# Income inequality: What is the correlation between education and income of men and women in Finland.

SL AA Mathematics Exploration

Table of Contents

## 1. Introduction:

Gender equality remains a rather acute issue for the whole modern world and persists in many areas of people's lives. Despite the fact that some progress has been made over many decades, women still earn between 50% and 96% of men's wages in the labor market around the world[1].

Occupational segregation based on gender, as a result of which women are concentrated in low-paid sectors of the economy, is an important source of the gender wage gap worldwide. Low wages directly affect retirement savings, therefore, there is also a wealth gap in this. According to the ILO report on wages for 2016, in most countries of the world, the gender gap is high for the most paid workers, who, accordingly, occupy high positions. For CIOs, this gap reaches a level of about 50%. [2]

There is discrimination in the vertical, the higher the position, the fewer women we see there. A woman needs to be several times stronger by will, by pressure to break into the top. There are not many women who head the largest companies. And even in the positions of company managers, women will lose out in salary. What is the situation with gender pay inequality in Finland? Finland became the first country in the world to give women full political rights, such as the right to vote and the right to stand for election. There are many women in leadership positions in the Finnish Government. The government is also headed by a woman, the youngest prime minister in the world, Sanna Marin.

I assume that Finland and other countries are experiencing some distortions in the field of remuneration. I set a task to determine whether there is a gap in the income of men and women with equal education. My interest is aroused by the relevance of this topic, as I

[1] https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_008091/lang--en/index.htm

[2] https://www.ilo.org/global/research/global-reports/global-wage-report/2016/WCMS_537982/lang--en/index.htm

believe that one of the main tasks of the modern world is to overcome gender inequality and ensure equal access to resources and prospects for everyone.

I devoted my essay to this very issue of concern to me and chose Finland to study, a country in which I will not only study, but also work in the future.

## 2. Methodology:

At the very beginning of the study, data were collected on the average monthly income of Finnish citizens, men and women. To do this, the median was used as a sample. The median was a very convenient solution because it is the median value of a set of numbers. The median is more accurate than the average because it better reflects the incomes of men and women. The income data of men and women in Finland were obtained from Statistics Finland, so the data are presented in euros, this is the local volute[3].

In order to use further data in statistics, I had to translate qualitative data in the form of names of degrees of education into numerical data. Degrees were formed in the number of years how many people studied to get a particular degree. Thus, people who graduated from high school exactly 0 years old, bachelor's degree - 3 years, master's degree - 6 years, doctorate - 11 years. In order to designate people who did not graduate from high school, - 1 number of years was used. This was done because after a person graduates from high school, the number of years of study before entering university is not taken into account in obtaining higher education.

Then, in order to establish the correlation and its strength between the statistical data on the average monthly income of men and women in Finland, I performed the Pearson correlation test. The Pearson correlation coefficient (r) measures the presence of statistical evidence of a

[3] https://www.stat.fi/til/pra/2020/pra_2020_2021-10-11_tau_001_en.html

linear relationship between pairs of variables or within sets of variables and between them[4]. The calculations were based on a formula from the article: “Thirteen Ways to Look at the Correlation Coefficient”[5].

$$r = \frac{\sum(X_i - \overline{X}) \times (Y_i - \overline{Y})}{\sqrt{\Sigma(X_i - \overline{X})^2 \times \Sigma(Y_i - \overline{Y})^2}}$$

Where,

r = Pearson correlation coefficient

x- values from the data set of years spent in education after high school graduation

y- values from the data set of median monthly income

To build linear regressions after analyzing the relevance of statistical data, I used the online charting program Desmos. However, to make my results more accurate, I constructed regression lines for various polynomials and calculated their coefficient of determination ($R^2$), which evaluates how well the regression model matches the real data. In other words, it is an indicator of the overall accuracy of the model. Regression lines were used because they were better suited for processing speed and differentiation data.

