Title of Your Report

Shuk Yin Chung (1003384964)

October 19, 2020

Average Years of Using Internet at Different Ages

#Shuk Yin Chung #October 19, 2020

Abstract

This paper investigates the usage of Internet in different age groups. By using the linear model and scatterplot, it will show and compare the relationship between ages and time being spent on Internet. It is assumed that older people would spend less time on the Internet.

Introduction

The goal of this study is to investigate how often do different age groups use the Internet. For example, I want to see which age groups use the Internet frequently, and which age groups do not. By using linear model, age is the explanatory variable (x variable) while average years of using Internet is the response variable (y variable). Then, I could analyze if there is any relationship or correlation between ages and using Internet. I chose this data because I notice that many elderly are lacking of technical skills. They do not know how to access the Internet and digital information. Especially during the COVID-19, a lot of information has been posted on online social media, so elderly may not able to access some useful articles and online communication. Since the internet was introduced in the mid-late 20th century, I am interested to see how common the internet was in the early years. As the trend of learning computer skills and using Internet at that time had placed a huge impact on today's people, especially on the elderly. This study also wants to encourage elderly to learn technology and computer skills.

Data

The data were taken from cycle 14 of the Canadian General Social Survey (GSS) in 2000, in which the title of the survey was "access to and use of information communications technology". The survey was held by the General social surveys and the Statistics Canada. It aimed to discover and monitor social trends, such as the frequency of using technology (i.e. computers, Internet, and email) in different locations (i.e. school and work), and the changes in Canadians' lifestyle under the effect of technology ("General Social Survey", 2001).

According to the GSS, the survey was randomly sampling individuals in 2000, and the data was collected through the computer assisted telephone interviewing ("General Social Survey", 2001). The frame population is all Canadians. The target population is people who are above the age of 15 and live in a private household in Newfoundland, Prince Edward Island, Saskatchewan, Alberta, Quebec, Ontario, Nova Scotia, New Brunswick, Manitoba, or British Columbia, Canada ("General Social Survey", 2001). It contained 25090 records in total ("General Social Survey", 2001), so the sample population is 25090.

"Age" is the explanatory variable (x variable) because it should be independent to determine how the length of using the internet changes at different ages. The data of age was drawn from the survey question "Age group of the respondent" ("General Social Survey", 2001). On the other hand, the average years of using the Internet is the response variable (y variable) because it is dependent on and varies based on ages. Its data was drawn from the survey question "How long have you been using Internet?" ("General Social Survey", 2001).

The pro of the survey is that the sample population is large and involves different questions. The con is that it included a lot of "Not stated" or "Not asked" responses. Although it contained 25090 records, it remained 11976 records after cleaning the "NA" responses. The sample sizes were largely reduced and many variables were not useful in my study. Also, the age and years of using Internet were grouped by different ranges, for example, ages group: "15 to 17" and years: "1 to 3 years". They were not convenient for calculation and producing scatterplot. Therefore, I decided to take the average value of each age and year groups. For example, ages group: "15 to 17" becomes "16" and years "1 to 3 years" becomes "2". Then, I group the data based on ages and find its average years. It helps to demonstrate the plot clearly.

Model

This study uses the linear model to illustrate the correlation between ages and average years of using Internet. Generally, the plot of a linear model is a scatterplot with dots and a regression line. The dots represent the data points. It shows how the values of y variable change when the values of x variable increase. The regression line shows the correlation of data, which could be positive (y increases while x increases), no correlation (y remains constant while x increases), and negative (y decreases while x increases).

To estimate the model, we follow the function: $y_i = beta0 + beta1 x_i + e_i$, where y_i is the outcome value of the ith observation, and x_i is the vector of the ith observation. beta0 represents the intercept while beta1 represents the slope. If slope is positive, then the data is positive correlation. If slope is negative, then the data is negative correlation.

Results

Using the table to create a scatterplot with linear regression line:

Average Years of Using Internet in 2000
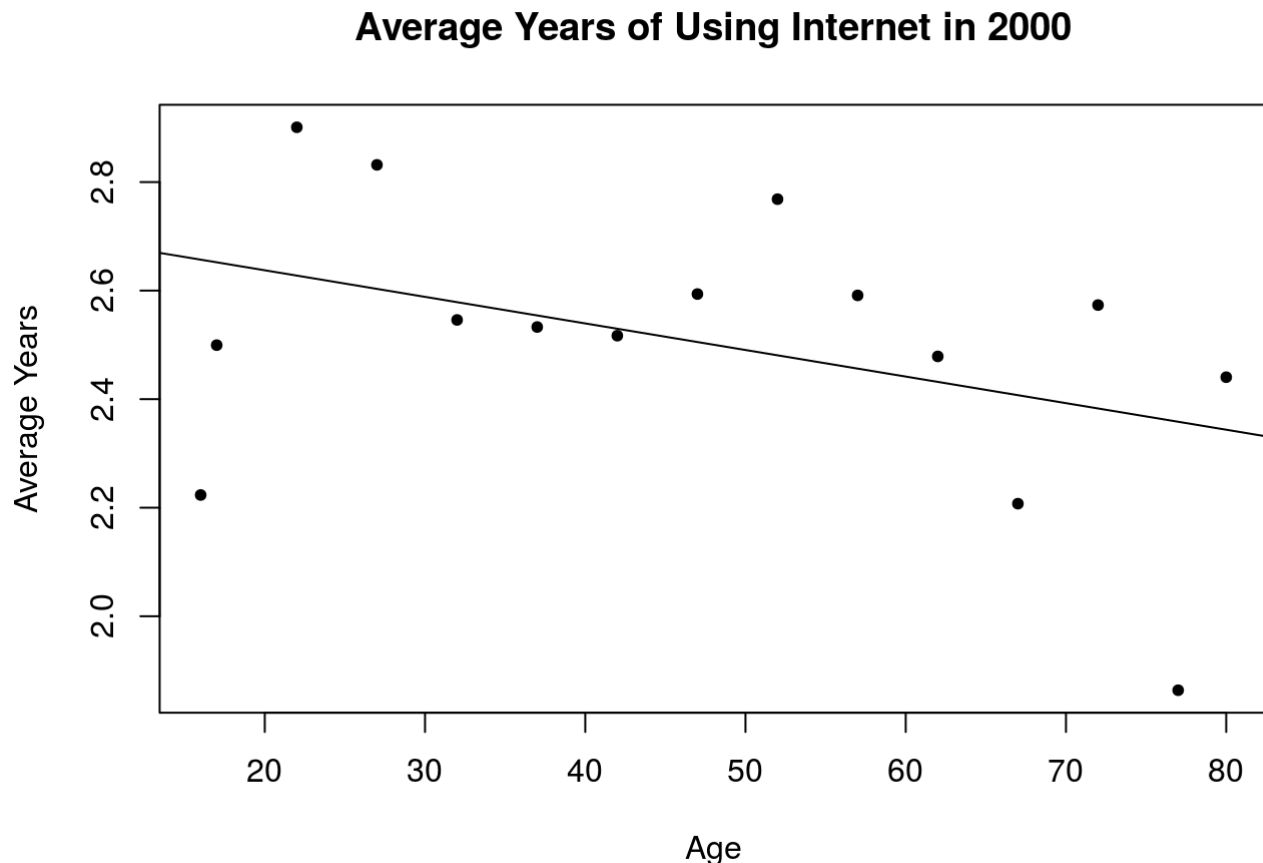
The plot shows a weak negative correlation. Meanwhile, ages between 20 and 30 have the highest two data points, so they are the most often users of Internet. However, ages between 70 and 80 have the least average years of using Internet.

Discussion

In conclusion, older people are less often to use the Internet than the younger people.

We can use lm function to find beta0 and beta1 in this study:

```
## 
## Call:
## lm(formula = average_years ~ age)
## 
## Coefficients:
## (Intercept)          age
##     2.735361     -0.004897
```

Hence, beta0 = intercept = 2.735361 and beta1 = coefficient of age = -0.004897 ei is zero in this case Then, the linear regression estimation is yi = 2.735361 - 0.004897xi It reveals a negative correlation because the slope is negative.

Here is the table of average years grouped by ages:

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

age <dbl>	average_years <dbl>
16	2.223368
17	2.499517
22	2.900954
27	2.831711
32	2.545965
37	2.532914
42	2.516946
47	2.593671
52	2.768487
57	2.591187

1-10 of 15 rows   Previous   **1**   2   Next

The result is the same to my assumption that the elderly had spent less time on using Internet. As the data points decrease from the age of 60 to 80.

Here is the five number summary:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.864   2.460   2.533   2.505   2.592   2.901
```

It shows the time that most people with ages of 60 to 80 had spend on Internet was below the average of 2.505 hours. Hence, I encourage to provide computer training for the elderly, so they can access digital information and online communication.

Weaknesses

The study takes the average of each "age group" and "year of using internet", so the data may not be precise. The sample size is not large enough to represent the entire population. The data were collected in 2000, it might not represent the usage of internet nowadays. The survey question for asking respondents' age and length of using Internet should let the respondents to fill in numeric. Meanwhile, there are some possible factors that affect the result. The potential lurking variable could be country of birth. It is because some respondents could be immigrants who came from countries that had limited internet access in 2000. Also, another potential lurking variables could be income because some respondents might not be affordable to own a computer.

Next Steps

Here you discuss subsequent work to be done after this report. This can include next steps in terms of statistical analysis (perhaps there is a more efficient algorithm available, or perhaps there is a caveat in the data that would allow for some new technique). Future steps should also be specified in terms of the study setting (eg. including a follow-up survey on something, or a subsequent study that would complement the conclusions of your report).

The further study could hold a survey for collecting new data, so the data can represent the trend of using technology nowadays. The further study can also focus on the purposes of using Internet. Here is a possible analysis: If computer skills are required at job or something, would more older people spend more time on Internet?

References

1. "CHASS Microdata Analysis and Subsetting with SDA, Canadian general social surveys." Retrieved from https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/html/gss.htm (https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/html/gss.htm)
2. "General social survey, cycle 14 - access to and use of information communications technology, 2000." General social surveys & Statistics Canada. University of Toronto Data Library Service, 2001. Retrieved from http://meta.chass.utoronto.ca.myaccess.library.utoronto.ca/inventory/1000/1851.htm#doc (http://meta.chass.utoronto.ca.myaccess.library.utoronto.ca/inventory/1000/1851.htm#doc)