Income Trap Cycle[*]

The United Kingdom Edition

Yu Sun   Ariel Xu   Qiao Zhou

24 March 2022

Abstract

Research has shown that Income has a positive correlation to happiness. This leads us to think about maximizing our income to achieve greater satisfaction. This paper explores the possible correlation between income and various variables gathered from the European Social Survey(Research Data, 2002) for the United Kingdom. We have concluded the correlation between the variables chosen by the backward selection method and income.

Introduction

There are many decisions to make in life: career path, family, education... However, when we dig into the career path's decision, there are two main reasons that we hear the most: possible income and interest. We cannot talk about how interest affects career path decisions because it is a very biased topic; people will never come up with an agreement. Thus, we will be shifting our focus to the income factor. Research on the correlation between family income and happiness has been the study of the century. A small positive correlation is shown between concurrent happiness and family income. (North et al. (2008)) With acknowledging the positive relationship between happiness and income. We grew interested in factors contributing to income growth to maximize our happiness and satisfaction level.

This research will discuss the possible correlation between factors such as education level, gender, and income. We aim to explore the data to determine uncommon correlations. Our initial hypothesis is that all of the selected interest variables show a correlation to income. The result shows that women have low income compared to men and that higher education levels tend to have a higher income, and last (age...); However, our result has shown some interesting factors determining income in the United Kingdom, such as being more republican on the political spectrum will tend to negatively affect your income base on the data available.

The rest of the paper is organized as follows: In the Data section(section 2), we explain the data background and analysis regarding the potential factors influencing income in the UK. In the Result section(section 3), we show the impact of the chosen variable, age, education level, and gender corresponding to the income. In the discussion(section 4), we made comments on the potential bias in paper such as data, p-hacking, other potential variables not included in the dataset and we felt it will greatly impact income.

Data

The research uses `R` language(R Core Team 2020) as its foundation, where we have used packages such as `tidyverse`(Wickham et al. 2019), `haven`(Wickham and Miller 2022), `here`(Müller 2020), `Rstata`(Braglia 2016),`gridExtra`(Auguie 2015),`readr`(Wickham, Hester, and Bryan 2022),`hrbrthemes`(Bob Rudis ORCID iD [aut 2020),`car`(Fox and Weisberg 2019),`dplyr`(Wickham et al. 2021),`knitr`(Xie 2021),`readr`(Wickham, Hester, and Bryan 2022),`grid`(murrell 2005) to prepare data for this project.Where we then used `Kable`(Zhu 2020) to generate table and `ggplot`(Wickham 2016) to generate plots.

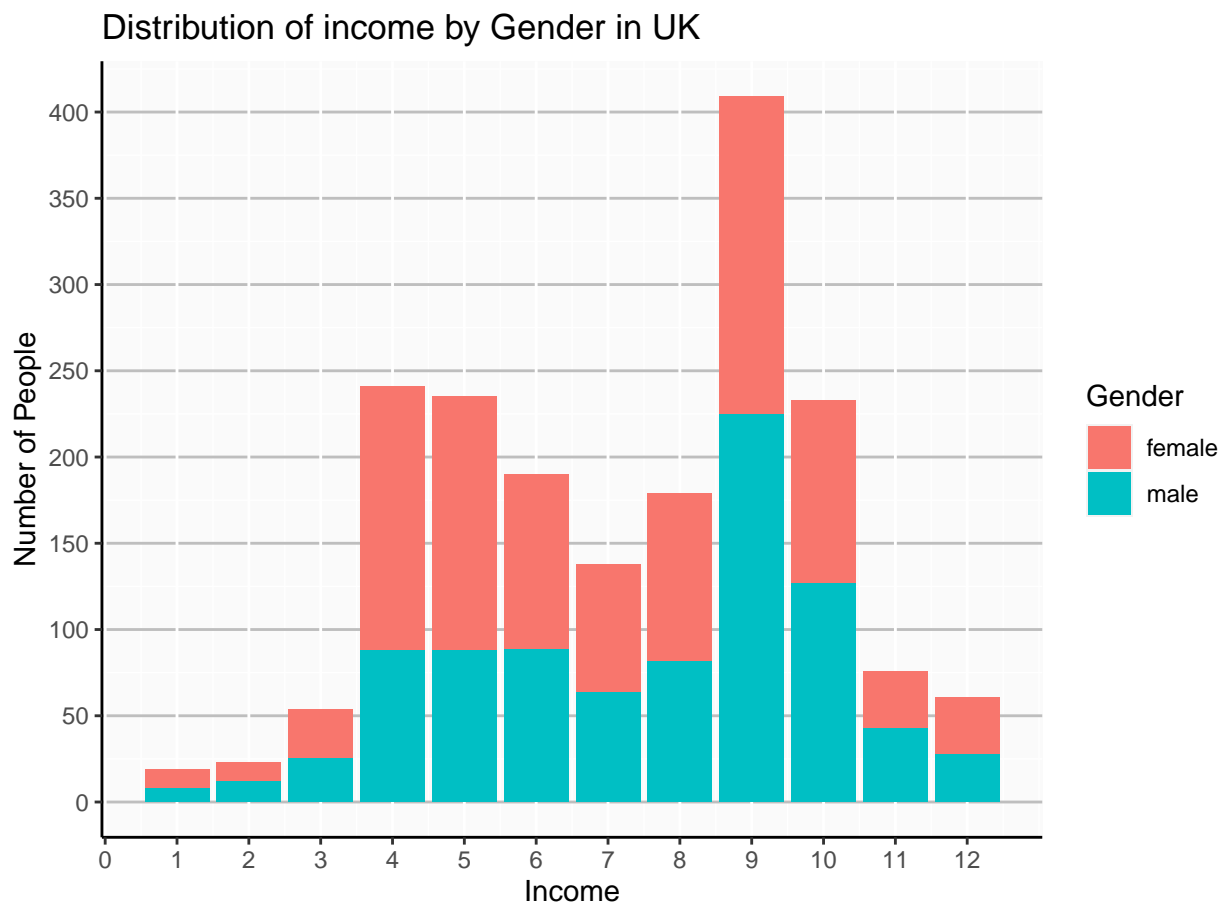[*]Code and data are available at: https://github.com/sunyu30/Income-Trap-Cycle

