Are Colby students who exercise regularly happier than those who don't?

Nico Hillison

Introduction

An observational study published in the International Journal of Environmental Research and Public Health found that of 2,345 healthy adults, those who exercised more rated themselves higher in life satisfaction and general happiness. It also described a baseline correlation between age and happiness. The study found that individuals with moderate ($p<0.001$) and high ($p<0.001$) activity levels were significantly happier than those with low levels. More physical activity was significantly related to happiness and life satisfaction for all demographic groups (An).

Another study by the Association for Psychological Science examined the relationship between the use of time, happiness, and success and incorporated exercise as one of several contributors to happiness and success level. (Mogilner). It was concluded that as the level of physical activity can indicate personal physical satisfaction, those individuals who exercised more frequently considered themselves to be more successful. Believing in one's success then resulted in increased happiness. It is important to note that this study involved middle-aged men and women, while our project focused on a younger group: students attending Colby College.

Happiness has benefits extending beyond surface-level effects, and it is commonly reported as a method of ranking world nations. However, the reasons behind this may not immediately be apparent. An analysis published in the Psychological Bulletin of 102 happiness studies reported that positive moods lead to increased creativity as measured through cognitive functions such as fluency, originality, flexibility, and insight (Baas). Creativity and cognitive function are essential tools for people in general and Colby students specifically. Therefore, it is important to investigate what factors might lead to an increase in mood.

Methods

We collected data from 48 students through a survey by selecting individuals randomly from class lists. The main variables of the dataset are the explanatory variable self-described activity level on a scale from Active to Not Active and the response variable happiness levels on a scale of 1 to 10. A secondary explanatory variable is hours spent per week exercising, which may be a more accurate measure of exercise than self-categorization. Two potential confounders were included: daily hours spent sleeping and hours spent studying. The general percentage of time participants are in a good mood is the final response variable. The survey was purposefully short and straightforward to entice individuals to participate.

Results

The most important variables we gathered in the survey were the Weekly Hours Spent Exercising (independent variable) and the Percentage of Time Respondents Spent in a Good Mood (dependent variable). Our data showed that individuals who exercised between 15 and 19 hours a week were in a good mood the highest percent of the time ($\overline{x}$ = 85.0), while those who exercised between zero and four hours were happy the least amount of the time ($\overline{x}$ = 59.35) (Table 1). The standard deviations are all relatively similar.

General Results
Time Spent in a Good Mood (Percentage)

Range of Hours Spent Exercising Weekly	# of Observations	Mean	Maximum	Minimum	Std. Dev
0 - 4	17	59.35	80	25	22.31
5 - 9	22	71.59	90	15	19.2
10 - 14	5	66.0	85	20	25.32
15 - 19	3	85.0	90	80	13.87

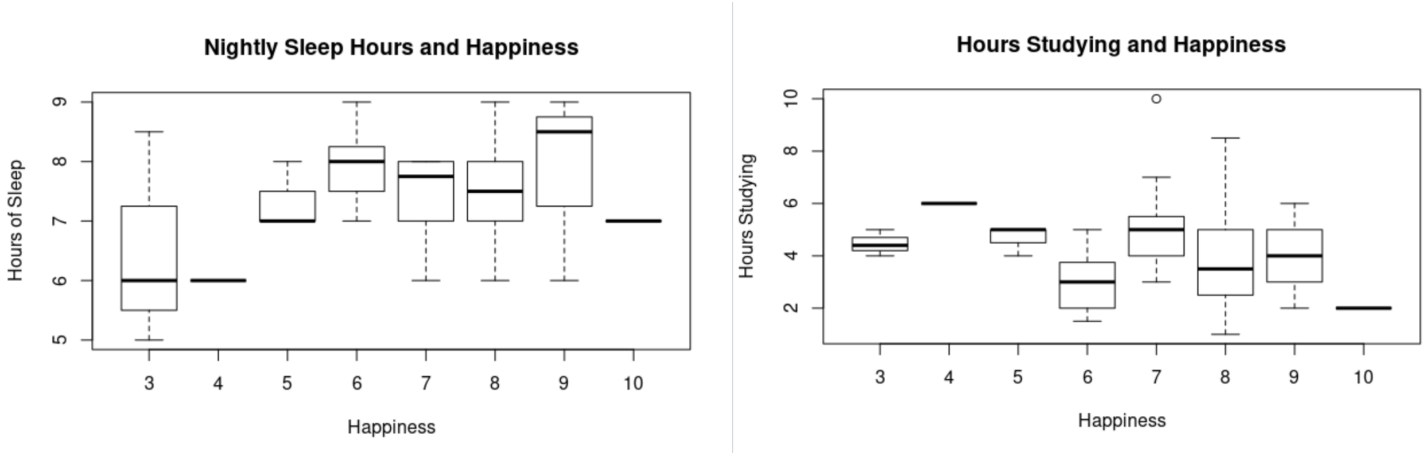
Table 1. Summary of descriptive statistics for weekly exercise hours and time in a good mood.

