Effect of Pillow Type and Age Group on Sleep Quality
Final Project Submission

Apoorvi Mittal, Asta Adhira Anggono, Ayushi Khasnobis, Joy Shyu, Junzhe Gong
School of Professional Studies, Columbia University
APAN 5300: Research Design
Dr. Nicole DiCrecchio
December 03, 2022

Part 1: Research Proposal

Executive Summary/Abstract

In this study, an experiment has been conducted to understand how the quality of sleep depends on the type of pillow that we sleep on. Getting a good night's sleep lets us start our day on the right foot with a positive attitude and a well rested body. A lot of factors contribute to the quality of our sleep, one of them being the type of pillow that is used by individuals. The purpose of this study is to help Mattress Firm understand how to create a marketing campaign for the different types of pillows that they have and which segment of the population would be their target. The conclusion of the study suggests that the mean sleep score for the latex pillow is significantly higher than for the standard feather pillow. It also suggests that the age group 21-40 have a significantly lower mean sleep score than the age group 41-60. Therefore Mattress Firm can make strategic decisions and create an advertising campaign for the latex pillow to target the younger population suggesting that using latex pillows will help them get a better quality of sleep.

The two research questions that this study attempts to answer are the first: “Is the mean sleep score measured by the Fitbit Charge 5 higher for the latex pillow (treatment) than for the standard feather pillow (control)?” and the second: “Is the mean sleep score as measured by the Fitbit Charge 5 for the age group 21-40 lower than for the age group 41-60?”. The total sample size of the study is 400 people between the age range 21-60 years old. People with pre-existing neck and back problems and people with insomnia have been excluded from this study with the help of a survey that was collected when participants were being recruited. Gender is a variable that has been controlled for in the study by having an equal distribution of gender within each age group. 100 people from each combination of age group (21-40 and 41-60) and gender (male and female) were randomly assigned to treatment and control pillow groups. The division was equal: 50 in treatment and 50 in control. Participants were asked to wear a fitbit and sleep on only their assigned pillow for two weeks. Post which the data collected in the fitbit was loaded and processed for statistical analysis. A two sample one sided t-test was performed to analyze the results for both research questions.

There are five research papers that were reviewed as a basis for selecting the independent variables, dependent variables, control variables, sample size, ways to measure sleep scores etc. The five studies essentially explore the effects of pillow types on sleep quality, age and gender on sleep quality, and the types of devices used to track sleep. These studies have been used as a reference or a benchmark for designing the experiment in order to answer the two research questions.

Statement of Problem

According to Jeon et. al (2014) sleep maintains homeostasis, immunity and cellular repair. In addition, sleep regulates the cognitive abilities, memory, and judgment required for daily life and work (Schutz et al. 2009). Kyle et al. (2010) found that sleep quality is directly related to people's health and living standards. A 2016 study from the Centers for Disease Control (CDC) and Pention's Morbidity and Mortality Weekly Report shows that more than a third of American adults don't get enough sleep on a regular basis. According to Jeon et. al (2014), one third of a human beings' life is spent sleeping. Therefore, we believe that it is important to have good sleep in order to improve quality of life over the long term. Nighttime sleep quality is closely related to sleep posture. Choosing the appropriate pillow can help optimize sleep posture and promote high-quality sleep (Bernateck et al. 2008). Since sleep is an important factor for human beings, we have decided to experiment how different types of pillows have an impact on the quality of sleep. Mattress Firm wants to understand how to advertise their different pillow types, and potentially which age group in the population they should target.They have hired us to carry out this research study for them.

Research Questions, Hypotheses and Effects

Research Question 1:

Is the mean sleep score, measured by the Fitbit Charge 5, higher for the latex pillow (treatment) than for the standard feather pillow (control)?

Hypothesis:

Null hypothesis: The mean sleep score for the latex pillow (treatment) is not significantly higher than that for the standard feather pillow (control).

Alternative hypothesis: The mean sleep score for the latex pillow (treatment) is significantly higher than for the standard feather pillow (control).

Effect Size:

We will consider a difference in the mean sleep score of 5 between the treatment (latex pillow) and the control (standard feather pillow) groups to be a meaningful effect size. This will be a difference large enough for the company to consider marketing their latex pillow as improving sleep quality over the standard feather pillow.

Research Question 2:

Is the mean sleep score, measured by the Fitbit Charge 5, for the age group 21-40 lower than that for the age group 41-60?

Hypothesis:

Null hypothesis: The mean sleep score for the age group 21-40 is not significantly lower than that for the age group 41-60.

Alternative hypothesis: The mean sleep score for the age group 21-40 is significantly lower than that for the age group 41-60.

Effect Size:

We will consider a difference in the mean sleep score of 8 between the age group 21-40 and the age group 41-60 to be a meaningful effect size. This will be a difference large enough for the company to consider targeting the age group with the lower sleep quality, as their target market.

Importance of the Study and Social Impact

As humans, on average, we spend a third of our life sleeping. The importance of sleep has been emphasized throughout our lives, as the benefits of quality sleep and the drawbacks of the lack of sleep have repeatedly been proven in scientific research. Quality sleep allows our bodies and minds to recharge and help prevent certain medical conditions, such as type 2 diabetes, heart diseases, and poor mental health (Pacheco, D. & Singh, A., 2022). We will conduct this study in partnership with Mattress Firm, one of the top mattress and bedding companies in the US. Mattress Firm is interested in marketing its latex pillows, so this study will also allow the company to have a scientifically-backed advertising campaign, if latex pillows improve sleep quality. Moreover, they are interested in targeting their marketing efforts towards an age group which has lower sleep quality as they also care about the wellness of their customers and want to help them get better sleep. Lastly, if this study is successful in showing a significant increase in people's sleep quality and improves the wellness of the people in the United States, this positive phenomenon can potentially spread to other parts of the world. If study's results are significant and people start buying latex pillows instead of their regular one, then this could improve the wellness of people in the US, and potentially of those in other parts of the world.

