method provides superior performance and user experience compared to alternative methods, addressing the constructs of efficiency, accuracy, and user satisfaction.

1.1.4 How would you operationalise the constructs in the above hypotheses?

To operationalize the constructs in the hypotheses, we need to define measurable variables and specific methods for data collection. Here's how we can do that:

1. Efficiency

Time Taken for Tool Delivery (H1a)

Measurement: Stopwatch or automated time-tracking system to record the time from when the plumber requests a tool to when they receive it.
Units: Seconds or minutes.

Total Task Completion Time (H1b)

Measurement: Stopwatch or automated system to record the time from the start of the task to its completion.
Units: Minutes or hours.

2. Accuracy

Error Rate in Tool Delivery (H1c)

Measurement: Counting the number of instances where the robot delivers the wrong tool or drops the tool.
Units: Number of errors per task.

Correct Tool Delivery Rate (H1d)

Measurement: Counting the number of times the robot delivers the correct tool as requested by the plumber.
Units: Percentage of correct deliveries out of total requests.

3. User Satisfaction

Comfort Level (H1e)

Measurement: Plumbers will complete a survey after each task, rating their comfort level with each handover method on a Likert scale (e.g., 1-10, where 1 = very uncomfortable and 10 = very comfortable).

Units: Numerical rating (1-10).

Overall Satisfaction (H1f)

Measurement: Plumbers will complete a survey after each task, rating their overall satisfaction with each handover method on a Likert scale (e.g., 1-10, where 1 = very dissatisfied and 10 = very satisfied).
Units: Numerical rating (1-10).

Ease of Use (H1g)

Measurement: Plumbers will complete a survey after each task, rating the ease of use of each handover method on a Likert scale (e.g., 1-10, where 1 = very difficult and 10 = very easy).
Units: Numerical rating (1-10).

Safety Perception (H1h)

Measurement: Plumbers will complete a survey after each task, rating their perception of safety with each handover method on a Likert scale (e.g., 1-10, where 1 = very unsafe and 10 = very safe).
Units: Numerical rating (1-10).

Additional Operationalization Steps

Pilot Testing

Purpose: Conduct a pilot study with a small number of participants to ensure that the measurement tools and survey questions are clear and that they effectively capture the constructs.

Randomization

Purpose: Randomly assign plumbers to use different handover methods in a randomized order to control for order effects.

Control Variables

Purpose: Record additional information such as task complexity and participant experience level to control for their potential influence on the outcomes.

Conclusion: By clearly defining how each construct will be measured, this approach ensures that the data collected will be reliable and valid for testing the hypotheses.

1.1.5 How would you design your study (context, conditions, allocation of subjects)? Justify your study design.

Study Design

Setting: A controlled, simulated worksite that mimics a typical plumbing environment, complete with necessary tools, pipes, sinks, and other plumbing fixtures.

Participants: 20 experienced plumbers with varying levels of experience, age, and physical capabilities to ensure diversity.

Duration: Each participant will be involved in the study for one day, with each handover method being tested multiple times.

Conditions

Handover Methods: Four distinct methods:

1. Direct handover (robot hands tool directly to plumber).
2. Placing tool on a tray (robot places tool on a tray for plumber to pick up).
3. Placing tool on the ground (robot places tool on the ground for plumber to pick up).
4. Using a conveyor belt (robot places tool on a conveyor belt that brings the tool to the plumber).

Tasks: Three typical plumbing tasks that require different tools:

1. Fixing a pipe.
2. Installing a sink.
3. Unclogging a drain.

Allocation of Subjects

Random Assignment:

- Participants will be randomly assigned to complete each of the three tasks.
- The order in which each participant uses the different handover methods will be randomized to control for order effects.

Procedure

Briefing:

- Participants will be briefed about the study's purpose, procedures, and safety measures.

- Demonstration of each handover method.

Task Execution:

- Each participant will perform the three plumbing tasks using all four handover methods.
- The order of tasks and handover methods will be randomized.

Data Collection:

- Time taken for each handover and task completion will be recorded.
- Errors in tool delivery will be noted.
- Participants will complete a survey after each task to rate their comfort, satisfaction, ease of use, and safety perception with each handover method.

Debriefing:

- Participants will provide open-ended feedback on their experiences.