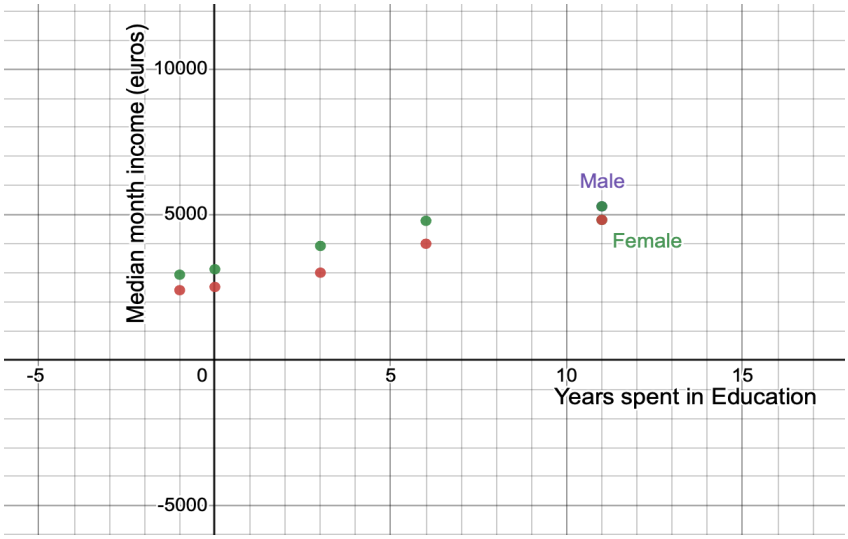
To explore the differences more deeply, I calculated the rate of return using differentiation, which represents the marginal benefits of further education for both genders.

[4] https://libguides.library.kent.edu/spss/pearsoncorr

[5] https://www.jstor.org/stable/2685263

| Level of education (years spent in education) | Finland | |
|---|---|---|
| | Male | Female |
| | Median month income (euros) | |
| -1 | 2933 | 2401 |
| 0 | 3120 | 2512 |
| 3 | 3923 | 3003 |
| 6 | 4785 | 3997 |
| 11 | 5283 | 4818 |

*Table 1: Median month income for male and female in Finland*




*Graph 1: Scattered diagram of median month income for years spent in education in Finland*

According to the graph 1, it can be concluded that there is a correlation between the level of education and monthly income, and for both genders. The correlation is positive, that is, as the number of years spent on education increases, the average monthly income also increases. Then I analyzed the strength, correlations, using the Pearson coefficient.

## 3. Processed data and calculations:

| $X$ | $Y$ | $(X_i - \overline{X})$ | $(Y_i - \overline{Y})$ | $(X_i - \overline{X}) \times (Y_i - \overline{Y})$ | $(X_i - \overline{X})^2$ | $(Y_i - \overline{Y})^2$ |
|---|---|---|---|---|---|---|
| -1 | 2933 | -4.8 | -1075.8 | 5163.84 | 23.04 | 1157345.64 |
| 0 | 3120 | -3.8 | -888.8 | 3377.44 | 14.44 | 789965.44 |
| 3 | 3923 | -0.8 | -85.8 | 68.64 | 0.64 | 7361.64 |
| 6 | 4785 | 2.2 | 776.2 | 1707.64 | 4.84 | 602486.44 |
| 11 | 5283 | 7.2 | 1274.2 | 9174.24 | 51.84 | 1623585.64 |

*Table 2: Processed data for male population in Finland*

First, I find the mean of the X data = $\frac{\text{sum}}{\text{number of terms}}$

$$\text{Mean } (\overline{X}) = \frac{-1+0+3+6+11}{5} = 3.8$$

Similarly, I find the mean of the Y data:

$$\text{Mean } (\overline{Y}) = \frac{2933+3120+3923+4785+5283}{5} = 4008.8$$

The difference between the X value and the mean value:

$$(X_i - \overline{X}) = -1 - 3.8 = -4.8$$

The difference between the Y value and the mean value:

$$(Y_i - \overline{Y}) = 2933 - 4008.8 = -1075.8$$

Then we find the cross product of X and Y:

$$(X_i - \overline{X}) \times (Y_i - \overline{Y}) = -4.8 \times -1075.8 = 5163.84$$

