

EDUCATION AND LIFE OUTLOOK

ANALYSIS

Intro to Data Analysis

Section 112

The #1 Team



Marist College

School of Computer Science and Mathematics

Submitted to:

Reza Sadeghi

Table of Contents

Table of Figures.....	3
Introduction.....	4
Github Repository Address.....	4
Project Description.....	5
Review of Related Work.....	6
Project Merits.....	8
Descriptive Statistics.....	9
Data Visualization.....	10
Advanced Descriptive Statistics.....	15
Bivariate analysis.....	16
Null Hypothesis Significance Testing.....	18
Standardized Residual.....	21
Effect Size.....	24
References.....	25

Table of Figures

Education Level Bar (Figure 1.1).....	10
Perceived Fairness Bar (Figure 1.2).....	11
Overall Happiness Bar (Figure 1.3).....	12
Helpfulness of Others Bar (Figure 1.4).....	12
Excitement for Life Bar (Figure 1.5).....	13
Trustworthiness of Others Bar (Figure 1.6).....	14

Project Report of Education and Life Outlook

Analysis

Team Name: The #1 Team

Team Members:

Beck Towne..... beck.towne1@marist.edu (Team Head)

My name is Beck Towne. I am a Sophomore at Marist University majoring in Computer Science and Data Science. I chose to work with Rich as I know him to be a reliable and intelligent person. We have worked together on class projects before and we were easily able to collaborate and create a presentation both of us were proud of. We decided I would be team head since I created our report document and made the classroom announcement declaring our group's goal. Despite this declaration, we do intend for the project work to be 50/50.

Rich Vultaggio..... richard.vultaggio1@marist.edu (Team Member)

My name is Richard Vultaggio, I go by Rich and I am a Sophomore From Long Island. I recently just switched from Computer Science to Data Science and I have a minor in Spanish. I am looking forward to working with Beck as she and I have worked on projects in the past in a chemistry class we took together and we gave amazing presentations. Beck was decided as the team leader as she took lead in creating the document and putting our team together.

Github Repository Address:

https://github.com/RichVult/112_Education-and-Life-Outlook-Analysis_The-1-Team.git

Project Overview

The goal of this project is to investigate whether or not there is any correlation between a person's education level and their overall outlook on life. The data used in this project will be sourced from the General Social Survey's results from 2022. Using this data, we will examine the respondents' education level and compare this to their answers regarding their perceived overall happiness, willingness to trust others, the fairness of life, and more. The specific variables we intend to use are EDUC, HAPPY, LIFE, HELPFUL, FAIR, and TRUST. The details of these variables are as follows:

EDUC- The respondent's level of education

HAPPY- Taken all together, how would you say things are these days--would you say that you are very happy, pretty happy, or not too happy?

Possible answers: very happy, pretty happy, not too happy

LIFE- In general, do you find life exciting, pretty routine, or dull?

Possible answers: exciting, routine, dull

HELPFUL- Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?

Possible answers: helpful, self-centered

FAIR- Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?

Possible answers: advantage, fair

TRUST- Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?

Possible answers: yes, no

Our hypothesis at this time is that higher levels of education correlate with a more positive and optimistic outlook on life.

Related Work

IMPACT OF EDUCATION ON OPTIMISM/PESSIMISM

By Dr. D.L. Pareek & Ms. Neerja Sharma

This experiment looks at the relationship between one's optimism and outlook on life and the highest education their parents reached.

This could be very helpful to us as it provides us a lot of insight to how education can affect the optimism or pessimism of the people around you. Along with its relatively large sample size of 250.

This might not be perfect as it is not exactly what we are asking but only closely related.

Does Higher Education Increase Hedonic and Eudaimonic Happiness?

By Boris Nikolaev

This study looked at an individual's general life happiness and satisfaction, and whether or not they had received a college degree. This is a very useful study as it directly asks the same question as we do, and gives evidence towards our claim that receiving a college degree does lead to an increase in general happiness.

Intro to Data Analysis 220L112_Project Progress Report_Phase #2_The#1Team
Education, wages and job satisfaction

By Cecilia Albert and María Angeles Davia

This study looks at life/job satisfaction and its correlation with higher education.

This study is good as it gives insight into how the two things are related, but also points out some possible confounding variables, such as wage.

