The Influence of Internet Habits on Life Satisfaction: An Analysis Among Youth

Eric Geisler

John Carroll University

May 09, 2024

## Introduction

With the ever-growing development of digital media that has characterized the modern world for the last few decades, an unavoidable question has inevitably developed: how much is too much? There is no doubt—the internet has connected people on a level never seen before in human history. However, has this ceaseless connection gone too far? The constant societal pressure placed on individuals in contemporary life is immense. People are now incessantly comparing themselves to other people, as the highlights of everyone's lives are posted everywhere. Humanity centers around a screen.

To analyze this issue among youth, a Cochran–Mantel–Haenszel (CMH) Test was conducted examining the association between self-reported internet usage levels and overall happiness with life. In this test, a significant linear association between the variables was discovered. Significant results were also found when partitioning the tables to compare individual levels of internet usage.

All group members contributed equally to both the analysis and presentation.

## Problem

What is the relationship between self-reported internet usage habits and happiness with youth? The general issue of the increasing presence of digital life on individuals' well-being will be analyzed through this research question in this work. Ultimately, the trend that this research is trying to uncover is a lowering of overall happiness based on more and more internet usage. At least, that is the working hypothesis before conducting any exploratory analysis.

## Data

data is linked here: https://www.kaggle.com/datasets/miroslavsabo/young-people-survey

The data with which this research will be carried out was gathered from a sample of students in the United Kingdom, ranging from ages fifteen to thirty. The original copy of the data contained 1010 rows. Each participant was asked a long series of 150 questions in either written or electronic form. The breakdown of questions was as follows: nineteen questions about music preferences, twelve questions about movie preferences, thirty-two questions about hobbies and interests, ten questions about phobias, three questions about health habits, fifty-seven questions about personality traits, views on life, and general opinions, seven questions about spending habits, and ten questions about demographic information. Evidently, this was a comprehensive study. Specifically, this research will pull the questions regarding self-reported internet usage and happiness with life. Regarding internet usage, each participant was asked, "How much time do you spend online?" Responses were categorized into four buckets: *No time at all*, *Less than an*

*hour a day*, *Few hours a day*, and *Most of the day*. These responses were stored in a column named *Internet usage*. Regarding happiness with life, each participant was asked to respond to the following statement, "I am 100% happy with my life." Responses were recorded on a five-point scale: *strongly disagree, disagree, neither agree nor disagree, agree,* and *strongly agree*. These responses were stored in a column named *Happiness in life*.

## Data Cleaning and Preprocessing

Before any data cleaning or feature engineering in regards to conducting statistical analysis was done, all missing data from the dataset was removed. Then, a thorough cleaning process began. A new column for each variable was created to store the eventual clean data. Starting with the *Internet usage* column, the four original categories of responses were converted to three for simplicity. As there was almost no data with the *No time at all* response, the *No time at all* responses were combined with the *Less than an hour a day* responses. These values were then transformed into a new category named *Low Usage.* Additionally, the *Few hours a day* responses from the original column were transformed into a new category named *Moderate Usage*. Lastly, the *Most of the day* responses from the original column were transformed into a new category named *High Usage*. The *Low Usage*, *Moderate Usage*, *and High Usage* response categories were stored in a new column named *InternetUsage*. Moving onto the *Happiness in life* column, more data transformations were done. Each level of agreement or disagreement in the original set of responses was textually translated to terms of actual happiness of a given participant. All *strongly disagree* responses were categorized as *Very Unhappy.* All *disagree* responses were categorized as *Unhappy. All neither agree nor disagree* categorized as *Neither Happy nor Unhappy*. All *agree* responses were categorized as *Happy.* All *strongly agree* responses were categorized as *Very Happy.* These new categories were stored in a new column named *LifeHappiness*.