Literature Review

A research study called “Your Pillow May Not Guarantee a Good Night's Sleep or Symptom-Free Waking” conducted by Gordon and Grimmer-Somers collected data on the performance of five different types of experimental pillows with respect to retiring and waking symptoms such as pain, stiffness, headache and scapula pain compared to participants’ usual pillows (Gordon & Grimmer-Somers, 2011). The sample size of this study included 106 randomly selected participants above the age of 18 which were then systematically divided into their age groups and asked to record their experience using the assigned pillow in the mornings and evenings in a tick box format for 7 consecutive days. The boxes contained options that helped obtain information about any sleeping or waking symptoms as well as sleep quality and pillow comfort. Waking symptoms were compared for six gender and age groups. Association between interval measures of pillow comfort and sleep quality was established through linear regression models. Logistic regression models with confidence intervals of 95% were used to establish relationships between binary forms of pillow comfort, sleep quality and waking symptoms. The result shows consistently poor sleeping comfort and quality for those using feather pillows, and consistently high sleeping comfort and quality for those using latex pillows. The result also shows that age and gender had little to no influence on sleep quality and comfort. Overall considering the sample size, method of data collection, analysis, and results obtained, it seems to be a reasonable study with minimal biases. Due to the study’s very minimal limitations, we can use this study as a starting point in optimizing our research on the effectiveness of different pillows on sleep quality and better refine our research plan.

Another study titled “Improving the Quality of Sleep with an Optimal Pillow: A Randomized Comparative Study” hypothesized that the pillow types would differ in the support for cervical lordosis, change of temperature as well as comfort scores (Jeon et al., 2014). Referring to other studies, the authors mention that a pillow that provides sufficient support for cervical lordosis could improve the quality of sleep (Persson & Moritz, 1998), and that which reduces head temperatures could be critical to deep sleep (Liu et al., 2011) . The sample size of their study consisted of 20 participants, 10 men and 10 women between the ages of 21-30, who did not have certain neurological and spinal conditions. The independent variables of the study were the three pillows (feather pillow, memory foam pillow, orthopedic pillow), and the dependent variables were cervical curve, pillow temperature, and pillow comfort. This study’s first result showed a significant increase in cervical curve for the orthopedic pillow compared to the other two pillows. The second result showed that the degree of increase in pillow temperature was significantly lesser for the orthopedic pillow than the other two pillows. The third result showed a significant difference in the pillow comfort scores. This study’s shortcomings are that the long-term effects of each pillow were not tracked and sample size was small. This study is relevant to the one we are carrying out as it shows that the pillow type can affect the cervical curve, pillow temperature and pillow comfort, which ultimately impact the sleep quality. Should we reject our null hypothesis for the first research question, the results of this study may explain the reason for the same.

Another research paper titled “Effect of different pillow designs on promoting sleep comfort, quality, spinal alignment: A systematic review” takes into account various peer-reviewed research papers and examines what characteristics make a quality pillow. It uses Grading of Recommendation, Assessment, Development, and Evaluation (GRADE) to evaluate the quality of evidence of these papers (Radwan, A., et al., 2020). The study aims to find the best parameter of a pillow, including material, height, shape, and thermal properties, that would enhance the best sleep quality. Many studies suggested that having a good posture during sleep, which means having an aligned spinal, would promote better sleep quality. In the paper, the data points the researcher chose for this paper included participants between the ages of 24 and 64. There were 304 participants across the 11 studies, with 225 females, 54 males, and the rest not recorded. The study's results showed that various peer-reviewed studies indicated that their participants experienced the least amount of scapular and arm pain and thus had the highest sleep quality from using latex pillows. This is another reason why we adapted latex pillows in our own experimental design. The downside of this research is that most, if not all, of the peer studies, were conducted for the short term. Therefore, it is hard to conclude how these pillow attributes would affect people in the long run .

“Age and gender variations of sleep in subjects without sleep disorders” by Luca et al. (2015) explored the influence of age and gender on sleep quality. The author divided 1147 subjects into four age groups of 40-50 years old, 60 years old, 60-70 years old, and 70-80 years old, and each age group was also divided into males and females. Eight indicators including electrograms were analyzed for nighttime dynamic polymorphism. The main result of the study was that sleep efficiency decreased with age in both genders, but men were more affected due to a decrease in spectral power density within slow waves. Moreover, it was found that sleep quality reported by participants was also enhanced with increase in age. A limitation of this study is that the time interval between questionnaire administration and somnography may have an impact on the subjective and objective measurements and lead to inaccurate results. The results of this study is going to provide backing for our second research question, where we are exploring whether people from a lower age group have a lower mean sleep score as compared to those from the higher age group.

