# Analysis of Library Impact

Pauline Fittanto

## Introduction

The inspiration for this project is an article written by Samantha Becker published in the *Library Journal* in 2015[1]. In her article, "Outcomes, Impacts, and Indicators," she discusses the impact that librarians have on their communities and how they show that impact. She writes that persuading legislators to expand funding for their libraries is often a struggle for librarians and library program managers because they have to demonstrate concrete evidence of an impact that is often described in abstract terms. Libraries can be said to be pillars of learning, knowledge, and wisdom, offer a place where children can learn and grow, and establish programs that give residents of all ages an outlet to learn, create, socialize, and gain important life skills. This all sounds very nice, but when it comes to receiving funding, most lawmakers want hard evidence that libraries are worth the investment. To begin, I looked at three indicators of crime and observed their individual relationships to library usage, before moving on to educational impact, and perception of library impact on smaller communities. I predicted that library usage will be associated with positive societal factors.

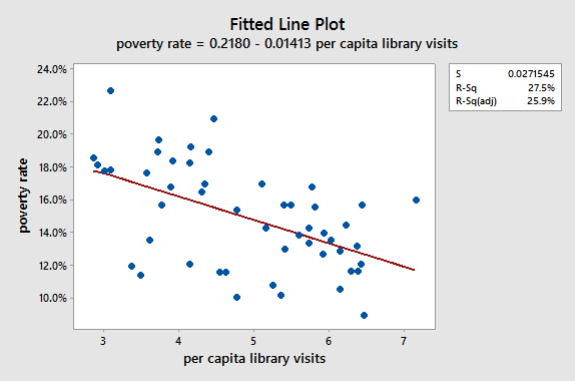
## Poverty Rate as an Indicator of Crime

The first relationship to be examined is that between library usage and poverty rate. Many United States citizens often strongly associate a high poverty rate with a high crime rate. While there are certainly a profuse amount of other factors affecting crime, there is some truth in these claims. A study done by Blake Taylor in his work, "Poverty and Crime," in 2006 found that about 14.21% of the error in crime rate can be explained by the poverty rate. This may seem like a very low number, but the regression equation for this data (y = 3397.1 + 134.85x) tells us that a one percent decrease in poverty rate would be associated with 135 less crimes. Even if there is a large amount of variability around that line, this would be a significant number for policy makers and voters.[2]

## Library Usage and Poverty Rate

The first step to associating libraries to reduced crime, therefore, is examining the relationship between library usage and poverty rate. For data on library usage, I used library visits per capita for all fifty states collected from the institute of museum and library services.[3] I also collected my data on poverty rate for all fifty states from the same source. The scatter plot created from these data pairs (in the form (library visits per capita, poverty)) is shown below.

There is no obvious curve, so the

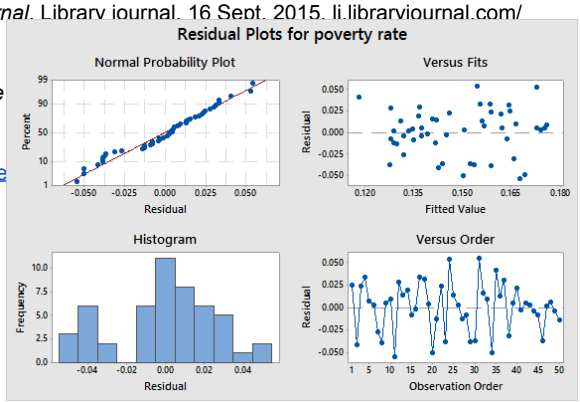



relationship between the two variables appears to be fairly linear. The coefficient of determination, however, is .278633, which tells us that only approximately 28% of the variation in poverty rate can be explained by library visits per capita. While this low number suggests that there is no meaningful relationship between the two variables, a model utility test on the slope of the equation, $b_1$ does suggest that linear modeling can be used. If linear modeling can be used, the null hypothesis (β is equal

[1]Becker, Samantha. "Outcomes, Impacts, and Indicators." *Library Journal*. Library journal, 16 Sept. 2015. lj.libraryjournal.com/2015/09/managing-libraries/outcomes-impacts-and-indicators/.