Justification of Study Design

Randomization:

- **Purpose:** To control for potential confounding variables and ensure that the effects observed are due to the handover methods rather than the order of tasks or inherent differences between participants.
- **Method:** Randomly assigning the order of tasks and handover methods helps to distribute any learning effects, fatigue, or biases evenly across conditions.

Controlled Environment:

- **Purpose:** To ensure consistency and control over external variables that could affect the outcomes.
- **Method:** Conducting the study in a simulated worksite ensures that all participants face the same conditions, making the results more reliable and valid.

Diversity of Tasks:

- **Purpose:** To test the handover methods in a range of realistic scenarios that plumbers encounter.
- **Method:** Using three different tasks provides a comprehensive evaluation of the handover methods' effectiveness across various plumbing activities.

Comprehensive Data Collection:

- **Purpose:** To gather both quantitative and qualitative data for a thorough analysis.
- **Method:** Measuring time, accuracy, and subjective ratings ensures that both objective performance metrics and user experiences are considered.

Pilot Testing:

- **Purpose:** To refine the study design and address any potential issues before the main study.
- **Method:** Conducting a pilot study with a small group helps to ensure that the procedures, measurements, and survey questions are clear and effective.

Conclusion: This study design is robust, controlling for potential biases and ensuring that the findings are generalizable and relevant to real-world plumbing scenarios.

1.1.6 What would your null hypothesis look like, for each of the hypotheses that you formulated above?

H1a (Time Taken)

There is no significant difference in the time taken for plumbers to receive requested tools between the direct handover method and other handover methods.

$$H_{0a} : \mu_{\text{direct}} = \mu_{\text{tray}} = \mu_{\text{ground}} = \mu_{\text{conveyor}}$$

H1b (Task Completion Time)

There is no significant difference in the total task completion time between the direct handover method and other handover methods.

$$H_{0b} : \mu_{\text{direct\_task}} = \mu_{\text{tray\_task}} = \mu_{\text{ground\_task}} = \mu_{\text{conveyor\_task}}$$

2. Accuracy

H1c (Error Rate)

There is no significant difference in the error rate of tool delivery between the direct handover method and other handover methods.

$$H_{0c} : \mu_{\text{direct\_errors}} = \mu_{\text{tray\_errors}} = \mu_{\text{ground\_errors}} = \mu_{\text{conveyor\_errors}}$$

H1d (Correct Tool Delivery Rate)

There is no significant difference in the correct tool delivery rate between the direct handover method and other handover methods.

$$H_{0d} : \mu_{\text{direct\_correct}} = \mu_{\text{tray\_correct}} = \mu_{\text{ground\_correct}} = \mu_{\text{conveyor\_correct}}$$

3. User Satisfaction

H1e (Comfort Level)

There is no significant difference in the comfort level reported by plumbers between the direct handover method and other handover methods.

$$H_{0e} : \mu_{\text{direct\_comfort}} = \mu_{\text{tray\_comfort}} = \mu_{\text{ground\_comfort}} = \mu_{\text{conveyor\_comfort}}$$

H1f (Overall Satisfaction)

There is no significant difference in the overall satisfaction reported by plumbers between the direct handover method and other handover methods.

$$H_{0f} : \mu_{\text{direct\_satisfaction}} = \mu_{\text{tray\_satisfaction}} = \mu_{\text{ground\_satisfaction}} = \mu_{\text{conveyor\_satisfaction}}$$

H1g (Ease of Use)

There is no significant difference in the ease of use reported by plumbers between the direct handover method and other handover methods.

$$H_{0g} : \mu_{\text{direct\_ease}} = \mu_{\text{tray\_ease}} = \mu_{\text{ground\_ease}} = \mu_{\text{conveyor\_ease}}$$

H1h (Safety Perception)

There is no significant difference in the safety perception reported by plumbers between the direct handover method and other handover methods.

$$H_{0h} : \mu_{\text{direct\_safety}} = \mu_{\text{tray\_safety}} = \mu_{\text{ground\_safety}} = \mu_{\text{conveyor\_safety}}$$

Guy Hoffman and Xuan Zhao. 2020. A Primer for Conducting Experiments in Human Robot Interaction ACM Trans. Hum. Robot Interact . 10, 1, Article 6 (October 2020), 31 pages `https://doi.org/10.1145/3412374`

2

Task 2. Understanding Power Analysis for t-tests (25%)

2.1 With the help of statsmodels1 library, create plots showing the relationship between alpha level, effect size, sample size and statistical power for (i) two independent sample t-test, and (ii) paired sample t-test. Explain your observations based on the plots.