## Data Visualization

After data cleaning was complete, the distributions of the new variables were visualized in order to provide a cursory examination of possible patterns or effects. First the data was put into a contingency table, seen below:

|  | Very Unhappy | Unhappy | Neither Happy nor Unhappy | Happy | Very Happy |
|---|---|---|---|---|---|
| Low Usage | 2 | 1 |  | 24 | 54 | 14 |
| Moderate Usage | 3 | 29 |  | 141 | 254 | 75 |
| High Usage | 3 | 12 |  | 22 | 37 | 3 |

Upon immediate observation, it is clear that the majority of participants fell into the *Moderate Usage* category for internet usage, followed by *Low Usage*. Somewhat surprisingly, the *High Usage* category had the least amount of responses. Shifting focus to happiness, most of the responses seemed to be concentrated in the *Neither Happy nor Unhappy* and *Happy* categories, with a lesser amount in the *Very Happy* category. The categories measuring levels of unhappiness

did not obtain large amounts of data. Next, the data was visualized in a bar chart to further examine the distribution of levels of happiness across levels of internet usage. This bar chart can be seen below:
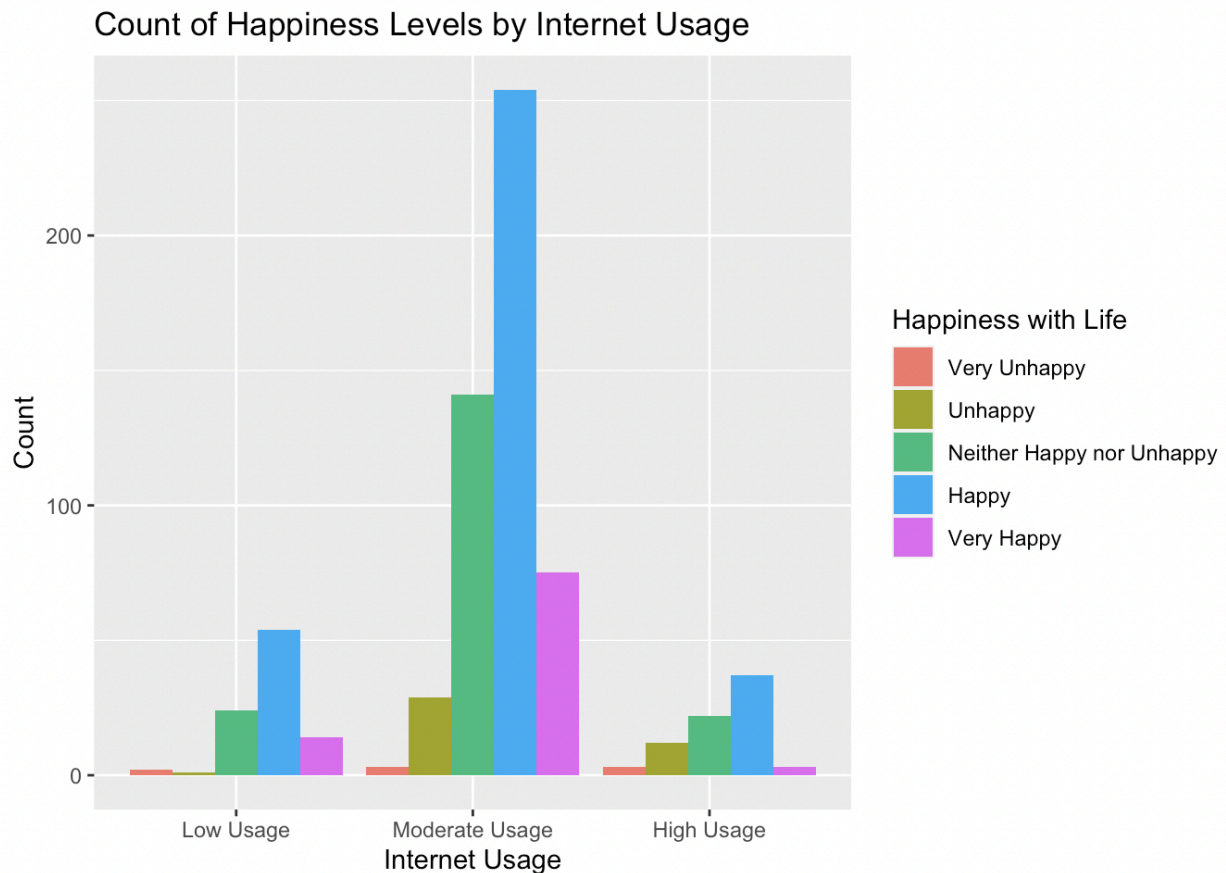