Project Merits

The results of this study are relevant to a variety of individuals and organizations. Firstly, many individuals considering the pursuit of higher education hope that doing so will lead to a happier, more fulfilling life. Should we find that educated people have a more positive outlook on life, this could sway these individuals' decisions on the subject. The inverse would therefore have the opposite effect. In the same respect, our findings may be relevant to educational institutions. If this study shows that people with more education are more positive, this could be used to the advantage of college recruiters. The inverse is true for those attempting to recruit for trade schools or other alternatives to college. Finally, the interpretation of this data could be a gateway to further study. Should the results show a correlation (positive or negative) between education and life outlook, further investigation could be done to discern the reason for this correlation, leading humanity to a greater understanding of what makes a person happy.

Descriptive Statistics

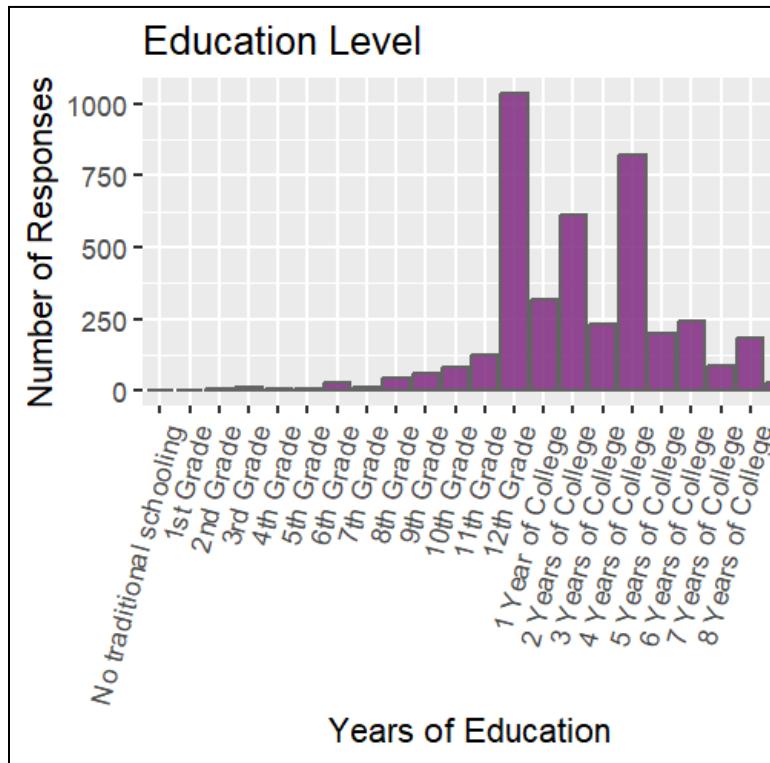
Intro to Data Analysis 220L112_Project Progress Report_Phase #2_The#1Team

<u>Variable</u>	<u>Mean</u>	<u>Median</u>	<u>SD</u>	<u>Min</u>	<u>Max</u>
EDUC	14.152913	14	2.9075612	0	20
HAPPY	2.023775	2	0.6703174	1	3
LIFE	1.638057	2	0.5754427	1	3
HELPFUL	1.700000	2	0.6757650	1	3
FAIR	1.619403	2	0.6430645	1	3
TRUST	1.836483	2	0.5593554	1	3

Interpretation:

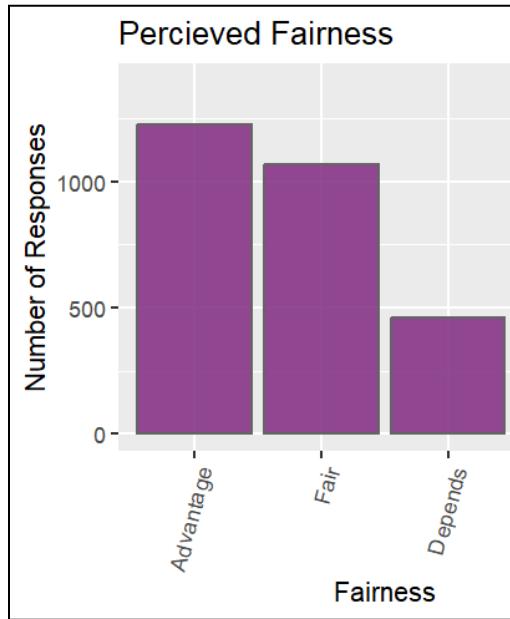
These values will be very useful for our study. First is EDUC, which is how many years of education an individual had taken, this ranged from 0 (no schooling) to 20(Grad school). It had a mean of around 14 and a median of 14, and it had a SD of 2.9. So EDUC is a bit right skewed and has a relatively large amount of deviation, it could be our most important variable as it is at the core of our question. The other variables, HAPPY (happiness), LIFE (life outlook), HELPFUL (how likely a person thinks someone else will help), FAIR (how fair a person thinks other people are), and TRUST (how much a person trusts others). These all are on a one to three scale, 3 being the most optimistic/best outlook. They all have a median of two and an SD around .6. We are very interested in any relationship or correlation between these other variables and our EDUC variable.