[2] Taylor, Blake. "Economics.fundamentalfinance.com." *Poverty & Crime* economics.fundamentalfinance.com/povertycrime.php

[3] "Public Libraries." *Public Libraries Hot Report*, 2014, https://thedatawe pls_children.hrml?STATE=2&

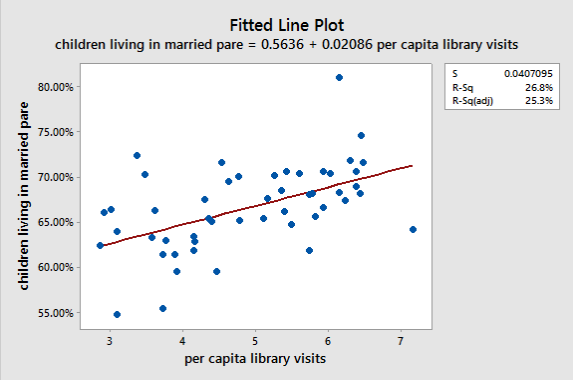


to zero) should be rejected. Fortunately, our p-value for this test was less than 0.01. This poverty rate can be related to a change in library visits per capita. The residuals for the set of poverty rate data can be shown to be normal in the plot below, as the residuals versus fits show no pattern and the histogram of the residuals exhibits close to a normal distribution.

Because of this, we are able to perform regression analysis and calculate a 95% confidence interval for the true slope of the equation. We are 95% confident that the true expected change in percent of poverty rate when library visits per capita is increased by one more visit per capita is between 0.75% and 2.08%. Even at the low end of our interval, these numbers hold significance. Take California for example. The state has an approximate population of 38,680,810 people, and 0.75% of the population is still 290,106 people. When library visits per capita are raised by one, 290,106 less people would be below the poverty line!. This is a good time to remind the reader that the library visits, while related to the poverty rate, do not necessarily cause the poverty rate to decline, but rather they are associated with the decline in poverty. Both phenomenons may be caused by a different underlying factor. For example, although there are more ice cream stands when it is warm, increasing the amount of ice cream stands will not change the weather. We can only say that increased amounts of ice cream stands are associated with warmer weather. Likewise, we may state that more library visitation is associated with a lower poverty rate.

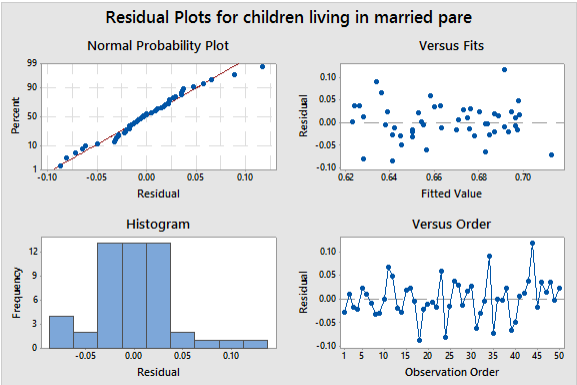
**Library Usage vs. Intact Families**

The next relationship we will explore in this study is that between library usage and family dynamics, specifically the percentage of children living in married family homes. Again, we will use library visits per capita for each state compared with the percentage by state of children living in married homes. The scatter plot showing the relationship between these two data sets is shown below.




