Hypothesis Testing

December 2, 2023

1 Hypothesis Testing

The observed correlations from the data analysis and visualizations suggest several hypotheses that could be explored through further analysis:

1. **Impact of Sleep Disturbances on Quality:** Given the strong negative correlation between sleep disturbances and quality, we can hypothesize that increased sleep disturbances are likely to negatively impact the quality of sleep.

2. **Age in Relation to Sleep Duration:** The negative correlation between age and calculated night sleep duration leads to the hypothesis that sleep duration may decrease with age.

3. **Relationship Between Sleep Onset Time and Quality:** The moderate negative correlation observed between sleep onset time and quality suggests that a longer time to fall asleep might be associated with poorer sleep quality.

4. **Influence of Exercise on Sleep Duration and Quality:** The slight positive correlation between exercise days per week and sleep duration hints at a potential hypothesis that increased physical activity could contribute to longer and possibly better quality sleep.

5. **Nap Duration's Effect on Nighttime Sleep Duration and Quality:** Although the correlation is weak, we could investigate whether the duration of naps has any effect on the duration of nighttime sleep.

1.1 Hypothese 1 - Increased sleep disturbances negatively impact the quality of sleep.

Null Hypothesis (H_0): The level of Sleep Disturbances has no impact on Sleep Quality.

Alternative Hypothesis (H_1): The level of Sleep Disturbances has a negative impact on Sleep Quality.

We will a 1-tail negative *Spearman correlation test* by setting `alternative='less'` to measure the correlation

```
[12]: import scipy.stats as stats

correlation, p_value = stats.spearmanr(df['Sleep Quality'].to_numpy(),␣
  ↪df['Sleep Disturbances Ordinal'].to_numpy().astype(float),␣
  ↪alternative='less')

print(f'Correlation: {correlation:.3f}')
```

```
print(f'P-value: {p_value}')
```

```
Correlation: -0.453
P-value: 4.198014093688787e-07
```

With the given results of a Spearman correlation coefficient ( ) of approximately -0.453 and a p-value of approximately 4.2e-07, we can draw the following conclusions about the relationship between sleep disturbances and sleep quality:

- **Strength and Direction of Correlation:** The Spearman correlation coefficient of -0.453 indicates a moderate negative correlation between sleep disturbances and sleep quality. This means that as sleep disturbances increase (become more frequent), sleep quality tends to decrease (gets worse).

- **Statistical Significance:** The p-value is a measure of the probability that the observed correlation occurred by chance if there were no actual relationship in the population. A p-value of 4.2e-07 is extremely small, far below the common alpha level of 0.05 used to determine statistical significance. This means that the negative correlation observed is highly unlikely to be due to random variation in the sample; it's statistically significant.

Conclusion: Based on the Spearman correlation test, we can confidently reject the null hypothesis that there is no correlation between sleep disturbances and sleep quality. The data supports the alternative hypothesis that sleep disturbances do affect sleep quality, with more disturbances associated with worse sleep quality. This result aligns with what might be expected intuitively: that individuals who experience more disturbances during sleep tend to report lower overall sleep quality.

1.2 Hypothese 3 - The longer it takes to fall as sleep, the worse Sleep Quality becomes

Null Hypothesis (H_0): The increase of Sleep onset time has no impact on Sleep Quality.

Alternative Hypothesis (H_1): The increase of Sleep onset time leads to the decline on Sleep Quality.

```python
[13]: onset_mapping = {
          "<15 Minutes": 7.5,
          "30-60 Minutes": 45,
          "15-30 Minutes": 20,
          ">60 Minutes": 60,
      }

      df["Sleep Onset Time Ordinal"] = df["Sleep Onset Time"].map(onset_mapping)
```

```python
[15]: import scipy.stats as stats

      # Perform Spearman correlation test
      correlation, p_value = stats.spearmanr(df['Sleep Quality'].to_numpy(),␣
       ↪df['Sleep Onset Time Ordinal'].to_numpy().astype(float), alternative='less')
```

```
print(f'Correlation: {correlation}')
print(f'P-value: {p_value}')
```

```
Correlation: -0.37764117513828727
P-value: 2.7990152273389178e-05
```

- **Correlation Coefficient:** The negative correlation coefficient indicates an inverse relationship between sleep onset time and sleep quality. This suggests that longer times to fall asleep (indicating difficulty initiating sleep) are associated with lower sleep quality ratings.

- **Statistical Significance:** The p-value measures the probability that the observed correlation is due to random chance. A p-value of 2.80e-05 is very small and well below the conventional alpha level of 0.05, which is commonly used to assess statistical significance. This indicates that the observed correlation is highly unlikely to have occurred by chance.

Conclusion:

Based on the Spearman correlation test, we can reject the null hypothesis (H) that sleep onset time has no impact on sleep quality. Instead, we accept the alternative hypothesis (H) that there is a statistically significant negative relationship between sleep onset time and sleep quality. In practical terms, this result suggests that interventions aimed at reducing sleep onset time might be beneficial for improving overall sleep quality.