Technical Report
Group 20

1. Introduction:

As university students, most of our main goals are to succeed academically. However, before we are students, we are also humans; we thrive for happiness. The goal in this analysis is to find out whether there exists a relationship between time spent on hobbies and time spent studying that contributes to students' happiness level. We also understand that each and every students' self-measured happiness levels depend on what priorities they have.

1.1 Objectives:

The result(s) from this study can help students efficiently allocate time to studying and hobbies to achieve a higher level of happiness.

There are two main objectives in this study:

1. Find out whether a relationship between time spent on hobbies and happiness level exists.
2. Find out whether a relationship between time spent on studies and happiness level exists.

As stated in the introduction, assuming that every participant has different priorities that lead to higher levels of self-measured happiness, we also plan to explore the proportions of students who believe spending time on hobbies or studies contribute to higher levels of happiness.

1.2 Hypotheses:

We asked students to rate their Happiness on a discrete scale from 1 to 10, 10 being very happy. As a result, we are unable to use linear regression, since linear regression requires the data to be continuous.

Under the assumption that participants' happiness is dependent on time spent on hobbies and time spent on studies, we plan to use methods for ordinal data since the data for happiness is ordinal. To achieve the first objective: finding out whether a relationship between time spent on hobbies and happiness level exists, we plan to investigate the association between the two variables using Kendall tau rank correlation, because our dependent variable: Happiness is ordinal with many ties.

Letting τ_1= Kendall tau rank correlation coefficient for time_hobby and Happiness, our hypothesis is:

$$H_o: \tau_1 = 0 \ vs \ H_a: \tau_1 \neq 0.$$

Our null hypothesis is that there exists no relationship between the time spent on hobbies and happiness, and the alternate hypothesis is that there exists a relationship between the two.

To achieve our second objective, let τ_2 =Kendall tau rank correlation coefficient for time_studies and Happiness:

$$H_o: \tau_2 = 0 \ vs \ H_a: \tau_2 \neq 0.$$

Our null hypothesis is that there exists no relationship between the time spent on studies and happiness, and the alternate hypothesis is that there exists a relationship between the two.

Lastly, to see if participants truly have different views on how to increase their happiness, we investigate the proportion of students who think that allocating more time to hobbies contributes more to their happiness. From the data we collected, contribute_happiness was a binary question that returned 1: the time they spend on

hobbies contributes more to their happiness and 2: the time they spend on studies contributes more to their happiness.

Letting $p = \frac{the\ number\ of\ entry\ "1"\ in\ contribute\_happiness}{n}$,

$H_o: p = 0.5\ vs\ H_a: p \neq 0.5$

Our null hypothesis is that the proportion of students who think that spending time on hobbies contributes more to their happiness is $\frac{1}{2} = 0.5$, and our alternate hypothesis is that it is not 0.5. It could either be greater than or less than 0.5.

2. Methodology:

From October 2023 to early November 2023, our survey to analyze a relationship between time spent on hobbies and studies to the happiness level of 2023 STA304H5F students was uploaded on the course piazza. Our survey was conducted anonymously by random students in the course, therefore using SRS to conduct our survey. To ensure randomness and anonymity while avoiding duplicate data, we gathered the participants' nicknames. During that time, we received 29 valid responses from random participants, and 1 invalid response.

3. Sample Size

For our study, we are looking at 2023 STA304H5F students as our population with population size N= 200.

To estimate $p = \frac{the\ number\ of\ entry\ "1"\ in\ contribute\_happiness}{n}$, we need to determine the sample size n with a bound of error B = 0.175. From the 3rd hypothesis, we are unsure of the proportion of students who think the time they spend on hobbies contributes more to happiness, so assume p = 0.5. Note that our data collection method was SRS, so the sample size formula is:

$$n = \frac{Npq}{(N-1)D + pq} = \frac{200 * 0.5 * 0.5}{(200-1)*\frac{(0.175)^2}{4} + (0.5*0.5)} = 28.19134878 \approx 29.$$

We have collected 30 responses: 15 from LEC0101, 15 from LEC0102 but a participant from LEC0101 had physically impossible responses(i.e. Spending 34 hours a day on hobbies and 53 hours studying), we removed the response from all

calculations. Hence, we have 29 responses in total: 14 from LEC0101, 15 from LEC0102, which is sufficient.