Although the relationship is not steep (less change in family dynamic when library usage per capita is changed), the relationship still appears fairly linear, as there is no noticeable curve in the scatter plot. Similar to poverty rate, the coefficient of determination is 0.26, so only 26% of the variation in percentage of children living in married homes can be explained by library visits per capita, and so there is seemingly no valuable linear relationship. Also like poverty rate, however, the model utility test suggests otherwise. The p-value for the test between the null hypothesis ($\beta_1$ is equal to zero) and the alternative hypothesis ($\beta_1$ is not equal to zero) was less than 0.01. This tells us that there is sufficient evidence that $\beta_1$ is not equal to zero and that linear modeling can be used to describe library visits per capita as a function of the percentage of children living in married parent homes. The residuals for the set of poverty rate data are shown to be normal in the set of graphs below.

The plot of residuals versus fits is scattered evenly and not in any particular pattern, and the histogram resembles the normal curve. Therefore, we are able to calculate a confidence interval. In fact, we are 95% confident that the true change in the percentage of children living in married homes when library visits per capita are increased by one more visit is between 1.1% and 3.1%. Like the previous example, small percentages are still significant because even a small percent of a population is a lot of people to be positively or negatively impacted. We can conclude that a larger level of library visits per capita is associated with a larger percentage of children living in households with both parents.




**Educational Attainment as an Indicator of Crime Level**

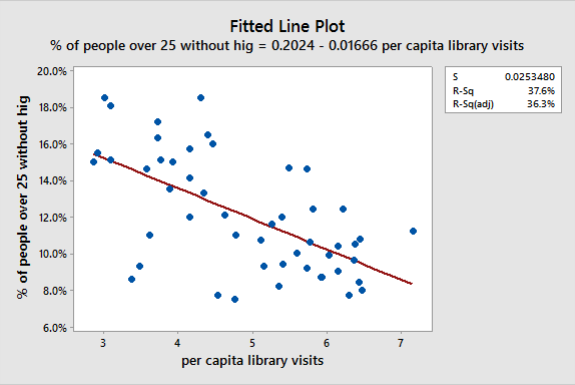
Parents and teachers constantly stress the importance of staying in school to their kids or

students. For many of us in the area I grew up, it was not an option to not attend college and it was not permitted to even think about not graduating high school. Many parents in America associate lack of education with violence and crime. A study performed by students at the Georgia Institute of Technology shows a coefficient of determination of 0.114, meaning 11.4% of the variation in crime can be described by educational attainment. Again, this number seems fairly small, but education is another one of the many factors associated with crime rate, so only some of the variability in crime would be expected to be explained by this factor. Furthermore, a little can go a long way. With only a one percent increase in the members of the population with a high school diploma, there is an associated 8.9% decrease in crime, again, a significant number.[4]

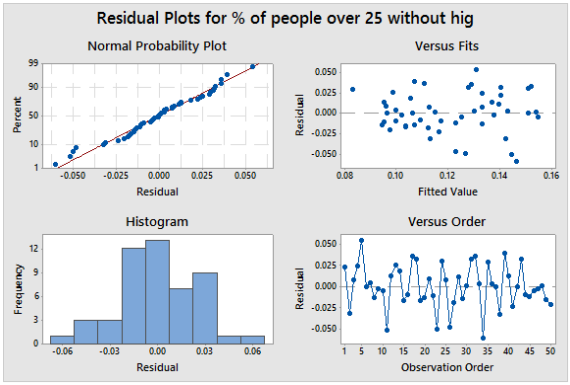
**Library Usage and Educational Attainment**

The final relevant relationship to be studied is that between library usage and educational attainment. Once again for library usage, I used data for each of the fifty states showing library visits per capita. For educational attainment, I looked specifically at the percent of the population over twenty-five years of age that had failed to earn a high school diploma. The scatter plot showing the relationship between these two sets of data is shown below.

There is no obvious curve in the plot, and




the coefficient of determination is 0.37996. Therefore, we can say that 38% of the variation in the percent of residents over twenty-five without a highschool diploma can be explained by library visits per capita. The correlation coefficient is -0.614, showing the negative relationship between the two variables. A higher rate of library visits per capita is associated with a lower percentage of the population that does not have a high school diploma. Again, despite the low coefficient of determination, I ran a model utility test to ensure that modeling could be used. The p-value was less than 0.01. Therefore, the null hypothesis was rejected, suggesting that educational attainment does indeed depend on per capita library visits. The residuals are normally distributed as shown below, so we are able to perform regression analysis and calculate a confidence interval.