Data Source

All of the data we use in this research was collected from the website of the `European social survey`(Research Data 2002). Data in the European social survey was collected via face-to-face CAPI interviews in all participating countries. And in each country, the national funding agency appoints a National Coordinator (NC) and a survey organization to implement the survey according to the common ESS Specification. We can see from the website that the response rate targets 70 percent, and the non-contact rate targets 3 percent at maximum. These standards ensure data accuracy in each country and optimize the comparability of data across countries.

Data overview

In this research, we want to investigate what factors affect the income level in the United Kingdom. Since it is known that the Great Recession (a period of marked general decline observed in national economies globally) occurred between 2007 and 2009. We decided to mainly investigate 2006(a year between the great recession) and 2010(a year after the great recession). We can further check if the great recession has an overall impact on the income level in the United Kingdom. After searching through the database, we finally chose six variables that we think are highly correlated with the income level. They are gender, age, the highest level of education, placement on the left-right scale, worry about work problems when not working (how often), and plan for the future or take each day as it comes.

Descriptive analysis

Income level and gender

Distribution of income by Gender in UK

Income data collected from European Social Survay in 2006(Figure 1)

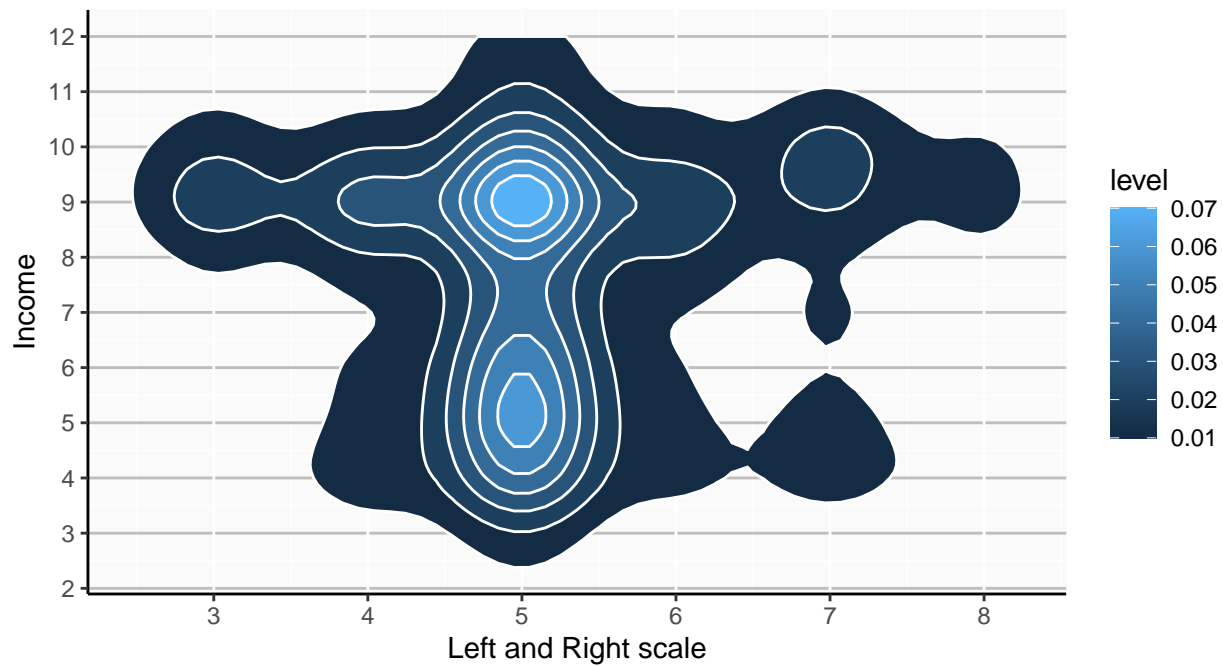# Distribution of income by Gender in UK

Income data collected from European Social Survay in 2010(Figure 2)