2.1.1 Two independent sample t-test

Python Code

```python
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.stats.power import TTestIndPower

# Initialize the power analysis class
ttest_ind_power = TTestIndPower()

# Define the ranges for sample size and alpha
sample_sizes = np.arange(5, 101, 5)
alphas = [0.05, 0.1, 0.5]
fixed_effect_size = 0.5

# Function to plot power vs. sample size for fixed alpha level and varying effect sizes
def plot_power_vs_sample_size_fixed_alpha(power_analysis, alpha):
    plt.figure(figsize=(10, 6))
    for effect_size in [0.2, 0.5, 0.8, 1.0]:
        power = power_analysis.power(effect_size=effect_size, nobs1=sample_sizes, alpha=alpha)
        plt.plot(sample_sizes, power, label=f'Effect Size = {effect_size}')
    plt.xlabel('Sample Size')
    plt.ylabel('Power')
    plt.title(f'Power vs. Sample Size (alpha = {alpha})')
```

```python
    plt.legend()
    plt.grid(True)
    plt.show()

# Function to plot power vs. sample size for fixed effect size and varying alpha values
def plot_power_vs_sample_size_varying_alpha(power_analysis, effect_size):
    plt.figure(figsize=(10, 6))
    for alpha in alphas:
        power = power_analysis.power(effect_size=effect_size, nobs1=sample_sizes, alpha=alpha)
        plt.plot(sample_sizes, power, label=f'Alpha = {alpha}')
    plt.xlabel('Sample Size')
    plt.ylabel('Power')
    plt.title(f'Power vs. Sample Size (Effect Size = {effect_size})')
    plt.legend()
    plt.grid(True)
    plt.show()

# Plotting for two independent sample t-test with fixed alpha = 0.05
plot_power_vs_sample_size_fixed_alpha(ttest_ind_power, 0.05)

# Plotting for two independent sample t-test with fixed effect size = 0.5 and varying alpha values
plot_power_vs_sample_size_varying_alpha(ttest_ind_power, fixed_effect_size)
```