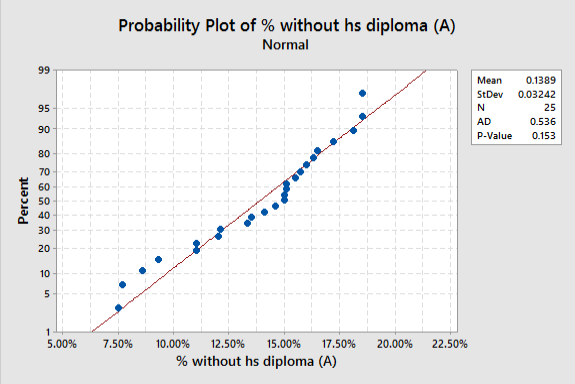

We are 95 % confident that the true expected change in percent of the population over 25 without a high school diploma when library visits per capita is increased by one more visit per capita is between 1.04% and 2.29%. Even at the low end of the interval, the percentage describes a significant amount of affected people. We can conclude that increased library usage is associated with a larger percentage of the population of residents over the age of twenty-five graduating high school.

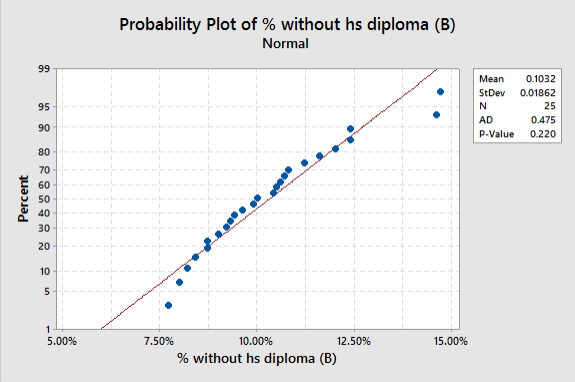
In order to further solidify my claim that library usage has a positive impact on education, I looked at educational attainment in states with high and low levels of library visits per capita to answer the following question: Is there a difference in the average percentage of residents over the age of 25 without a highschool diploma in states with lower levels of library visitation per capita vs. states with higher levels of library visitation per capita? First, I sorted the fifty states from lowest to highest level of library visits per capita and numbered the lowest half one to twenty-five and the highest half one to twenty-

[4] Gentry, Brian, et al. *Impact of Educational Attainment on Crime in the United States: A Cross-Metropolitan Analysis* . 17 Nov. 2016, https://smartech.gatech.edu/bitstream/handle/1853/56029/gentry_mokkapati_rampersad_-_educational_attainment_and_crime.pdf

five. Two sets of ten numbers were then randomly chosen between one and twenty-five using a random number generator. The states corresponding to those numbers were chosen for the study. The percentage of residents of each state over the age of twenty five without high school diplomas in states with low levels of library visits per capita (population A) and the percentage of residents over the age of twenty-five without high school diplomas in states with high levels of library visits per capita (population B) are normal populations, as shown by the normal probability plots below.

We use *A* to represent the population of states with low levels library visits per capita and *B* to represent the population of states with high levels of library visits per capita. Both p-values suggest that at an alpha level of 0.10, we would fail to reject the null hypothesis, so both populations are normal.




Because we have independent randomly collected samples from normal populations, we can perform a two-sample t-test for the true difference between the average percentage of residents over the age of 25 without a highschool diploma. The two samples chosen are as follows:

*A* (low library visits per capita):

- Oklahoma (13.3%), North Carolina (14.60%), Hawaii (9.3%), Montana (7.50%). West Virginia (15.5%), Texas (18.50%), California (18.50%), Maryland (11%), Delaware (12%), Virginia (12.10%)

*B* (high library visits per capita):

- Kansas (10%), New Jersey (11.6%), Ohio (11.20%), Missouri (12%), Wisconsin (9.2%), South Dakota (9.30%), New York (14.70%), Vermont(8.40%), Colorado (9.60%), Washington (9.90%)

The Minitab test results are shown below.