Square the difference for X:

$$(X_i - \overline{X})^2 = (-4.8)^2 = 23.04$$

Square the difference for Y:

$$(Y_i - \overline{Y})^2 = (-1075.8)^2 = 1157345.64$$

Sum of difference squared for X:

$$\Sigma\left(X_i - \overline{X}\right)^2 = 23.04+14.44+0.64+4.84+51.84 = 94.8$$

Sum of difference squared for Y:

$$\Sigma\left(Y_i - \overline{Y}\right)^2 = 4180744.8$$

Sum of cross product of X and Y

$$\Sigma(X_i - \overline{X}) \times \Sigma\left(Y_i - \overline{Y}\right) = 19491.8$$

$$r = \frac{\sum(X_i-\overline{X})\times(Y_i-\overline{Y})}{\sqrt{\Sigma(X_i-\overline{X})^2 \times \Sigma(Y_i-\overline{Y})^2}} = \frac{19491.8}{\sqrt{94.8\times4180744.8}} = 0.979086 = 0.979 \text{ (3 s.f)}$$

| $X$ | $Y$ | $(X_i - \overline{X})$ | $\left(Y_i - \overline{Y}\right)$ | $\left(X_i - \overline{X}\right) \times \left(Y_i - \overline{Y}\right)$ | $\left(X_i - \overline{X}\right)^2$ | $\left(Y_i - \overline{Y}\right)^2$ |
|---|---|---|---|---|---|---|
| -1 | 2401 | -4.8 | -945.2 | 4536.96 | 23.04 | 893403.04 |
| 0 | 2512 | -3.8 | -834.2 | 3169.96 | 14.44 | 695889.64 |
| 3 | 3003 | -0.8 | -343.2 | 274.56 | 0.64 | 117786.24 |
| 6 | 3997 | 2.2 | 650.8 | 1431.76 | 4.84 | 423540.64 |
| 11 | 4818 | 7.2 | 1471.8 | 10596.96 | 51.84 | 2166195.24 |

*Table 3: Processed data for female population in Finland*

First, I find the mean of the X data = $\frac{\text{sum}}{\text{number of terms}}$

$$\text{Mean } (\overline{X}) = \frac{-1+0+3+6+11}{5} = 3.8$$

Similarly, I find the mean of the Y data:

$$\text{Mean } (\overline{Y}) = \frac{2401+2512+3003+3997+4818}{5} = 3346.2$$

The difference between the X value and the mean value:

$$(X_i - \overline{X}) = -1 - 3.8 = -4.8$$

The difference between the Y value and the mean value:

$$\left(Y_i - \overline{Y}\right) = 2401 - 3346.2 = -945.2$$

Then we find the cross product of X and Y:

$$\left(X_i - \overline{X}\right) \times \left(Y_i - \overline{Y}\right) = -4.8 \times (-945.2) = 4536.96$$

Square the difference for X:

$$\left(X_i - \overline{X}\right)^2 = (-4.8)^2 = 23.04$$

Square the difference for Y:

$$\left(Y_i - \overline{Y}\right)^2 = (-945.2)^2 = 893403.04$$

Sum of difference squared for X:

$$\Sigma\left(X_i - \overline{X}\right)^2 = 23.04+14.44+0.64+4.84+51.84 = 94.8$$

Sum of difference squared for Y:

$$\Sigma\left(Y_i - \overline{Y}\right)^2 = 893403.04+695889.64+117786.24+423540.64+2166195.24 = 4296814.8$$

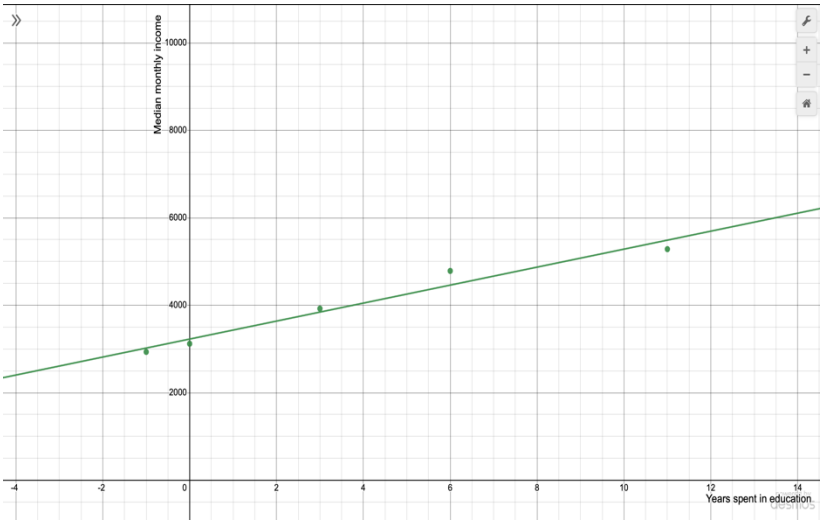
Sum of cross product of X and Y

$$\Sigma\left(X_i - \overline{X}\right) \times \left(Y_i - \overline{Y}\right) = 4536.96+3169.96+274.56+1431.76+10596.96 = 20010.2$$