Listing 2.1: Python Code for Power Analysis

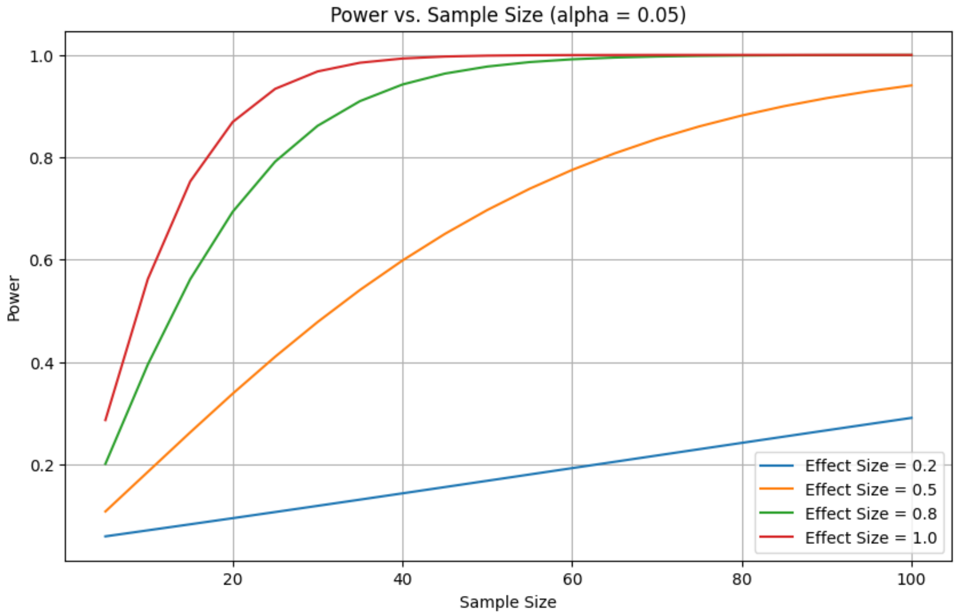

Figure 2.1: Power vs Sample Size with same Alpha=0.05

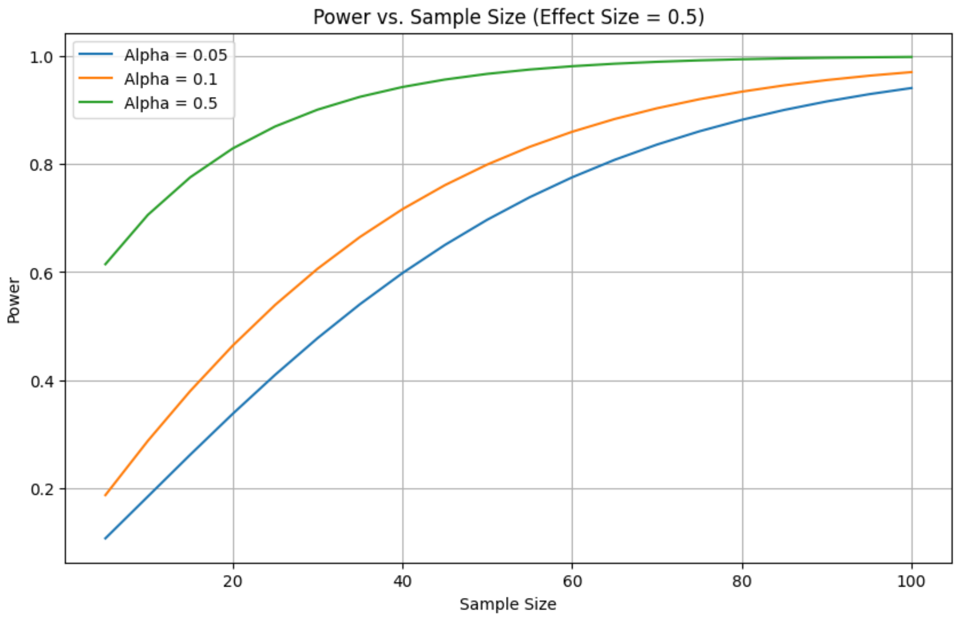

Figure 2.2: Power vs Sample Size with same Effect Size=0.05

2.1.2 Explanations Based on Plots

Power vs. Sample Size for Fixed Alpha Level (0.05) and Varying Effect Sizes

1. The power increases with an increase in sample size.
2. Larger effect sizes lead to higher power with fewer samples needed to achieve high power.
3. For smaller effect sizes (e.g., 0.2), a larger sample size is needed to achieve the same power compared to larger effect sizes.

Power vs. Sample Size for Fixed Effect Size (0.5) and Varying Alpha Levels (0.05, 0.1, 0.5)

1. The power increases with an increase in sample size.
2. Higher alpha levels (e.g., 0.5) result in higher power compared to lower alpha levels (e.g., 0.05) for the same sample size.
3. This demonstrates that increasing the alpha level increases the power, but at the cost of an increased likelihood of Type I errors

Summary

These plots illustrate the relationships between effect size, sample size, alpha level, and statistical power for an independent sample t-test. They show that larger effect sizes and larger sample sizes contribute to higher statistical power. Additionally, increasing the alpha level also increases the power, but at the expense of a higher probability of committing Type I errors.