4. Data Exploration

4.1 Estimators

Our point estimator $\hat{p}$, for the proportion of students who think time spent on hobbies contributes more to their happiness is calculated as follows:

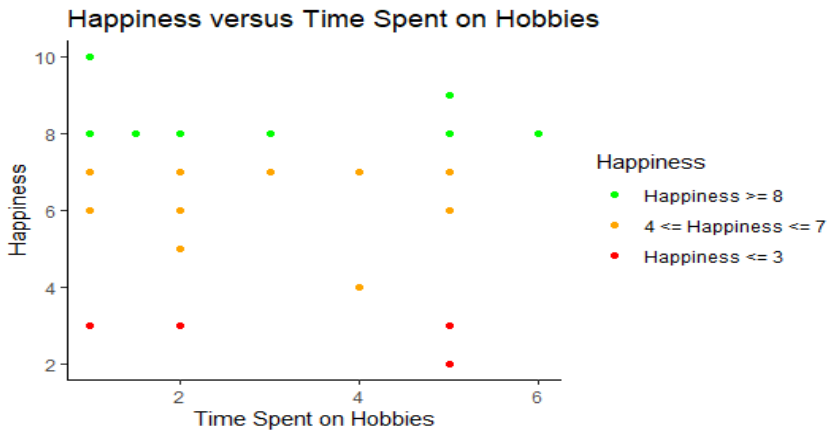
$$\hat{p} = \frac{the\ number\ of\ entry\ "1"\ in\ contribute\_happiness}{n} = \frac{25}{29} = 0.86201$$

Here is a table containing the means of some of the other columns in our data:

Variable	**Mean**
Happiness	6.689655
Time Spent on Hobbies per day	2.810345
Time Spent on Studies per day	4.672414

4.2 Visualizations

Let us visualize the relationship between Happiness and the Time Spent on Hobbies (per day).



Visually, there does not seem to be any correlation. There are students who spend the same amount of time on their hobbies, yet their happiness levels completely differ.

Next, let us visualize the relationship between Happiness and the Time Spent on Studies (per day).

5. Analysis

5.1 Assumptions

In order to use Kendall's tau, we need to see if a few assumptions for our data are satisfied.

1. Random Sampling
2. Existence of monotonic association
3. Variables are at least ordinal
4. Paired Samples
5. Independence of observations

We used SRS to collect our data, so 1 is met. We assume a monotonic relationship exists between time_hobby and Happiness, time_studies and Happiness. We are using the correlation test to see if it truly exists, so 2 is also met. Our dependent variable: Happiness is an ordinal data, and our independent variables are also not nominal which satisfy 3. All of our observations are paired, and independent of each other, so 4 and 5 are also satisfied.

For our third hypothesis, we will use the binomial test to find out the true proportion of participants who think that spending time on hobbies would increase their happiness.

There are 4 assumptions that need to be checked when using the binomial test.

1. Binary
2. Independence of observations
3. Fixed sample size *n*
4. Same probability

Our variable *contribute_happiness* is essentially a binary data that returns *1* if the participant agrees that spending time on hobbies contributes more to their level of happiness, *2* if the participant disagrees. Therefore, it satisfies the first assumption. Also, we utilized SRS when collecting this data, so all the participants' responses were independent of each other and n =30 was fixed. We assumed p = 0.5 since all participants have different opinions.

5.2 Computations

Using R, we can easily compute Kendall's tau correlation to see if there exists an association between two ordinal variables.

For our first hypothesis:

$H_o: \tau_1 = 0 \ vs \ H_a: \tau_1 \neq 0.$

We initiate a correlation test between time_hobby and Happiness to find out whether a correlation between them exists with 95% confidence. Upon executing the correlation test, our z = -0.67888, *p-value* = 0.4972 and

τ_1 = -0.1025904. Since our p-value > 0.05, we have no evidence to reject the null hypothesis, so we conclude that no correlation exists between time spent on hobbies and happiness with 95% confidence.