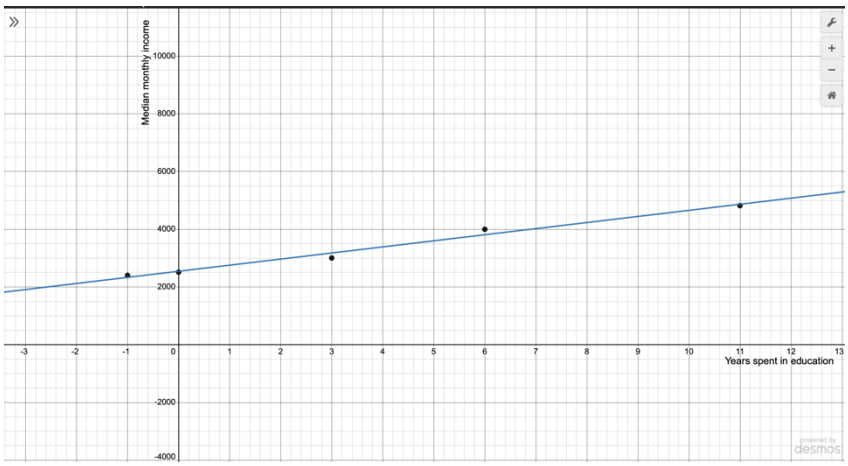
$$r = \frac{\Sigma\left(X_i - \overline{X}\right) \times \left(Y_i - \overline{Y}\right)}{\sqrt{\Sigma\left(X_i - \overline{X}\right)^2 \times \Sigma\left(Y_i - \overline{Y}\right)^2}} = \frac{20010.2}{\sqrt{94.8 \times 4296814.8}} = 0.99145713 = 0.991 \text{ (3 s.f.)}$$

Based on the results of calculations, it is possible to characterize the r-values that were obtained in the results of the analysis. Both Pearson correlation coefficients for men (0.979) and women (0.991) are positive numbers. This confirms the initial assumptions and observations made above that both genders demonstrate a positive correlation. Also, according to calculations, we can say a few words about the strength of correlation. The correlations for men and women are very strong, since their r-values are greater than 0.7. However, the female correlation is stronger than the male by 0.012. This may mean that

women can expect a higher or lower income, given their level of education. Then I entered the results into the Desmos program [6] .




*Graph 2: Linear regression graph for median month income in the Finland for male*




*Graph 3: Linear regression graph for median month income in the Finland for female*

Although linear regressions for men and women in visualize a high correlation, some values on the graph are not transmitted by the function. That is, the relationship between the two variables is roughly linear, so a simple linear regression would probably work well with this data. However, for a more accurate result, I decided to use polynomial regression (curves of