Now, let's investigate the possible correlation between gender and income level in 2006 and 2010 in the United Kingdom. We plot two histograms; the first one (figure1) is for 2006 while the second one (figure2) is for 2010. The x-axis represents the income level for each histogram, and the y axis represents the number of people at a certain income level. Also, the number of people was grouped by gender Let's take a look at the year 2006 (figure1) first; we can see that the number of females and the number of males is quite balanced at each income level, there is no clear sign that gender has a significant effect on the income level. As for the histogram for the year 2010 (figure2), we can notice that the number of females is significantly greater than the males for low-income groups. For the medium to high-income level, the number of females does not differ much from the number of males. What's more, if we look at the histogram separately by gender, we can find out that the histogram for the female is right-skew while the histogram for men is relatively stable along the axis of income level. Thus, we can conclude that in 2010, the income level was correlated with gender identity. Those workers with low-income levels are mainly composed of women, and female workers are less likely to have high-income levels.
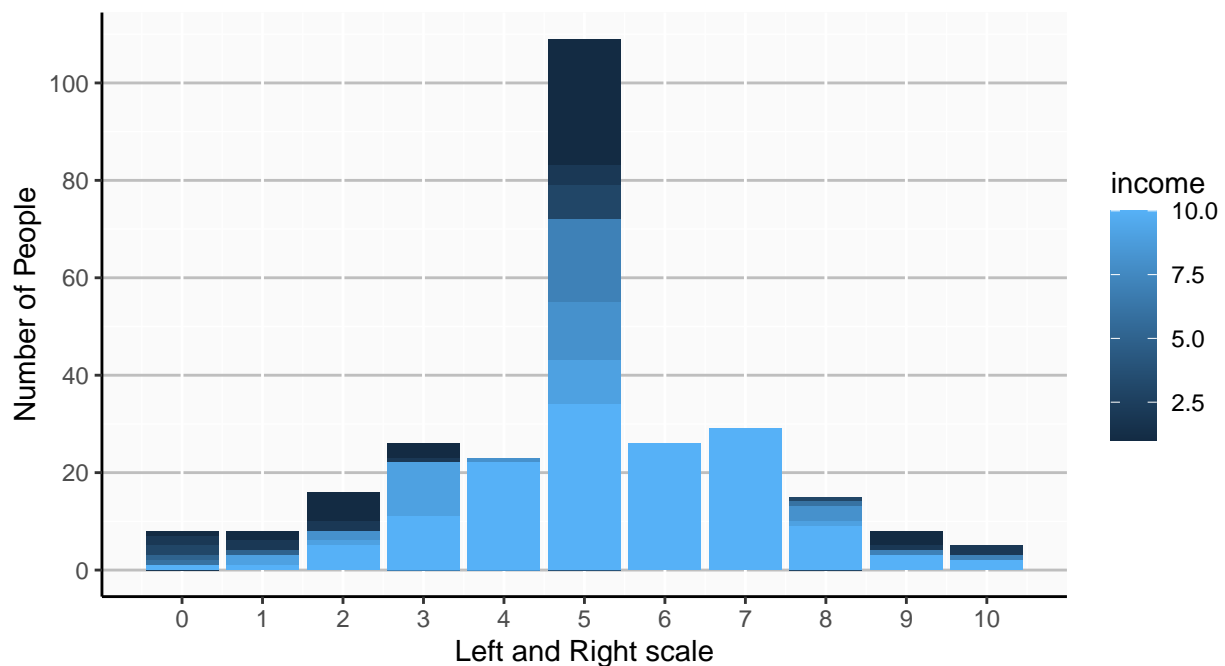
Distribution of income by the political stand in UK

Income data collected from European Social Survay in 2010(Figure 3)

## Distribution of income by the political stand in UK

Income data collected from European Social Survay in 2010(Figure 4)

We plot two graphs to investigate the possible correlation between income and placement on the left-right scale. The density plot (figure3) is made using the data in 2006, and the histogram (figure 4) is drawn using the data in 2010. Let's look at the density plot (figure3) first; we can observe a few facts for this plot. First, most people in the United Kingdom in 2006 had centrism and the range of their income level range from the lowest to the highest. Second, for the people who are more supportive of the left-wing democracy,

most of them had a very high-income level in 2006. As for those who are more supportive of the right-wing democracy, the range of their income level is much broader than those who support left-wing democracy. Next, we look at the histogram (figure4); this histogram is a unimodal one. The people who support centrism have the most significant count, which is much higher than the others. This histogram has no clear sign that the placement on the left-right scale significantly impacts the income level.