Based on the mean results, the amount of time spent in a good mood appears to increase with the number of hours of weekly exercise. However, the mean outcome does not reflect this trend in the 10 - 14 hour range. This result appears to be heavily influenced by a low outlier value of 20. The other values given from respondents in this range were: 75, 80, 85, and 70. Due to only five observations in this range, the value of 20 heavily affects this mean. If the value of 20 were removed, the mean of the 10-14 range would be 77.5 and be consistent with the trend of a positive relationship between hours spent exercising and the percentage of time in a good mood. This outlier also affects the standard deviation. These values appear to decrease as time exercising increases, but the 10-14 range, with the inclusion of the outlier, interrupts this trend. Broadly assessed, the data results of the two main variables indicate a positive relationship between the two.

We performed a linear regression to explore whether there is a correlation between 'Hours of Exercise' and 'Time in a Good Mood .' we performed a linear regression. Since LINES conditions were not met, we used a randomization test for correlation. The test produced a p-value significantly less than a 5% significance level (p-value = 0.0090), meaning there is evidence for a non-zero correlation. We found a weak positive correlation of r = 0.344.

Figure 1. Correlation of exercise hours and time in a good mood.

We also performed an analysis of variance (ANOVA) test to determine whether there is an association between sleep or study hours and happiness level (Figures 2 & 3). In the sleep test, conditions for ANOVA for means were not met, so we used randomization for ANOVA. The test revealed no significant evidence that there is variability in mean sleep hours between happiness levels (p-value = 0.189). For the study hours test, the conditions were again not met, and there was no evidence of a difference in mean hours between levels (p-value = 0.176).



Figures 2 & 3. The distribution of hours of sleep and studying for each level of happiness.

Finally, we investigated the association between activity level and happiness (Figure 3). We used a randomization test since the conditions for Chi-square were not met. The results of the test showed no significant evidence of association (p=0.073).

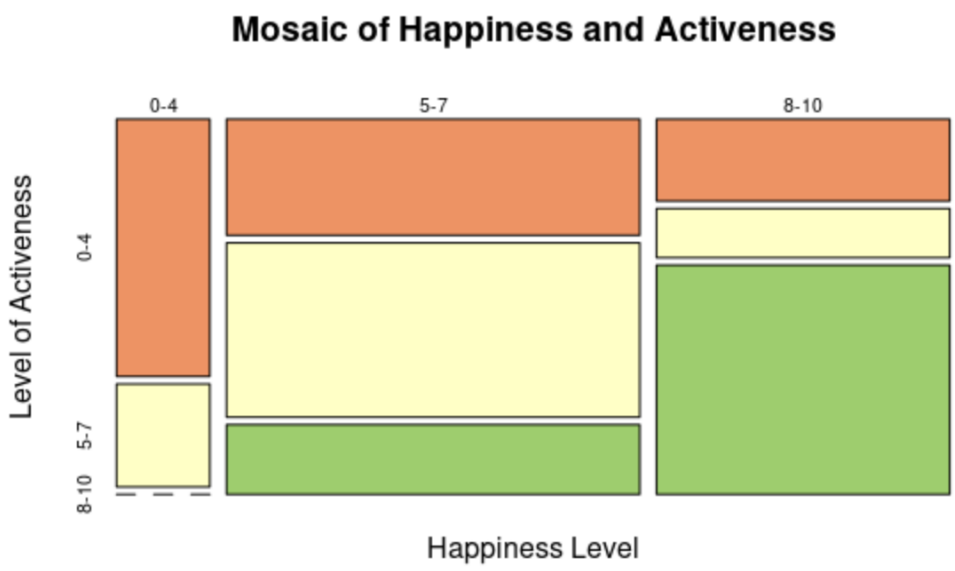


Figure 4. Correlation between reported activity level and happiness.

