Erasmus University Rotterdam

Erasmus School Of Economics

Bachelor Thesis Economics & Business Economics

Specialization: Financial Economics

Financial literacy and stock market participation among young people

Author:
Andrew Milovanov
Student number:
595343

Supervisor:
Dr. D. Karpati
Second assessor:
Dr. H. Zhu

July 1, 2024

Abstract

We have devised a survey to measure financial literacy and study its relationship to stock market participation, as well as, diversification level of portfolio amongst young people. We find that majority of respondents display basic financial and economics knowledge and have an understanding of concepts such as interest compounding, inflation, and the time value of money. Most importantly, we find that financial literacy affects financial decision-making: those with low literacy are much less likely to invest in stocks and are more likely to hold poorly diversified portfolios.

Keywords: financial literacy, diversified portfolio, stock market, decision making

Contents

1 Introduction

For many years Dutch pension system was among the best in the world (Toledo, 2023), however on average Dutch people live longer meaning they need to be paid more than before. At the same time low interest rates in the past years mean that pension funds have to hold higher levels of capital in order to meet their current and future obligations, lowering their interest earnings and resulting in pension reductions (CiteDrive, 2023). Market liberalization and structural reforms to social security and pensions have caused an ongoing shift from employer-sponsored defined benefit (DB) pension plans and are swiftly giving way to private defined contribution (DC) plans, shifting the responsibility for retirement saving and investing from employers to employees. Skills are becoming more critical, leading to divergence in wages between those with a college education or higher and those with lower levels of education. At the same time, financial markets are rapidly changing with developments in technology and more complex financial products. From student loans to mortgages, credit cards, mutual funds, and annuities, the range of financial products people must choose from is very different from what it was in the past. Decisions relating to these financial products have implications for individual well-being. There are concerns that individuals are not saving enough for retirement, are accumulating excessive debt and are not taking advantage of financial innovation (Lusardi and Mitchell, 2007). Lusardi and Mitchell (2011) found that financial illiteracy is widespread and particularly acute among specific groups of the population, such as women, elderly and those with low educational attainment. Stock market participation is much lower among women than men (Bucher-Koenen et al., 2021) and consistent with the sharp differences in literacy between women and men (Lusardi and Mitchell, 2011).

We will analyse whether individuals are well-equipped and possess adequate financial literacy and knowledge to make financial decisions. In this context it is important to look directly at their financial behaviour by examining whether financial literacy has an effect on stock market participation. In addition we will examine whether financial literacy affects portfolio diversification level of individuals.

Firstly, we find that individuals with higher financial literacy levels invest more in stocks and, secondly, those respondents who are more knowledgeable on the topic of finance and economics have a more well-diversified portfolio. This paper is organized as follows: In Section 2 we provide a review of the current literature on financial literacy and stock market participation. Section 3 is divided in two parts data and methodology. In the former we describe the gathered data set and in the later we introduce our measures of financial literacy and describe the problems of measuring literacy. In Section 4 we report the results of our empirical work. Lastly, in Section 5 we conclude our results and provide several extensions including examination of areas for future research.

2 Literature review

There has been little research on the topic that provide information on both financial literacy and variables related to financial decision-making who specifically analyze young people. Previous researchers Lusardi and Mitchell (2011) came up with an indicator for basic financial literacy. Their questions aimed at testing basic understanding of interest compounding and portfolio diversification. Their findings indicate that financial illiteracy is widespread and individuals lack basic economic knowledge, cannot perform simple calculations and do not know the difference between stock and bonds and how inflation works (Lusardi, 2015). Their results are surprising because their questionnaire was very simple and the sample mainly consisted of 50 year old plus respondents. Most of the participants of the same age group already have credit cards, checking accounts and took out several mortgages. Similar results were found by Hilgert et al. (2003) in the research of financial literacy among all age groups including students and adult population. Existing studies have also shown that those who are not financially literate are less

likely to plan for retirement and to accumulate wealth (Lusardi and Mitchell, 2011). Illiterate individuals are more likely to take up high-interest mortgages (Lusardi, 2011) and have problems with debt (Lusardi and Tufano, 2015). Lusardi and Tufano (2015) find that those who are not able to correctly calculate interest rates out of a stream of payments end up borrowing more and accumulating lower amounts of wealth. Calvet et al. (2009) construct an index of financial sophistication for Swedish household dataset and find that poorer, less educated and immigrant households demographic characteristics that are strongly associated with low financial literacy are more likely to make financial mistakes.

In our research we link financial literacy to participation in the stock market. While extensive research on this topic exists, it is still a puzzle why so many households do not hold stocks. Some explanations that have been offered are short sale constraints, income risk, inertia, and departures from expected utility maximization Haliassos and Bertaut (1995). More recent papers have incorporated other reasons, such as trust and culture (Guiso et al., 2008) and the influence of neighbors and peers (Brown et al., 2008). Previous research also found that lack of understanding of economics and finance is a significant deterrent to stock ownership (Van Rooij et al., 2011).

Work of Van Rooij et al. (2011) improves substantially upon these studies by considering more refined indices of financial literacy and financial sophistication that have been explicitly designed for a survey on Dutch households. Van Rooij et al. (2011) test financial literacy in two parts where first covers basic financial questions from the workings of interest rates and interest compounding to the effect of inflation, discounting, and nominal versus real values. The second set of questions aims to measure more advanced financial knowledge and covers topics such as the difference between stocks and bonds, the function of the stock market, the workings of risk diversification, and the relationship between bond prices and interest rates. Van Rooij et al. (2011) put gathered data into financial literacy index which is used to test whether higher literacy results in more stock market participation. In addition the different measures of financial knowledge employed by Van Rooij et al. (2011) all show that lack of literacy prevents individuals from participating in the stock market.

We will use the previous researches as inspiration in order to see how financial literacy effect financial decision specifically amongst young people and students.

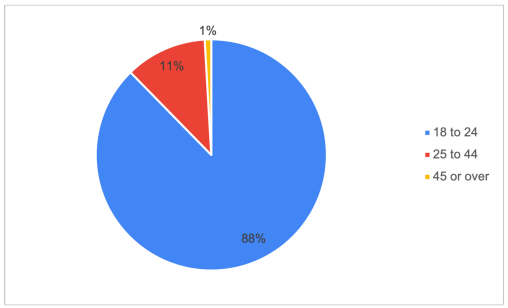
3 Methodology and Data

3.1 Data

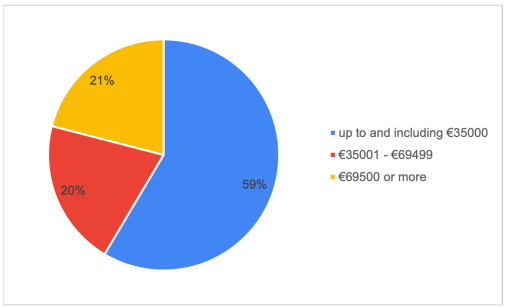
We gather the data through an online questionnaire we created. The share of Dutch households with internet access is among the biggest in Europe with 98.86 percent (Department, 2024). Survey participants have to answer 16 mandatory questions which are divided into three parts (Appendix A). The first part consists of personal questions about the individuals to better understand their backgrounds. The second part tests their basic financial literacy. Finally, we ask whether participants invest their savings into stocks and if their portfolio is diversified (if applicable).