For our second hypothesis:

$H_o: \tau_2 = 0 \ vs \ H_a: \tau_2 \neq 0.$

We initiate a correlation test between time_studies and Happiness to find out whether a correlation between them exists with 95% confidence. Upon executing the correlation test, our z = -0.57363, *p-value* = 0.5662 and

τ_2 = -0.08491301 . Since our p-value > 0.05, we have no evidence to reject the null hypothesis, so we conclude that no correlation exists between time spent on studies and happiness with 95% confidence.

For our third hypothesis:

$H_o : p = 0.5 \; vs \; H_a : p \neq 0.$

During our data exploration, we found that 25 out of 29 participants think that spending time on hobbies contributes more to their happiness. Using R, we can execute proportion tests to see if the true p=0.5 or not with 95% confidence. Binomial testing in R, *p-value* = 0.0001037<0.05, and our 95% confidence interval for the true proportion is: (0.6833594 0.9611052). Our hypothesized proportion, p = 0.5 does not lie in the confidence interval, so we have sufficient evidence to reject the null hypothesis and claim that the true proportion is not 0.5.

6. Conclusion & Limitations

In conclusion, our main goal of finding a correlation between time spent on hobbies, time spent on studies and happiness was unsuccessful. From the collected data, there is no evidence to suggest that relationships exist between happiness and time spent on either hobbies or studies.

From our last hypothesis we found out that the true proportion of students who think spending time on their hobbies contributes more to their happiness is not 0.5, but a rather interestingly high number.

This means that the majority of the students believe that the time they spend on hobbies contributes more to their happiness. Interestingly our other results show that the time a student spends on hobbies does not correlate with happiness.

Upon conducting this research, the result of no correlation between time spent on either hobbies and studies to happiness made us question if our hypothesis was wrong or if our method of research was wrong; we believed that there were a few limitations to our research that returned the above results:

1. Low number of samples

We were only able to receive 30 responses from 2023 STA304H5F students. If we had more samples, we believe that the result could be different.

2. When asking students whether the time they spend on hobbies or the time they spend on studies contributes more to happiness, we didn't give them an option to say both equally contribute.
3. Existence of "Ties" when using Kendall's tau rank coefficient

Having ties in our observations can impact the precision and validity of the tests.

7. R Code

Starts on next page.

Importing Dataset:

```
library("readxl")
```

```
## Warning: package 'readxl' was built under R version 4.1.3
```

```
library(ggplot2)
```

```
raw_data <- read_excel("C:/Users/dhire/Downloads/STA304 group 20 dataset.xlsx")
```

Cleaning Data and for now removing an entry that doesn't make sense

```
data <- raw_data[-c(31:33),]
data <- subset(data, Nickname != "crest")
```

```
happiness <- data$Happiness
hobbytime <- data$time_hobby
studytime <- data$time_studies
```

Data Exploration

```
table(data$contribute_happiness)
```

```
##
##  1  2
## 25  4
```