Income level with respect to non-working hour work concerns

Distribution of income by worry about work in UK
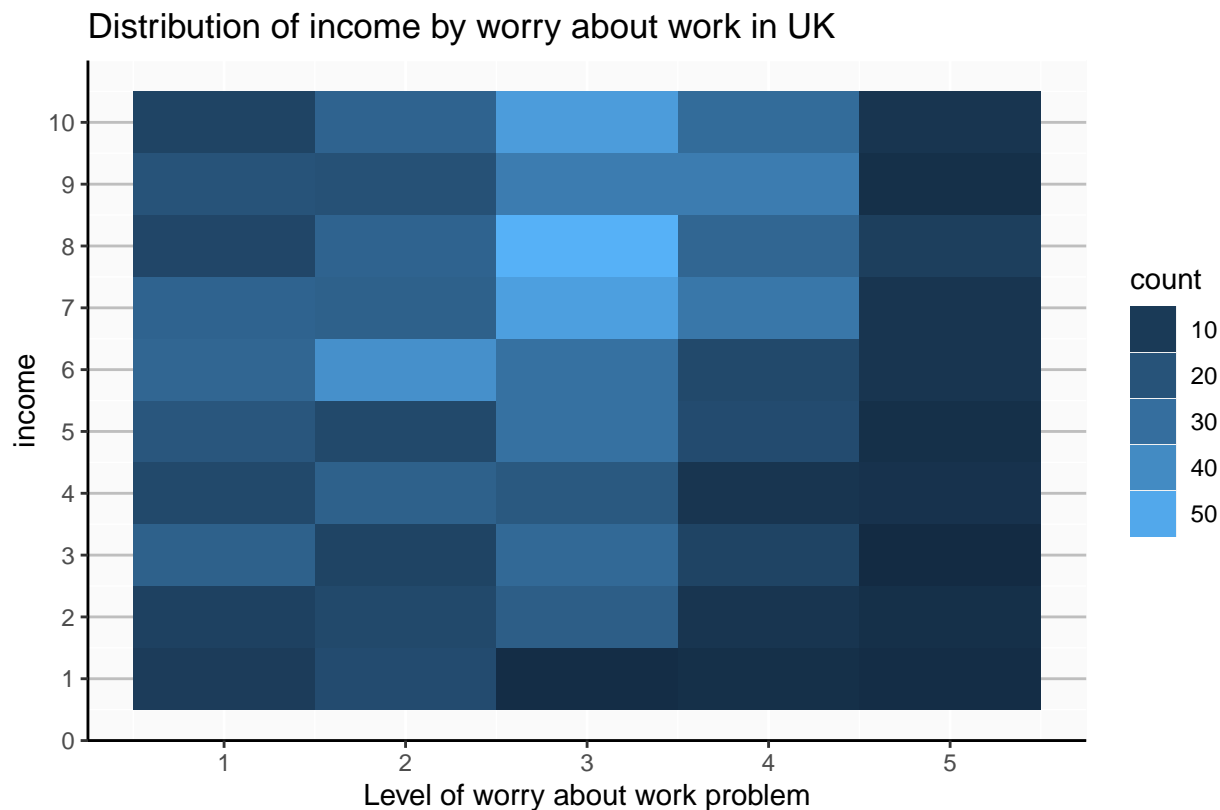
Income data collected from European Social Survey in 2010(Figure 5)

Now, let's look at the correlation between the income level and worry about work problems when not working. We only use the data in 2010 to plot this density graph (figure5) because this variable was not included in the 2006 dataset. First, look at the top row of the figure; we can see that at the highest income level, there are much fewer people who worry too much or don't worry at all than those people who have a reasonable frequency of worrying at level 3. This observation may imply that worrying too much or too less hurt the income level. However, in our data, the people with worrying level 3 have the most significant count, which might be a reason for the above observation. Thus, we need further investigation to check if this variable impacts income level at all. So, we next look into the column where the worrying level is 4 (Note: we don't look at level 5 because there are only 70 people in this group, which is not representative at all), we found out that at this level of worrying, people tend to have medium to high-income level. And at the column where the worrying level is 1 and 2, people tend to have medium to low income. Then we conclude that worrying more about work impacts the income level positively to some extent.

The great recession and income level

Trend in UK Income
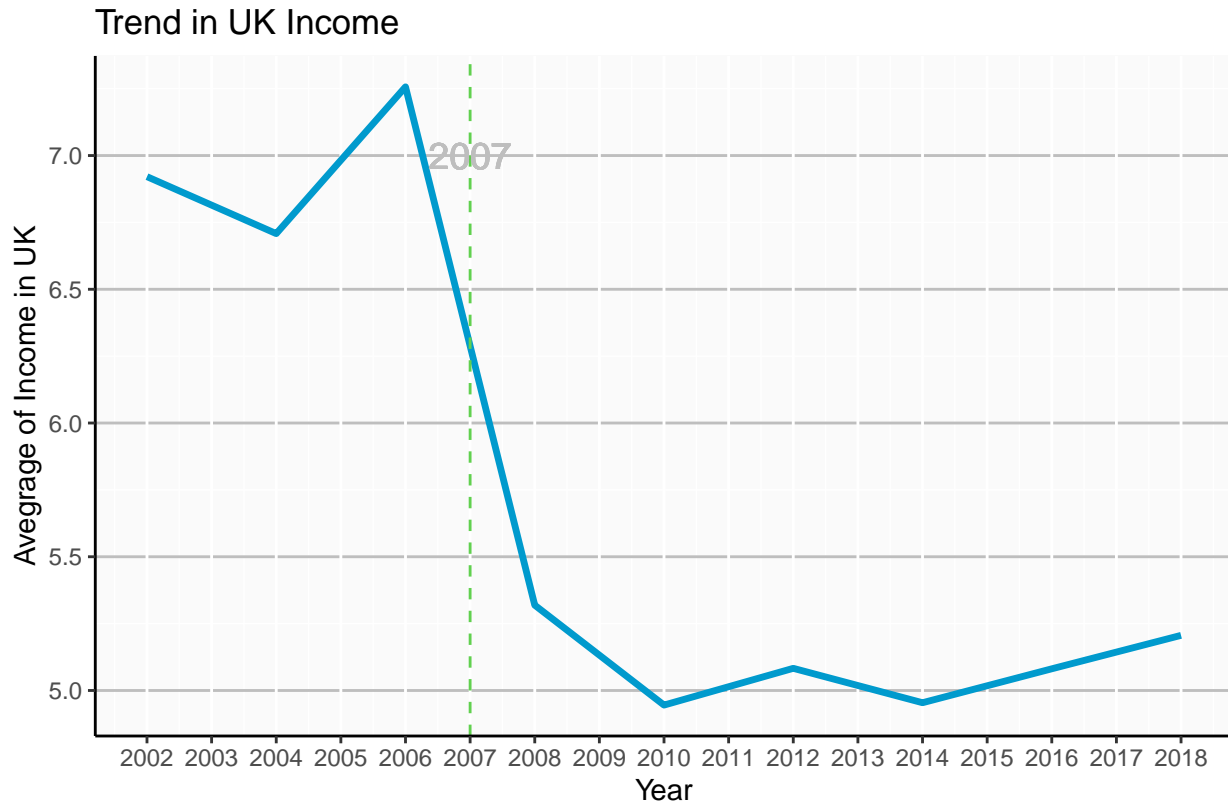
Income data collected from European Social Survey(Figure 6)

