

# STA304 - Fall 2022

## Assignment 1

Yihan Wang - 1007499789

2022 September 17

### Part 1

#### Goal

Approximately 40% of the population has difficulty initiating or maintaining sleep, as well as being sleepy during the daytime (Dugas et al., 2017). More than a third of adults do not meet the current recommendation of 7–9 hours of sleep per night (Hirshkowitz et al., 2015). It is therefore necessary to identify the determinants of poor sleep to improve sleep quality. There is a possibility that nicotine, a key chemical compound in tobacco that leads to compulsive drug use or addiction (Centers for Disease Control and Prevention (US) et al., 2010), may affect the central nervous system pharmacologically by stimulating neurotransmitters involved in sleep-wake cycles, which could result in a rise in sleep latency and insomnia (Jaehne et al., 2012).

In this report, a survey is designed and implemented in order to collect sleep data from adults. To observe how smoking can affect sleep quality, several variables are collected, including sleep duration, time spent falling asleep, sleep medication use, and poor sleep frequency.

#### Procedure

From September 23-26, 2022, the survey was posted through the social networking platform called WeChat, which is similar to Facebook but most of its users are Chinese people (Zhang et al., 2017). The target population of the survey is all adults whose age is no less than 18. Therefore, the geological bias exists since most people who can access this survey is only from one country and it is not representative of the target population.

The survey was shared to the WeChat moments of the author's personal account, the survey can be opened by clicking the link and a brief description of the survey is provided along with the link. The frame population is the list of social accounts which are among the WeChat connections with the author (approximately 400 people). The sampled population is the people who filled out this survey.

Since the sampling method is non-probability, and taking this survey is totally voluntary based, people who response this survey may have specific interest about sleep quality and are also more likely to experience sleep disturbances (Cheung et al., 2017). Additionally, since most people among the author's WeChat friends are undergraduate students, the age group collected would lack variety. However, even though bias existed, the volunteer-based method is efficient to collect enough responses in a short time, it is convenient to operate, and the survey can just be posted to the moments in WeChat, which is also highly cost-effective.

## Showcasing the survey

<https://forms.gle/S7ETuao5y7q82oL59>

### Question1

Have you ever smoked cigarettes or other tobacco in the last month? (including e-cigarettes)

- Yes
- No

The purpose of this question is to distinguish smokers from non-smokers, because we are analyzing the effect of smoking on sleep quality. The time period is specified during the past one month, so people won't be confused about which choice they should make if they smoked before but quit smoking. However, since the question is designed to be limited to one month, smokers who have been trying to quit during the last month, are likely to experience withdrawal effects (Patterson et al., 2019), which could lead to different results from those who don't smoke. Moreover, non-response bias may exist since the participants can only choose between the smoking and non-smoking.

### Question2

How many hours of sleep do you typically obtain at night in the last month? (Please only enter numbers)

Participants are expected to enter only numbers in the units of hours for this short answer question. This question measures sleep duration, which will be compared later between smokers and non-smokers. As the question specifies the format of the answer, only numbers can be submitted, which will facilitate data processing. There are however some special cases, for example, when the participant was on holiday during the month or busy preparing for a test, which may cause them to sleep longer or shorter than normal. Limiting the measurement to only a month may cause bias of the sleep duration.

### Question3

Do you need sleep medications to help you fall asleep? (During the past one month)

- Yes
- No

This question is aimed to reflect the ability to fall asleep by observing the use of sleep medications. This question is clear and since the time period is also specified, it will be straightforward for people to answer whether they have taken sleep medications or not during the past one month. However, there are also other ways to help fall asleep such as drinking hot milk and reading instead of taking sleep medications, therefore, some people who have trouble in falling asleep may be ignored if they choose other ways rather than taking medications (Breedlove, 2021).