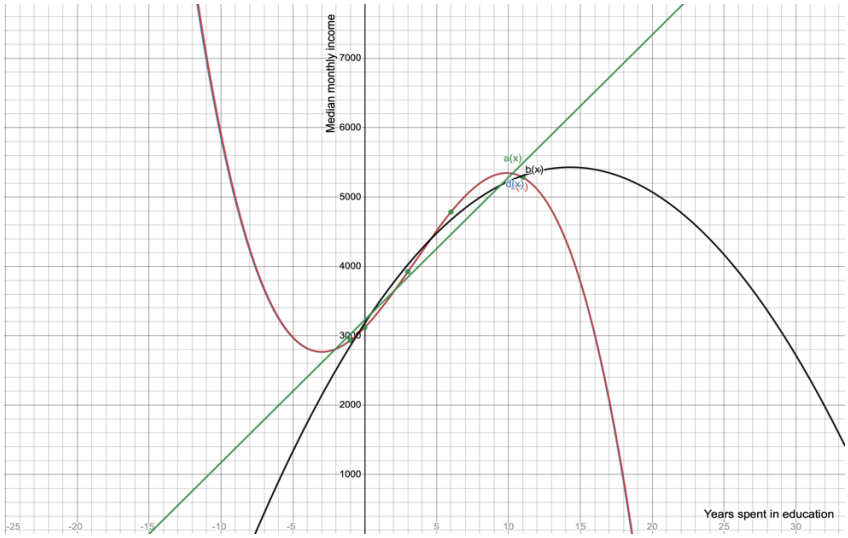
[6] https://www.desmos.com

the best fit: linear regression, quadratic, cubic, quadrangular). I decided to use polynomial regression (best fit curves: linear regression, quadratic, cubic, quadrangular). All calculations are done using the Desmos application [7].

| Name of the regression | Equation | |
|---|---|---|
| | Male | Female |
| Linear regression a(x) | y = 205.61$x$+3227.48<br>$r^2$ = 0.9586 | y = 211.078$x$+2544.1<br>$r^2$= 0.983 |
| Quadratic regression b(x) | y = -10.9967$x^2$ + 314.463$x$ + 3181.13<br>$R^2$ = 0.9919 | y = -0.514699$x^2$+216.173$x$+2541.93<br>$R^2$ = 0.9831 |
| Cubic regression c(x) | y = -2.42841$x^3$ + 25.1095$x^2$ + 214.284$x$+3119.84<br>$R^2$=1 | y = -3.41645$x^3$ + 50.2821$x^2$ + 75.2332$x$ + 2455.71<br>$R^2$= 0.9986 |
| Quartic regression d(x) | y = -0.00155123$x^4$ − 2.40029$x^3$ + 24.9781$x^2$ + 214.377$x$ + 3120<br>$R^2$ = 1 | y = -0.545896$x^4$ + 6.47828$x^3$ + 4.03138$x^2$ + 108.007$x$ + 2512<br>$R^2$=1 |

*Table 4: Regression equations*

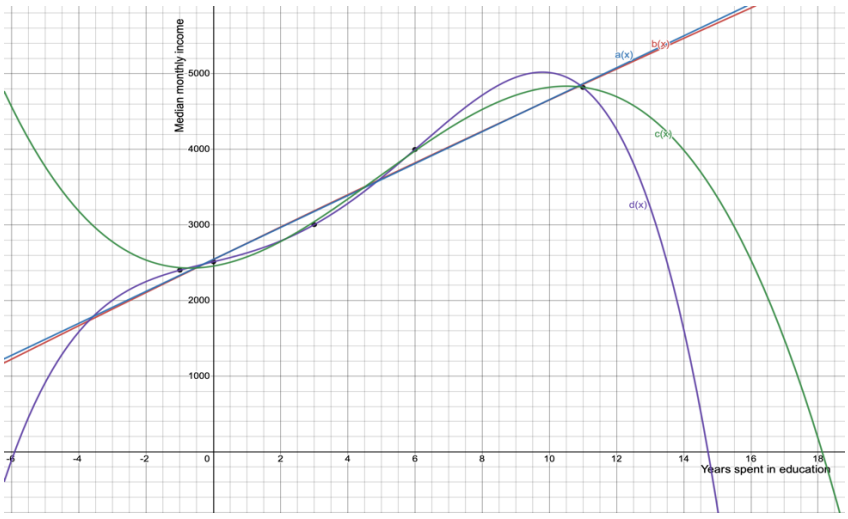
[7] https://www.desmos.com




*Graph 4: Regression function for female values*

Graph 4 shows that of all the curves, the quarterly regression (d(x)) corresponds closest to the values from the dataset. Linear regression (a(x)) and quadratic regression (b(x)) are also closely related to the values, so they overlap each other. Moreover, the quadratic regression gave the largest value of $R^2$ equal to 1, and other models gave values less than 1. The value of the R-square will always be in the range from 0 to 1.

A value of 1 indicates that the independent variables can fully explain the variance of the response variable, and a value of 0 indicates that the independent variables cannot explain the variance of the response variable.