There are many studies that evaluated sleep patterns using various types of wearable sleep monitors like wristband fitbit models. One such study (Haghayegh, 2019) “Accuracy of

Wristband Fitbit Models in Assessing Sleep: Systematic Review and Meta-Analysis" conducted a systematic review of other published research studies that utilized fit bits to monitor sleep. The method in which the research papers were selected were using keyword 'fitbit' in various databases such as Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Cochrane, Embase, MEDLINE, PubMed, PsycINFO, and Web of Science databases and pre-set inclusion and exclusion criteria. Totally there were 22 articles that qualified for the systematic review out of which 8 had quantitative data for performing meta analysis. The inclusion criteria included valid sleep data from any fitbit, containing Polysomnography (PSG), actigraphy, home EEG, sleep diary, or survey method as a reference. The exclusion criteria included sample size less than 5 participants, review of a paper, absent or incorrect use of statistical analysis, and duplications of same data and findings. The sleep parameters that were evaluated for precision included total sleep time (TST), sleep onset latency (SOL), wake after sleep onset (WASO), and sleep efficiency (SE), and the sensitivity, specificity, and accuracy of detecting both sleep epochs and sleep stages. The objective of the study was to determine how accurate fitbits or consumer wristband devices are in measuring sleep parameters. The findings of the study was that Polysomnography (PSG) was the best for the assessment of sleep parameters. When compared to non-sleep staging, fitbit models overestimate TST and SE, underestimate WASO, but determine SOL as well as PSG. This is why we chose to use fitbits for our study. It was important for us to understand the best way to measure sleep scores and this study evaluates various methods of measuring sleep quality. We also believe using the fitbit to measure sleep quality might also reduce the influence of subjective degrees.

Research Plan

<u>Population of Interest:</u>

For our research, our population of interest includes participants in the age range of 21-60. As found in the study conducted by Luca et al 2015, sleep quality reported by participants increased with age. Therefore, age has an impact on sleep and we want to study its effect. Since we also studied that sleep efficiency of men was more affected (Luca et al., 2015), we would like to include an equal number of males and females, to make sure our result is not biased due to an imbalance. We are also only interested in participants who do not have pre-existing conditions, such as those with neck injuries, insomnia or other cervical and back problems, Participants in the study conducted by Jeon et. al (2014) also did not have any of certain neurological and spinal conditions. Therefore, their participation in our study might impact the participants further negatively, and as such, will not be ethical. Moreover, it might also impact the results of the study. Since Mattress Firm is an American mattress brand, we will be conducting this study only on participants based in the United States of America.

Sample Selection:

We will recruit participants from Dynata, by first sending out preliminary surveys to potential participants within the United States of America. These surveys will ask the following questions:

1. Do you have any existing back and/or neck conditions or insomnia?
2. What is your age?
3. What is your gender?

If they respond yes to the first question, we will automatically exclude them from the study. Further, if they are below the age of 21, or above the age of 60, we will exclude them as well. Lastly, we are recording information on gender to prevent imbalance between treatment and control groups, as well as for potential future studies and analyses.

Once we have the preliminary survey data, we will randomly select 100 people from each combination of age group (21-40 and 41-60) and gender (male and female). Then we will randomly assign 50 to treatment and 50 to control as below. The reason for doing so is to make sure that age groups and gender are equally represented, and to minimize the attribution of results to confounding variables. We will then reach out to the qualifying participants. We will assume that participants are independent from each other and do not know each other in advance, so that they may not exchange pillows and there is no interaction between treatment and control groups. Reasons for dividing by age and gender have been described under the population of interest.

Thus we want to conduct a randomized controlled trial, using stratified random sampling, while also adjusting for other variables.

Age Group	**Gender**	**Experimental Group**	**Number of people**
21-40	Male	Treatment	50
21-40	Male	Control	50
21-40	Female	Treatment	50
21-40	Female	Control	50
41-60	Male	Treatment	50
41-60	Male	Control	50

41-60	Female	Treatment	50
41-60	Female	Control	50

Operational Procedures:

We will spend the first two weeks recruiting, as described in the brief schedule section. We will host virtual information sessions for the two groups (treatment and control) on different occasions to run them through the basic guidelines for participating in our research. The basic guidelines are:

1. Always wear a Fitbit that charges more than 20% to bed. A regular Fitbit's battery life lasts 5-7 days so it shouldn't be much of an inconvenience to the participants.
2. Follow the same sleep schedule as one had before volunteering for this study. We want the most realistic effect of a pillow change, so we want the participants to have the same routine.
3. Contact us immediately if they observe any irregularities with Fitbit. We will get in touch and replace any malfunctioning Fitbit as soon as possible to ensure the documentation and proceed.

In addition to running through the guidelines, we will collect the participants' gender and address information. We will also address any questions they may have regarding data privacy and how we're planning on storing their information during the session (more on data privacy in a later section of this paper). We will also designate the participants with individual IDs to protect their privacy. Before the distribution process, we would set up the Fitbit profile for each device we're handing out and have the devices labeled with serial numbers. We will hand the designated pillow and a Fitbit Charge 5 with a charger to the participants, by mailing it to their home addresses. The study will be conducted for both the control and treatment groups simultaneously. Two weeks later, we would regroup to either collect the equipment or have the participants ship back the equipment assigned to them. Finally, we would export their sleep scores from the web profiles we have created onto the AWS database.We would later take around three weeks to a month to gather our findings and put together a report for the company. To incentivize participants to participate in the study, we will give a $25 Amazon gift card at the end of the study. Mattress Firm will fund the entire research including, the cost of Fitbits and pillows, shipping costs, and the gift cards to incentivize the participants.

Brief Schedule:

This experiment will take around 2 months to complete and will be split into three different phases:

Phase 1 Preparation: This phase would take approximately 2 weeks to complete. During this phase, we would be recruiting participants, gathering equipment and setting up the Fitbits.

Phase 2 Exercise Experiment: This phase would take 2 weeks to complete. During this phase, the participants would be wearing the Fitbit to sleep and we will closely monitor any irregularities from the recordings through the profiles that we've created. The reason for choosing a 2 week period is to address the limitation of a study by Jeon et. al (2014) reviewed above which was carried out only over two days. This was also a limitation mentioned in Radwan, A., et al. (2020) reviewed above, which has multiple tests carried out for different time periods but all shorter than 2 weeks.