$H_0$: $\mu_A - \mu_B = 0$

$H_A$: $\mu_A - \mu_B \neq 0$

Let A = sample of population of states with little library visits per capita

Let B = sample of population of states with a lot of library visits per capita

**Two-sample T for A vs B**

| | N | Mean | StDev | SE Mean |
|---|---|---|---|---|
| A | 10 | 0.1323 | 0.0363 | 0.011 |
| B | 10 | 0.1059 | 0.0184 | 0.0058 |

**Difference** = μ (A) - μ (B)
**Estimate for difference:** 0.0264
**95% CI for difference:** (-0.0014, 0.0542)
**T-Test of difference = 0 (vs ≠):**
**T-Value** = 2.05
**P-Value** = 0.061
**DF** = 13

As you can see, our p-value was fairly small. Therefore, we reject the null hypothesis as there is sufficient evidence to conclude that there is a difference in the true average percentage of the population over the age of twenty-five without a highschool diploma in states with high library visits per capita vs. states with low library visits per capita.

In fact, we are 95% confident that the true difference in percentage of the population over the age of twenty-five without a highschool diploma is between -0.14% and 5.4%. In other words, we are 95% confident that states with low library visits per capita have between 0.14% less and 5.4% more of a population over 25 without a highschool diploma than states with high library visits per capita.

This confidence interval appears to be weighted unevenly around zero. That is, the true average difference appears to be more positive than negative. Because of this, I also ran an identical test to the one just described, but I changed the alternative hypothesis to $H_A$: $\mu_1 - \mu_2 > 0$. The results of this new test are shown below.

$H_0$: $\mu_A - \mu_B = 0$

$H_A$: $\mu_A - \mu_B > 0$

Let A = sample of population of states with little library visits per capita

Let B = sample of population of states with a lot of library visits per capita

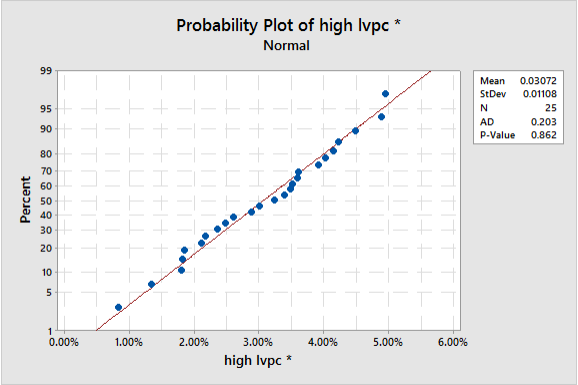
**Two-sample T for A vs. B**

| | N | Mean | StDev | SE Mean |
|---|---|---|---|---|
| A | 10 | 0.1323 | 0.0363 | 0.011 |
| B | 10 | 0.1059 | 0.0184 | 0.0058 |

**Difference** = μ (A) - μ (B)
**Estimate for difference:** 0.0264
**95% lower bound for difference:** 0.0036
**T-Test of difference = 0 (vs >):**
**T-Value** = 2.05
**P-Value** = 0.030
**DF** = 13

