

# ENGR 371 Report

April 14, 2024

## 1 Notes and Details

Notes and details of your experiment. Dates that data was collected, information not clear from experimental design (for example weather that day if it was relevant to your experiment.)

## 2 Analysis

To begin with calculate the sample mean, the sample variance and the sample median. Describe the distribution of the sample mean. Generate a confidence interval on the sample mean (decide on an appropriate confidence bound). Also set up a hypothesis test for your hypothesis and test its va

```
[ ]: # Setup Project Environment
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

plt.style.use("ggplot")
pd.set_option(
    "display.max_columns", 200
) # display up to 200 columns (removes '...' hidden columns)

[ ]: # Read in Data and Display Shape (rows, columns)
df = pd.read_csv("survey_data.csv")
df.shape

[ ]: # Drop the columns that don't agree with the Confidentiality Agreement
if "Agreement of Confidentiality" in df.columns:
    df = df[
        df["Agreement of Confidentiality"]
        == "I agree that the information submitted will be kept confidential, and_
↪ abide by the terms of confidentiality and anonymity."
    ].reset_index(drop=True)

# Drop the Timestamp, Agreement of Confidentiality and Feedback columns
try:
    df = df.drop(
```

```

        columns=[
            "Timestamp",
            "Agreement of Confidentiality",
            "Any thoughts you would like to add? (Optional)",
        ],
        axis=1,
    )
except KeyError:
    pass

# Rename Columns
df = df.rename(
    columns={
        "How long before you go to sleep do you usually stop using your phone, on a
        ↪regular basis?": "end_of_phone_usage",
        "What do you do on your phone before going to sleep?":
        ↪"end_of_phone_usage_activity",
        "Where do you usually put your phone right before you go to sleep?":
        ↪"phone_placement",
        "How long does it take you to fall asleep?": "time_to_fall_asleep",
        "How often do you wake up during the night?": "night_wakeups",
        "How many hours of sleep do you get per night on average?": "hours_of_sleep",
        "How would you rate your average sleep quality? \n(ex: waking up feeling
        ↪refreshed and ready to start the day, or often feel tired and like you didn't
        ↪sleep well)?": "sleep_quality",
        "Do you have any underlying conditions that may affect the quality of your
        ↪sleep? (ex: Anxiety, ADHD, etc.)": "underlying_conditions",
        "Please check all of the following that apply to your typical sleep
        ↪environment": "sleep_environment",
        "How often do you exercise?": "exercise_frequency",
    }
)
df.head()

```

```

[ ]: # Convert the 'underlying_conditions' column to a boolean
df["underlying_conditions"] = df["underlying_conditions"].apply(
    lambda x: False if x == "No" else True
)

# One-hot encode certain columns.
df_encoded = pd.get_dummies(
    df,
    columns=[
        "end_of_phone_usage",
        "end_of_phone_usage_activity",
        "phone_placement",
        "time_to_fall_asleep",
    ]
)

```

```

        "night_wakeups",
        "exercise_frequency",
    ],
)

# TODO: Handle sleep_environment later (currently being dropped)
df_encoded = df_encoded.drop(columns=["sleep_environment"], axis=1)

# replace spaces with underscores and make lowercase
df_encoded.columns = df_encoded.columns.str.replace(" ", "_").str.lower()

# Rename Columns
df_encoded = df_encoded.rename(
    columns={
        "end_of_phone_usage_30_minutes_-_1_hour": ␣
        ↪ "end_of_phone_usage_between_30_minutes_and_1_hour",
        "end_of_phone_usage_<_30_minutes": "end_of_phone_usage_less_than_30_minutes",
        "end_of_phone_usage_>_1_hour": "end_of_phone_usage_more_than_1_hour",
        "end_of_phone_usage_activity_browsing_social_media_(ex:_instagram,_tiktok...)": "end_of_phone_usage_activity_browsing_social_media",
        "time_to_fall_asleep_30_minutes_-_1_hour": ␣
        ↪ "time_to_fall_asleep_between_30_minutes_and_1_hour",
        "time_to_fall_asleep_<_30_minutes": ␣
        ↪ "time_to_fall_asleep_less_than_30_minutes",
        "time_to_fall_asleep_>_1_hour": "time_to_fall_asleep_more_than_1_hour",
        "night_wakeups_never_(0_times)": "night_wakeups_never_(0)",
        "night_wakeups_rarely_(1_time)": "night_wakeups_rarely_(1)",
        "night_wakeups_sometimes_(2-3_times)": "night_wakeups_sometimes_(2-3)",
        "exercise_frequency_i_don't_exercise": "exercise_frequency_never",
    }
)

```

```

[ ]: # Heatmap of correlations between hours of sleep, sleep quality, and phone usage ␣
    ↪ before bed
df_corr = df_encoded[
    [
        "hours_of_sleep",
        "sleep_quality",
        "end_of_phone_usage_between_30_minutes_and_1_hour",
        "end_of_phone_usage_less_than_30_minutes",
        "end_of_phone_usage_more_than_1_hour",
    ]
].corr()

heatmap = sns.heatmap(df_corr, annot=True, fmt=".2f")

```

### 3 Conclusion

A conclusion on whether your hypothesis was reasonable, and justification for your conclusion

### 4 Comments

- 4.1 Did your method of sampling result in a random sample?
- 4.2 If your sample was not a random sample, what sorts of measures could you take if you were to do this project again, to get a random sample?
- 4.3 Based on the experiment, would it be appropriate to write a revised hypothesis (“in Hampstead, 5 cars go through each yellow light.”)
- 4.4 Comment on whether you think your results can be extrapolated to draw more general conclusions, perhaps on wider populations. State your opinion and then back it up with well-reasoned arguments.