Phase 3 Analyze Data: This phase would take 3 weeks to a month to complete. During this phase, we will analyze the data collected from the Fitbits and produce an in-depth report at the end of the phase.

Data Collection:

We will collect the sleep score data via the Fitbit web profiles that we create prior to distributing the Fitbit watches. During the two week experimental process, we will keep track of the data input every morning as their Fitbit watches will be synced to the web profiles. After the two weeks, we would retrieve the sleep score data from the Fitbit web profiles and upload them into our Amazon Web Services cloud database securely. We will later download and use this data for our analyses after the experimental process concludes.

Data Security:

It will be important to store data for this study in an efficient and secure manner. Therefore, the data obtained from this study is going to be stored in the Amazon web services (AWS) cloud. We believe that AWS has the best data storage and provides the greatest data security in the market, since it ensures that data privacy for participants will be protected. Moreover, they are also the most cost efficient. Mattress Firm has previously employed an AWS certified engineer, whom they are willing to deploy for the purposes of this project. Moreover, in order to protect participants' privacy, we will be assigning participants with unique identifiers so that they may not be identified by their names.

Outcomes (Dependent Variables):

Our dependent variable is the sleep score that we measure on the participants' Fitbit device. We believe that the sleep scores will be an objective indicator of sleep quality. The sleep score will be measured for each participant of the study on the Fitbit charge 5 watch assigned to them, and at the end of the study, the mean sleep scores will be calculated as needed. To answer our first research question, we will measure the mean sleep score for each of the treatment and the control groups. To answer our second research question, we will measure the mean sleep score for each of the two age groups in the study.

Treatments (Independent Variables):

The independent variable for the first research question is the type of pillow that we are testing. The control is the standard feather pillow, and the treatment is the latex pillow. We have chosen these two pillows because the study conducted by Gordon and Grimmer-Somers 2011 shows that feather and latex pillow shows the biggest difference in sleep quality and comfort. Our independent variable for the second research question is age and we plan to have an equal number of participants in two different age cohorts; 21-40 and 41-60. We are studying the impact of age as the study by Luca et al 2015 showed that sleep quality reported by participants increased with age.

Other Variables:

Gender is an important variable, and we will collect gender information of the subjects from the preliminary surveys in Phase 1, in case future research requires this information. In our study we are considering gender as a binary variable (only male and female). Moreover, even though we are not studying its impact, we need to control for gender, so we have included an equal number of each gender in both the treatment and control groups.

Statistical Analysis Plan

First research question plan:

To analyze whether the mean sleep score for latex pillow is greater than that for standard pillows, we will utilize the two sample one-sided t test of means method. First we need to calculate the average sleep score for two pillow types. We have assumed a mean of 65 for the treatment group and 60 for the control group, and a standard deviation of 10 for both the groups. With an assumed effect size of 5, and an alpha of 0.05 as a significance level. If $p < 0.05$, it suggests that there is significant evidence stating that mean sleep score for latex pillows is greater than that for standard pillows. If $p > 0.05$, it indicates that there is no statistical evidence that mean sleep score for latex pillows is greater than that for standard pillows.

Results:

```
data:  bp.dat[Group == "Treatment", Score] and bp.dat[Group == "Control", Score]
t = 4.806, df = 397.85, p-value = 1.093e-06
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 3.052863      Inf
sample estimates:
mean of x mean of y
   64.256    59.609
```

Analysis:

The above test result shows that the p-value is less than our significance level of 0.05. Therefore, based on this finding, we reject the null hypothesis. There is enough evidence to conclude that the mean sleep score for latex pillows is significantly higher than that of standard feather pillows.

Second research question:

The second research question will analyze the difference between mean sleep scores of the age group, 21-40 and 41-60. In order to analyze whether the sleep score for age group 21-40 is lower than that for age group 41-60, we will use a two sample one-sided t test of means method as well. We have assumed a mean of 52 for the age group 21-40, 60 for 41-60 and a standard deviation of 10 for both groups. With the assumed effect size of 8 and utilizing an alpha level of 0.05 as a significance level, we will then be able to conclude findings from the experiment. If $p<0.05$, it suggests that there is significant evidence stating that mean sleep score for age group 21-40 is lower than the mean sleep score for age group 41-60. On the other hand, if $p>0.05$, we do not have enough evidence to reject the null hypothesis, therefore the mean sleep score for age group 21-40 is not significantly lower than the mean sleep score for age group 41-60.

Results:

```
data:  bp.dat[Group == "Age21-40", Score] and bp.dat[Group == "Age41-60", Score]
t = -7.9087, df = 397.85, p-value = 1.297e-14
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf -6.052863
sample estimates:
mean of x mean of y
   51.609    59.256
```

Analysis:

The above test result shows that the p-value is less than our significance level of 0.05. Therefore, based on this finding, we reject the null hypothesis. There is enough evidence to conclude that the mean sleep score of age group 21-40 is significantly lower than the mean sleep score of age group 41-60.

(The R code for the t-tests of both questions has been included in the appendix)

Sample Size and Statistical Power

<u>Sample size</u>

After the screening, we plan to have 400 participants in total. The specific division has been mentioned under sample selection. There will be 200 people in the age group 21-40, and 200 people in the age group 41-60. Of the 200 people in each age group, we will have an equal number of participants of each gender, which implies 100 people in each combination of age group and gender. We will then randomly assign 50 to treatment and 50 to control in each subgroup. We are ensuring a balance of gender to adjust for confounding variables. We will use our assumed sample size and effect size to determine the statistical power of our test. The reason for choosing a larger sample size is to address the limitation of a study by Jeon et. al (2014) which had a small sample size. The company's (Mattress Firm) budget for this research allows for a larger number of participants.