2.1.3 Paired sample t-test

```python
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.stats.power import TTestPower

# Initialize the power analysis class for paired t-test
ttest_power = TTestPower()

# Define the ranges for sample size and alpha
sample_sizes = np.arange(5, 101, 5)
alphas = [0.05, 0.1, 0.5]
fixed_effect_size = 0.5

# Function to plot power vs. sample size for fixed alpha level and varying effect sizes
def plot_power_vs_sample_size_fixed_alpha_paired(power_analysis, alpha):
    plt.figure(figsize=(10, 6))
    for effect_size in [0.2, 0.5, 0.8,1.0]:
        power = power_analysis.power(effect_size=effect_size, nobs=sample_sizes, alpha=
    alpha)
        plt.plot(sample_sizes, power, label=f'Effect Size = {effect_size}')
    plt.xlabel('Sample Size')
    plt.ylabel('Power')
    plt.title(f'Power vs. Sample Size (Paired, alpha = {alpha})')
    plt.legend()
    plt.grid(True)
    plt.show()

# Function to plot power vs. sample size for fixed effect size and varying alpha values
def plot_power_vs_sample_size_varying_alpha_paired(power_analysis, effect_size):
    plt.figure(figsize=(10, 6))
    for alpha in alphas:
        power = power_analysis.power(effect_size=effect_size, nobs=sample_sizes, alpha=
    alpha)
        plt.plot(sample_sizes, power, label=f'Alpha = {alpha}')
    plt.xlabel('Sample Size')
    plt.ylabel('Power')
    plt.title(f'Power vs. Sample Size (Paired, Effect Size = {effect_size})')
    plt.legend()
    plt.grid(True)
    plt.show()

# Plotting for paired sample t-test with fixed alpha = 0.05
plot_power_vs_sample_size_fixed_alpha_paired(ttest_power, 0.05)

# Plotting for paired sample t-test with fixed effect size = 0.5 and varying alpha values
plot_power_vs_sample_size_varying_alpha_paired(ttest_power, fixed_effect_size)
```

