Power Analysis for Final Project

Cynthia, Emanuel, Jonathan, and Rina

02/14/2024

1 Introduction

Our main research question is: Does wearing sunglasses before bedtime improve sleep quality? Where sleep quality is measured via wearable technology, such as a smart watch. Sleep quality itself can be operationalized in various ways. We will be looking specifically at sleep quality as measured by

1. Sleep Score. This is a proprietary metric whose calculation may vary from app to app but we are taking these differences as approximately the same.
2. Duration spent in REM and deep sleep. Time spent in specific stages of sleep, particularly REM and deep sleep, have been found to be essential for high quality sleep. (*Stages of sleep: What happens in a sleep cycle*)

Our assumptions for each power scenario are:

- Population is normally distribution.
- Significance level = 0.05.

We simulate the experiment at different sample sizes, from 10-100 subjects by increments of 10. We run 1,000 simulations at each sample size to calculate the power. We use t-test for hypothesis testing. Note: Because we are running a paired test, each subject has a control and treatment value; for example, 10 subjects yields 10 control and 10 treatment sleep measurements.

2 Scenario 1 - Measurement: Sleep Score, ATE: 5, SD: 5

For this scenario we assume:

- The average non-intervention sleep score is 75 (for scenarios 1, 2, & 3). This average is based off of data provided by Garmin (*Garmin Blog*).
- True $ATE_{sc1} = 5$. This is based on a small trial run of our experiment conducted with ourselves.
- $\sigma_{sc1} = 5$, regardless of intervention

```
## Average sleep score in control phase:  73.33333

##
## Average sleep score in treatment phase:  81.33333

##
## Jon pilot average treatment effect:  8
```

```
##     sample_sizes sample_powers
##            <num>         <num>
##  1:           10         0.345
##  2:           20         0.799
##  3:           30         0.980
##  4:           40         0.996
##  5:           50         1.000
##  6:           60         1.000
##  7:           70         1.000
##  8:           80         1.000
##  9:           90         1.000
## 10:          100         1.000
```

In this scenario, we are able to achieve a power of 0.799 with a sample size of 20.

3 Scenario 2 - Measurement: Sleep Score, ATE: 3, SD: 3

For this scenario we assume:

- The average non-intervention sleep score is 75.
- True $ATE_{sc2} = 3$
- $\sigma_{sc2} = 3$, regardless of intervention

```
##     sample_sizes sample_powers
##            <num>         <num>
##  1:           10         0.346
##  2:           20         0.802
##  3:           30         0.965
##  4:           40         0.995
##  5:           50         0.999
##  6:           60         1.000
##  7:           70         1.000
##  8:           80         1.000
##  9:           90         1.000
## 10:          100         1.000
```

In this scenario, we are able to achieve a power of 0.802 with a sample size of 20.

4 Scenario 3 - Measurement: Sleep Score, ATE: 5, SD: 7

For this scenario we assume:

- The average non-intervention sleep score is 75.
- True $ATE_{sc3} = 5$
- $\sigma_{sc3} = 7$, regardless of intervention

```
##     sample_sizes sample_powers
##            <num>         <num>
##  1:           10         0.142
```

```
##  2:           20          0.386
##  3:           30          0.657
##  4:           40          0.827
##  5:           50          0.933
##  6:           60          0.982
##  7:           70          0.989
##  8:           80          0.994
##  9:           90          1.000
## 10:          100          1.000
```

5 Scenario 4 - Measurement: Deep & REM sleep

We are interest in potentially looking at other sleep metrics as measurement of our outcome of sleep quality. We know that sleep quality, and not just duration, particularly the amount of time spent in the deep and REM phases of sleep can be strong indicators of sleep quality (*Suni, E.*). We attempted to used the effect size calculated by Hedges' g in the study by *Chinoy, E, et al.* in which they compared several different consumer sleep tracking devices to the laboratory standard of polysomnography.

For this scenario we use values from the *Chinoy, E, et al.* study

- The average non-intervention deep sleep phase as a percentage of total sleep is approximately 15.19%.
- The average non-intervention REM sleep phase as a percentage of total sleep is approximately 26.27%.
- The true ATE of deep sleep is 0.185 (minutes) or 0.04%.
- The true ATE of REM sleep is 0.25 (minutes) or 0.05%..
- Standard deviation of deep sleep = 6.02%, regardless of intervention.
- Standard deviation of REM sleep = 8.36%, regardless of intervention.

```
##     sample_sizes sample_powers
##            <num>         <num>
##  1:           10         0.005
##  2:           20         0.001
##  3:           30         0.001
##  4:           40         0.000
##  5:           50         0.000
##  6:           60         0.001
##  7:           70         0.003
##  8:           80         0.000
##  9:           90         0.000
## 10:          100         0.000
```

```
##     sample_sizes sample_powers
##            <num>         <num>
##  1:           10         0.001
##  2:           20         0.001
##  3:           30         0.002
##  4:           40         0.001
##  5:           50         0.000
##  6:           60         0.001
##  7:           70         0.000
##  8:           80         0.000
##  9:           90         0.001
## 10:          100         0.000
```

We had some difficulty in interpreting the units of the Hedges' g effect size and how it factors into our power calculation. From these power results, we can see we have extremely low statistical power. We also looked at the deep & REM phases of our self-conducted trial study and found that the difference in means between treatment and control for time spent in deep & REM phases of sleep was also extremely small (not even one-tenth of a percentage point). It is therefore unlikely we can feasibly use these metrics for our outcome.