Data Visualization



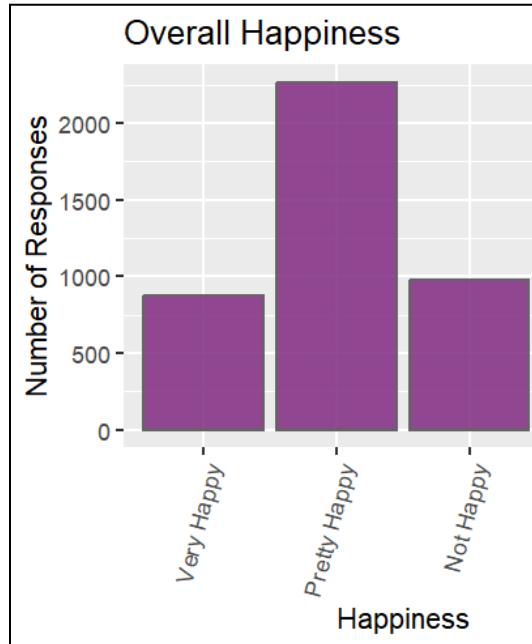
(Figure 1.1)

Interpretation: This graph shows the education level of each of the respondents. The category with the most responses is, by far, the category for respondents that have completed 12th grade but not completed any years of college. However, there are many more respondents that have completed more than a high school education than there are respondents that did not complete high school. This causes a heavy right skew in the distribution. The extremely light left tail suggests that anyone who has completed less than a 6th grade is statistically an outlier, and should perhaps be removed from the dataset before further analysis is conducted.



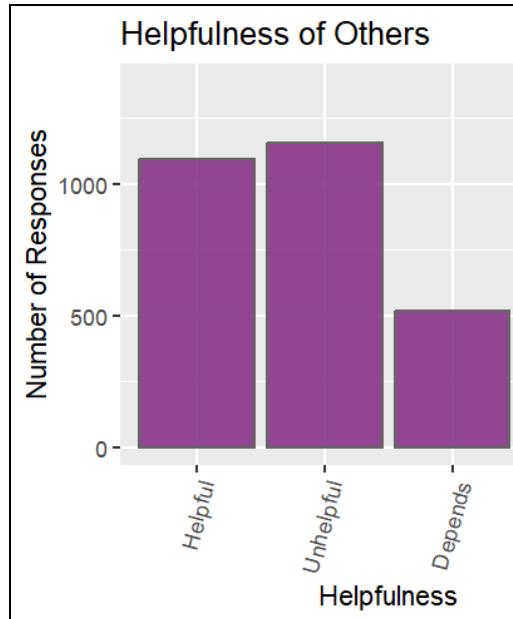
(Figure 1.2)

Interpretation: This graph depicts the number of respondents that either thought people would take advantage of them, that others would try to be fair, or that the fairness of others depends. The graph makes it clear that the highest percentage of respondents thought that people would try to take advantage of them if they had the chance. Still, this is less than half of the respondents in total, considering the number of responses in the “fair” and “depends” categories.



(Figure 1.3)

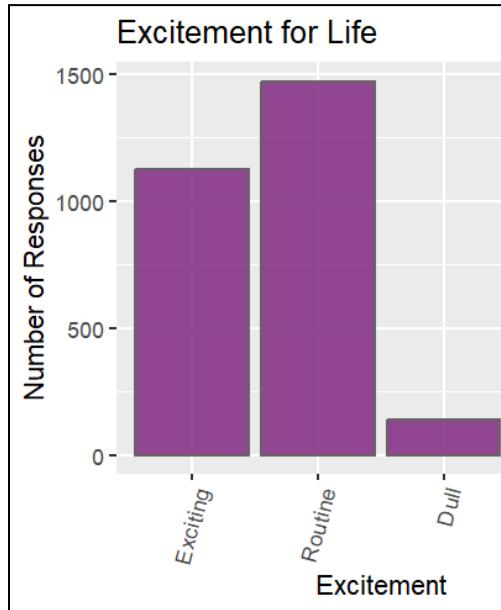
Interpretation: This graph depicts respondent's answers in respect to their overall happiness in life. There were more respondents that answered "pretty happy" than respondents that answered "very happy" and "not happy" combined.