## Part 2

### Data

- Data Collection Process

The survey was posted to the WeChat moments of the author’s account, participants can have access to the survey by clicking the link in the moments shared. Using Google Forms to design the survey and collect the responses, the platform can also automatically output the responses into a spreadsheet, which would be uploaded to R studio to conduct data analysis.

Although it is convenient to collect the data by sharing the survey link through WeChat moments, non-response bias exists since the people who voluntarily reply to this survey may have problems in sleep quality and show more attention to this topic. Moreover, aiming to simplify data processing, there is no open-ended questions in the survey, which may restrain the information we can get from the participants, for example, whether they have sleep disorders or whether they drink alcohol or exercise before sleeping, those are the confounding variables which will affect sleep quality as well.

- Data Showcasing

There are in total 57 responses received, the data contain 8 variables, including the background variables such as gender, marital status and age group, the categorical variable which we mainly focus on in this study is the smoking status, the numerical variables that represent the sleep quality are the sleep duration and time spent falling asleep, and the other two categorical variables which can reflect sleep disturbances are the use of sleep medication and the frequency of poor sleep.

The data is reformatted before further analysis, the original data contains the questions which make it inconvenient to refer to variables, so those questions are renamed and the updated version of the name of variables is shown in Table 1. Furthermore, the responses with missing answers are removed from the data in order to eliminate invalid results. However, removing those data may cause bias, since in this study the interest is in smoking status and sleep quality, if there are missing answers in gender or other variables that are not the main interest, deleting those responses may be unnecessary and the data may lose the information from those responses.