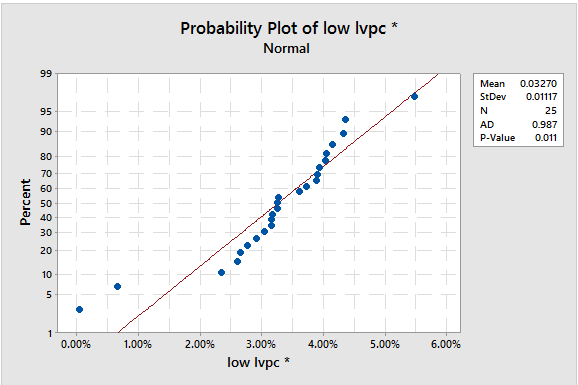
As you can see, our p-value is low (even lower than previously). Therefore, we reject our null hypothesis. There is sufficient evidence to suggest that states with lower levels of library visits per capita have a larger percentage of their populations over the age of twenty-five without a highschool diploma. In fact, we are 95% confident that there is at least a difference of 0.36%. That is, we are 95% confident that states with lower levels of library visits have at least 0.36% larger of a percentage of the population over twenty-five without a highschool diploma than states with higher levels of library visits per capita.

I was interested in performing a very similar test using crime rate as opposed to educational attainment data. However, when the states were split into upper and lower halves, the states with higher library levels per capita were shown to be a normal population, but the states with lower library levels per capita were very far from being a normal population. These results are shown below. Running the test, therefore, would not yield useful information.




**Library Usage and Illinois State Students**

I have shown so far that library usage is not only associated with decreased poverty rate, increased percentage of children living in married family homes, and decreased percentage of the population without a high school diploma, but also that the average educational attainment levels are higher in states with more library usage. Although these are interesting results, and certainly helpful to proving a library's worth in its community, they do



not prove any direct impact from the library on society. To further explore this concept, I looked at library impact at a very local level: the Illinois State University campus.

In order to measure the impact of libraries on the campus, I randomly distributed forty copies of the survey in the Appendix.

Because I wanted my data to be as random as possible, I collected data from a variety of places, including: Milner Library, Watterson Subway, Stevenson Hall, and Schroeder Hall. I also tried to find results from a mixture of both students and professors. Even so, I could only give them to people who actually wanted to answer them, and more students were available than professors. Furthermore, students in certain areas were more willing to answer my surveys than others. For example, students in Stevenson (the math building) were often sympathetic to the fact that I needed the answers for a math project, while students in Schroeder seemed very annoyed, so fewer of them agreed to complete the survey. Regardless, I believe the sample I collected accurately represents the population of the Illinois State University campus. Firstly, although slightly more of my surveys did come from Stevenson Hall and the Milner library, the gap was not large and a fairly even amount of surveys were still collected from each building. Secondly, there are significantly more students on campus than professors, so while I would have liked to have more of an adult opinion, my sample population is not grossly different from the original in that regard. Thirdly, Milner Library is filled with many different types of students of various studying habits and library visitation habits. For example, even students who have never opened a book for fun at home and hate libraries in general, still tend to spend a lot of time at Milner library. I have reason to believe, therefore, that the surveys I collected from Milner were not only filled out by students who have a strong positive bias toward libraries. My results were as follows.

Out of 40 individuals…

- 35/40 (87.5%) said they have been positively impacted or helped by having access to library programs, materials, and resources
    - 95% CI = (0.772511, 0.977489)
- 30/40 (75%) said they would feel negatively impacted if they had not in the past or would not in the future receive access to library programs, materials, and resources
    - 95% CI = (0.615810, 0.884190)
- 14/40 (35%) said they utilize on-site resources more than electronic resources
    - 95% CI = (0.202188, 0.497812)
- 24/40 (60%) said they utilize electronic resources more than on-site resources
    - 95% CI = (0.448182, 0.751818)
- 17/40 (42.5%) said they visit the library often
    - 95% Confidence interval = (0.270429, 0.591099)
- 28/40 (70%) said they visit the library for enjoyment
    - 95% CI = (0.534684, 0.834373)

Because a sample size of forty is large enough, we can apply the Central Limit Theorem to calculate the above confidence intervals, as well as run some two proportion z-tests. I intend for my tests to show that the true population proportion of people who have been positively impacted by having access to library programs and the true population proportion of people who feel they would be negatively impacted if they had not or will not have access to those programs and resources are not statistically different. The results are shown below.