The recruitment of participants was done partially via Erasmus ESE university WhatsApp groups. The audience was asked to fill in the survey honestly without seeking advice from friends or relatives. We also asked to not look up additional information from the internet or books. Usage of a calculator was also prohibited. In addition we went on campus to gather extra responses. Survey participants were chosen at random. We also used Dutch Facebook groups to target specifically local population. The group publications asked to participate in a thesis survey in English.

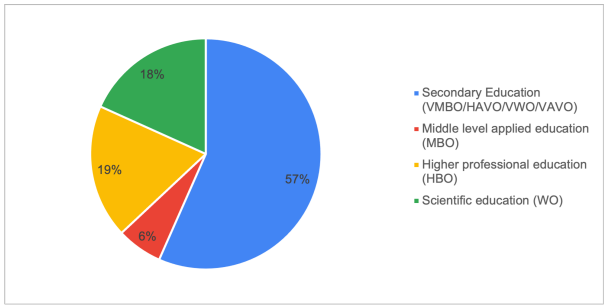
The total number of respondents is 244 people. As the first step we perform a data cleaning process, in order to remove from our sample individuals who presumably answered randomly and did not pay attention to our questions. This is done by analyzing the responses to the last two questions accessing



(a) Age



(b) Income



(c) Education

Figure 1: Demographics distributions

investment behaviour. We make an assumption that if a person answers Question 15 (Do you invest in stocks?) with a "No", they would have to answer the following question with a "Non applicable", "Neither agree nor disagree", "Disagree" or "Strongly disagree". When filtering out the aforementioned combinations, i.e. keeping only the ones answering "No" to (Question 15) and "Agree" or "Strongly agree" to (Question 16), we find that 25 responses do not match our criteria and most likely have been answered randomly without paying attention to the question. We remove these 25 individuals from our sample leaving us with a total of 219 responses.

The gathered data shows a fairly equal distribution between male and female participants: females accounting for 53% and males for 47%. The age varies from 18 to over 45 years old, where 192 respondents or 88% are aged 18 to 24 (Figure 4). 124 respondents have completed at least secondary level of education, while 41 and 40 individuals attained HBO and MBO respectively (Figure 4). 81% of the respondents are single which can be explained by the fact that most of the respondents are still students. Overall the vast majority of the individuals do not own either a house or a car. Only 43 out of 219 individuals own a house and 41 people do own a car. 59% of the respondents make up to €35000 and the other 41% are divided equally between "€35001 to €69499" and "€69500 or over" groups (Figure 4). This can be explained by the fact that most of the responses were gathered on the territory of Woudestein campus of the Erasmus University Rotterdam. From the information mentioned above we can conclude that the sample is representative because majority of respondents fall under the category of young people.

3.2 Methodology

3.2.1 Linear Regression

The linear model has been a mainstay of statistics for the past 30 years and remains one of our most important tools (Hastie et al., 2009). Linear regression is a statistical model which estimates the linear relationship between a scalar response and one or more explanatory variables also known as a dependent and independent variables (Hastie et al., 2009). The case where there are more than one explanatory

variables is called a multiple linear regression model. The linear regression model is represented by the following general formula:

$$f(x) = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \epsilon, \tag{1}$$

where f(x) is the dependent variable, β_0 represent the intercept, β_i is the slope for every given variable i, X_i is the independent variable also known as control variable and, lastly, ϵ represents the error term. The error term, also called the noise, is a variable that captures all other factors which influence the dependent variable y that can not be explained by X_i. To demonstrate linear regression in a simple and easy to understand fashion we build an example regression model (Figure 2). Figure 2 consists of 50 observations and shows the speed of cars measured in miles per hour and the distances taken for them to stop in feet. The scatter plot (Figure 2) suggests a linearly increasing relationship between the dependent variable "distance" and the independent variable "speed".

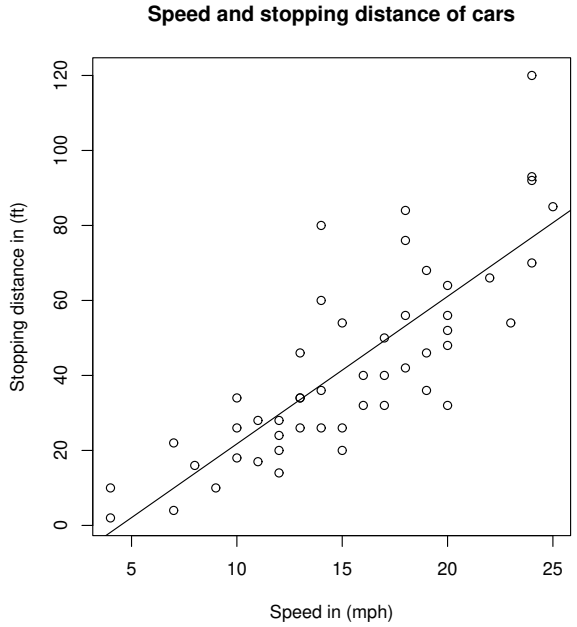


Figure 2: Example linear regression model

We use the same principle as mentioned above to to test how well the participant's portfolios are diversified. As a dependent variable we use respondent's stock portfolio diversification level, while financial literacy index acts as the independent variable. In addition we add control variables such as education, gender, income, wealth, marital status, household size and age in order to enhance the internal validity. Adding control variables helps establishing a causal effect and eliminates possibility of omitted variable bias. Omitted variable bias occurs when a statistical model fails to include one or more relevant variables and results in inconsistent coefficients making the model unreliable (Nikolopoulou, 2022). The formula for our linear regression looks as follows:

$$\begin{aligned} f(x) = \beta_0 &+ \beta_1 X_{Female} + \beta_2 X_{AgeMid} + \beta_3 X_{AgeOld} + \\ &+ \beta_4 X_{Educ2} + \beta_5 X_{Educ3} + \beta_6 X_{educ4} + \beta_7 X_{MaritalMar} + \beta_8 X_{MaritalDiv} \\ &+ \beta_9 X_{HH2} + \beta_{10} X_{HH3} + \beta_{11} X_{HH4Plus} + \beta_{12} X_{House} + \beta_{13} X_{Car} \\ &+ \beta_{14} X_{Income2} + \beta_{15} X_{Income3} + \beta_{16} X_{Pass} + \beta_{17} X_{Diversified}. \end{aligned}$$