<u>Statistical power</u>

We have used the t-test library in the (pwr) package that is included in R to calculate our statistical power. These are the results based on the different research questions.

Research Question 1
Difference in Mean = 5
N1 = 200
N2 = 200
SD = 10
Significance Level = 0.05
Power Result = 0.99

```
    t test power calculation

              n1 = 200
              n2 = 200
               d = 0.5
       sig.level = 0.05
           power = 0.999591
     alternative = greater
```

Research Question 2
Difference in Mean = 8
N1 = 200
N2 = 200
SD = 10
Significance Level = 0.05
Power Result = 3.33e-16

```
  t test power calculation

          n1 = 200
          n2 = 200
           d = 0.8
   sig.level = 0.05
       power = 3.330669e-16
 alternative = less
```

(The R code for the calculation of powers has been included in the appendix)

Possible Recommendations

The results of the experiment suggest that the mean sleep score for the latex pillow is significantly higher than for the standard feather pillow. It also suggests that the age group 21-40 have a significantly lower mean sleep score than the age group 41-60. This suggests that Mattress Firm can use the results of this research study to create an alluring advertisement campaign for latex pillows as improving sleep quality. Moreover, their target audience can be the population in age group 21-40, which not only might help bring in more revenue for the company, but also given that Mattress Firm cares about its customers, it can help in improving sleep quality of their customers who fall under that age group. Lastly, this gives Mattress Firm the opportunity to explore other demographic variables in further research studies that could have a potential impact on sleep quality. This way, the advertisement can be catered to that section of the population. For instance, they could see if sleep quality differs by gender or income group. Creating an even more targeted advertising campaign in the long run will help them have a higher customer conversion rate which can increase revenue.

Limitations and Uncertainties

Although this study has been designed to reduce the limitations, there are a few. Since we are randomly assigning to treatment and control, a small part of the differences in mean sleep scores for the two groups could be attributed to differences between individuals. However, we have tried to control for this by dividing into age groups and gender. Moreover, although we have tried to control for the impact of gender by including an equal number of participants in each age group and gender, we have included only binary genders. So, this study may exclude participants who identify as non-binary. Secondly, participants could pass their fitbit to someone else just to get the reward for the study. They could also use their own everyday pillows to sleep on so there is no way to actually monitor that and whether they are using their own pillow or pillow assigned to them. Thirdly, blinding might not be possible that participants might know which pillow they are assigned to based on pillow shape and other characteristics. This study may have a participation bias, as only those who respond to the preliminary survey or have an interest in the study will participate. Lastly, although we tried to perform sensitivity analyses by changing certain parameters while simulating the data, we got a low power (using pwr.t.test) for our second research question in Part 1 of the study. This could potentially be addressed by increasing the sample size in a future study.

Part 2: Simulation of Effects

Research Question	Scenario	Mean Effect in Simulated data	95% in Confidence interval of mean effect	% of false positive	% of true positive	% of false negative	% of true negative
Question 1	No effect	0.2496	(1.3999,Inf)	0.05	0.078	0.922	0.95
Question 1	Effect	5.05	(3.4001,Inf)	0.05	1	0	0.95
Question 2	No effect	-0.1496	(-Inf ,1.4999)	0.05	0.061	0.939	0.95
Question 2	Effect	-8.05	(-Inf,-6.4)	0.05	1	0	0.95

We followed the logic below for each of the columns as follows:

1. Mean effect in simulated data: mean of effect size across the 1000 experiments
2. 95% confidence interval of mean effect: means of lower and upper bound across the 1000 experiments
3. % of false positive: alpha (significance level)
4. % of true negative: 1 - alpha (confidence level)
5. % of false negative: beta
6. % of true positive: 1 - beta (power)

According to the results provided above, the power (shown by % of true positives) is 1, for the effect scenarios of question 1 and question 2. In addition, the absolute mean effect in the effect scenarios is large for both questions (5.05 and 8.05), and in the no effect scenarios it is small for both questions (0.25 and 0.15). This indicates that there is a meaningful difference between the mean sleep score of the treatment and control groups and in the mean sleep scores of age groups 21-40 and 41-60, in the effect scenarios. There is no meaningful difference for both the questions in the no effect scenarios. This demonstrates that for both research questions, there is convincing evidence that we are correctly rejecting the null hypothesis when it is false. For the 95% confidence interval of mean effect, we found that all of them have one side as infinity.

In conclusion, we can say that our alternative hypotheses have been consolidated after testing our assumptions in the 1000 simulated experiments.

(The R code for the simulation of 1000 experiments has been included in the appendix)

References

Bernateck, M., Karst, M., Merkesdal, S., Fischer, M.J. & Gutenbrunner, C. (2008). Sustained effects of comprehensive inpatient rehabilitative treatment and sleeping neck support in patients with chronic cervicobrachialgia: a prospective and randomized clinical trial. International Journal of Rehabilitation Research, 31(4), 342-346.

Best mattress prices-top brands-same day delivery. Mattress Firm. (n.d.). Retrieved November 25, 2022, from https://www.mattressfirm.com/

Centers for Disease Control and Prevention. (2016, February 16). 1 in 3 adults don't get enough sleep. Centers for Disease Control and Prevention. Retrieved October 18, 2022, from https://www.cdc.gov/media/releases/2016/p0215-enough-sleep.html

Cloud computing services - amazon web services (AWS). (n.d.). Retrieved December 2, 2022, from https://aws.amazon.com/

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Gordon, S. J., & Grimmer-Somers, K. (2011). Your Pillow May Not Guarantee a Good Night's Sleep or Symptom-Free Waking. Physiotherapy Canada. Physiotherapie Canada, 63(2), 183–190. https://doi.org/10.3138/ptc.2010-13

Haghayegh, Shahab, et al. “Accuracy of Wristband Fitbit Models in Assessing Sleep: Systematic Review and Meta-Analysis.” Journal of Medical Internet Research, U.S. National Library of Medicine, 28 Nov. 2019, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6908975/.