Discussion/Conclusion

We found that there is a positive correlation between hours spent exercising and the percent of time individuals are in a good mood. Though the correlation is not very strong, it is still significant. The correlation may be slightly different in the population, but we still believe it will be positive. Through ANOVA, we concluded that neither hours of sleep nor hours studying have an association with happiness levels. Finally, we found that reported activity level is not associated with happiness level. We conclude that the more one spends exercising, the more they will be in a happy mood. Future analysis could include an exploration of the role of sleep and study hours as confounding variables through multiple linear regression.

Bibliography

An, Hsin-Yu et al. "The Relationships between Physical Activity and Life Satisfaction and Happiness among Young, Middle-Aged, and Older Adults." International Journal of Environmental Research and Public Health 17.13 (July 2020): 4817. 14 October 2021 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7369812/

Baas M, De Dreu CK, Nijstad BA. "A meta-analysis of 25 years of mood-creativity research: hedonic tone, activation, or regulatory focus?" Psychological Bulletin. 134.6 (Nov. 2008): 779-806. 14 October 2021 https://pubmed.ncbi.nlm.nih.gov/18954157/

Mogilner, Cassie. "The Pursuit of Happiness: Time, Money, and Social Connection." Psychological Science 21.9 (2010): 1348–54. 14 October 2021 http://www.jstor.org/stable/41062376.

Appendix

Test for correlation between exercise hours and time in a good mood:

1. Parameters: x = daily hours of exercise, y = % of time being in a good mood
2. Hypothesis:
    i. Ho : $\rho = 0$
    ii. Ha : $\rho \neq 0$

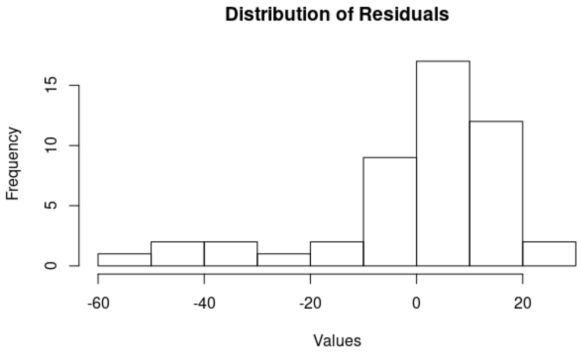
2.1. Check conditions: LINES

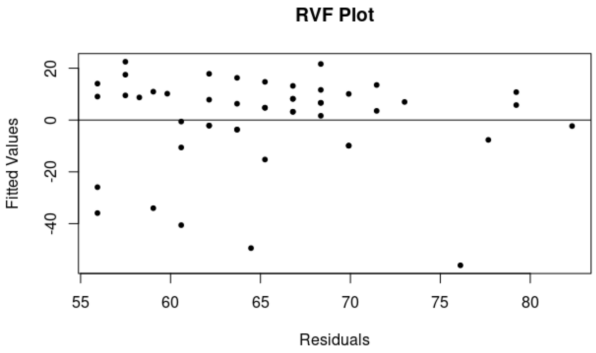
- ➔ Linearity: test can be performed if the relationship is approximately linear. A linear relationship is the best approximation for the two variables (Figure 1).
- ➔ Independence: The y values in our data are independent of each other.

➔ Normality: The residuals are not normally distributed, as the distribution is strongly skewed left, so LINES fails (see distribution of residuals).

➔ Equal Variance: The residuals of the linear regression are not fanned or curved, but also not evenly distributed above and below the line (rvf plot).

➔ S (Randomization): The data was collected using a SRS.