*Graph 5: Regression function for male value*

Graph 5 shows that practically all regression lines of male dataset correspond very closely to the values that overlap each other. However, despite this, there are differences. The closest to the points is the red curve (d(x)). In contrast, linear regression (a(x)) and quadratic regression (b(x)) are further away. Although both cubic and quartile regressions gave a value of $R^2$equal to 1 after calculations in Desmos, quartile regression is used in further plotting. This is done because it more closely matches the values from the male dataset.

Further, to represent data sets, models of quartic equations were used, which have more accurate values, unlike other equations. According to the data table, we can conclude about the difference in education and income for both genders. To do this, it is necessary to find the difference between the incomes of the lowest level of education and the highest. For these men, a characteristic feature is that their income does not depend so much on their level of education, unlike women (5,283 - 2,933 = 2,350 euros for men versus 4818 - 2401 = 2417 for women). That is, the level of education is very important for women, their income will depend on it. The graph also shows that the largest income gap is between the last two levels of education (master and doctorate), and the lowest gap is between the absence of a high school diploma and its availability. It is also worth mentioning that the functions represent only the values of 11 years of study, these functions cannot predict the future, even though the graph shows a downward trend after the last point (doctorate). The study also found that education plays a greater role for women than for men.

For a more accurate analysis of income for both genders in Finland, one can resort to using differentiation. «Differentiation is the process of finding the derivative or rate of change of a function»[8]. This theory is often applied in practice in various fields of science for example, in

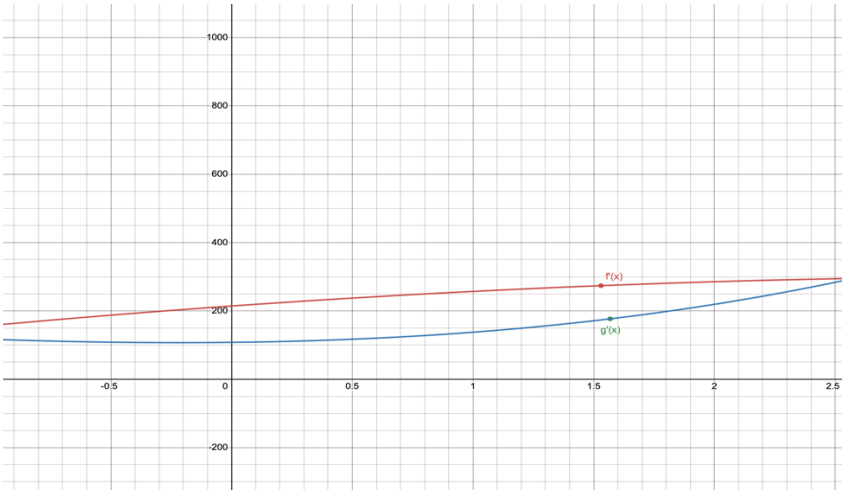
[8] Britannica, The Editors of Encyclopaedia. "differentiation". *Encyclopedia Britannica*, 23 Dec. 2022, https://www.britannica.com/science/differentiation-mathematics.

economics. According to this, it is possible to trace the dynamics of changes in income after the time spent on education. Also of great interest is further education for both genders. The differentiation formula was taken from the textbook Mathematics Analysis and Approaches SL:

*If f(x)= u(x) + v(x), then f'(x)= u'(x) + v'(x)*

| Rate of profit | Differentiation |
|---|---|
| Male | $f(x) = -0.00155123x^4 - 2.40029x^3 + 24.9781x^2 + 214.377x + 3120$<br>$f'(x) = \frac{d}{dx}(-0.00155123x^4 - 2.40029x^3 + 24.9781x^2 + 214.377x + 3120) = -0.00155123 \times 4x^3 - 2.400293 \times 3x^2 + 24.97812x + 214.377 + 0$<br>$= -0.00620492x^3 - 7.20087x^2 + 49.9562x + 214.377$ |
| Female | $g(x) = -0.545896x^4 + 6.47828x^3 + 4.03138x^2 + 108.007x + 2512$<br>$g'(x) = \frac{d}{dx}(-0.545896x^4 + 6.47828x^3 + 4.03138x^2 + 108.007x + 2512) = -0.545896 \times 4x^3 + 6.47828 \times 3x^2 + 4.03138 \times 3x + 108.007 + 0 = 2.183584x^3 + 19.43484x^2 + 8.06276x + 108.007$ |