Jeon, M. Y., Jeong, H. C., Lee, S. W., Choi, W., Park, J. H., Tak, S. J., Choi, D. H., & Yim, J. (2014). Improving the quality of sleep with an optimal pillow: A randomized, comparative study. The Tohoku Journal of Experimental Medicine, 233(3), 183–188.
https://doi.org/10.1620/tjem.233.183

Kyle, S.D., Morgan, K. & Espie, C.A. (2010). Insomnia and health-related quality of life. Sleep Med. Rev., 14(1), 69-82.

Liu, S.-F., Lee, Y.-L., & Liang, J.-C. (2011). Shape design of an optimal comfortable pillow based on the analytic hierarchy process method. Journal of Chiropractic Medicine, 10(4), 229–239. https://doi.org/10.1016/j.jcm.2011.04.002

Luca, G., Rubio, J. H., Andries, D., Tobback, N., Vollenweider, P., Waeber, G., Marques, P., Preisig, V. M., Heinzer, R., Tafti, M. (2015, July 29). Age and gender variations of sleep in subjects without sleep disorders. Taylor & Francis. Retrieved November 20, 2022, from https://www.tandfonline.com/doi/full/10.3109/07853890.2015.1074271

Persson, L., & Moritz, U. (1998). Neck support pillows: a comparative study. Journal of Manipulative and Physiological Therapeutics, 21(4), 237–240. https://pubmed.ncbi.nlm.nih.gov/9608378/

Pacheco, D., Singh, A. (Eds.). (2022, August 10). Why do we need sleep? Sleep Foundation Retrieved November 20, 2022, from https://www.sleepfoundation.org/how-sleep-works/why-do-we-need-sleep

Radwan, A., Ashton, N., Gates, T., Kilmer, A., VanFleet, M. (2020, December 26). Effect of different pillow designs on promoting sleep comfort, quality, spinal alignment: A systematic review. European Journal of Integrative Medicine. Retrieved November 11, 2022, from https://www.sciencedirect.com/science/article/pii/S1876382020314505?casa_token=RyRgzBl7IREAAAAA%3AjI282fGvyBmanfkXLni3ZPWUa-tbbktHwPwmTW2wK_aAPMGLnDHsyv10O8Wdn1sCOhPagaDQ

Schutz, T.C., Andersen, M.L. & Tufik, S. (2009). The influence of orofacial pain on sleep pattern: a review of theory, animal models and future directions. *Sleep Medicine* 10, no. 8 (2009): 822-828.

World's largest first party data platform. Dynata. (2022, September 27). Retrieved November 23, 2022, from https://www.dynata.com/