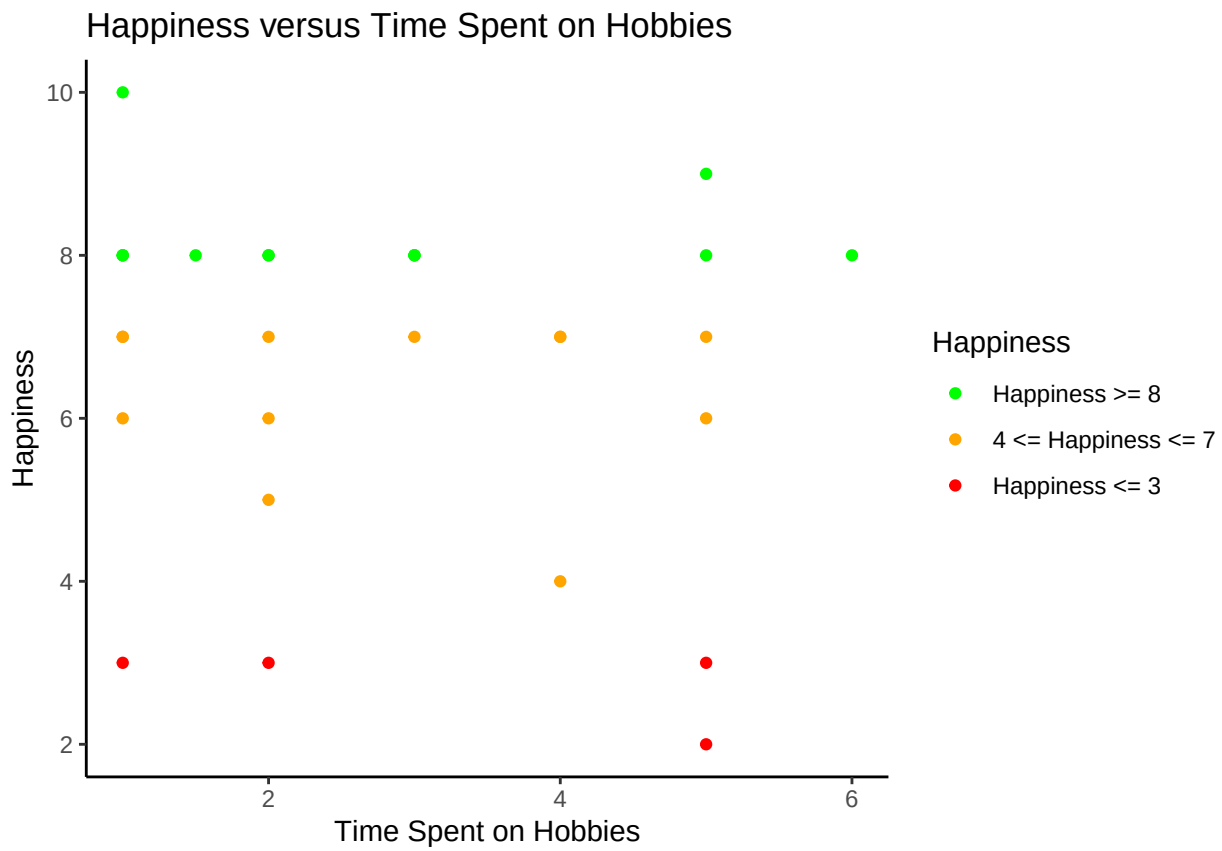
```
colMeans(data[, 4:6])
```

```
##   time_hobby time_studies    Happiness
##     2.810345     4.672414     6.689655
```