After examining the bar chart, some patterns can be observed at a cursory level. Firstly, the distributions of levels of happiness across levels of internet usage seems to be slightly similar. However, noticeably, the difference between *Neither Unhappy nor Happy* and *Happy* seems to be far greater with *Moderate Usage* and, to a lesser extent, *Low Usage*, than with the *High Usage* category of internet usage. This same principle additionally appears to be true for the difference between the responses involving unhappiness and the *Neither Unhappy nor Happy* category. Thus, the bar graph provides preliminary evidence that there might be a lessening of overall happiness that comes with a high level of internet usage. To examine this pattern on a deeper level, the data must be tested.

## CMH Test

Now, the data is ready to be tested. Again, this data required the usage of a CMH Test, as both variables being studied presented ordinal data. The null hypothesis for this test was that

internet usage is independent of happiness with life, and that there is no linear association between the two variables. The results from the test are seen below:

```
        Cochran-Mantel-Haenszel Statistics for   by


                        AltHypothesis  Chisq Df        Prob
        cor           Nonzero correlation 13.318  1 0.00026289
        rmeans  Row mean scores differ 18.555  2 0.00009350
        cmeans  Col mean scores differ 18.440  4 0.00101226
        general     General association 29.015  8 0.00031524
```

After conducting a CMH Test on *InternetUsage* and *LifeHappiness*, there is evidence to reject the null hypothesis that there is no linear association between the variables at a .05 significance level. This evidence comes from the p-value of .00026289, which is below this threshold of .05. To draw conclusions about a possible direction of this association, the residuals must be analyzed. These residuals are seen below:

```
                Very Unhappy    Unhappy Neither Happy nor Unhappy      Happy Very Happy
Low Usage          0.8917109 -2.2529949             -0.5828643  1.1897341  0.3329528
Moderate Usage    -2.4135534 -0.8340403              0.3396148 -0.5229155  1.6670253
High Usage         2.3323871  3.6075637              0.1721351 -0.5847598 -2.6488914
```

After examining the residuals, a pattern is apparent. As the *High Usage* category for internet usage contains less than expected counts for *Happy and Very Happy* and much lower than expected counts for *Very Unhappy* and *Unhappy*, a conclusion can be made that higher internet usage results in less happiness in life. This is additionally reinforced by the higher than expected counts in *Happy and Very Happy* for *Low Usage* and *Very Happy* for *Moderate Usage.* Interestly, there is a slight deviation from this trend in the count of *Very Unhappy* responses for *Low Usage*, which is lower than expected. As this column includes individuals with zero internet usage, this could suggest that a total lack of internet usage has a negative effect on happiness with life. Again, as the world becomes increasingly digital, being completely excluded from this online connection might have decreasing effects on happiness. In general, however, higher internet usage is associated with lower happiness.

**Partitioned Results: *Low Usage* and *Moderate Usage***

In order to examine this effect even further, levels of internet usage were partitioned to analyze deeper relationships. Firstly, *Low Usage* and *Moderate Usage* were compared. The partitioned table can be seen below:

```
                Very Unhappy Unhappy Neither Happy nor Unhappy Happy Very Happy
Low Usage                 2       1                         24    54          14
Moderate Usage            3      29                        141   254          75
```

The null hypothesis for this test was that low internet usage is independent of moderate internet usage with regards to their effect on the happiness of an individual. For the partitioned results, a Likelihood Ratio Test was utilized. The results of the test are seen below:

```
                        X^2 df P(> X^2)
        Likelihood Ratio 7.4481  4  0.11402
        Pearson          6.5525  4  0.16151

        Phi-Coefficient    : NA
        Contingency Coeff.: 0.104
        Cramer's V         : 0.105
```

After conducting the test, there is no evidence to reject the null hypothesis that low internet usage is independent of moderate usage with regards to their effect on the happiness of an individual at the .05 significance level. The p-value of .11402 is above this threshold of .05. Odds ratios were calculated with the partitioned table to put this comparison into words. For simplicity in the interpretation of the odds ratios, *Very Unhappy* and *Unhappy* were combined, as well as *Very Happy and Happy.* The interpretations are as follows:

The estimated odds of being unhappy with life while having low internet usage are about .48 times the estimated odds of being unhappy with life while having moderate internet usage.