6 Scenario 5 - Deep & REM sleep meaningful ATEs

In order to achieve power of around 60%-80% wot 20 to 30 people we would need the true ATE for each one of these to be:

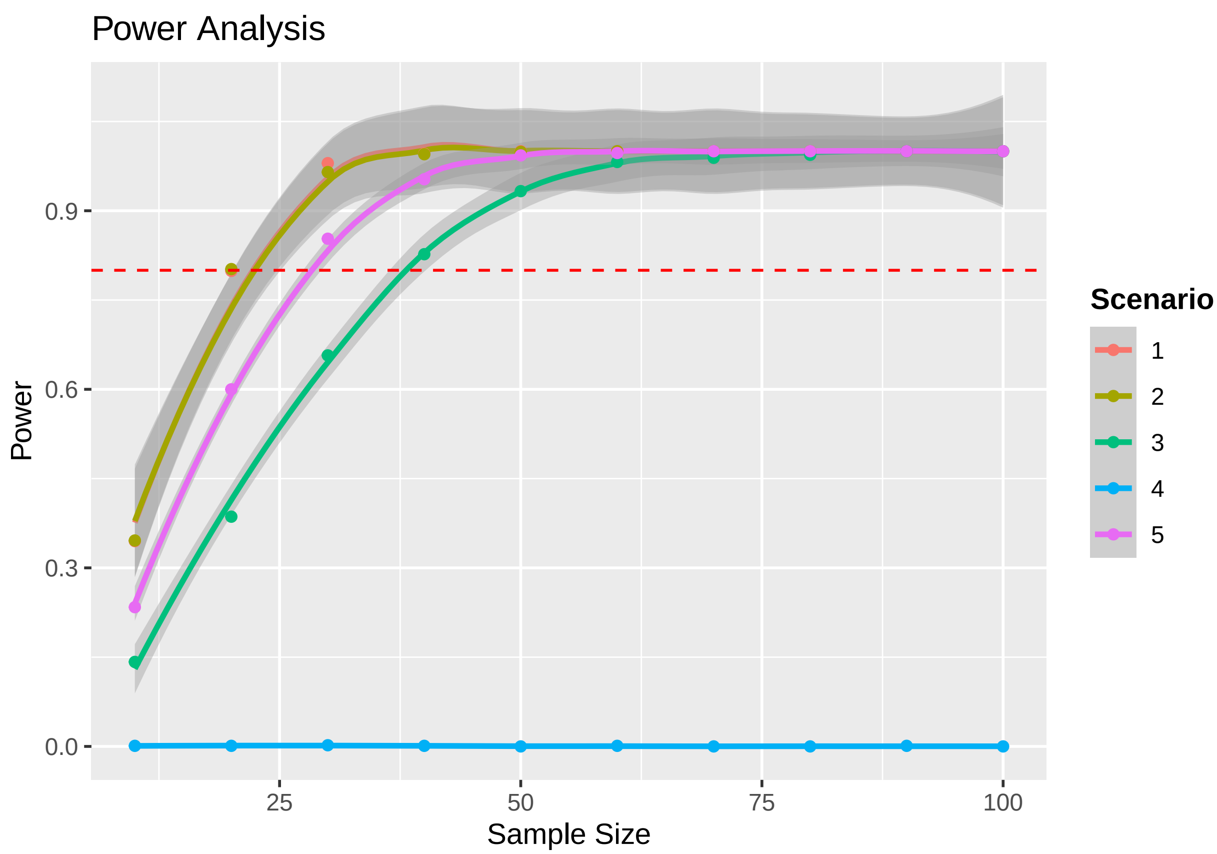
- $ATE_{deep} = 5\%$.
- $ATE_{REM} = 7\%$..

```
##     sample_sizes sample_powers
##            <num>         <num>
##  1:           10         0.211
##  2:           20         0.581
##  3:           30         0.871
##  4:           40         0.957
##  5:           50         0.994
##  6:           60         0.999
##  7:           70         1.000
##  8:           80         0.999
##  9:           90         1.000
## 10:          100         1.000
```

```
##     sample_sizes sample_powers
##            <num>         <num>
##  1:           10         0.234
##  2:           20         0.600
##  3:           30         0.853
##  4:           40         0.953
##  5:           50         0.993
##  6:           60         0.997
##  7:           70         1.000
##  8:           80         1.000
##  9:           90         1.000
## 10:          100         1.000
```

7 Plot of Achieved Power

Comparing the powers of each scenario, we plot:



8 References

Evan D Chinoy, Joseph A Cuellar, Kirbie E Huwa, Jason T Jameson, Catherine H Watson, Sara C Bessman, Dale A Hirsch, Adam D Cooper, Sean P A Drummond, Rachel R Markwald, Performance of seven consumer sleep-tracking devices compared with polysomnography, Sleep, Volume 44, Issue 5, May 2021, zsaa291, https://doi.org/10.1093/sleep/zsaa291

Miller DJ, Sargent C, Roach GD. A Validation of Six Wearable Devices for Estimating Sleep, Heart Rate and Heart Rate Variability in Healthy Adults. Sensors. 2022; 22(16):6317. https://doi.org/10.3390/s22166317

Suni, E., Singh, A. (2023, December 8). Stages of sleep: What happens in a sleep cycle. Sleep Foundation. https://www.sleepfoundation.org/stages-of-sleep

New data examines quality of Garmin users' sleep. Garmin Blog. (2023, August 30). https://www.garmin.com/en-US/blog/health/new-data-examines-quality-of-garmin-users-sleep/