Listing 2.2: Python Code for Power Analysis

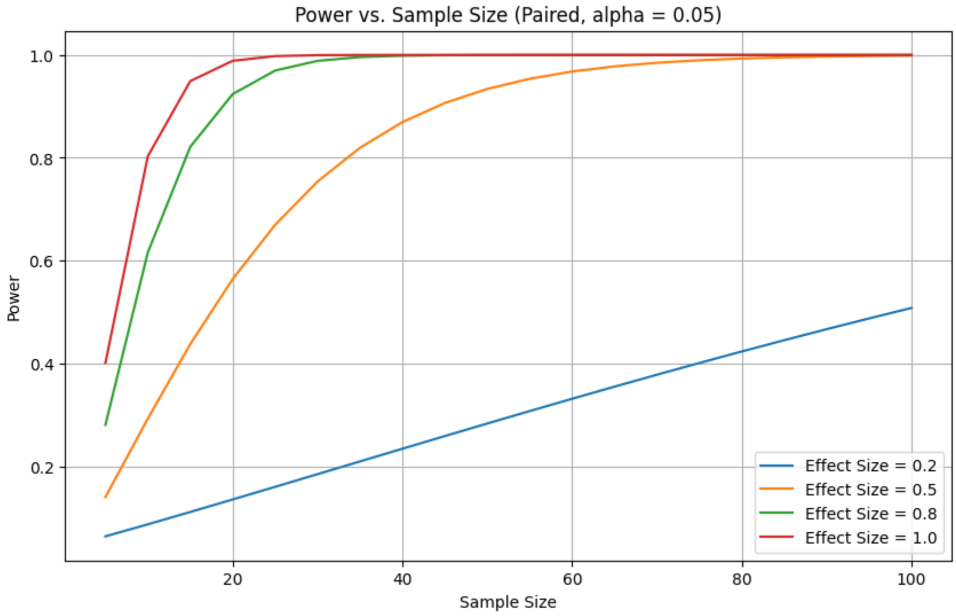

Figure 2.3: Power vs Sample Size with same Alpha=0.05

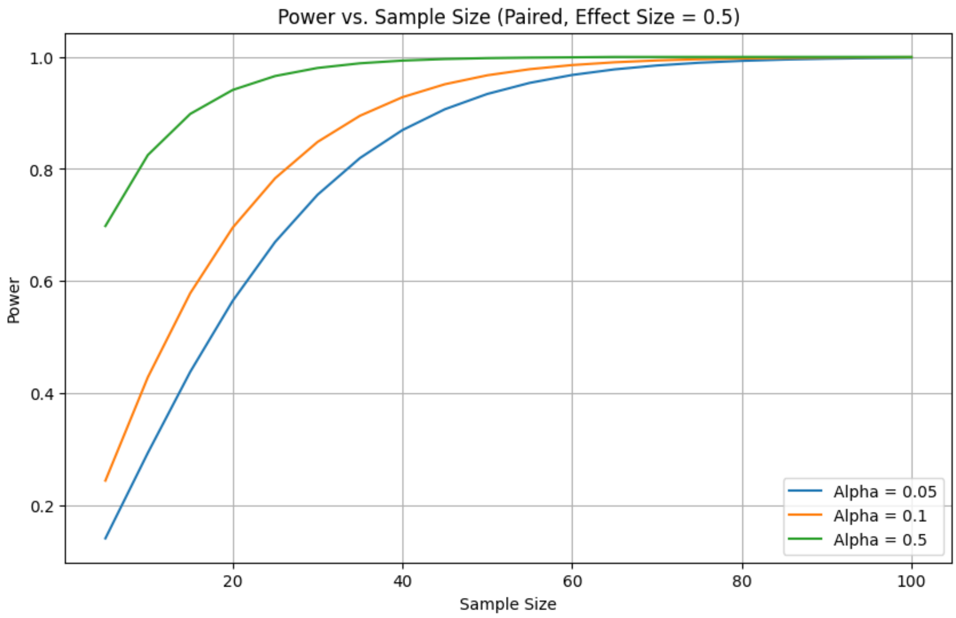

Figure 2.4: Power vs Sample Size with same Effect Size=0.05

2.1.4 Explanations Based on Plots

Power vs. Sample Size for Fixed Alpha Level (0.05) and Varying Effect Sizes (Paired t-test)

1. The power increases with an increase in sample size.
2. Larger effect sizes lead to higher power with fewer samples needed to achieve high power.
3. For smaller effect sizes (e.g., 0.2), a larger sample size is needed to achieve the same power compared to larger effect sizes.

Power vs. Sample Size for Fixed Effect Size (0.5) and Varying Alpha Levels (0.05, 0.1, 0.5) (Paired t-test)

1. The power increases with an increase in sample size.
2. Higher alpha levels (e.g., 0.5) result in higher power compared to lower alpha levels (e.g., 0.05) for the same sample size.
3. This demonstrates that increasing the alpha level increases the power, but at the cost of an increased likelihood of Type I errors.

Summary

These plots illustrate the relationships between effect size, sample size, alpha level, and statistical power for a paired sample t-test. They show that larger effect sizes and larger sample sizes contribute to higher statistical power. Additionally, increasing the alpha level also increases the power, but at the expense of a higher probability of committing Type I errors.

reference: `https://www.geeksforgeeks.org/introduction-to-power-analysis-in-python`

3

Task 3 Performing statistical tests (30%)

A robotics company has created a new social companion robot for elderly persons. Since the robot is designed for elderly men and women who are living alone, it has to appeal to both demographic groups equally. Before the company mass-produces the robot, it wants to make sure that the robot appeals equally to both elderly men and women. More specifically, the company wants to know whether men and women engage with the robot in the same way. To achieve this, 20 elderly men and 20 elderly women were recruited and asked to interact with the robot in a home setting. The participants were then asked to fill in a questionnaire that measures their engagement with the robot. An overall engagement score was computed based on the responses to the questionnaire. The scores for each person is provided in the file.

3.1 a) Plot appropriate graphs and comment on outliers, if any.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the data
data = pd.read_csv('advertising_agency_engagement.csv')

# Display the first few rows of the dataframe
print(data.head())

# Separate the data into two groups
male_scores = data[data['gender'] == 'Male']['engagement']
Female_scores = data[data['gender'] == 'Female']['engagement']

# Plot boxplots to visualize the data and check for outliers
plt.figure(figsize=(12, 6))
sns.boxplot(x='gender', y='engagement', data=data)
plt.title('Engagement Scores by gender')
plt.show()

# Plot histograms to visualize the distribution
plt.figure(figsize=(12, 6))
sns.histplot(male_scores, kde=True, label='Male', color='blue', bins=20)
```