**Test and CI for Two Proportions**

| Sample | X | N | Sample p |
|---|---|---|---|
| 1 | 35 | 40 | 0.875 |
| 2 | 30 | 40 | 0.75 |

**Difference** = p(1) - p(2)
**Estimate for difference:** 0.125
**95% CI for difference:** (-0.0438515, 0.293851)
**Test for difference = 0 (vs ≠ 0):**
**Z** = 1.43
**P-Value** = 0.152
**Fisher's exact test: P-Value** = 0.252

At an alpha level of 0.1, our p-value is large and so we fail to reject the null hypothesis. We can conclude that there is no difference in the proportion

of people who have felt positively impacted by a library and people who would feel negatively impacted if that library had not existed or had been taken away. This result added with the confidence intervals calculated earlier, show that a large fraction of the Illinois State community feel that libraries and the programs and resources that they offer have had a positive impact. Furthermore, we determined that we can be 95% confident that between 53.4% and 83.4% of our campus population actually goes to the library for their own enjoyment. Despite the wide interval, this shows that over half of the campus enjoys the library and its services. Both the proven impact on the students and the amount of students who enjoy it demonstrates the value of the library in our community. Furthermore, many students attending the university have grown up in different communities and also have been impacted by those libraries and not necessarily Milner. There is a chance, therefore, that the survey may have covered a larger scope than just our campus.

**In Conclusion…**

Despite being unable to prove causation in this manner, I believe the evidence collected here supports the claim that libraries improve and positively impact the communities in which they operate. First, we have shown that poverty rate, percentage of children living in married parent homes, and percentage of the population over the age of twenty-five without a highschool diploma can be written as a function of library usage. Secondly, we have shown that a lower poverty rate, a higher percentage of children living in married parent homes, and a lower percentage of the population over the age of twenty-five without a highschool diploma is associated with a higher level of library usage. Additionally, we have shown that states with lower levels of library visits per capita have a larger true average percentage of their populations that do not have high school diplomas than states with higher levels of library usage. Finally, we have shown that in the local population of the Illinois State University Campus, a significant amount students have felt a positive impact of libraries either at home, on campus, or both. I believe that further studies providing a deeper analysis into the points described here would only help my claim. In order to further investigate an causal link between positive societal impact and library programs, for example, We could collect a group of students having similar financial and family situations, as well as a similar intelligence level and measure grades and education they go on to receive, as well as any trouble they get into with the law. I hope I might have the chance to look into improving this study or to work on a similar study in the future.

# Appendix

## Survey of Library Usage

Have you ever been positively impacted or helped by having access to library programs, materials, and resources?
(a) Yes
(b) no

Do you feel that you would be negatively impacted if you had not in the past or would not in the future receive access to the aforementioned library programs, materials, and resources?
(c) Yes
(d) no

Of these library resources (including all functions a library offers) Which do you utilize more?
(a) Electronic resources
(b) On-site resources

Of these library resources (including all functions a library offers) which do you enjoy using more
(a) Electronic resources
(b) On-site resources

Do you visit the library often? (more than 3 times a week)
(e) Yes
(f) no

Do you ever visit the library when it is not necessary to do so? (for enjoyment)
(g) Yes
(h) no

Have you ever been positively impacted or helped by having access to library programs, materials, and resources?
(i) Yes
(j) no

Do you feel that you would be negatively impacted if you had not in the past or would not in the future receive access to the aforementioned library programs, materials, and resources?
(k) Yes
(l) no

Of these library resources (including all functions a library offers) Which do you utilize more?
(c) Electronic resources
(d) On-site resources

Of these library resources (including all functions a library offers) which do you enjoy using more
(c) Electronic resources
(d) On-site resources

Do you visit the library often? (more than 3 times a week)
(m) Yes
(n) no

Do you ever visit the library when it is not necessary to do so? (for enjoyment)
(o) Yes
(p) no