3.2.2 Logistic regression

In statistics the logistic regression is used to predict the probability of an outcome using binary classification (Hastie et al., 2009). Logistic regression analyzes the relationship between one or more independent variables and classifies data into discrete classes. The dependent variable can be either "Yes" or "No", "0" or "1", "True" or "False", etc. but instead of giving the exact value as 0 and 1, the logistic regression gives the probabilistic values which lie between 0 and 1 (Hastie et al., 2009). In logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1). Logistic regression finds the best possible fit between the predictor and target variables to predict the probability of the target variable belonging to a labeled class/category (Hastie et al., 2009). In logistic regression a logit transformation is applied on the odds, that is: the probability of success divided by the probability of failure (Hastie et al., 2009). This is also commonly known as the log odds. The logistic regression general function is represented by the following formula:

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{n} \beta_i X_i)}}, \tag{2}$$

where p(x) is the probability of an event (x) occurring, β_0 is the vertical intercept and X_i are the independent variables. The complex concepts are best understood when explained with examples: let's assume that the LR model is tasked to identify the probability of the engine being v-shaped vs inline given it's power (measured in horsepower). There are 32 observations in total and the scatter plot looks as follows: (Figure 3).

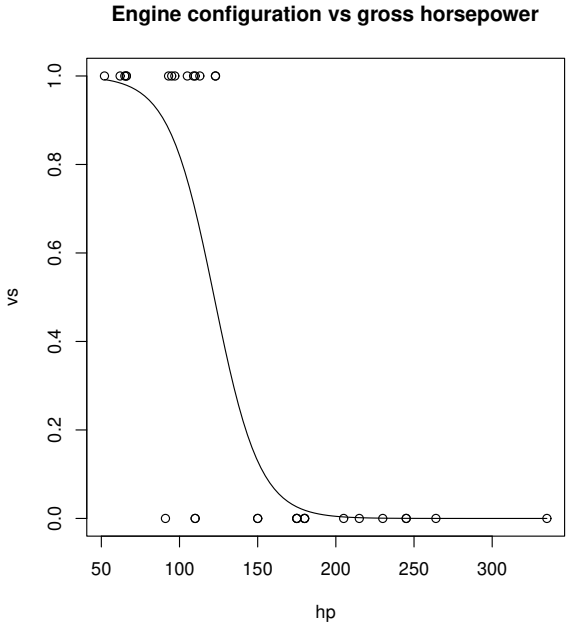


Figure 3: Example logistic regression

We use logistic regression to test the probability that an individual invests in stock, which is a binary question with a "Yes" or "No" answer, given the financial literacy level and control variables such as education, gender, income, wealth, marital status, household size and age. The formula for our logistic regression looks the following way:

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{Female} + \beta_2 X_{AgeMid} + \cdots + \beta_1 6 X_{Pass} + \beta_1 7 X_{Stockholder})}}. \tag{3}$$

3.2.3 Financial literacy index and finalising steps

As mentioned before, we designed a module made of six questions to test the basic financial literacy of individuals. This leads to an improvement in accuracy compared to the previous works by Lusardi and Mitchell (2011) who took into account only three questions. Our questionnaire tests both basic financial knowledge such as interest compounding and inflation and more advanced knowledge on difference between bonds and stocks, as well as understanding the principles of risk diversification. We took the questions from the previous work by Van Rooij et al. (2011) and modified them to make the survey shorter without hindering the overall quality of the assessment of the respondents financial literacy. We took four questions from Van Rooij et al. (2011) basic literacy test to measure the ability to perform simple calculations (Question 9), the understanding of how compound interest works (Question 10) and the effect of inflation (Question 11). We also designed the questions to assess the knowledge of time discounting and whether respondents suffer from money illusion (Question 12). These concepts lie at the foundation of basic financial transactions, financial planning and day-to-day financial decision-making. The extra two questions are taken from the advanced literacy test part by Van Rooij et al. (2011). They measure the key concepts for our study. Question 13 determines whether the respondents understand the difference between stocks and bonds and which one of these is a riskier investment. Question 14 tests if the individuals understand that portfolio diversification means spreading risk by holding several asset classes or indexes. By making a combination of the six questions we form an index.

In order to create a financial literacy index we first make a criteria that if an individual answers four or more out of six questions (67%) correctly, he "passes" the test and is considered "financially literate". This allows us to create a binary variable "financial literacy index" for each individuals where "1" stands for "financially literate" and "0" for the rest.

In order to test the validity of the research we will compare our questionnaire results to a similar study on financial literacy done by LISS in 2011 (Van Overveld, 2011). LISS panel is an independent non-profit research institute at Tilburg University, managed by Centerdata, which does high-quality online research, representative of Dutch population, making it a good benchmark. We use their 201 research consisting of four questions on financial literacy as a benchmark and compare their literacy level outcomes to that of ours. Their questions test individual's understanding of interest compounding and the effect of inflation on purchasing power. In addition their questions test the understanding interest rate has on bond prices and the difference in risk when investing in a single company's stocks compared to a fund. We combine LISS dataset on financial literacy with background variables using their respondents number allowing us to see the respondents age. We then filter the respondents to young people up to and including 24 years old, matching our dataset. After getting a filtered dataset we take the mean of the correct responses to see what the overall literacy rate. The mean correct answers for LISS dataset is 47.23% while for our dataset its 83.50%.

We move on to converting categorical explanatory variables into dummy variables because they cannot be directly entered into regression. First we convert the age column into two separate subgroups named "AgeMid" and "AgeOld" where individuals aged "18 to 24" are assigned a value of 0 while respondents aged "25 to 44" and "45 or older" correspond to a value of 1. We repeat the same process for the rest of the control variables "Gender", "Education", "Marital status", "Household size" and "Income" receiving a filtered data sample. We chose to measure individual's wealth by taking into account whether they own a real-estate property or not in combination with a car, because these are the main storages of household wealth in most countries (OECD, 2022).

After successfully creating a financial literacy index and assigning binary variables to the categorical variables we perform a logistic regression in order to confirm the validity of the control variables (Table 1). We run three logistic regressions with different dependent variables. The first logistic regression

tests the relationship between "Index" the (total number of correct answers) and the control variables. The second logistic regression tests relationship between "Pass4" (four or more correct answers) and the control variables. Lastly, the third logistic regression tests the relationship between "Pass6"(total of six correct answers) and the control variables. The majority of the control variables are significant meaning that they are related with our independent variables of interest and therefore it is reasonable to include them in the main part of our statistical analysis.