(Figure 1.4)

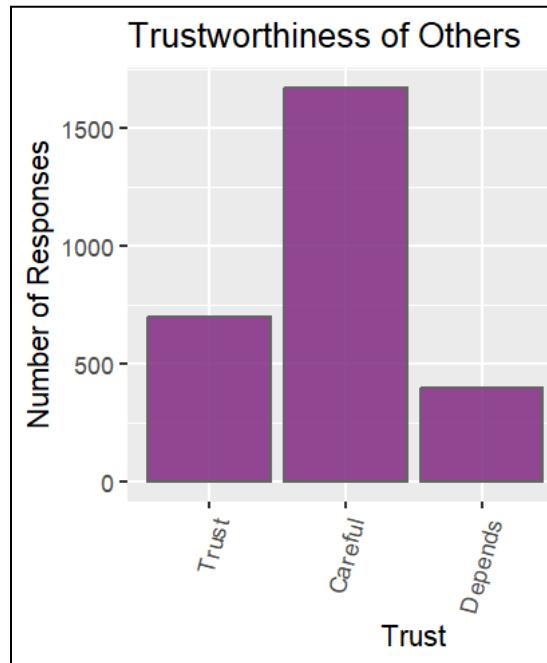
Intro to Data Analysis 220L112_Project Progress Report_Phase #2_The#1Team

Interpretation: This graph depicts how helpful the respondent views other people in general. It appears there is a trend towards people thinking people are unhelpful, but it is close.



(Figure 1.5)

Interpretation: This graph tracks a respondent's general excitement in life, most people view it as exciting or routine, while a small minority view it as dull.



(Figure 1.6)

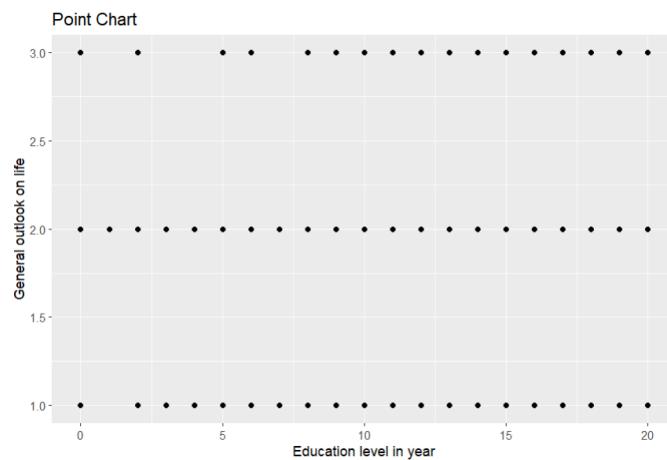
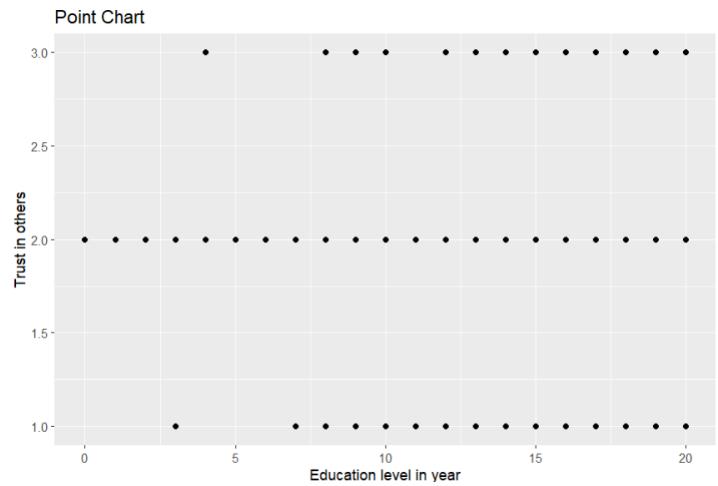
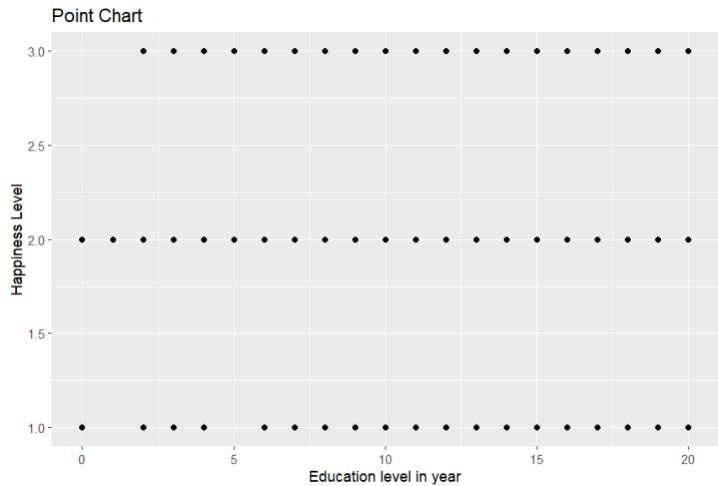
Interpretation: The graph tracks a respondent's trust of others, most people are careful around others, while some trust without question, or only trust depending on the context.