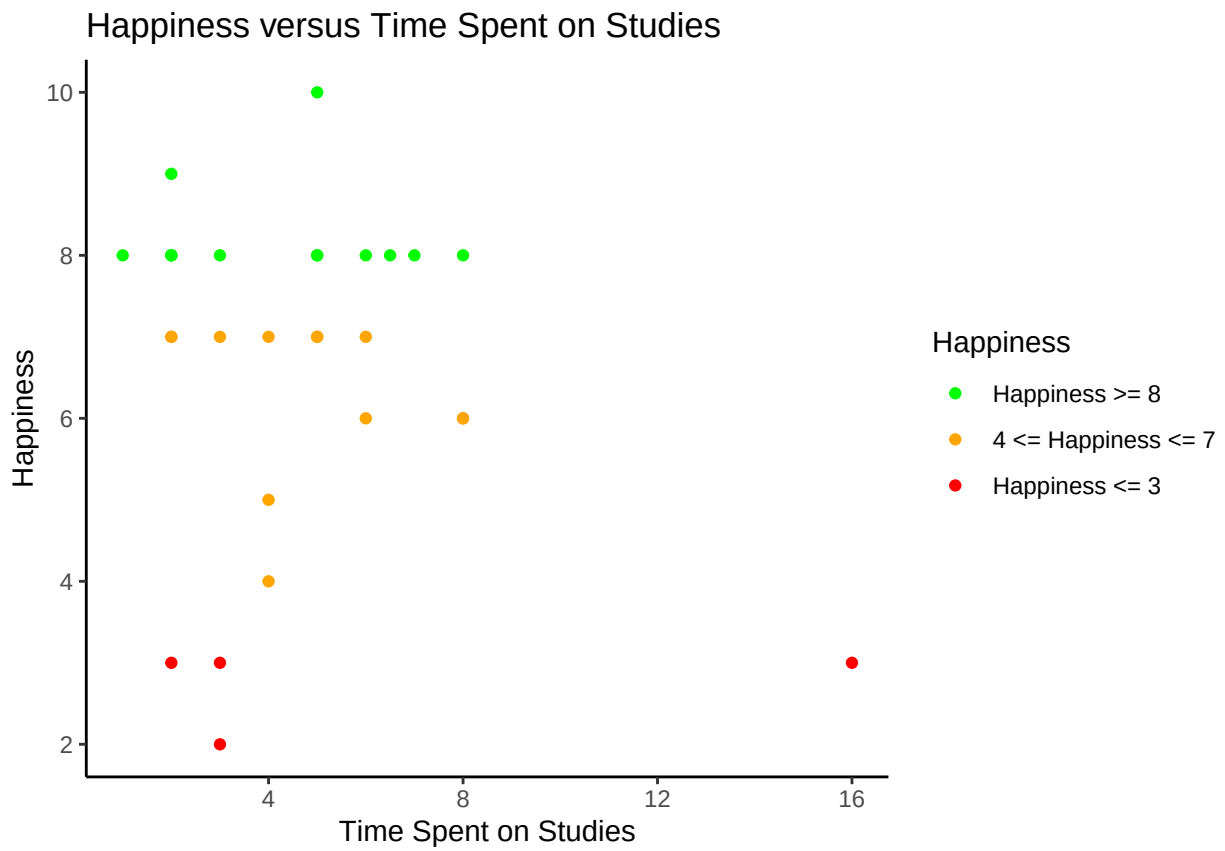
```
colors_happiness = happiness
colors_happiness[happiness >= 8] <- "green"
colors_happiness[happiness <= 7 & happiness >= 4] <- "orange"
colors_happiness[happiness < 4] <- "red"
colors_happiness
```

```
##  [1] "orange" "green"  "orange" "orange" "green"  "green"  "green"  "orange"
##  [9] "orange" "green"  "orange" "green"  "orange" "green"  "green"  "green"
## [17] "orange" "green"  "orange" "orange" "red"    "red"    "orange" "green"
## [25] "red"    "green"  "red"    "orange" "green"
```

```
ggplot(mapping = aes(x=hobbytime, y=happiness, color = colors_happiness)) +
  geom_point() +
  scale_color_manual(values = c("green" = "green", "orange" = "orange", "red" = "red"), labels=c("Happi
  theme_classic() + labs(title = "Happiness versus Time Spent on Hobbies", x = "Time Spent on Hobbies",
```




```
ggplot(mapping = aes(x=studytime, y=happiness, color = colors_happiness)) +
  geom_point() +
  scale_color_manual(values = c("green" = "green", "orange" = "orange", "red" = "red"), labels=c("Happi
  theme_classic() + labs(title = "Happiness versus Time Spent on Studies", x = "Time Spent on Studies",
```




```
cor.test(studytime, happiness, method = "kendall", exact = FALSE)
```

```
##
##  Kendall's rank correlation tau
##
## data:  studytime and happiness
## z = -0.57363, p-value = 0.5662
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##         tau
## -0.08491301
```

```
cor.test(hobbytime, happiness, method = "kendall", exact = FALSE)
```

```
##
##  Kendall's rank correlation tau
##
## data:  hobbytime and happiness
## z = -0.67888, p-value = 0.4972
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##        tau
## -0.1025904
```

```
binom.test(25, 29, p = 0.5, alternative = "two.sided")
```

```
##
##  Exact binomial test
```

```
##
## data:  25 and 29
## number of successes = 25, number of trials = 29, p-value = 0.0001037
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.6833594 0.9611052
## sample estimates:
## probability of success
##               0.862069
```