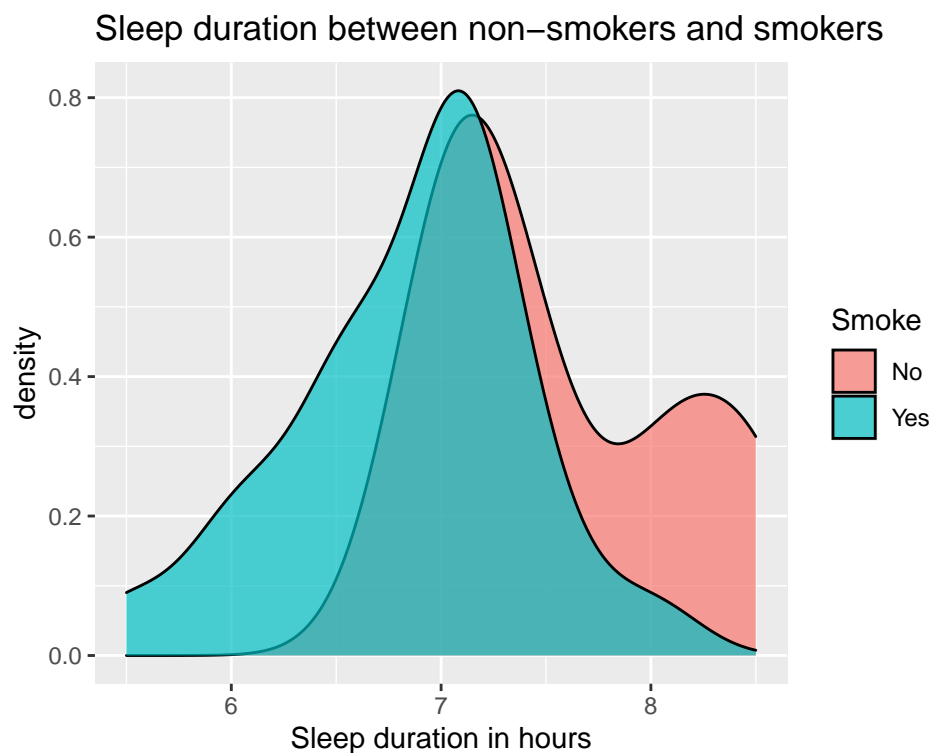
Table 1: Survey data summary

Smoke or not	No, N = 24	Yes, N = 23
<b>Gender</b>		
Man	12 / 24 (50%)	9 / 23 (39%)
Prefer not to answer	1 / 24 (4.2%)	0 / 23 (0%)
Woman	11 / 24 (46%)	14 / 23 (61%)
<b>Marital Status</b>		
Married	9 / 24 (38%)	14 / 23 (61%)
Single	15 / 24 (62%)	9 / 23 (39%)
<b>Age</b>		
18-44	19 / 24 (79%)	16 / 23 (70%)
45-64	5 / 24 (21%)	7 / 23 (30%)
<b>Sleep duration (hours)</b>	7.54(0.57)	6.85(0.57)
<b>Time used to fall asleep (minutes)</b>	16(7)	24(9)
<b>Whether take sleep medications</b>	13 / 24 (54%)	17 / 23 (74%)
<b>Frequency of poor sleep</b>		
Rarely	7 / 24 (29%)	3 / 23 (13%)
Sometimes	12 / 24 (50%)	15 / 23 (65%)
Very often	5 / 24 (21%)	5 / 23 (22%)

From Table1, after removing the responses containing missing answers, there are 47 observations, 24 of them are from non-smokers, 23 of them are from smokers. In the sample, the proportion of woman who smoke is much higher than the proportion of man who smoke, the proportion of married participants who smoke is higher than single participants, the proportion of participants among 18-44 years old who smoke is much higher than the other age groups, but it also result from the lack of data from other age groups. Compared to non-smokers, the average sleep duration of smokers is shorter, the average time spent falling asleep is longer, and they tend to experience a higher frequency of poor sleep.

Table 2: Sleep duration between non-smokers and smokers

Smoking Status	Number of people	Minimum sleep hour	Maximum sleep hour	Average sleep hour	Variance of sleep duration
No	24	6.8	8.5	7.537500	0.3267935
Yes	23	5.5	8.0	6.847826	0.3198814

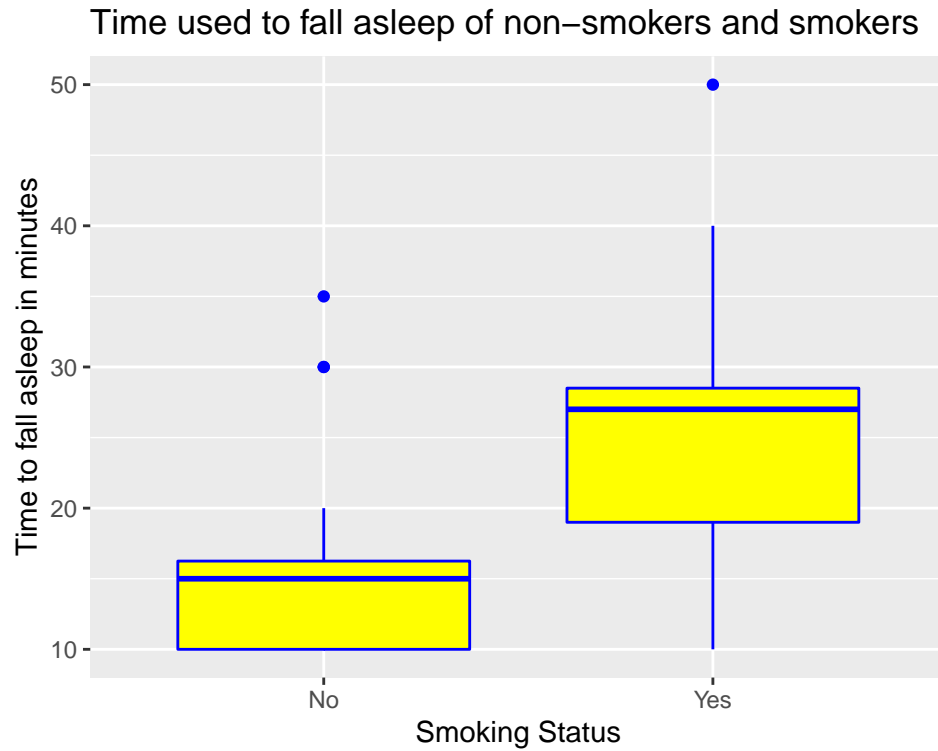


The numerical table for the sleep duration between non-smokers and smokers shows that in the sample, the average sleep hours of non-smokers is a little higher than smokers (7.54 hours vs. 6.85 hours). The sample variance of the two groups is approximately equal, which satisfies one of the assumptions of the two-sample t-test to be conducted later.