Firstly,if we look at the line plot (figure6), we can observe a huge decline in the income level in the United Kingdom since the Great Recession. And there is no sign of the income level recovering back to the level before the recession until now. Then if we look back at the graphs in the above part (figure1)(figure2), we can notice that the shapes of the two histograms are pretty different. The histogram for 2006 is bimodal, while the histogram for 2010 is a right-skew one. This fact clearly shows that the great recession has a significant negative impact on the labor market. After the great recession, the number of people who get very low-income levels increased significantly compared to 2006. This observation might be due to the high unemployment rate due to the recession. What's also need to notice is after the recession, the number of people who have very high-income level also has a significant increase compared to in 2006. This fact implies the increased disparities in wealth caused by the great recession.

Potential data limitations and ethical problems

For the gender variable in our data, there are only two female and male response options. With only two response options, various genders remain under- or unrepresented. For example, trans and non-binary people are not considered in our data, or they are just being considered their physiological gender. This fact may bring up potential bias for our research since gender identity might be a significant influencer for the income, especially for those gender identities minority. According to a study by LinkedIn published in 2019, transgender employees face a 14% income gap compared to non-trans employees.This gap equates to an annual income shortfall of £5,340.Also, they are experiencing verbal abuse in the office, and they have been made to feel uncomfortable by colleagues because of their gender identity. This discrimination could explain why 14% of LGBT+ respondents think their chances of promotion in their company would be hindered if they came out.(Heyndyk (2019)) However, our data cannot show the correlation between those minority gender identities and the income because it does not include complete options for the gender identities when designing the survey.

Result

We will present our main results by linear models. Firstly, we will introduce the methods we use: linear regression, hypothesis test, VIF values and BIC backward selection. Then, we will sketch a few plots to estimate the potential outcome. Lastly, we will fit models and analyze them to get the results. ### Method First, I would like to introduce the primary method called linear regression. Linear regression models the relationship between multi variables by fitting a linear equation to observed data. We can use the linear regression model to predict a particular variable. The variables we used for prediction care are called independent variables. Just like its name, we expect these variables have no repeated information. Then, the variable we want to predict is called the dependent variable. Then, I will explain the linear regression model.

Linear Regression   First, I would like to introduce the primary method called linear regression. Linear regression models the relationship between multi variables by fitting a linear equation to observed data. We can use the linear regression model to predict a particular variable. The variables we used for prediction care are called independent variables. Just like its name, we expect these variables have no repeated information. Then, the variable we want to predict is called the dependent variable. Then, I will explain the linear regression model.

Here is the general linear regression model `Yale`(Uni, n.d.):

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_i x_i + \epsilon_i$$

x_i : This represents the independent variable in the multi-linear regression. y_i : This represents the dependent variable in the multi-linear regression. β_0 : This presents the y-intercept of the multi-linear regression model. It shows the value of the dependent variable y_i are equal to zero. Sometimes this variable is meaningless. β_i : It is called regression coefficient. These values shows how much we expect the dependent value to change as x_i changes. ϵ_i : This is the residual. The residual represents the true deviations between the observed values to their means.

We want to use this dataset to find out the fitted value to use them to estimate the actual regression line. The coefficients also interpret the power of influence of each factor. Then, I would like to state a few assumptions for the linear regression model `CFI`(CFI 2022):

Linearity   We assume all the explanatory variable x_i have a linear relationship to the dependent variable y_i

Constant Variance   The error should have constant variance. In another words, this assumption suggests the amount of error in the residual for each point is the same and equal to σ^2

Normality   All the errors should follow a normal distribution. The mean value should be 0, and the variance is σ^2.

Independence   All the errors should be independent of each other. It should be independently and identically distributed. If errors are dependent, it will cause the predictive power of the model will be worse somewhere.

Independence of observations   The observations in our model should be independent. It means there should be any relation between observations. We cannot find a pattern or relationship between observations by simply looking at the data frame. Every observation cannot affect others.

VIF Values

The VIF is used to detect multicollinearity in a linear regression model. It will estimate how much a linear regression coefficient variance is inflated due to multicollinearity. The range of the variance inflation factors range starts from 1. The value of VIF interpreted what percentage the error is inflated for each coefficient.

Thus, we want the VIF as small as possible, and we will remove the variables with high VIF values. `Statistic how to`(Stephanie 2020)

BIC and backward selection

First, I will explain the process of backward selection:

1. Starts with the model that contains all the variables under previous consideration (also called the Full Model).

2. Removing the minor significant variables from the model.

3. Stop when the stopping rule is reached out. I use the BIC value to be my stopping rule in the Backward selection. I want to find the model with the best AIC value using the "step()" function.The BIC (Bayesian information criterion) evaluates whether the model fits the data. We use the Backward selection function to find the lowest possible BIC.