Advanced Descriptive Statistics

Variable	EDUC	HAPPY	LIFE	HELPFUL	FAIR	TRUST
IQR	4	0	1	1	1	1
Kurtosis (Difference)	1.393	-0.776	-.723	-.809	-.657	-.092
Range	0-20	1-3	1-3	1-3	1-3	1-3
Skew	-.411	-.027	.236	.446	.55	-.033

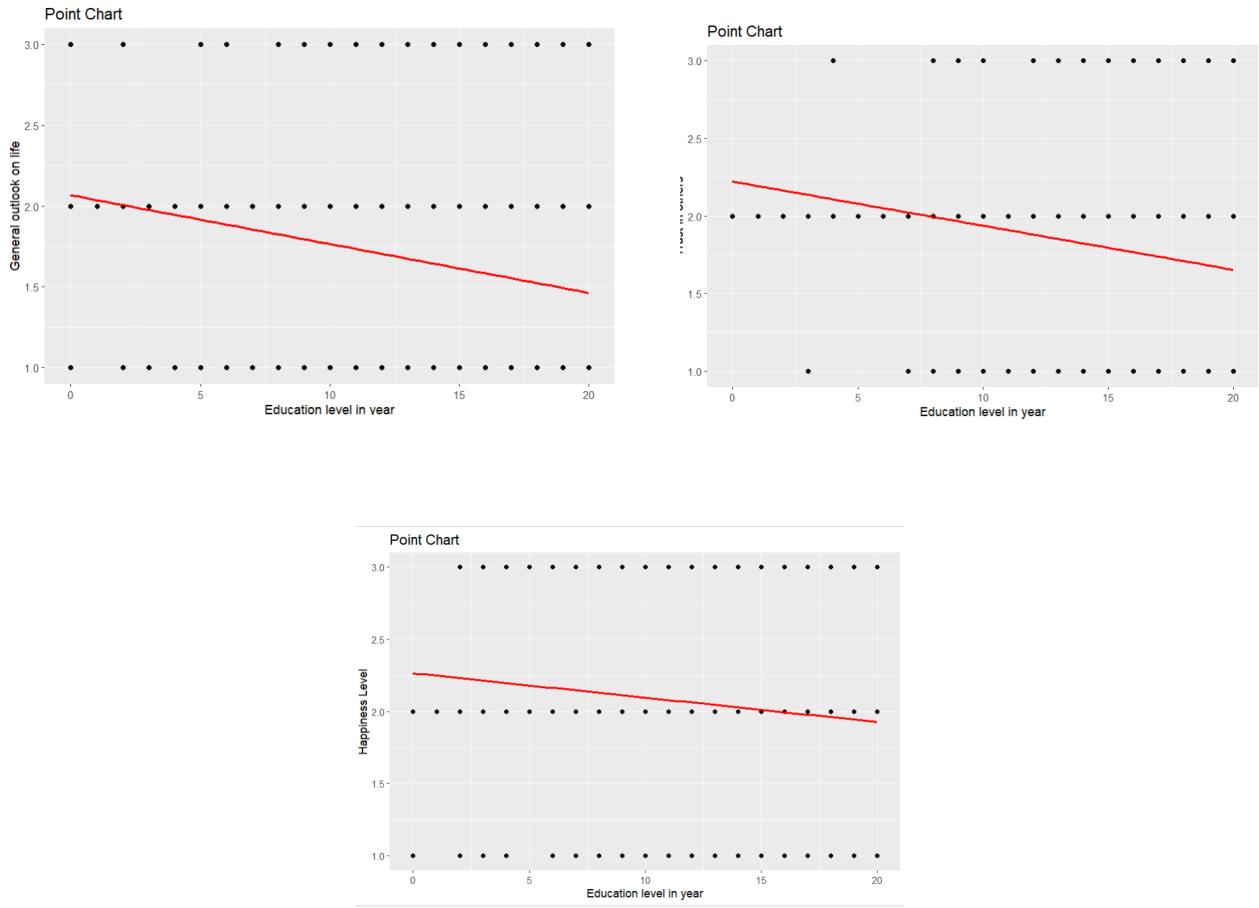
EDUC is the most interesting, it has the widest range of values from 0-20, being the range of years spent in school. It has an IQR or 4, and is leptokurtic, meaning to have steep increases, with long tails. It also is almost symmetrical with a low magnitude negative skew. All other variables have a range of 1-3, including HAPPY, which measures how happy a person is, it is platykurtic meaning the graph would be wider with short or no tail. It also has a negligible skew and an IQR of 0. The final four variables all had an IQR of 1, likely because of the smaller range. LIFE is also platykurtic, along with the other next 4 variables, and has a negligible skew, it represents a general outlook on life. HELPFUL represents how the respondent viewed the helpfulness of the average person, and has the same attributes, range of 1-3, negligible skew, platykurtic, and an IQR of one, as FAIR and TRUST, which looked at whether or not the respondent thought that life was fair, and if they trusted a random person, respectively.