The estimated odds of being happy with life while having low internet usage are about 1.32 times the estimated odds of being happy with life while having moderate internet usage.


### Partitioned Results: *Moderate Usage* and *High Usage*

Next, *Moderate Usage* and *High Usage* were compared. The partitioned table can be seen below:

| | Very Unhappy | Unhappy | Neither Happy nor Unhappy | Happy | Very Happy |
|---|---|---|---|---|---|
| Moderate Usage | 3 | 29 | 141 | 254 | 75 |
| High Usage | 3 | 12 | 22 | 37 | 3 |

The null hypothesis for this test was that moderate internet usage is independent of high internet usage with regards to their effect on the happiness of an individual. For the partitioned results, a Likelihood Ratio Test was utilized. The results of the test are seen below:

```
                              X^2 df    P(> X^2)
             Likelihood Ratio 19.939  4 0.00051333
             Pearson          22.214  4 0.00018170

             Phi-Coefficient    : NA
             Contingency Coeff.: 0.192
             Cramer's V         : 0.196
```

After conducting the test, there is evidence to reject the null hypothesis that moderate internet usage is independent of high internet usage with regards to their effect on the happiness of an individual at the .05 significance level. The p-value of .00051333 is below this threshold of .05. Odds ratios were calculated with the partitioned table to put this comparison into words. For simplicity in the interpretation of the odds ratios, *Very Unhappy* and *Unhappy* were combined, as well as *Very Happy and Happy.* The interpretations are as follows:

The estimated odds of being unhappy with life while having moderate internet usage are about .28 times the estimated odds of being unhappy with life while having high internet usage.

The estimated odds of being happy with life while having moderate internet usage are about 1.75 times the estimated odds of being happy with life while having high internet usage.

## Partitioned Results: *Low Usage* and *High Usage*

Finally, *Low Usage* and *High Usage* were compared. The partitioned table can be seen below:

```
             Very Unhappy Unhappy Neither Happy nor Unhappy Happy Very Happy
 Low Usage            2       1                          24    54         14
High Usage            3      12                          22    37          3
```

The null hypothesis for this test was that low internet usage is independent of high internet usage with regards to their effect on the happiness of an individual. For the partitioned results, a Likelihood Ratio Test was utilized. The results of the test are seen below:

```
                              X^2 df    P(> X^2)
             Likelihood Ratio 20.290  4 0.00043775
             Pearson          18.204  4 0.00112591

             Phi-Coefficient    : NA
             Contingency Coeff.: 0.309
             Cramer's V         : 0.325
```

After conducting the test, there is evidence to reject the null hypothesis that low internet usage is independent of high internet usage with regards to their effect on the happiness of an individual at the .05 significance level. The p-value of .00043775 is below this threshold of .05. Odds ratios were calculated with the partitioned table to put this comparison into words. For simplicity in the interpretation of the odds ratios, *Very Unhappy* and *Unhappy* were combined, as well as *Very Happy and Happy.* The interpretations are as follows:

The estimated odds of being unhappy with life while having low internet usage are about .13 times the estimated odds of being unhappy with life while having high internet usage.

The estimated odds of being happy with life while having low internet usage are about 2.33 times the estimated odds of being happy with life while having high internet usage.

## Conclusion

Overall, there is a linear association between internet usage and happiness. Generally, higher internet usage results in less happiness. While there is not a significant difference between happiness levels for low and moderate internet usage, there are significant differences between happiness levels for both moderate and high usage and low and high usage. This suggests that the most problematic effects of internet usage on happiness are seen when an individual's usage is exceedingly high.

## Future Work

In terms of future work, conducting research with new types of data would be beneficial to get more specific answers on this topic. For example, a regression analysis using a continuous measure of internet usage expressed as a time could provide interesting results. Regarding this specific analysis, surveying more people would help explain if these trends remain the same over larger groups of people. Additionally, examining the relationships between internet usage and other factors that contribute to happiness, such as sleep quality or general mental health, would help provide a better understanding of this association as well.