Table 1: Validity test

	Dependent variable:		
	Index	Pass4	Pass6
	(1)	(2)	(3)
Female	-0.606^{***}	-0.126^{***}	-0.244^{***}
	(0.169)	(0.047)	(0.067)
AgeMid	0.543^{*}	0.179^{**}	0.097
	(0.315)	(0.089)	(0.124)
AgeOld	-1.061	-0.090	-0.240
	(0.975)	(0.274)	(0.385)
Educ2	0.344	0.033	0.058
	(0.371)	(0.104)	(0.146)
Educ3	-0.436^{*}	-0.136^{**}	-0.121
	(0.235)	(0.066)	(0.093)
Educ4	0.166	0.025	0.173^{*}
	(0.254)	(0.071)	(0.100)
MaritalMar	0.0005	0.021	-0.055
	(0.265)	(0.074)	(0.104)
MaritalDiv	-1.099^{*}	-0.383^{**}	-0.287
	(0.621)	(0.174)	(0.245)
HH2	-0.306	-0.124^{*}	-0.058
	(0.265)	(0.075)	(0.105)
HH3	-0.466^{*}	-0.092	-0.069
	(0.243)	(0.068)	(0.096)
HH4Plus	-0.078	0.004	-0.082
	(0.241)	(0.068)	(0.095)
House	-0.477^{*}	-0.096	-0.185^{*}
	(0.260)	(0.073)	(0.102)
Car	-0.409	-0.083	-0.044
	(0.264)	(0.074)	(0.104)
Income2	-0.277	-0.129^{*}	0.067
	(0.237)	(0.066)	(0.093)
Income3	0.193	-0.027	0.119
	(0.252)	(0.071)	(0.100)
Constant	5.715^{***}	1.029^{***}	0.685^{***}
	(0.205)	(0.058)	(0.081)
Observations	219	219	219
R^2	0.202	0.171	0.143
Adjusted R^2	0.143	0.110	0.080
Residual Std. Error (df = 203)	1.219	0.342	0.481
F Statistic (df = 15; 203)	3.433^{***}	2.794^{***}	2.262^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01

4 Results

Table 2 reports the distributions of correct and wrong answers for financial literacy test in percentages across demographics such as: gender, age, education attained, marital status, household size, income and the wealth indicators (house and car). The results show that women are slightly less literate than men, this can be concluded by the fact that higher percentage of male answer all of the questions correctly. In all of the questions apart from (Question 13) the age group of 25 to 44 years old gives the most correct answers out of the three age groups. In the majority of the questions individuals who have attained middle level of applied education have more correct answers, however, they make only 14 out of 219 respondents. Respondents who are divorced have the least number of correct answers in all of the literacy questions. Household size and incomes subgroups do not have a visible correlation with correct responses. Lastly, respondents who own a car and a house have a slightly bigger percentage rate of correct answers to the literacy test.

The main hypothesis of our research paper is that financial literacy affects stock market participation. Therefore, we test whether more financially literate individuals invest more in stocks. In order to test this we will perform a logistic regression with stock market participation acting as the binary dependent variable and financial literacy being among the independent variables. We also add a "Pass4" variable (setting the passing rate to at least four out of six correct answers) and add the necessary control variables. The results can be seen in column 3 (Table 3). The regression shows that an estimated coefficient of the variable "Female" equals to -1.38 and is significant at a 1% significance level. This means that females are on average approximately 75% less likely to invest in stocks than men. This is in line with previous findings by (Van Rooij et al., 2011) who also found that females invest less in stocks. Jianakoplos and Bernasek (1998) earlier found that women exhibit relatively more risk aversion in financial decision making than single men. This can be the reason why women invest less in stocks as stocks are a risky asset class. Variable "Age" is insignificant so does not seem to exhibit an effect on stock market participation. On the other hand, "Educ2", representing the middle level applied education, shows to be significant at a 5% significance level. The coefficient for "Educ2" is -2.31 meaning the probability that individual invests in stocks is on average approximately 90% smaller compared to secondary education. Some researchers have previously argued that knowledge and cognitive ability may have an effect on preferences such as risk aversion and the rate of time preference Dohmen et al. (2010). Higher professional education and scientific education - "Educ3" and "Educ4" respectively - are insignificant and can not be interpreted. In addition neither of the marital statuses are significant and do not have an effect on the dependent variable. Control variable "Car" is significant at 10% significance level and has a coefficient of 0.96 which on average increases the probability of individual investing in stocks by approximately 162%. We can assume that this is due to the fact that cars are luxury goods and people who own a car most likely already have some sort of savings which they can allocate for investing in stocks. Income does not have a significant effect for our sample. The coefficient for the main independent variable of interest - "Pass4" equals -0.15 and is insignificant.

		Q9		Q10		Q11		Q12		Q13		Q14		
		0	1	0	1	0	1	0	1	0	1	0	1	Total
Gender	Female	10.3	89.7	25.9	74.1	29.3	70.7	24.1	75.9	22.4	77.6	16.4	83.6	116
	Male	3.9	96.1	15.5	84.5	11.7	88.3	12.6	87.4	16.5	83.5	5.8	94.2	103
Age	18 to 24	7.8	92.2	21.9	78.1	22.4	77.6	18.8	81.2	18.2	81.8	11.5	88.5	192
	25 to 44	0.0	100.0	12.0	88.0	8.0	92.0	12.0	88.0	32.0	68.0	8.0	92.0	25
	45 or over	50.0	50.0	50.0	50.0	50.0	50.0	100.0	0.0	0.0	100.0	50.0	50.0	2
Educ	Higher professional education (HBO)	9.8	90.2	36.6	63.4	39.0	61.0	19.5	80.5	31.7	68.3	12.2	87.8	41
	Middle level applied education (MBO)	7.1	92.9	0.0	100.0	14.3	85.7	28.6	71.4	14.3	85.7	7.1	92.9	14
	Scientific education (WO)	10.0	90.0	20.0	80.0	15.0	85.0	10.0	90.0	15.0	85.0	7.5	92.5	40
	Secondary Education (VMBO/HAVO/VWO/VAVO)	5.6	94.4	18.5	81.5	17.7	82.3	20.2	79.8	17.7	82.3	12.9	87.1	124
Marital	Divorced	20.0	80.0	40.0	60.0	60.0	40.0	60.0	40.0	20.0	80.0	40.0	60.0	5
	Married or domestic partnership	2.9	97.1	20.0	80.0	22.9	77.1	17.1	82.9	31.4	68.6	14.3	85.7	35
	Single	7.8	92.2	20.7	79.3	19.6	80.4	17.9	82.1	17.3	82.7	10.1	89.9	179
HH	1	3.6	96.4	19.6	80.4	14.3	85.7	12.5	87.5	23.2	76.8	5.4	94.6	56
	2	8.3	91.7	18.8	81.2	25.0	75.0	18.8	81.2	18.8	81.2	14.6	85.4	48
	3	12.2	87.8	24.5	75.5	30.6	69.4	18.4	81.6	26.5	73.5	18.4	81.6	49
	4 or more	6.1	93.9	21.2	78.8	16.7	83.3	24.2	75.8	12.1	87.9	9.1	90.9	66
Income	€35001 - €69499	11.1	88.9	26.7	73.3	22.2	77.8	28.9	71.1	13.3	86.7	13.3	86.7	45
	€69500 or more	4.4	95.6	24.4	75.6	17.8	82.2	17.8	82.2	24.4	75.6	15.6	84.4	45
	up to and including €35000	7.0	93.0	17.8	82.2	21.7	78.3	15.5	84.5	20.2	79.8	9.3	90.7	129
House	0	6.2	93.8	17.0	83.0	19.3	80.7	15.3	84.7	18.2	81.8	9.7	90.3	176
	1	11.6	88.4	37.2	62.8	27.9	72.1	32.6	67.4	25.6	74.4	18.6	81.4	43
Car	0	5.6	94.4	17.4	82.6	19.1	80.9	18.0	82.0	18.5	81.5	9.0	91.0	178
	1	14.6	85.4	36.6	63.4	29.3	70.7	22.0	78.0	24.4	75.6	22.0	78.0	41