Appendix

```
#Calculation of power
library(pwr)
pwr.t2n.test(n1 = 200, n2 = 200, d = 5/10, sig.level = 0.05,
        alternative = "greater")
pwr.t2n.test(n1 = 200, n2 = 200, d = 8/10, sig.level = 0.05,
        alternative = "less")
# Simulate the data
n <- 400
library(data.table)
library(DT)
set.seed(seed = 329)
bp.dat1 <- data.table(Age_Group = c(rep.int(x = "21-40", times = n/16)),
                Gender = c(rep.int(x = "Male", times = n/16)),
               Group = c(rep.int(x = "Treatment", times = n/16)))
bp.dat1[Age_Group == "21-40", Score := round(x = rnorm(n = .N, mean = 65, sd = 10), digits =
1)]
bp.dat1
bp.dat2 <- data.table(Age_Group = c(rep.int(x = "21-40", times = n/16)),
               Gender = c(rep.int(x = "Male", times = n/16)),
               Group = c(rep.int(x = "Control", times = n/16)))
bp.dat2[Age_Group == "21-40", Score := round(x = rnorm(n = .N, mean = 60, sd = 10), digits =
1)]
bp.dat2
bp.dat3 <- data.table(Age_Group = c(rep.int(x = "21-40", times = n/16)),
               Gender = c(rep.int(x = "Female", times = n/16)),
               Group = c(rep.int(x = "Treatment", times = n/16)))
bp.dat3[Age_Group == "21-40", Score := round(x = rnorm(n = .N, mean = 65, sd = 10), digits =
1)]
bp.dat3
bp.dat4 <- data.table(Age_Group = c(rep.int(x = "21-40", times = n/16)),
               Gender = c(rep.int(x = "Female", times = n/16)),
               Group = c(rep.int(x = "Control", times = n/16)))
bp.dat4[Age_Group == "21-40", Score := round(x = rnorm(n = .N, mean = 60, sd = 10), digits =
1)]
bp.dat4
bp.dat5 <- data.table(Age_Group = c(rep.int(x = "41-60", times = n/16)),
```

```
                   Gender = c(rep.int(x = "Male", times = n/16)),
                   Group = c(rep.int(x = "Treatment", times = n/16)))
bp.dat5[Age_Group == "41-60", Score := round(x = rnorm(n = .N, mean = 65, sd = 10), digits =
1)]
bp.dat5

bp.dat6 <- data.table(Age_Group = c(rep.int(x = "41-60", times = n/16)),
                   Gender = c(rep.int(x = "Male", times = n/16)),
                   Group = c(rep.int(x = "Control", times = n/16)))
bp.dat6[Age_Group == "41-60", Score := round(x = rnorm(n = .N, mean = 60, sd = 10), digits =
1)]
bp.dat6

bp.dat7 <- data.table(Age_Group = c(rep.int(x = "41-60", times = n/16)),
                   Gender = c(rep.int(x = "Female", times = n/16)),
                   Group = c(rep.int(x = "Treatment", times = n/16)))
bp.dat7[Age_Group == "41-60", Score := round(x = rnorm(n = .N, mean = 65, sd = 10), digits =
1)]
bp.dat7

bp.dat8 <- data.table(Age_Group = c(rep.int(x = "41-60", times = n/16)),
                   Gender = c(rep.int(x = "Female", times = n/16)),
                   Group = c(rep.int(x = "Control", times = n/16)))
bp.dat8[Age_Group == "41-60", Score := round(x = rnorm(n = .N, mean = 60, sd = 10), digits =
1)]
bp.dat8

bp.dat <- rbind(bp.dat1,bp.dat2,bp.dat3,bp.dat4,bp.dat5,bp.dat6,bp.dat7,bp.dat8)
datatable(data = bp.dat)

### 2 sample T-test for age group(effect)
n <- 400
library(data.table)
library(DT)
set.seed(seed = 329)
# Simulate the data
bp.dat <- data.table(Group = c(rep.int(x = "Age21-40", times = n/2), rep.int(x = "Age41-60",
times = n/2)))
```

```
bp.dat[Group == "Age21-40", Score := round(x = rnorm(n = .N, mean = 52, sd = 10), digits = 1)]
bp.dat[Group == "Age41-60", Score := round(x = rnorm(n = .N, mean = 60, sd = 10), digits = 1)]
datatable(data = bp.dat)

t.test(x = bp.dat[Group == "Age21-40",
         Score], y = bp.dat[Group == "Age41-60", Score], mu = 0, alternative = "less")

analyze.experiment <- function(the.dat) {
  require(data.table)
  setDT(the.dat)

  the.test <- t.test(x = the.dat[Group == "Age21-40",
                   Score], y = the.dat[Group == "Age41-60", Score], alternative = "less")

  the.effect <- the.test$estimate[1] - the.test$estimate[2]
  upper.bound <- the.test$conf.int[2]
  lower.bound <- the.test$conf.int[1]
  p <- the.test$p.value

  result <- data.table(effect = the.effect, upper_ci = upper.bound,lower_ci=lower.bound,
           p = p)

  return(result)
}

analyze.experiment(the.dat = bp.dat)

# Run the simulation 1000 experiments
library(dplyr)
library(readr)
library(stringr)
library(tidyverse)
B <- 1000
n <- 400
RNGversion(vstr = 3.6)
set.seed(seed = 4172)
Experiment <- 1:B
```

```
Group <- c(rep.int(x = "Age21-40", times = n/2), rep.int(x = "Age41-60", times = n/2))

sim.dat <- as.data.table(expand.grid(Experiment = Experiment, Group = Group))
setorderv(x = sim.dat, cols = c("Experiment", "Group"), order = c(1,1))
sim.dat[Group == "Age21-40", Score := round(x = rnorm(n = .N, mean = 52, sd = 10), digits =
1)]
sim.dat[Group == "Age41-60", Score := round(x = rnorm(n = .N, mean = 60, sd = 10), digits =
1)]
dim(sim.dat)

exp.results <- sim.dat[, analyze.experiment(the.dat = .SD),
                  keyby = "Experiment"]

DT::datatable(data = round(x = exp.results[1:100, ], digits = 100),
            rownames = F)

exp.results[, mean(p < 0.05)]
exp.results[, summary(effect)]
exp.results[, summary(upper_ci)]
exp.results[, summary(lower_ci)]
exp.results[, mean(p > 0.05)]

### 2 sample T-test for age group(no effect)
n <- 400
library(data.table)
library(DT)
set.seed(seed = 329)

bp.dat <- data.table(Group = c(rep.int(x = "Age21-40", times = n/2), rep.int(x = "Age41-60",
times = n/2)))

bp.dat[Group == "Age21-40", Score := round(x = rnorm(n = .N, mean = 52, sd = 10), digits = 1)]
bp.dat[Group == "Age41-60", Score := round(x = rnorm(n = .N, mean = 52.1, sd = 10), digits =
1)]
datatable(data = bp.dat)
t.test(x = bp.dat[Group == "Age21-40",
               Score], y = bp.dat[Group == "Age41-60", Score], mu = 0, alternative = "less")
```

```
analyze.experiment <- function(the.dat) {
  require(data.table)
  setDT(the.dat)

  the.test <- t.test(x = the.dat[Group == "Age21-40",