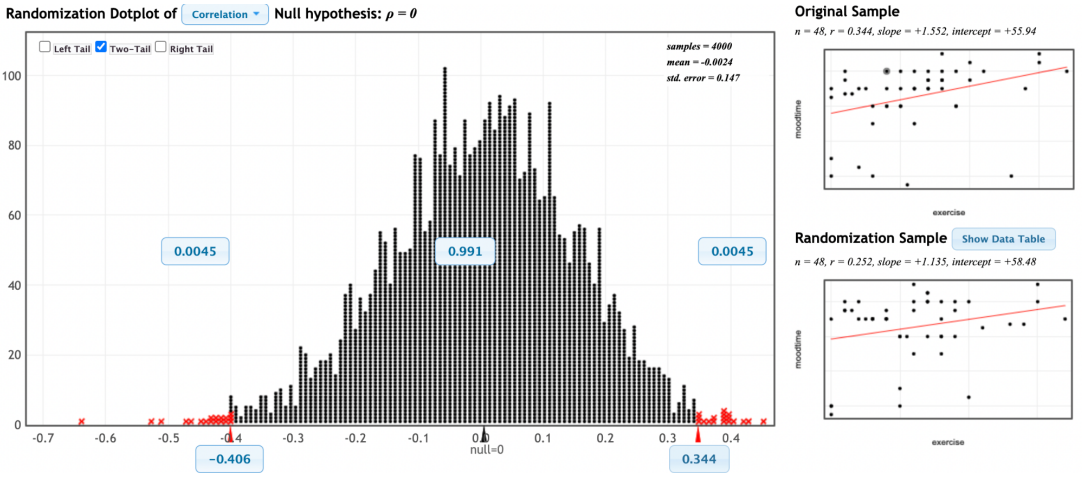


The LINES rules are not met, so we cannot use a t test and must use a randomization.

Randomization test:

2.2

r = 0.344



3. p-value = .0045*2 = 0.0090
4. We reject the null hypothesis at the 5% significance level. p < .05
5. There is evidence of a non-zero correlation between exercise hours and time in a good mood.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   55.936      4.608  12.138 6.06e-16 ***
E$exercise     1.552      0.624   2.488   0.0165 *
---
```

From the RStudio output, we get a regression formula where GMT represents the percent of time an individual is in a good mood and EX the hours spent exercising daily:

$$\widehat{GMT} = 55.936 + 1.552 * EX$$

ANOVA test for level of happiness and amount of sleep:

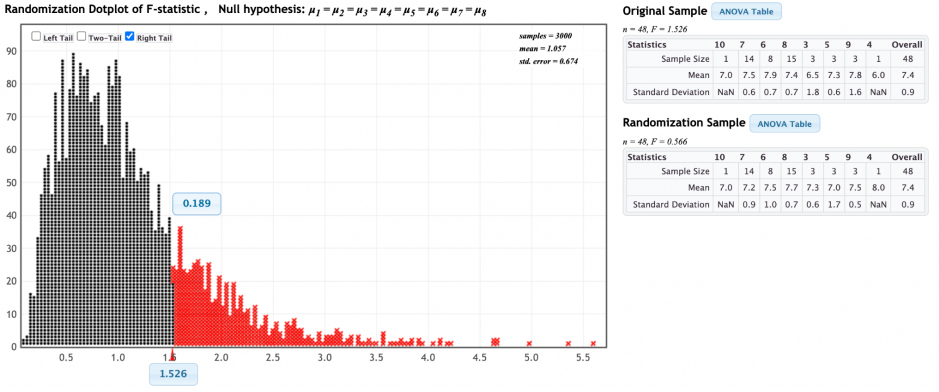
1. Parameters: $\mu_3 \ldots \mu_{10}$ = mean sleep hours for happiness levels, which range from 3 to 10 in the data.
2. Hypothesis:
   a. H0 : $\mu_3 = \mu_4 \ldots = \mu_{10}$
   b. Ha : some μ_i != μ_j

2.1 Check Conditions:

- ➢ SRS (Simple Random Sample)
- ➢ Groups are independent
- ➢ Categories do not have more than 30 samples
- ➢ Largest standard deviation is more than 2 times the smallest (1.8/2 = 0.9 > 0.58).

Since the conditions are not met, an ANOVA is not applicable. We must do a randomization test of ANOVA for difference in means.

2.2 F = 1.526



Original Sample ANOVA Table

$n = 48, F = 1.526$

Statistics	10	7	6	8	3	5	9	4	Overall
Sample Size	1	14	8	15	3	3	3	1	48
Mean	7.0	7.5	7.9	7.4	6.5	7.3	7.8	6.0	7.4
Standard Deviation	NaN	0.6	0.7	0.7	1.8	0.6	1.6	NaN	0.9

Randomization Sample ANOVA Table

$n = 48, F = 0.566$

Statistics	10	7	6	8	3	5	9	4	Overall
Sample Size	1	14	8	15	3	3	3	1	48
Mean	7.0	7.2	7.5	7.7	7.3	7.0	7.5	8.0	7.4
Standard Deviation	NaN	0.9	1.0	0.7	0.6	1.7	0.5	NaN	0.9

3. p-value = 0.189
4. We cannot reject the null hypothesis.
5. There is no evidence of a difference in mean hours of sleep between different reported levels of happiness.

ANOVA test for level of happiness and hours of studying:

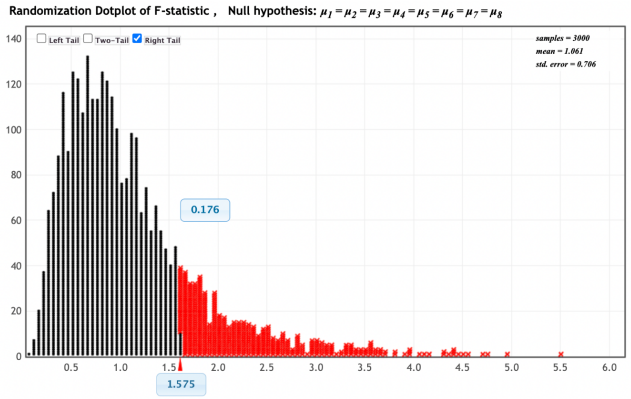
1. Parameters: $\mu_3 \ldots \mu_{10}$ = mean study hours for happiness levels, which range from 3 to 10 in the data.
2. Hypothesis:
    c. H0 : $\mu_3 = \mu_4 \ldots = \mu_{10}$
    d. Ha : some μ_i != μ_j

2.1 Check Conditions:

- ➢ SRS (Simple Random Sample)
- ➢ Groups are independent
- ➢ Categories do not have more than 30 samples
- ➢ Largest standard deviation is more than 2 times the smallest (2.1/2 = 1.05 > 0.5).

Since the conditions are not met, an ANOVA is not applicable. We must do a randomization test of ANOVA for difference in means.

2.2 F = 1.575



Original Sample ANOVA Table

$n = 48, F = 1.575$

Statistics	10	7	6	8	3	5	9	4	Overall
Sample Size	1	14	8	15	3	3	3	1	48
Mean	2.0	5.1	3.0	3.9	4.5	4.7	4.0	6.0	4.2
Standard Deviation	NaN	1.8	1.2	2.1	0.5	0.6	2.0	NaN	1.8

Randomization Sample ANOVA Table

$n = 48, F = 3.136$

Statistics	10	7	6	8	3	5	9	4	Overall
Sample Size	1	14	8	15	3	3	3	1	48
Mean	4.0	3.8	3.3	4.7	4.5	2.8	5.0	10.0	4.2
Standard Deviation	NaN	1.5	1.2	1.8	2.3	1.6	0.0	NaN	1.8

3. p-value = 0.176
4. We cannot reject the null hypothesis.
5. There is no evidence of difference in mean hours studying between happiness levels.

Chi-square test for activity level and happiness:

1. Parameters: activity level (1-10) and happiness level (1-10)
2. Ho: There is no association between activity and happiness.
   Ha: There is an association.

2.1 Check conditions: some of the expected counts are not greater than 5, so we must use a randomization Chi-square test

2.2 $\chi^2 = 77.496$

3. p-value = 0.073
4. At a 5% level, we cannot reject the null hypothesis.
5. There is not enough evidence that activity level and happiness are associated.