Table 2: Literacy across demographics

We make another assumption that, because the financial literacy test is basic, those individuals who are educated enough on the topic and invest in stocks will answer all of the question correctly. We perform a similar logistic regression with a new financial literacy passing rate of 100% as our new independent variable - "Pass6". We add the same control variables as before with stock market participation staying as the dependent variable. The new regression results (Table 3, column 4) show that variable "Female" coefficient equals to -1.15 and it is significant at a 1% significance level. From this coefficient we can infer that on average female are less likely to invest in stocks with a probability of approximately 68.27%. We once again find that variable "Educ2" is significant at a 1% significance level and has a coefficient of -2.45 which means that individuals who attained middle level applied education on average are approximately 91.39% less likely to invest in stock than those with secondary level of education. Variable "MaritalDiv" has a coefficient of 2.34 and is significant at 10% significance level meaning respondents who are divorced on average have approximately a 939% higher probability of investing in stocks compared to single individuals. The "Car" variable is once again significant at 5% significance and has a coefficient of 1.10, implying that on average individuals who posses a car are approximately 200% more likely to participate in the stock market. Finally and most importantly variable "Pass6" representing the financial literacy level of the respondents is significant at 1% significance level. The coefficient for variable "Pass6" equals 1.04 which means that individuals who gave the correct answers to all of the questions are on average approximately 182% more likely to invest in stocks. This supports the previous findings of (Van Rooij et al., 2011) who found that different measures of financial knowledge show that lack of literacy prevents households from participating in the stock market as opposed to those who display higher literacy.

The second hypothesis of our research is that financial literacy level affects the diversification level of the individual's portfolio. In order to test this hypothesis we perform an OLS regression where diversification level acts as the dependent variable and financial literacy represented by "Pass4" is the independent variable. We include a wide set of variables that were listed previously. The results can be seen in the first column of (Table 3). The variable "Female" is significant at a 1% significance level with a coefficient of -1.10, meaning that if the respondent is a female their portfolio is less diversified by on average approximately 1.10 units. Age variables are insignificant and can not be interpreted. The variable "Educ2" is significant at a 5% level and its coefficient equates to -1.29 units, meaning that if the respondent has attained an MBO their diversification level of portfolio decreases on average by approximately 1.29 units. "MaritalDiv" has a coefficient of 1.58 implying on average an equivalent approximate unit increase in diversification of a portfolio in case the individual is divorced. The p-values of the variables representing household sizes and house ownership exceed the significance threshold, implying that the sample data does not provide enough evidence of their relevance. At the same time "Car" variable is significant at a 5% level, meaning that car ownership will on average result in an approximately 0.91 unit increase in portfolio diversification. Lastly, neither income nor financial literacy level variables are significant, hence we do not make any conclusions based on their coefficients' values.

The index of literacy might have been measured with a substantial error. We then make another assumption: financially literate individuals are, most likely, able to answer all of the questions correctly. Hence, we create an identical OLS regression as before but change the passing rate to 100%. This way we make an assumption that the respondents who invest into stocks and are financially literate will answer all of the basic literacy questions correctly. The results can be seen in the second column of (Table 3). The constant is still significant at a 1% level but the new coefficient is 1.58 compared to previous 2.24. The variable "Female" is significant at a 1% significance level and indicates that if the respondent is female the diversification level of their portfolio decreases on average by approximately 0.90 units. For the variables "AgeMid" and "AgeOld" p-values are above the significance threshold of 10% which means that there unlikely is a relationship between the changes in the independent variable and shifts in these control variables. "Educ2" has a coefficient of -1.34 and is statistically significant. A unit

change in "Educ2" results in on average an approximately 1.34 units decrease in diversification of the portfolio. "Educ3" and "Educ4" are insignificant and, therefore, we can not make any inferences about their effect on portfolio diversification levels. Variable "MaritalDiv" is significant at a 5% significance level. As a result we can state that on average it has an effect of approximately 1.85 units of change of the dependent variable per unit of its own change. Variables representing household sizes and the variable "House" are insignificant so their regression coefficients can not be interpreted. When an individual has a "Car" the diversification level of their portfolio on average increases by 0.96 units. Finally, at a passing rate of 100% the variable "Pass6" is significant at 1% significance level and on average results in an approximately 0.73 unit increase in the portfolio diversification. This result supports the findings of Van Rooij et al. (2011) that unsophisticated investors are unlikely to make wise choices when investing in financial markets, while highly educated investors are more likely to hold well diversified portfolios.