Hypothesis test

The hypothesis test can test a claim about a proper parameter in a population. We will do the hypothesis test for every single coefficient for the multilinear regression model. Here is how I procedure:

1. First, We need to state the hypothesis, alternate hypothesis, and significance level. In this model, I am interested in the value of coefficients.

2. Then, we can calculate the p-value for every coefficient.

3. we can compare the p-value with a significant level and get our conclusion.

4. If the significant level is greater than the p-value, we can conclude that we have no evidence to support our null hypothesis.

Model and final conclusions

Trend of income by 3 Variables in 2006 and 2010(Figure 6)
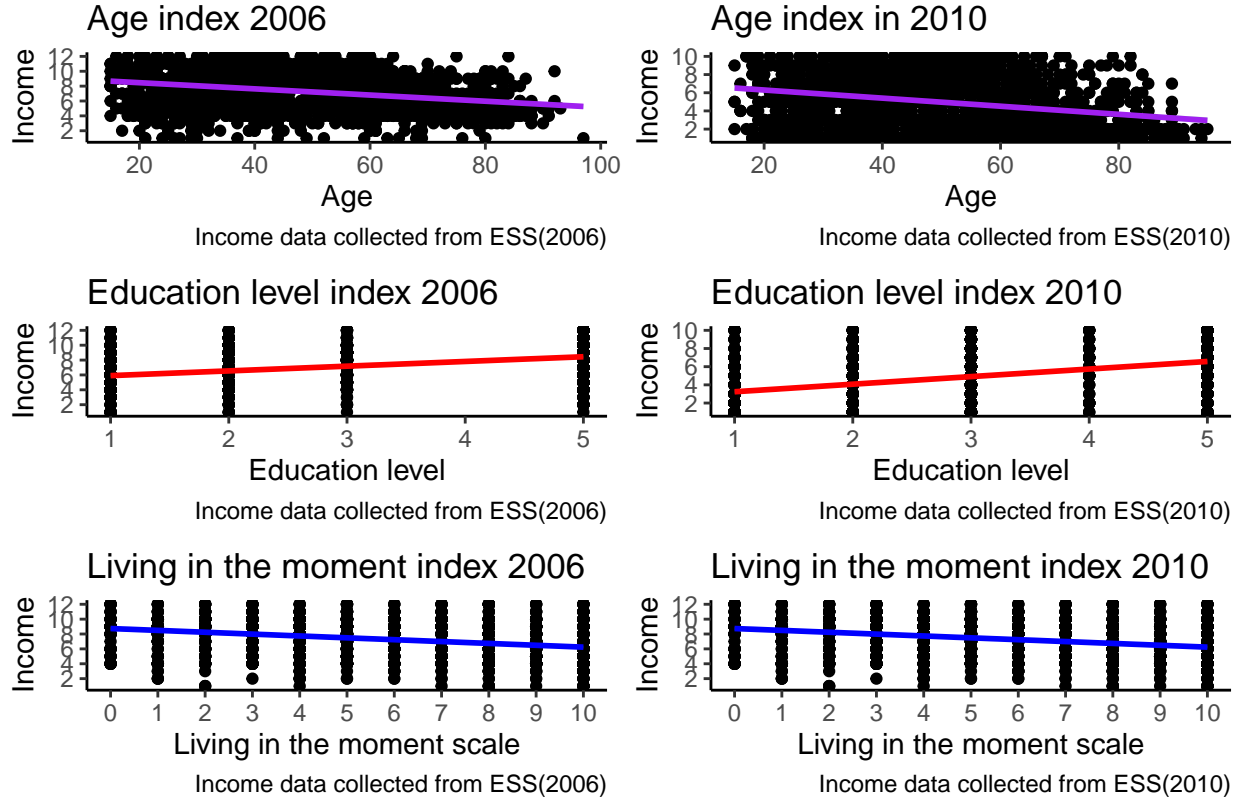
Before starting to fit the linear model, we would like to see the general relationship between the income level and the variables we didn't mention in the data part. We plot scatterplots for each variable in both years and fit the trend with a linear line.

For the scatterplot, we can observe there is a negative linear trend between income level and age, a positive linear trend between the income level and highest education, a negative linear trend between income and plan for the future(i.e. if you are more likely to take each day as it comes, you will be more likely to have lower income level)

Now, we start to fit the model to check if the information from the plots is correct.We fit two models using all of the variables we chose for 2006 and 2010. The variables are gender, age, the highest level of education, placement on the left-right scale, worry about work problems when not working (how often), and the plan for the future or take each day as it comes. (note: the 2006 dataset does not include variable worry about work when not working, and the 2010 dataset does not have the variable plan for the future)

Table 1: VIF value from the full linear model of 2006 and 2010

variables of 2006 linear model	vif value	variables of 2010 linear model	vif value
age	1.023343	age	1.053425
education level	1.007733	education level	1.001744
plan level	1.032080	worry level	1.052735
political stand	1.025262	political stand	1.023509
gender	1.019851	gender	1.017284

From the table 1, we can see that all of the variables for the two models have significant p-value, and the

VIF values are relatively small. So, we decided to keep all those variables in the full models.Then, I use the backward selection to help me select the best model. Our full model for 2006 contains five variables: gender, age, the highest level of education, placement on the left-right scale, and the plan for the future or take each day as it comes. The backward selection model removes the placement on the left-right scale from my full model. Thus, the backward selection model for 2006 has four variables; gender, age, the highest level of education, and the plan for the future or take each day as it comes. And those variables are concluded as having the most correlation with income level. Our full model for 2010 contains five variables: gender, age, the highest level of education, placement on the left-right scale, worry about work problems when not working (how often). The backward selection model removes the highest education level from my full model. Thus, the backward selection model for 2010 has four variables; gender, age, worry about work when not working, placement on the left-right scale. And those variables are concluded as having the most correlation with income level.