The density plot shows that the compared to smokers, the proportion of non-smokers who have more than 8 hours' sleep is higher. But the difference of the sleep duration between the two groups is not obvious, therefore, a hypothesis test would be applied to test whether the average sleep hours of the two groups are equal.

Table 3: Time used to fall asleep (in minutes) between non-smokers and smokers

Smoking Status	Number of people	Minimum Time	Maximum Time	Average Time	Variance of Time used to fall asleep
No	24	10	35	16.04167	47.7808
Yes	23	10	50	24.34783	84.6917



The numerical table for the time spent falling asleep between non-smokers and smokers shows that in the sample, smokers tend to use longer time to fall asleep on average.

The box-plot also presents that the range of time spent in the smoking group is much wider than non-smoking group, and some extreme values exist in the smoking group as we can see that there are some observations higher than 40 minutes.

All analysis for this report was programmed using R version 4.0.2.

## Methods

The hypothesis testing process evaluates the strength of evidence from the sample and provides a framework for making population-related determinations, it aims to determine whether the sample under study contains sufficient evidence to reject the null hypothesis (Davis & Mukamal, 2006). A hypothesis test is performed to determine if smokers and non-smokers have different sleep duration on average.

Using the two-sample t-test, a statistical test used to compare the means of two groups, we test the null hypothesis that the average sleep duration of smokers and non-smokers is the same. It assumes that the data are normal, independent and have equal variances (Kim & Park, 2019). Two-sample t-test assumptions are met since the observations of the data are independent, the sample size is large enough to approximate normality, and the variances of the two groups are approximately equal.

In this hypothesis test, the parameter of interest is the average sleep hours of the two groups (Smokers vs. Non-smokers). The null hypothesis is that there is no difference in the average of sleep duration between smokers and non-smokers, the alternative hypothesis is that the averages of sleep duration between smokers and non-smokers are different.

Assuming the null hypothesis is true, the p value, which is the probability of observing the test statistic, can be calculated. In this study, we set a significance level of 0.05, which is agreed upon by researchers in order to determine whether the null hypothesis is false (Tenny & Abdelgawad, 2021). If the output p-value is less than 0.05, the null hypothesis will be rejected.

Confidence interval is an interval estimate which specifies a range of values on either side of the sample statistic such as sample mean, within which the population parameter can be expected to fall with a chosen level of confidence (Sim & Reid, 1999). The level of confidence indicates how confident it is that the true parameter will fall within the interval range. In this study, we are interested in the means of time spent falling asleep by smokers and non-smokers. As each of the responses is independent in the data, and the sample size is greater than 30 to approximate normality, the assumptions of CI are met (Rosenblum & van der Laan, 2009).

The following formula will be used to calculate the 95% confidence intervals of the mean time spent falling asleep for non-smokers and smokers.

$$\text{Confidence interval} = \text{Sample mean} \pm \text{Critical value}(z) \times \text{Standard error of mean}$$

$$\text{Standard error of mean} = \frac{\text{Standard error of time spent falling asleep}}{\sqrt{\text{sample size}}}$$

The critical value or z value depends on the confidence level and is derived from the mathematics of the standard normal curve, for confidence level 95%, the z value is 1.96 (Rosenblum & van der Laan, 2009). Noted that the sample size for the non-smoking group is 24, for the smoking group is 23. The standard error of time spent falling asleep for the non-smoking group is 6.91 while the standard error of time spent falling asleep for the non-smoking group is 9.20.