Table 3: OLS and Logit regression results

	Dependent variable:			
	Diversified		Stockholder	
	OLS		*logistic*	
	(1)	(2)	(3)	(4)
Female	−1.095***	−0.898***	−1.379***	−1.148***
	(0.251)	(0.250)	(0.317)	(0.323)
AgeMid	−0.269	−0.367	−0.508	−0.661
	(0.467)	(0.454)	(0.635)	(0.638)
AgeOld	0.500	0.688	−0.311	0.033
	(1.430)	(1.404)	(2.153)	(2.189)
Educ2	−1.287**	−1.335**	−2.313**	−2.452***
	(0.544)	(0.534)	(0.917)	(0.938)
Educ3	−0.179	−0.069	−0.308	−0.142
	(0.349)	(0.340)	(0.430)	(0.437)
Educ4	−0.316	−0.445	−0.337	−0.524
	(0.372)	(0.368)	(0.468)	(0.480)
MaritalMar	0.440	0.477	0.620	0.713
	(0.388)	(0.381)	(0.486)	(0.494)
MaritalDiv	1.576*	1.845**	2.093	2.341*
	(0.920)	(0.896)	(1.384)	(1.332)
HH2	0.172	0.234	0.121	0.194
	(0.392)	(0.382)	(0.489)	(0.494)
HH3	−0.325	−0.260	−0.214	−0.153
	(0.358)	(0.350)	(0.450)	(0.460)
HH4Plus	−0.100	−0.041	0.047	0.117
	(0.354)	(0.348)	(0.438)	(0.455)
House	−0.003	0.146	−0.053	0.134
	(0.382)	(0.376)	(0.473)	(0.490)
Car	0.910**	0.955**	0.962*	1.098**
	(0.388)	(0.379)	(0.491)	(0.509)
Income2	−0.364	−0.393	−0.319	−0.404
	(0.350)	(0.341)	(0.453)	(0.462)
Income3	0.226	0.144	0.130	0.030
	(0.370)	(0.364)	(0.450)	(0.457)
Pass4	−0.158		−0.150	
	(0.366)		(0.474)	
Pass6		0.725***		1.037***
		(0.256)		(0.338)
Constant	2.238***	1.579***	0.514	−0.346
	(0.482)	(0.343)	(0.607)	(0.438)
Observations	219	219	219	219
R^2	0.195	0.225		
Adjusted R^2	0.131	0.164		
Log Likelihood			-128.400	-123.562
Akaike Inf. Crit.			290.801	281.123
Residual Std. Error (df = 202)	1.786	1.753		
F Statistic (df = 16; 202)	3.058***	3.665***		
Note:				*p<0.1; **p<0.05; ***p<0.01

5 Conclusion

In this paper, we show that lack of understanding of economics and finance is a significant deterrent to stock ownership. The basic measurement of financial knowledge shows that lack of knowledge prevents individuals from participating in the stock market. In addition we find that less literate individuals are more likely to hold poorly-diversified portfolios. We had two hypotheses. The first one states that financial literacy affects stock market participation and the other that financial literacy affects diversification level of a portfolio. The results show that hypothesis number one was sound as it is confirmed by our findings, namely the coefficient of the independent variable "Pass6" is significant at 1% significance level. Hypothesis number two is also confirmed because the independent variable "Pass6" is significant at 1% significance level. Our results are supported by the previous findings by (Calvet et al., 2009) who conclude that unsophisticated investors are unlikely to make wise choices when investing in financial markets. For example, low education investors are likely to hold poorly diversified portfolios. As more workers transition to a system where they have to decide how much to save for retirement and how to invest their retirement wealth, it is important to consider ways to enhance their financial sophistication or to guide them in their financial decisions via, for example, financial education programs.

References

Brown, J. R., Ivković, Z., Smith, P. A., and Weisbenner, S. (2008). Neighbors matter: Causal community effects and stock market participation. *The Journal of Finance*, 63(3):1509–1531.

Bucher-Koenen, T., Alessie, R. J., Lusardi, A., and Van Rooij, M. (2021). Fearless woman: Financial literacy and stock market participation. Technical report, National Bureau of Economic Research.

Calvet, L. E., Campbell, J. Y., and Sodini, P. (2009). Measuring the financial sophistication of households. *American Economic Review*, 99(2):393–398.

CiteDrive, I. (2023). The dutch pension system: highlights and characteristics. https://www.pensioenfederatie.nl/website/the-dutch-pension-system-highlights-and-characteristics [Accessed: (01.01.2024)].

Department, S. R. (2024). Share of households with internet access in the netherlands from 2010 to 2023. https://www.statista.com/statistics/377747/household-internet-access-in-the-netherlands/.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100(3):1238–1260.

Guiso, L., Sapienza, P., and Zingales, L. (2008). Trusting the stock market. *the Journal of Finance*, 63(6):2557–2600.

Haliassos, M. and Bertaut, C. C. (1995). Why do so few hold stocks? *the economic Journal*, 105(432):1110–1129.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Hilgert, M. A., Hogarth, J. M., and Beverly, S. G. (2003). Household financial management: The connection between knowledge and behavior. *Fed. Res. Bull.*, 89:309.

Jianakoplos, N. A. and Bernasek, A. (1998). Are women more risk averse? *Economic inquiry*, 36(4):620–630.

Lusardi, A. (2011). Americans' financial capability. Technical report, National Bureau of Economic Research.

Lusardi, A. (2015). Financial literacy: Do people know the abcs of finance? *Public understanding of science*, 24(3):260–271.

Lusardi, A. and Mitchell, O. S. (2007). Baby boomer retirement security: The roles of planning, financial literacy, and housing wealth. *Journal of monetary Economics*, 54(1):205–224.

Lusardi, A. and Mitchell, O. S. (2011). Financial literacy around the world: an overview. *Journal of pension economics & finance*, 10(4):497–508.

Lusardi, A. and Tufano, P. (2015). Debt literacy, financial experiences, and overindebtedness. *Journal of pension economics & finance*, 14(4):332–368.

Nikolopoulou, K. (2022). What is omitted variable bias? — definition examples. https://www.scribbr.com/research-bias/omitted-variable-bias/.

OECD (2022). Housing is a central component of households' wealth and debt. https://www.oecd.org/housing/policy-toolkit/data-dashboard/wealth-distribution/.

Toledo, M. (2023). Netherlands ranks 1st among 47 global pension systems. https://www.ai-cio.com/news/netherlands-ranks-1st-among-47-global-pension-systems/ [Accessed: (01.01.2024)].

Van Overveld, M. (2011). Financial literacy. (3).

Van Rooij, M., Lusardi, A., and Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial economics*, 101(2):449–472.

