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

For this scenario we assume:

- The average non-intervention sleep score is 75 (for scenarios 1 & 2). This is based on a small trial run of our experiment conducted with ourselves.
- True ATE = 5. This is also based on the trial run we conducted.
- Standard deviation = 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         0.381
##  2:           20         0.832
##  3:           30         0.969
##  4:           40         0.999
##  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 0.832 with a sample size of 20.

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

For this scenario we assume:

- The average non-intervention sleep score is 75 (for scenarios 1 & 2). This is based on a small trial run of our experiment conducted with ourselves.
- True ATE = 3
- Standard deviation = 3, regardless of intervention

```
##     sample_sizes sample_powers
##            <num>         <num>
##  1:           10         0.375
##  2:           20         0.829
##  3:           30         0.956
##  4:           40         0.999
##  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 0.829 with a sample size of 20.

4 Scenario 3 - Measurement: Deep & REM sleep, ATE: 0.15 & 0.26, respectively

For this scenario we assume:

- The average non-intervention deep sleep phase as a percentage of total sleep is approximately 15%.
- The average non-intervention REM sleep phase as a percentage of total sleep is approximately 26%.
- The true ATE of deep sleep is 0.185.
- The true ATE of REM sleep is 0.25.
- Standard deviation of deep sleep = 6%, regardless of intervention.

- Standard deviation of REM sleep = 8%, regardless of intervention.

All of these values are based on a study conducted that compared consumer sleep-tracking devices to laboratory standards *Chinoy, E, et al.*.

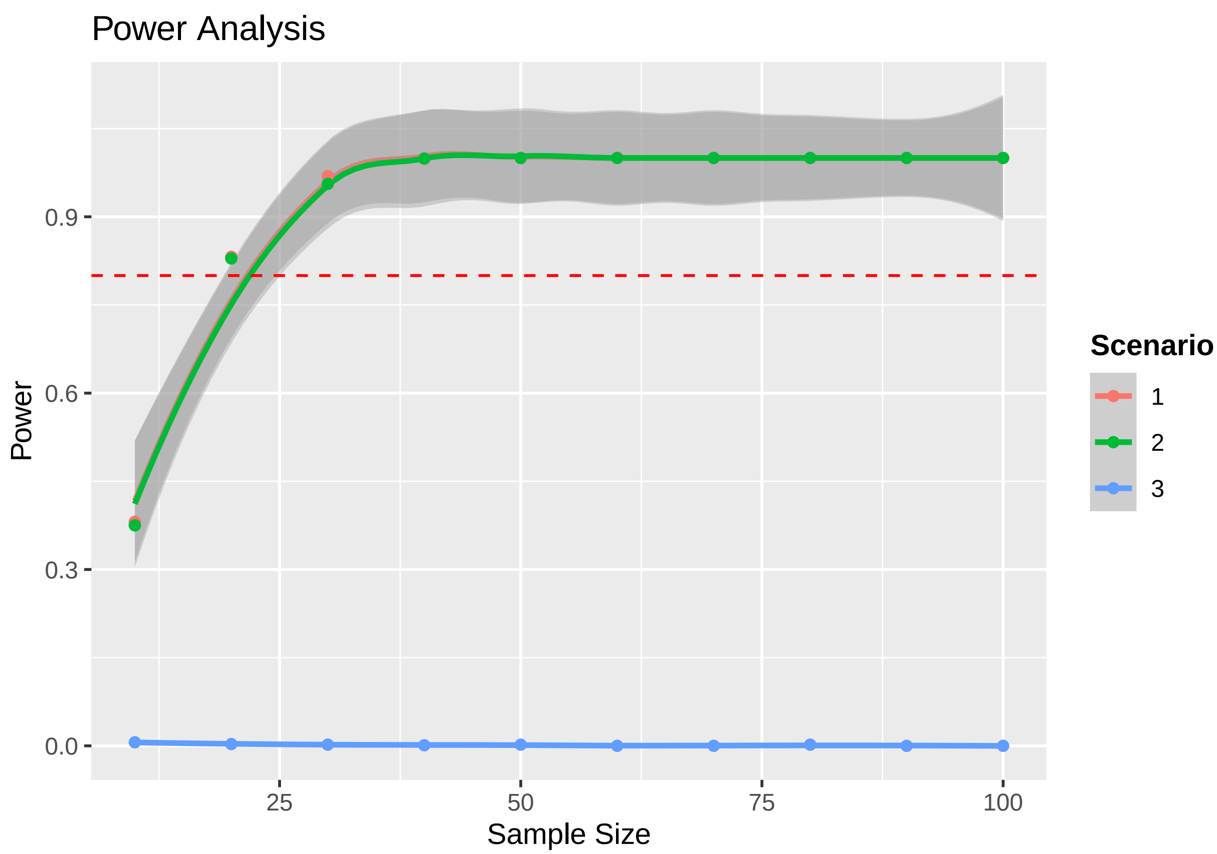
```
##     sample_sizes sample_powers
##            <num>         <num>
##  1:           10         0.003
##  2:           20         0.000
##  3:           30         0.000
##  4:           40         0.001
##  5:           50         0.002
##  6:           60         0.001
##  7:           70         0.000
##  8:           80         0.001
##  9:           90         0.000
## 10:          100         0.001
```

```
##     sample_sizes sample_powers
##            <num>         <num>
##  1:           10         0.006
##  2:           20         0.003
##  3:           30         0.002
##  4:           40         0.001
##  5:           50         0.002
##  6:           60         0.000
##  7:           70         0.000
##  8:           80         0.002
##  9:           90         0.000
## 10:          100         0.000
```

Unfortunately, we have extremely low statistical power if using percentage of REM or deep sleep phases as our main readout metric. It is unlikely we can feasibly go this route.

5 Plot of Achieved Power

Comparing the powers of each scenario, we plot:



6 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