Bivariate Analysis



These graphs can help demonstrate the relationship between the selected variables, HAPPY, LIFE, and TRUST against years of education, due to the limited range of HAPPY,LIFE, and TRUST, the graphs can be confusing/hard to read, but inserting a trend line can help the viewer understand what is being shown.

Intro to Data Analysis 220L112_Project Progress Report_Phase #2_The#1Team



The trend lines help show that there is at least a slight negative relationship between the variables and education level.

Null Hypothesis Significance Testing

As we are studying the effects of education level on multiple variables, the null hypothesis significance testing must be done on each variable as they compare to the education level.

Happiness:

Null hypothesis : A person's level of education has no effect on their happiness.

The proportions for education level are the same for different values of happiness.

Alternative hypothesis: A person's level of education has an effect on their happiness. The proportions for education level are not the same for different values of happiness.

Test statistic (chi squared) : 84.156

Df: $(21-1)*(4-1) = 60$

A chi square distribution table shows that for 60 degrees of freedom and a significance level of 0.05, chi squared must be greater than 79.082. Since our test statistic, 84.156, is greater than 79.082, we can reject the null hypothesis.

Fairness:

Null hypothesis : A person's level of education has no effect on their perception of how fair others are. The proportions for education level are the same for different values of fairness.

Alternative hypothesis: A person's level of education has an effect on their perception of the fairness of others. The proportions for education level are not the same for different values of fairness.

Intro to Data Analysis 220L112_Project Progress Report_Phase #2_The#1Team
Fairness (cont.)

Test statistic (chi squared): 158.8

Df: $(21-1)*(4-1) = 60$

A chi square distribution table shows that for 60 degrees of freedom and a significance level of 0.05, chi squared must be greater than 79.082. Since our test statistic, 158.8, is greater than 79.082, we can reject the null hypothesis.

Helpful:

Null hypothesis : A person's level of education has no effect on their perception of how helpful others are. The proportions for education level are the same for different values of helpfulness.

Alternative hypothesis: A person's level of education has an effect on their perception of how helpful others are. The proportions for education level are not the same for different values of helpfulness.

Test statistic (chi squared): 80.462

Df: $(21-1)*(4-1) = 60$

A chi square distribution table shows that for 60 degrees of freedom and a significance level of 0.05, chi squared must be greater than 79.082. Since our test statistic, 80.462, is greater than 79.082, we can reject the null hypothesis.

**Intro to Data Analysis 220L112_Project Progress Report_Phase #2_The#1Team
Life:**

Null hypothesis : A person's level of education has no effect on their excitement for life. The proportions for education level are the same for different values of excitement.

Alternative hypothesis: A person's level of education has an effect on their excitement for life. The proportions for education level are not the same for different values of excitement.

Test statistic (chi squared): 127.48

Df: $(21-1)*(4-1) = 60$

A chi square distribution table shows that for 60 degrees of freedom and a significance level of 0.05, chi squared must be greater than 79.082. Since our test statistic, 127.48, is greater than 79.082, we can reject the null hypothesis.

Trust:

Null hypothesis : A person's level of education has no effect on their willingness to trust others. The proportions for education level are the same for different values of trust.

Alternative hypothesis: A person's level of education has an effect on their willingness to trust others. The proportions for education level are not the same for different values of trust.

Test statistic (chi squared): 189.56

Df: $(21-1)*(4-1) = 60$

A chi square distribution table shows that for 60 degrees of freedom and a significance level of 0.05, chi squared must be greater than 79.082. Since our test statistic, 189.56, is greater than 79.082, we can reject the null hypothesis.