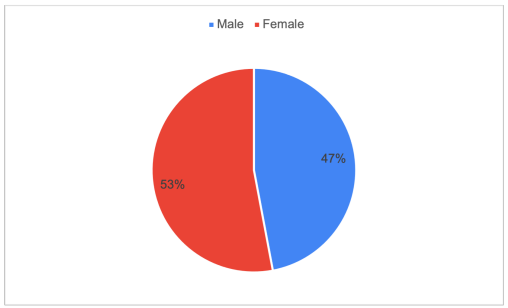
Appendix A Questionnaire

1. What is your gender?
    - Male
    - Female
2. What is your age?
    - 18 to 24
    - 25 to 44
    - 45 or over
3. What is your highest level of education completed
    - Secondary Education(VMBO/HAVO/VWO/VAVO)
    - Middle level applied education(MBO)
    - Higher professional education(HBO)
    - Scientific education(WO)
4. What is your marital status?
    - Single
    - Married or domestic partnership
    - Divorced
5. How many people are in your household including yourself?
    - 1
    - 2
    - 3
    - 4 or more
6. Do you own a house/apartment?
    - Yes
    - No
7. Do you own a car?
    - Yes
    - No
8. What is the annual gross income of your household?
    - up to and including €35000
    - €35001 - €69499
    - €69500 or more
9. Suppose you had 100€ in a savings account and the interest rate was 2% per year. After 5 years, how much would you have in the account if you left the money to grow?

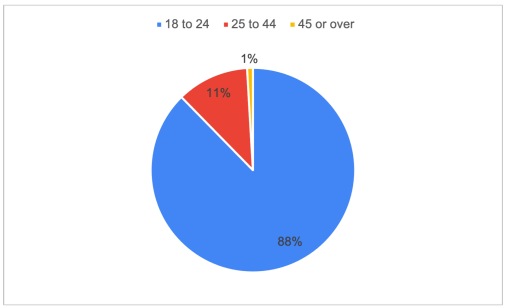
- More than 102€
- Exactly 102€
- Less than 102€

10. Suppose you had 100€ in a savings account, the interest rate is 20% per year and you never withdraw any money. How much would you have in the account after 5 years?
    - More than 200€
    - Exactly 200€
    - Less than 200€
11. Imagine that the interest rate on your savings account was 1% per year and the inflation was 2% per year. After 1 year, how much would you be able to buy with the money in this account?
    - More than today
    - Exactly the same
    - Less than today
12. Suppose that in the year 2025 your income has doubled and the prices of all goods have doubled too. In 2025, how much will you be able to buy with your income?
    - More than today
    - The same
    - Less than today
13. Stocks are normally riskier than bonds. Is this statement true or false?
    - True
    - False
14. When an investor spreads his money among different assets what happens to the risk of losing money?
    - Risk increases
    - Risk decreases
    - Risk stays the same
15. Do you invest in stocks?
    - Yes
    - No
16. Is your portfolio well diversified?
    - Strongly agree
    - Agree
    - Neither agree nor disagree
    - Disagree
    - Strongly disagree
    - Non applicable

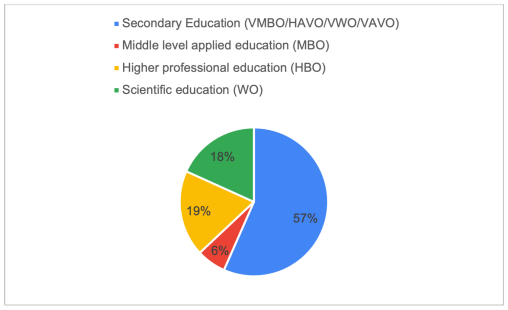
Appendix B Descriptive statistics



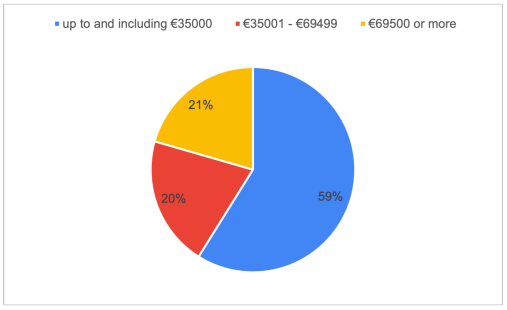
(a) Gender distribution



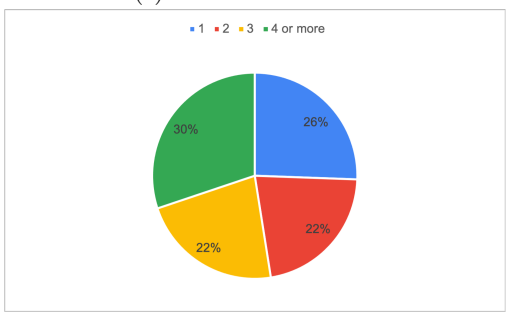
(b) Age distribution



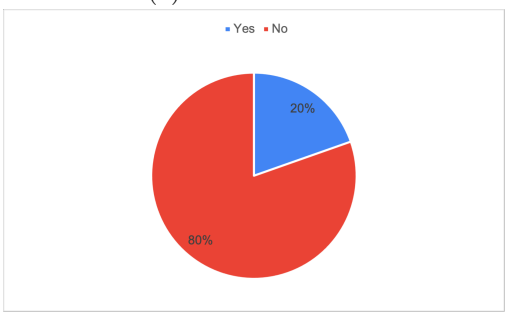
(c) Education distribution



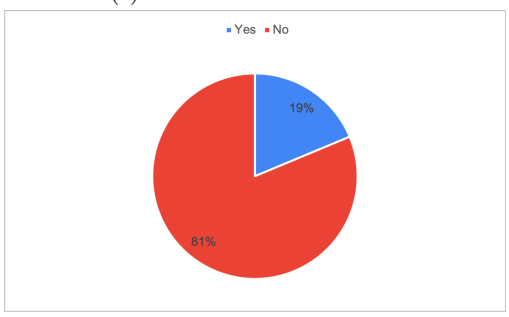
(d) Income distribution



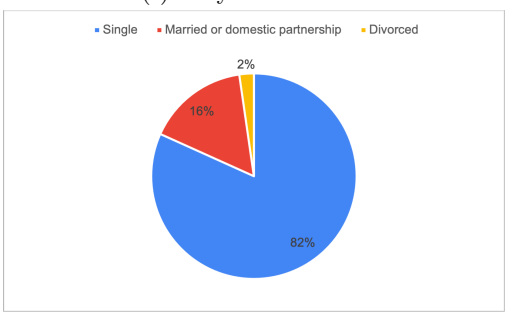
(e) Household size distribution



(f) Do you own a house?



(g) Do you own a car?