Final model(2006): $\quad \text{Income} = \hat{\beta}_0 + \hat{\beta}_1 * age + \hat{\beta}_2 * edulvla + \hat{\beta}_3 * plnftr + \hat{\beta}_4 * gndr + \epsilon$

Final model(2010): $\quad \text{Income} = \hat{\beta}_0 + \hat{\beta}_1 * age + \hat{\beta}_2 * wrywprb + \hat{\beta}_3 * lrscale + \hat{\beta}_4 * gndr + \epsilon$

Here is the estimated result:

Table 2: Final linear model of the year 2006 and 2010

variables of 2006 linear model	coefficients	p-value	variables of 2010 linear model	coefficients	p-value
intercept	8.649245	< 2e-16	intercept	8.649245	< 2e-16
age	-0.008536	1.32e-15	age	-0.008536	2.12e-06
education level	-0.550107	0.000869	worry level	-0.550107	< 2e-16
plan level	-0.010612	< 2e-16	political stand	-0.010612	2.82e-06
gender	-0.476784	0.000105	gender	-0.476784	0.000162

Now, we can finally analyze our linear regression model. Let us first look at the betahat_0. This value is the intercept of our linear regression models. It means when the other variable equals 0, Accordingly, our income level will equal 8.64 and 9.55, which is a relatively high-income level. This result is meaningless in an actual situation. Then, we can find out that the coefficients for the age are both negative, which matched the result we found through the scatterplot before we fit the models. The coefficient for the education level is positive, and the coefficient for the plan to the future is negative, which matched our assumption when analyzing the scatterplots. The estimates for gender in both years are negative; this further proves that being a female has a negative impact on the income level. As for the coefficient for worry for work when not working, this result contradicts what we assume. The more you worry about the work problems when not working, the lower your income level you may have. Lastly, for the coefficient of the placement on the left-right scale, it means that the tendency toward the right-wing democracy might result in a lower income level.

Then, let us look at the p-value. In the method part, we have made assumptions for our hypothesis tests. By observing the result, we can find that p-values are smaller than the significant level, 0.05. Therefore, we can conclude that we have evidence to support our null hypothesis; the income level has a relationship with the selected variables in the above final models. And our proof is vital since we have relatively small p-values.

Discussion

The divider of recession

In the data section, we have explored different factors affecting income, such as gender, age, political beliefs, etc... I wish not to focus on these known factors in the discussion section; there is something else we have

noticed when comparing data by year. When we compare our data, 2008, the financial recession and the typical year 2010, We can see that the proportion of high-income and low-income has significantly differed. Many Opportunities have been found during the recession; new industries raise on the economy's ruins. Approximately 9 million American workers lost their jobs during the recession, and the unemployment rate peaked at 10% in late 2009(Andres 2018). The recession made new college grads a decade behind. Graduating into the recession made newly graduates much hard to find an entry-level job, not to say that the salary is much lower than the non-recession period. It also delayed their career advancement, forcing high-educated capital with a mediocre company that does not pay well. The salaries for those graduates tend to fall behind 10-20 years which is equivalent to $60000 to $100000 loss in their life spend.(Andres 2018) Another issue caused by the recession is that the majority of new jobs are contracted and temporary; this will cause a personal financial issue when it comes to an end. Bring up uncertainty to the contracted workers. Thirdly, the price index has been high; this is reflected through the purchasing power. The money today doxes not value as much as before; the economy is advancing, technologies are advancing, yet people cannot enjoy the benefits due to limited purchasing power. The 2018 average hourly wage holds somewhat the same purchasing power as in 1978.(Andres 2018) Lastly, the recession caused a shortage of qualified workers. During the recession period, companies cut back on investing in current employees and hiring a new employees. The cutback on the training creates a discrepancy between the original middle class: those who have received appropriate training and further advancement and those who do not acquire these skills.

Weakness

Our research primarily focuses on the different factors affecting the income, thus interesting factors as explanatory variable and income as the response variable. In our data, we have chosen the region of the European survey to be exclusively England. Observing the available data, we have noticed that gender's effect on income is minimal. The concern regarding data raises, we have noticed that the survey participants are predominantly male. The limitation to the female data might affect our outcome and ruin our expectations. Also, our data and research have been founded on surveys. There exist many disadvantages of survey research. The risk of people being dishonest when answering the survey issue has been commonly known. The level of social desirability bias will affect the data between the participants and the questions presented. We have filtered out many unanswered responses in this data, which greatly limited the amount of data available. Thirdly, there might exist an issue with interpreting the problem. One of our chosen variables happens to be political spectrum-related. It is difficult for some people to understand their belongs on the spectrum and can cause potential problems. Surveys do not convey emotions into numerical data well. For inference, in the European survey, many questions were related to personal questions such as relationship and feel of balance between work and family. The survey does not collect this data since it is designed to be short and concise.