Standardized Residual

Happiness:

	1	2	3
0	1.41123236	-0.20393539	-1.11667954
1	-0.73332839	1.27771808	-0.78941864
2	0.18694791	-0.20393539	0.05887842
3	-0.86650359	-0.32268700	1.20921467
4	1.70979941	-0.64151934	-0.89185945
5	-1.55695703	2.04191311	-0.89185945
6	-1.20734034	-1.31238434	2.69337118
7	0.32412050	-0.93476617	0.78145395
8	1.83528353	-2.36696023	1.00456706
9	0.02483761	-2.48771309	2.88410869
10	-1.64310989	-0.46663358	2.12317527
11	-2.25285289	0.78487449	1.24572754
12	-1.16181834	0.12909377	0.96467666
13	0.26496476	-2.06556559	2.16007852
14	-0.90110620	0.55546448	0.21594416
15	-0.08454683	-1.38224888	1.69693145
16	0.45220246	2.32050273	-3.14670569
17	1.79338562	-0.73782843	-0.85954052
18	2.39042340	0.36391538	-2.72067553
19	1.07164302	0.92346368	-2.10845432
20	-0.03455232	0.92590618	-1.04914063

The standardized residuals can help us find exactly where the data changes, for example some years that stand out are 9 years of education, with a residual of -2.49 of happy, and 2.88 of not happy. 18 years also stands out, with a very happy residual of 2.39, and a not happy residual of -2.72, highlighting a general positive correlation between happiness and education.

Intro to Data Analysis 220L112_Project Progress Report_Phase #2_The#1Team Fairness:

	1	2	3
0	1.06580306	-0.89145442	-0.31272713
1	-0.93904088	1.12269636	-0.31272713
2	-0.93904088	1.12269636	-0.31272713
3	-1.20517440	0.71009765	0.87326699
4	0.12692080	-0.77680426	1.13249382
5	1.84756404	-1.54533160	-0.54211085
6	0.23844433	-0.64750878	0.71155626
7	0.54794092	-0.07492177	-0.82947276
8	2.27260484	-2.06816043	-0.37499124
9	1.69823398	-1.15610408	-0.95932895
10	2.36303425	-2.46208800	0.15363559
11	3.14646332	-3.80655584	1.12585159
12	2.66033987	-2.14500236	-0.92038790
13	1.83044306	-2.03636750	0.34435092
14	1.78963541	-1.27951970	-0.90423012
15	-0.33723739	0.61021200	-0.47338955
16	-3.56803920	4.12654335	-0.94524852
17	-1.49523301	0.65849413	1.47153988
18	-4.88382698	4.21995187	1.19746925
19	-2.42559467	2.79112778	-0.61792019
20	-4.20478564	2.83252814	2.42752599

Some years that stand out are 20 years of education, with a residual of -4.2 of advantage, and 2.83 of fair. 11 years also stands out, with an advantage residual of 3.15, and a fair residual of -3.8, highlighting a general positive correlation between the perception of fairness of others and education.

Helpful:

	1	2	3
0	-0.8584421	1.0972348	-0.3729032
1	-0.8584421	1.0972348	-0.3729032
2	1.1658685	-0.9121389	-0.3729032
3	-1.7190271	1.1912678	0.7835662
4	-0.7056085	1.1912678	-0.7467373
5	-0.7056085	1.1912678	-0.7467373
6	-1.7670110	1.6651914	0.1350600
7	-0.7428425	0.6261569	0.1691715
8	-1.4183205	1.8890260	-0.7319889
9	-0.3381150	1.6475029	-1.9957224
10	-0.2066979	0.8902765	-1.0422249
11	-1.8407753	1.3942208	0.6586653
12	-1.6614339	1.0863281	0.8562392
13	-0.3507806	0.1513800	0.2994037
14	0.8081016	0.9209469	-2.6212713
15	0.3467676	0.7089472	-1.6021304
16	2.0348545	-2.5042423	0.7369049
17	0.6229058	-1.8933716	1.9397083
18	1.6908269	-2.3215015	0.9784033
19	0.9330549	-0.6777481	-0.3779159
20	1.1617785	-2.4293078	1.9412899

Some years that stand out are 20 years of education, with a residual of -2.43 of unhelpful, and 1.16 of helpful. 16 years also stands out, with a helpful residual of 2.03, and an unhelpful residual of -2.5, highlighting a general positive correlation between the perception of helpfulness of others and education.