```
sns.histplot(Female_scores, kde=True, label='Female', color='red', bins=20)
plt.title('Distribution of Engagement Scores')
plt.legend()
plt.show()
```

Listing 3.1: Python code for data analysis and plotting

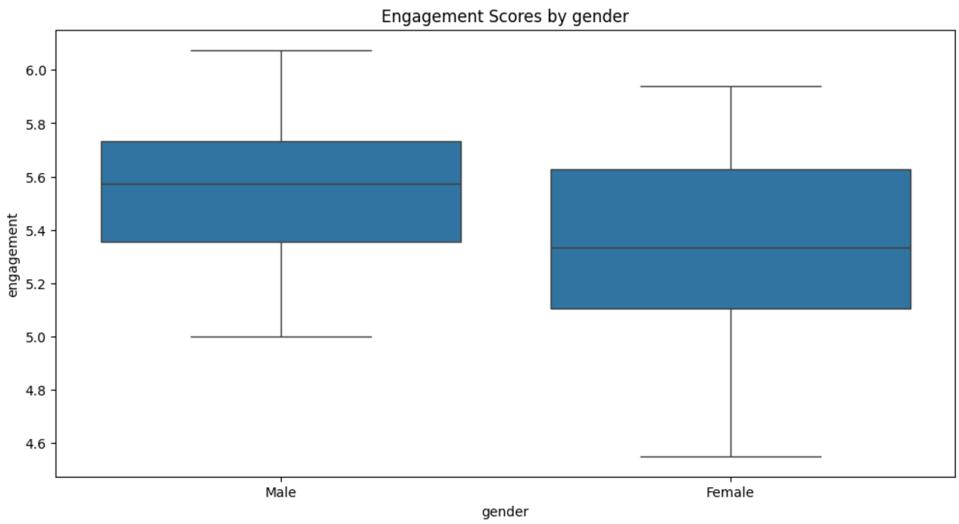

Figure 3.1: Box plot of Engagement Scores

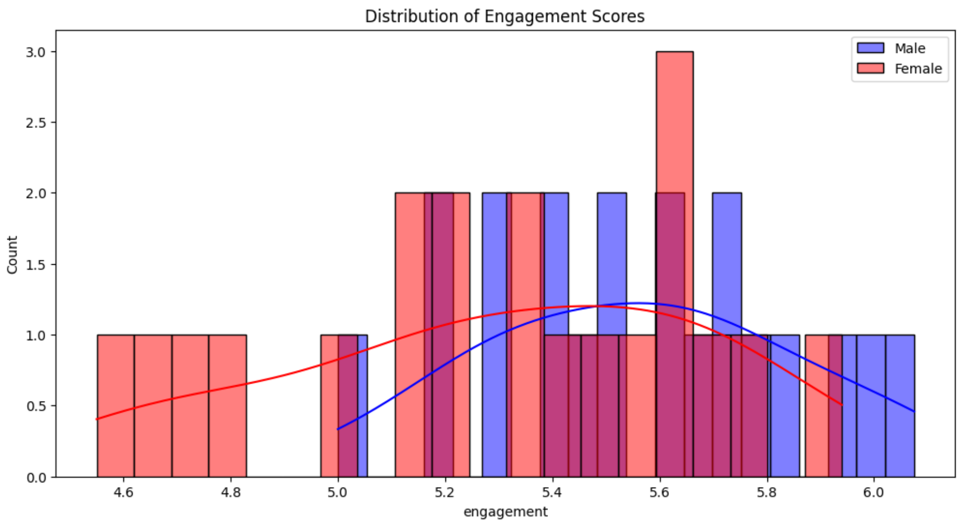

Figure 3.2: Histogram plot of Engagement Scores

3.2 b) Determine whether the data is normally distributed.

```python
from scipy.stats import shapiro
import scipy.stats as stats
import matplotlib.pyplot as plt

# Shapiro-Wilk test for normality
shapiro_male = shapiro(male_scores)
shapiro_female = shapiro(Female_scores)

print(f"Shapiro-Wilk Test for Male: W={shapiro_male[0]}, p-value={shapiro_male[1]}")
print(f"Shapiro-Wilk Test for Female: W={shapiro_female[0]}, p-value={shapiro_female[1]}")

# Q-Q plots
plt.figure(figsize=(12, 6))

plt.subplot(1, 2, 1)
stats.probplot(male_scores, dist="norm", plot=plt)
plt.title('Q-Q Plot for Male')

plt.subplot(1, 2, 2)
stats.probplot(Female_scores, dist="norm", plot=plt)
plt.title('Q-Q Plot for Female')