## Results

Table 4: Hypothesis Test: the means of sleep duration between non-smokers and smokers

Sample mean of non-smokers	Sample mean of smokers	P value
16.04167	24.34783	0.0001428

Table 5: 95% CI of Time used to fall asleep (in minutes) between non-smokers and smokers

Lower Bound of Confidence interval	Higher Bound of Confidence interval	Smoking Status
11.2517297609606	20.8316035723727	Not smoke
17.9707217628313	30.7249304110818	Smoke

From the two-sample t-test, the p-value is 0.0001428, which is much smaller than the significance level 0.05, we can reject the null hypothesis that there is no difference in the average hours of sleep duration between smokers and non-smokers. The result of the hypothesis test shows that smoking is a significant factor which can influence the sleep duration.

According to the results of the confidence interval, we are 95% confident that the population mean of time spent falling asleep of non-smokers is between 11.25 and 20.83 minutes. We are 95% confident that the population mean of time spent falling asleep of smokers is between 17.97 and 32.72 minutes. It can be observed that on average, non-smokers tend to fall asleep faster than smokers. According to the hypothesis test and confidence interval, smoking may interfere with sleep duration and prolong the time taken to fall asleep, which may lead to sleep disturbances and disorders such as insomnia and depression (Alicia Nuñez et al., 2021). Because smoking may negatively impact sleep quality, it is recommended to prevent or lessen it in order to improve sleep quality.

## Limitations

There are some limitations in this study that need to be considered when interpreting the results. Since smoking status is the only variable considered as a predictor of sleep quality in this study, confounding variables such as exercise, alcohol, and anxiety level are not controlled, which may have an impact on the outcome (Arbinaga, 2020). Additionally, smoking time and frequency can also affect sleep quality, which needs to be further evaluated. Since the data from participants are measured only during the past one month, it is not representative of the real condition, as non-smokers may experience withdrawal effects during that month if they quit smoking (Cohrs et al., 2014). Thus, further studies are needed to examine the effects of smoking on sleep quality by taking into account its interactions with other variables and the time period.

## Bibliography

1. Alicia Nuñez, Rhee, J. U., Haynes, P., Chakravorty, S., Patterson, F., Killgore, W. D. S., Gallagher, R. A., Hale, L., Branas, C., Carrasco, N., Alfonso-Miller, P., Gehrels, J. A., & Grandner, M. A. (2021). Smoke at night and sleep worse? The associations between cigarette smoking with insomnia severity and sleep duration. *Sleep Health*, 7(2), 177–182. <https://doi.org/10.1016/J.SLEH.2020.10.006>
2. Allaire, J.J., et. el. References: Introduction to R Markdown. RStudio. <https://rmarkdown.rstudio.com/docs/>. (Last Accessed: May 5, 2021)
3. Arbinaga, F. (2020). The relationship between subjective sleep quality and smoking in university students. *Sleep and Hypnosis*, 21(4). <https://doi.org/10.37133/Sleep.Hypn.2019.21.0200>