Intro to Data Analysis 220L112_Project Progress Report_Phase #2_The#1Team Life:

	1	2	3
0	-0.65338731	-0.16104909	1.85543567
1	-1.18079820	1.30546977	-0.32216695
2	-0.65338731	-0.16104909	1.85543567
3	-0.95812264	1.16712531	-0.50967224
4	-1.64446008	1.90338976	-0.64504679
5	-0.04820743	-1.52730838	3.62551922
6	-1.66258403	0.98625165	1.50745646
7	0.09679963	0.16650956	-0.60327473
8	-0.42701093	0.16191274	0.59755959
9	-1.43243544	0.44613751	2.22787400
10	0.01712694	-1.85646191	4.23479624
11	-2.49481207	1.48738939	2.24486758
12	-3.92450434	2.68061572	2.74665412
13	-1.98688088	2.16289501	-0.46436617
14	0.69743926	0.26027145	-2.18395197
15	0.51065863	-0.76541344	0.60167178
16	2.05238213	-0.79660648	-2.82977002
17	2.71101611	-2.72803394	0.11990930
18	3.72725336	-2.67495815	-2.31146588
19	-0.04811087	0.23013663	-0.42046959
20	2.43059980	-1.83879338	-1.29000018

Some years that stand out are 20 years of education, with a residual of 2.43 of exciting, and -1.84 of routine. 18 years also stands out, with an exciting residual of 3.73, and a routine residual of -2.73, highlighting a general positive correlation between excitement of life, and education.

Trust:

	1	2	3
0	-0.58140110	0.71793124	-0.30970956
1	-0.58140110	0.71793124	-0.30970956
2	-0.58140110	0.71793124	-0.30970956
3	-0.01141417	0.38034765	-0.62018778
4	-1.16424519	0.38034765	1.15261315
5	-1.16424519	1.43764432	-0.62018778
6	-2.26488669	2.79674892	-1.20649419
7	-0.66952516	1.10395517	-0.82145197
8	-1.18154260	0.76450197	0.53508973
9	-1.35087315	0.61280269	1.04984164
10	-2.50703341	2.60135379	-0.50650085
11	-1.73991343	2.97027556	-2.30474508
12	-2.31141941	2.24719139	-0.21347754
13	-2.00101243	1.82569683	0.01591722
14	-1.95765912	1.92277116	-0.21351808
15	1.53114495	-0.27837333	-1.88780778
16	3.22098625	-4.21429317	2.11306185
17	2.10057002	-2.20585739	0.46841164
18	4.06063733	-3.95233583	0.38263119
19	3.39398577	-2.72392492	-0.65191840
20	2.79727946	-3.37524069	1.35776547

Some years that stand out are 18 years of education, with a residual of 4.06 for trust, and -3.95 for careful. 16 years also stands out, with a trust residual of 3.22, and a careful residual of -4.21, highlighting a general positive correlation between trust of others, and education.

Effect Size - Cramer's V Testing

Happiness:

Cramer V = 0.1014

A Cramer V of .10414 shows a weak relationship between happiness and education.

Fairness:

Cramer V = 0.2221

This Cramer V shows a weak correlation between perceived fairness and education.

Helpful:

Cramer V = 0.1591

This Cramer V shows a weak correlation between education and the perceived helpfulness of others.

Life:

Cramer V = 0.1531

This Cramer V shows a weak correlation between education and excitement for life.

Trust:

Cramer V = 0.2142

This Cramer V shows a strong correlation between education and trust of others.

References

Pareek, D. L., & Sharma, N. (2020). *Impact of Education on Optimism/Pessimism.* <https://www.shahsatnamjieducation.com/wp-content/uploads/2023/05/Book-PDF-After-Correction-31-12-22-compressed.pdf#page=93>

Davern, Michael; Bautista, Rene; Freese, Jeremy; Herd, Pamela; and Morgan, Stephen L.; General Social Survey 1972-2022. [Machine-readable data file]. Principal Investigator, Michael Davern; Co-Principal Investigators, Rene Bautista, Jeremy Freese, Pamela Herd, and Stephen L. Morgan. NORC ed. Chicago, 2024. 1 datafile (Release 4) and 1 codebook (2022 Release 4).

Nikolaev, B. (2016). Does higher education increase hedonic and eudaimonic happiness? *Journal of Happiness Studies.* <https://doi.org/10.1007/s10902-016-9833-y>

Albert, C., & Davia, M. A. (2005, January). Education, wages and job satisfaction. In *Epubnet conference.* Citeseer.

<https://www.geeksforgeeks.org/kurtosis-in-r-programming/>

<https://www.geeksforgeeks.org/chi-square-test-in-r/>

<https://www.geeksforgeeks.org/how-to-calculate-cramers-v-in-r/>

https://www.w3schools.com/r/r_functions.asp

Help with R commands