plt.show()
```

Listing 3.2: Python code for Shapiro-Wilk test and Q-Q plots

```
Shapiro-Wilk Test for Male: W=0.9834404587745667, p-value=0.9704522490501404
Shapiro-Wilk Test for Female: W=0.9607778191566467, p-value=0.559502899646759
```

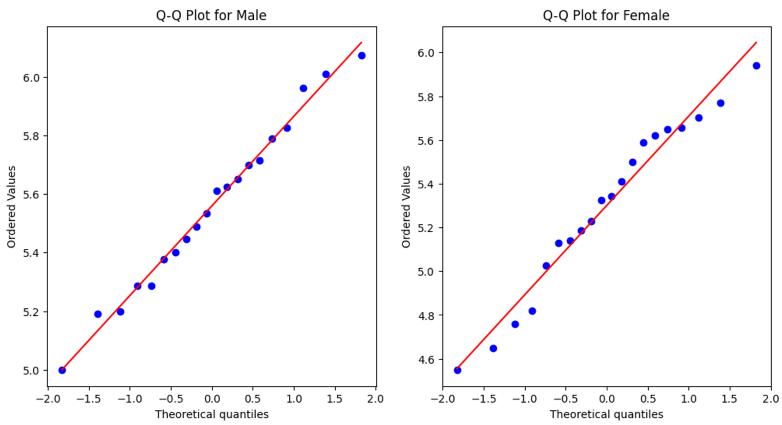

Figure 3.3: Box plot of Engagement Scores

3.3 c) Mention the statistical test that you would use to validate the hypothesis given below. Justify your answer. Hypothesis H1: Elderly men and elderly women engage differently with the robot.

```
from scipy.stats import ttest_ind, mannwhitneyu

# Choose the appropriate statistical test and conduct it
if shapiro_male[1] > 0.05 and shapiro_female[1] > 0.05:
    # Both distributions are normal
    stat, p_value = ttest_ind(male_scores, Female_scores)
    test_name = 'Independent t-test'
else:
    # At least one distribution is not normal
    stat, p_value = mannwhitneyu(male_scores, Female_scores)
    test_name = 'Mann-Whitney U test'

print(f"{test_name}: Statistic={stat}, p-value={p_value}")

# Conclusion
if p_value < 0.05:
    print("There is a significant difference in engagement between Male and Female.")
else:
    print("There is no significant difference in engagement between Male and Female.")
```

Listing 3.3: Python code for statistical test and conclusion

3.4 d) Conduct the statistical test and report the results.

3.4.1 Hypothesis

- **Null Hypothesis (H_0)**: There is no significant difference in engagement between elderly men and women.
- **Alternative Hypothesis (H_1)**: Elderly men and women engage differently with the robot.

3.4.2 Statistical Test Results

- **Test Used**: Independent t-test
- **Test Statistic**: 2.3645
- **p-value**: 0.0233

3.4.3 Interpretation

The p-value from the Independent t-test is 0.0233. In hypothesis testing, we compare the p-value with a significance level (α), typically set at 0.05.

- **p-value $<$ 0.05**: This means we reject the null hypothesis (H_0).
- **p-value $\geq$ 0.05**: This means we fail to reject the null hypothesis (H_0).

In this case, since the p-value (0.0233) is less than the significance level (0.05), we reject the null hypothesis. This indicates that there is a statistically significant difference in engagement scores between elderly men and women.

3.5 e) Explain your conclusion about the hypothesis.

3.5.1 Conclusion

1. **Statistical Significance**:
   - The low p-value (0.0233) indicates that the observed difference in engagement scores between men and women is unlikely to have occurred by random chance alone. Therefore, we conclude that there is a real difference in how elderly men and women engage with the robot.
2. **Practical Implications**:
   - **Product Development**: The company should consider that elderly men and women interact with the robot differently. Features, user interfaces, and functionalities might need to be tailored to better meet the needs and preferences of each gender.
   - **Marketing Strategies**: Marketing efforts could be designed to address the specific aspects of the robot that appeal to each gender group, thereby enhancing user satisfaction and engagement.
3. **Further Research**:
   - It would be beneficial to conduct further studies to understand why men and women engage differently with the robot. Qualitative research, such as interviews or focus groups, could provide deeper insights into the preferences and expectations of each group.
   - The company could also explore other demographic factors, such as age, socioeconomic status, or cultural background, to see if they influence engagement in a similar way.

3.5.2 Summary

The Independent t-test results indicate a significant difference in engagement between elderly men and women (p-value = 0.0233). This finding suggests that gender plays a role in how elderly individuals interact with the robot. Understanding and addressing these differences can help the company improve the robot's design and user experience, ultimately leading to better outcomes for the target demographic.