The other weakness I wish to disgust is related top-value. It has become more aware as we enter a higher level of courses. Science as the path of truth and reality has been damaged. Data scientists bent data into their need to reach their desired end. Scientist achievements are founded on manipulating variables in the needed way. Unfortunately, A scientist's success is determined by the paper he/she has published, and the predominant method for verifying their hypotheses is the p-value. Thus, their accomplishment is dependent on the significant level of 0.05. we are all "falling prey to natural human biases that lead us tip the scales and set up studies to produce false-positive results" (Aschwanden 2015) As an advancement of humankind contributing in all fields, more paper is being released as well; however, they are published at a remarkable speed and most heavily relied on the method of p-hacking. Although, we have avoided as much of the technic as possible. There were variables removed due to the insignificance relationship to the income and change in p-value.

Appendix

Survey link: https://forms.gle/PVdbds9bR3DQ265v9

Level of Satisfaction with the current earning

On January 1st, 2022, the Ontario general minimum wage has been raised to $15.00 per hour, a raise from $14.35. Although it is only 65 cents different from before, 8.8 percent of the workers in Ontario rely on minimum wage. How satisfied are you with your current income/earnings?

By proceed with this survey you understands your response to this petit poll will be use to better understand the income with respect to happiness. No information will be public, The survey is voluntary, meaning that if you decide to participate, you may choose not to under any questions without penalties.

For further information regarding privacy or question regard the poll, please contact yu at dy.sun@mail.utoronto.ca.

Thank you!

davidsun01111@gmail.com (not shared) Switch accounts

What is your age

Your answer

what is your gender?

◯ Male

◯ Female

◯ Prefer not to answer

◯ Other:

What is the highest level of education you have completed

○ No certificate, degree, diploma

○ Highschool diploma

○ Apprenticeship or trade certificate

○ College

○ Bachelor's Degree

○ Master Degree

○ Phd

○ Prefer not to answer

○ Other: _____

How do you value your current salaries

	1	2	3	4	5	
unsatisfied	○	○	○	○	○	really satisfied

where do you value yourself at the political spectrum?

	1	2	3	4	5	6	7	
Left	○	○	○	○	○	○	○	Right

What is your source of income

○ Employee

○ Self-Employed

○ Dependent on others

○ Prefer not to say

○ Other: _____

Which of these describes your personal income last year?

○ $0

○ $1 to $999

○ $10000 to $24999

○ $25000 to $49999

○ $50000 to $74999

○ $75000 to $99999

○ $100000 to $149999

○ $150000+

○ Prefer not to answer

○ Other: _____

How often does your job prevent you from your family?

	1	2	3	4	5	
Not Often	○	○	○	○	○	Often

How often do you concern regarding work outside of work?

	1	2	3	4	5	
Not Often	○	○	○	○	○	Often

Do you have a clear vision of future?

	1	2	3	4	5	
Not Often	○	○	○	○	○	Often

Do you have any sources other than your base salary

◯ Yes

◯ No

◯ Prefer not to say

did you receive any social security benefits in the last year

◯ Yes

◯ No

◯ Pref not to say

Submit   Clear form

Level of Satisfaction with the current earning

Thank you for your participation in this survey1

Submit another response

Google Forms

References

Andres, Tommy. 2018. *Divided Decade: How the Financial Crisis Changed Jobs. Marketplace.* https://www.marketplace.org/2018/12/19/what-we-learned-jobs/.

Aschwanden, Christie. 2015. *Science Isn't Broken. FiveThirtyEight.* FiveThirtyEight. https://fivethirtyeight.com/features/science-isnt-broken/.

Auguie, Baptiste. 2015. *gridExtra: Miscellaneous Functions for "Grid" Graphics.* http://CRAN.R-project.org/package=gridExtra.

Bob Rudis ORCID iD [aut, Patrick Kennedy [ctb], cre]. 2020. *Hrbrthemes: Additional Themes, Theme Components and Utilities for 'Ggplot2'.* http://github.com/hrbrmstr/hrbrthemes.

Braglia, Luca. 2016. *RStata: A Bit of Glue Between r and Stata.*

CFI, Education inc. 2022. *Multiple Linear Regression. Corporate Finance Institute.* https://corporatefinanceinstitute.com/resources/knowledge/other/multiple-linear-regression/.

Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression.* Third. Thousand Oaks CA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Heyndyk, Rachel Muller. 2019. *Transgender Employees Face 14. HR Magazine.* https://www.hrmagazine.co.uk/content/news/transgender-employees-face-14-income-gap.

murrell, paul. 2005. *Grid-Package.*

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files.* https://CRAN.R-project.org/package=here.

North, Rebecca J, Charles J Holahan, Rudolf H Moos, and Ruth C Cronkite. 2008. *Family Support, Family Income, and Happiness: A 10-Year Perspective.* https://doi.org/10.1037/0893-3200.22.3.475.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Research Data, Norwegain Centre for. 2002. *European Social Survey(ESS). European Social Survey Website.* https://www.europeansocialsurvey.org/about/.

Stephanie. 2020. *Variance Inflation Factor. Statistics How To.* https://www.statisticshowto.com/variance-inflation-factor/.

Uni, Yale. n.d. *Linear Regression.* http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2022. *Readr: Read Rectangular Text Data.*

Wickham, Hadley, and Evan Miller. 2022. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files.*

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://yihui.org/knitr/.

Zhu, Hao. 2020. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://CRAN.R-project.org/package=kableExtra.