(h) Marital Status distribution

Figure 4: Descriptive statistics

Appendix C R code

```r
#Part 1 ####
#Get data
df_raw <- fread('ThesisData.csv')

#Convert answers and create index
df_inter <- df_raw %>%
  mutate(Q9 = ifelse(Q9 == "More than 102€", 1, 0),
         Q10 = ifelse(Q10 == "More than 200€", 1, 0),
         Q11 = ifelse(Q11 == "Less than today", 1, 0),
         Q12 = ifelse(Q12 == "The same", 1, 0),
         Q13 = ifelse(Q13 == TRUE, 1, 0),
         Q14 = ifelse(Q14 == "Risk decreases", 1, 0)) %>%
  mutate(Index = Q9 + Q10 + Q11 + Q12 + Q13 + Q14) %>%
  relocate(Index, .after = Q14)
#Creating mean correct answers
mean_Andrew <- mean(colMeans(df_inter[, 9:14], na.rm = TRUE))

#Create literacy level Pass or Fail
df_inter <- df_inter %>%
  mutate(Pass4 = ifelse(Index >= 4, 1, 0),
         Pass6 = ifelse(Index >= 6, 1, 0)) %>%
  relocate(c(Pass4, Pass6), .after = Index)

#Convert explanatory variables
#Convert age categories into dummy variables where 0-youngest group
df_inter <- df_inter %>%
  mutate(AgeMid = ifelse(Age == "25 to 44", 1, 0),
         AgeOld = ifelse(Age == "45 or over", 1, 0)) %>%
  relocate(c(AgeMid, AgeOld), .after = Age) %>%
  mutate(Educ2 = ifelse(Educ == "Middle level applied education (MBO)", 1, 0),
         Educ3 = ifelse(Educ == "Higher professional education (HBO)", 1, 0),
         Educ4 = ifelse(Educ == "Scientific education (WO)", 1, 0)) %>%
  relocate(c(Educ2, Educ3, Educ4), .after = Educ) %>%
  mutate(MaritalMar = ifelse(Marital == "Married or domestic partnership", 1, 0),
         MaritalDiv = ifelse(Marital == "Divorced", 1 , 0)) %>%
  relocate(c(MaritalMar, MaritalDiv), .after = Marital) %>%
  mutate(Income2 = ifelse(Income == "€35001 - €69499", 1 ,0),
         Income3 = ifelse(Income == "€69500 or more", 1, 0)) %>%
  relocate(c(Income2, Income3), .after = Income) %>%
  mutate(Female = ifelse(Gender == "Female", 1 , 0)) %>%
  relocate(c(Female), .after = Gender) %>%
  mutate(HH2 = ifelse(HH == "2", 1, 0),
         HH3 = ifelse(HH == "3", 1, 0),
         HH4Plus = ifelse(HH == "4 or more", 1, 0)) %>%
  relocate(c(HH2, HH3, HH4Plus), .after = HH) %>%
```

```r
  mutate(House = ifelse(House == 'Yes', 1, 0),
         Car = ifelse(Car == 'Yes', 1, 0))

#Convert dependent variables
df_inter <- df_inter %>%
  mutate(Stockholder = ifelse(Stockholder == "Yes", 1, 0),
         Diversified = case_when(Diversified == "Strongly agree" ~ 5,
                                 Diversified == "Agree" ~ 4,
                                 Diversified == "Neither agree nor disagree" ~ 3,
                                 Diversified == "Disagree" ~ 2,
                                 Diversified == "Strongly disagree" ~ 1,
                                 TRUE ~ 0)) %>%
  mutate(Diversified = Diversified * Stockholder)

#Select columns for regression
df <- df_inter %>%
  select(Female, AgeMid, AgeOld, Educ2, Educ3,
         Educ4, MaritalMar, MaritalDiv, HH2, HH3, HH4Plus,
         House, Car, Income2, Income3, Pass4, Pass6, Stockholder, Diversified, Index)

#Test internal validity by making linear regression where independent variable is
↪  financial literacy index
validity_literacy <- lm(Index ~ ., family = binominal(link='logit'), data = df[,
↪  -c('Stockholder', 'Diversified', 'Pass6','Pass4')])
validity_pass4 <- lm(Pass4 ~ ., data = df[, -c('Stockholder', 'Diversified', 'Pass6',
↪  'Index')])
validity_pass6 <- lm(Pass6 ~ ., data = df[, -c('Stockholder', 'Diversified', 'Pass4',
↪  'Index')])

stargazer(validity_literacy, validity_pass4, validity_pass6, title = 'Validity test',
↪  align=TRUE, no.space= TRUE)

#Linear regression
#Pass4
linear_pass4 <- lm(Diversified ~ ., data = df[, -c('Stockholder', 'Pass6')])
summary(linear_pass4)
#Pass6
linear_pass6 <- lm(Diversified ~ ., data = df[, -c('Stockholder', 'Pass4')])
summary(linear_pass6)



#Logistic regression
#Pass4
logistic_pass4 <- glm(Stockholder ~., family = binomial(link='logit'), data = df[,
↪  -c('Diversified', 'Pass6')])
```

```r
summary(logistic_pass4)
#Pass6
logistic_pass6 <- glm(Stockholder ~., family = binomial(link='logit'), data = df[, -c('Diversified', 'Pass4')])
summary(logistic_pass6)

stargazer(linear_pass4, linear_pass6, logistic_pass4, logistic_pass6, title = "Results", align=TRUE, no.space = TRUE)

#Regression
fit <- lm(dist ~ speed, data = cars)
#Plot
plot(cars, xlab = "Speed in (mph)", ylab = "Stopping distance in (ft)", main= "Speed and stopping distance of cars")
abline(fit)

#Logistic regression on mtcars
#Model
fit = glm(vs ~ hp, data = mtcars, family = binomial)
#100 x values between min and max mtcars£hp
sigmoid <- data.frame(hp = seq(min(mtcars$hp), max(mtcars$hp), len = 100))
#y values based on the estimated model
sigmoid$vs = predict(fit, newdata = sigmoid, type="response")
#Plot
plot(vs ~ hp, data = mtcars, main = "Engine configuration vs gross horsepower")
lines(sigmoid)

#Part 2 ####
#Get data
liss_background <- read_dta('background.dta')
liss_literacy <- read_dta('Liss_literacy_data.dta')

#Keep ID and Age
age <- liss_background %>%
  select(nomem_encr, leeftijd)

#Join literacy and age %>% filter young
liss_literacy <- liss_literacy %>%
  left_join(age, by = 'nomem_encr') %>%
  filter(leeftijd <= 24)

#Recode answers
liss_literacy <- liss_literacy %>%
  mutate(ew11a002 = ifelse(ew11a002 == 1, 1, 0),
         ew11a003 = ifelse(ew11a003 == 3, 1, 0),
         ew11a004 = ifelse(ew11a004 == 2, 1, 0),
         ew11a005 = ifelse(ew11a005 == 2, 1, 0))
```

```r

#Percentage of correct answers
mean_Liss <- mean(colMeans(liss_literacy[, 5:8], na.rm = TRUE))
```