4. Breedlove, J. (2021). Easy tips to help you fall asleep. UWIRE Text, 1–1. <https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=&v=2.1&it=r&id=GALE%7CA655079211&sid=googleScholar&linkaccess=fulltext>
5. Cheung, K. L., ten Klooster, P. M., Smit, C., de Vries, H., & Pieterse, M. E. (2017). The impact of non-response bias due to sampling in public health studies: A comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. *BMC Public Health*, 17(1). <https://doi.org/10.1186/S12889-017-4189-8>
6. Cohrs, S., Rodenbeck, A., Riemann, D., Szagun, B., Jaehne, A., Brinkmeyer, J., Gründer, G., Wienker, T., Diaz-Lacava, A., Mobascher, A., Dahmen, N., Thuerauf, N., Kornhuber, J., Kiefer, F., Gallinat, J., Wagner, M., Kunz, D., Grittner, U., & Winterer, G. (2014). Impaired sleep quality and sleep duration in smokers—results from the German Multicenter Study on Nicotine Dependence. *Addiction Biology*, 19(3), 486–496. <https://doi.org/10.1111/J.1369-1600.2012.00487.X>
7. Davis, R. B., & Mukamal, K. J. (2006). Hypothesis testing: Means. *Circulation*, 114(10), 1078–1082. <https://doi.org/10.1161/CIRCULATIONAHA.105.586461>
8. Dugas, E. N., Sylvestre, M. P., O’Loughlin, E. K., Brunet, J., Kakinami, L., Constantin, E., & O’Loughlin, J. (2017). Nicotine dependence and sleep quality in young adults. *Addictive Behaviors*, 65, 154–160. <https://doi.org/10.1016/J.ADDBEH.2016.10.020>
9. Grolemond, G. (2014, July 16) *Introduction to R Markdown*. RStudio. [https://rmarkdown.rstudio.com/articles\\_intro.html](https://rmarkdown.rstudio.com/articles_intro.html). (Last Accessed: May 5, 2021)
10. Jaehne, A., Unbehauen, T., Feige, B., Lutz, U. C., Batra, A., & Riemann, D. (2012). How smoking affects sleep: A polysomnographical analysis. *Sleep Medicine*, 13(10), 1286–1292. <https://doi.org/10.1016/J.SLEEP.2012.06.026>
11. Kim, T. K., & Park, J. H. (2019). More about the basic assumptions of t-test: normality and sample size. *Korean Journal of Anesthesiology*, 72(4), 331–335. <https://doi.org/10.4097/KJA.D.18.00292>
12. Patterson, F., Grandner, M. A., Malone, S. K., Rizzo, A., Davey, A., & Edwards, D. G. (2019). Sleep as a Target for Optimized Response to Smoking Cessation Treatment. *Nicotine & Tobacco Research*, 21(2), 139–148. <https://doi.org/10.1093/NTR/NTX236>
13. Rosenblum, M. A., & van der Laan, M. J. (2009). Confidence Intervals for the Population Mean Tailored to Small Sample Sizes, with Applications to Survey Sampling. *The International Journal of Biostatistics*, 5(1). <https://doi.org/10.2202/1557-4679.1118>
14. Sim, J., & Reid, N. (1999). Statistical Inference by Confidence Intervals: Issues of Interpretation and Utilization. *Physical Therapy*, 79(2), 186–195. <https://doi.org/10.1093/PTJ/79.2.186>
15. Sjöberg, D. D., Whiting, K., Curry, M., Lavery, J. A., & Larmarange, J. (2021). Reproducible Summary Tables with the gtsummary Package. *R Journal*, 13(1), 570–580. <https://doi.org/10.32614/RJ-2021-053>
16. Tenny, S., & Abdelgawad, I. (2021). Statistical Significance. *Annual Review of Statistics and Its Application*, 7, 1–10. <https://doi.org/10.1146/annurev-statistics-031219-041051>
17. Zhang, X., Wen, D., Liang, J., & Lei, J. (2017). How the public uses social media wechat to obtain health information in china: a survey study. *BMC Medical Informatics and Decision Making*, 17(Suppl 2). <https://doi.org/10.1186/S12911-017-0470-0>



## Appendix

Here is a glimpse of the data set simulated/surveyed:

```
## Rows: 57
## Columns: 8
## $ `Gender identify`
## $ `Marital status`
## $ `Choose your age group`
## $ `How many hours of sleep do you typically obtain at night in the last month? (Please only enter nu
## $ `Have you ever smoked cigarettes or other tobacco in the last month? (including e-cigarettes)`
## $ `How long does it take you to fall asleep in the last month on average? (Please only enter numbers
## $ `Do you need sleep medications to help you fall asleep? (During the past one month)`
## $ `How often has poor sleep troubled you in the last month?`
```