                     Score], y = the.dat[Group == "Age41-60", Score], alternative = "less")

  the.effect <- the.test$estimate[1] - the.test$estimate[2]
  upper.bound <- the.test$conf.int[2]
  lower.bound <- the.test$conf.int[1]
  p <- the.test$p.value

  result <- data.table(effect = the.effect, upper_ci = upper.bound, lower_ci=lower.bound,
             p = p)

  return(result)
}

analyze.experiment(the.dat = bp.dat)

# Run the simulation 1000 times
library(dplyr)
library(readr)
library(stringr)
library(tidyverse)
B <- 1000
n <- 400
RNGversion(vstr = 3.6)
set.seed(seed = 4172)
Experiment <- 1:B
Group <- c(rep.int(x = "Age21-40", times = n/2), rep.int(x = "Age41-60", times = n/2))

sim.dat <- as.data.table(expand.grid(Experiment = Experiment, Group = Group))
setorderv(x = sim.dat, cols = c("Experiment", "Group"), order = c(1,1))
sim.dat[Group == "Age21-40", Score := round(x = rnorm(n = .N, mean = 52, sd = 10), digits =
1)]
sim.dat[Group == "Age41-60", Score := round(x = rnorm(n = .N, mean = 52.1, sd = 10), digits =
```

```
1)]
dim(sim.dat)

exp.results <- sim.dat[, analyze.experiment(the.dat = .SD),
                       keyby = "Experiment"]

DT::datatable(data = round(x = exp.results[1:100, ], digits = 100),
          rownames = F)

exp.results[, mean(p < 0.05)]
exp.results[, summary(effect)]
exp.results[, summary(upper_ci)]
exp.results[, summary(lower_ci)]
exp.results[, mean(p > 0.05)]


### 2 sample T-test for treatment/control (effect)
n <- 400
library(data.table)
library(DT)
set.seed(seed = 329)

bp.dat <- data.table(Group = c(rep.int(x = "Treatment", times = n/2), rep.int(x = "Control", times
= n/2)))

bp.dat[Group == "Control", Score := round(x = rnorm(n = .N, mean = 60, sd = 10), digits = 1)]
bp.dat[Group == "Treatment", Score := round(x = rnorm(n = .N, mean = 65, sd = 10), digits = 1)]
datatable(data = bp.dat)

t.test(x = bp.dat[Group == "Treatment",
              Score], y = bp.dat[Group == "Control", Score], mu = 0, alternative = "greater")

analyze.experiment <- function(the.dat) {
  require(data.table)
  setDT(the.dat)

  the.test <- t.test(x = the.dat[Group == "Treatment",
                          Score], y = the.dat[Group == "Control", Score], alternative = "greater")
```

```
  the.effect <- the.test$estimate[1] - the.test$estimate[2]
  upper.bound <- the.test$conf.int[2]
  lower.bound <- the.test$conf.int[1]
  p <- the.test$p.value

  result <- data.table(effect = the.effect, upper_ci = upper.bound,lower_ci= lower.bound,
                       p = p)

  return(result)
}

analyze.experiment(the.dat = bp.dat)

# Run the simulation 1000 times
library(dplyr)
library(readr)
library(stringr)
library(tidyverse)
B <- 1000
n <- 400
RNGversion(vstr = 3.6)
set.seed(seed = 4172)
Experiment <- 1:B
Group <- c(rep.int(x = "Treatment", times = n/2), rep.int(x = "Control", times = n/2))

sim.dat <- as.data.table(expand.grid(Experiment = Experiment, Group = Group))
setorderv(x = sim.dat, cols = c("Experiment", "Group"), order = c(1,1))
sim.dat[Group == "Control", Score := round(x = rnorm(n = .N, mean = 60, sd = 10), digits = 1)]
sim.dat[Group == "Treatment", Score := round(x = rnorm(n = .N, mean = 65, sd = 10), digits =
1)]
dim(sim.dat)

exp.results <- sim.dat[, analyze.experiment(the.dat = .SD),
                       keyby = "Experiment"]

DT::datatable(data = round(x = exp.results[1:100, ], digits = 20),
              rownames = F)
```

```
exp.results[, mean(p < 0.05)]
exp.results[, summary(effect)]
exp.results[, summary(upper_ci)]
exp.results[, summary(lower_ci)]
exp.results[, mean(p > 0.05)]

### 2 sample T-test for treatment/control (no effect)
n <- 400
library(data.table)
library(DT)
set.seed(seed = 329)

##Simulate the data
bp.dat <- data.table(Group = c(rep.int(x = "Treatment", times = n/2), rep.int(x = "Control", times
= n/2)))

bp.dat[Group == "Control", Score := round(x = rnorm(n = .N, mean = 60, sd = 10), digits = 1)]
bp.dat[Group == "Treatment", Score := round(x = rnorm(n = .N, mean = 60.2, sd = 10), digits =
1)]
datatable(data = bp.dat)

t.test(x = bp.dat[Group == "Treatment",
            Score], y = bp.dat[Group == "Control", Score], mu = 0, alternative = "greater")

analyze.experiment <- function(the.dat) {
  require(data.table)
  setDT(the.dat)

  the.test <- t.test(x = the.dat[Group == "Treatment",
                  Score], y = the.dat[Group == "Control", Score], alternative = "greater")

  the.effect <- the.test$estimate[1] - the.test$estimate[2]
  upper.bound <- the.test$conf.int[2]
  lower.bound <- the.test$conf.int[1]
  p <- the.test$p.value

  result <- data.table(effect = the.effect, upper_ci = upper.bound, lower_ci=lower.bound,
```

```
                p = p)

  return(result)
}

analyze.experiment(the.dat = bp.dat)

### Run the simulation 1000 times
library(dplyr)
library(readr)
library(stringr)
library(tidyverse)
B <- 1000
n <- 400
RNGversion(vstr = 3.6)
set.seed(seed = 4172)
Experiment <- 1:B
Group <- c(rep.int(x = "Treatment", times = n/2), rep.int(x = "Control", times = n/2))

sim.dat <- as.data.table(expand.grid(Experiment = Experiment, Group = Group))
setorderv(x = sim.dat, cols = c("Experiment", "Group"), order = c(1,1))
sim.dat[Group == "Control", Score := round(x = rnorm(n = .N, mean = 60, sd = 10), digits = 1)]
sim.dat[Group == "Treatment", Score := round(x = rnorm(n = .N, mean = 60.2, sd = 10), digits =
1)]
dim(sim.dat)

exp.results <- sim.dat[, analyze.experiment(the.dat = .SD),
                keyby = "Experiment"]

DT::datatable(data = round(x = exp.results[1:100, ], digits = 3),
          rownames = F)

exp.results[, mean(p < 0.05)]
exp.results[, summary(effect)]
exp.results[, summary(upper_ci)]
exp.results[, summary(lower_ci)]
exp.results[, mean(p > 0.05)]
```