*Table 5: Derivative calculations*




*Graph 6: Derivative f'(x) and g'(x) for male and female population of Finland*

Observing these graphs, you can see that the derivative for these women is lower than the derivative for women, which indicates a noticeable difference in income in the range between (-1< x <11). However, the further the functions move, the more it becomes flatter. This suggests that there will be no difference in the salaries of men and women after 11 years of study. Consequently, further education will not bring an increase in income after the last stage of education (doctorate). However, this graph cannot predict how it will actually be, due to the fact that many factors, such as the economy and geopolitics of Finland, are not taken into account by me.

# 4. Evaluation

## 4.1 Strengths:

This study has strengths and can be considered a success. The correlation between the level of education and the monthly income of men and women in Finland has been clearly demonstrated. Also, according to graphs and mathematical methods, comparisons were presented between the visualization of comparisons in income between men and women. In addition, it can be considered a strength that these samples were presented quite recently, in 2020, which makes this study closer to the current economic situation and more relevant.

### 4.2 Limitations:

High values of the coefficient of determination do not always indicate the presence of a causal relationship between the variables (as well as in the case of the correlation coefficient). Since, if the explained variable and factors that have no connection with the explained change have an increasing dynamic, then the coefficient of determination will be quite high. Therefore, the logical and semantic part of the model is of paramount importance.

It is also worth mentioning that in the statistical data there is a different sample for men and women. Somewhere the number of men with education is greater (for example, a high school diploma), and somewhere less (bachelor). It could also have affected the data, making it higher or lower due to fewer workers in the sample.

## 5. Extensions:

To expand the results of the study and provide more information on the topic of the study, other indicators can be additionally considered. It is possible to analyze several countries with the same high level of income of the population as Finland. You can add countries with lower income levels, perhaps the wage gap between the groups will be even more significant. Perhaps this would help interested persons decide where to look for a job after obtaining a certain degree.

In addition to other countries that could be analyzed, additional groups of people, such as children and the elderly, could be included. These groups are considered the most vulnerable and are subject to greater discrimination in remuneration. For example, in countries where child labor is exploited, the pay for performing the same labor functions has a huge gap between the pay of children and adult men.

It is also possible to study who was most affected by the loss of income during the 2020 pandemic, and in the years following the pandemic. It is also possible to consider data on which groups of people are more likely to be cut in a certain industry with equal education.

Since not many studies have been conducted on this topic, it would be possible to obtain data using anonymous questionnaires among men and women, as well as to organize a survey of additional groups of people. In principle, this simulates how the data I used in this study was obtained, but it should be considered that such a data collection process will take a very long time and will require significant resources.

## 6. Conclusions

It follows from the above graphs, men with equal data on education receive a significantly higher income than women with the same degree of education. Although the difference in performance for people with a diploma of secondary education and below secondary is not so noticeable. In contrast to the indicators of the above groups, the indicators of people with higher education show clear differences between the incomes of men and women.

When calculating the Pearson correlation coefficient, one can notice that there is a very strong correlation between the median incomes of men and women.

Comparison of the resulting graphs and correlation calculations can provide valuable information for people who still doubt that gender wage discrimination exists in the world. From this it becomes clear that even a woman receiving a higher education at the same level as a man and performing the same work tasks does not at all guarantee equal pay.

This up-to-date study confirms that society has not yet achieved gender equality. And achieving it requires more equitable laws and regulations, the expansion of safe education and health services, the necessary infrastructure, and social norms so that women can make the necessary choices for themselves and their families.

It is very important to me that, despite the existence of problems, Finnish society does not shy away from addressing these sensitive issues. Gender equality is one of the fundamental values of Finnish society. Since 2003, Equality Day, known as "Minna Kant Day", has been officially celebrated in Finland on 19 March.

## 7. References:


Britannica, The Editors of Encyclopaedia, (2022) “differentiation" [online], Accessed December 15 https://www.britannica.com/science/differentiation-mathematics. Accessed 14 February 2023.

Desmos Studio (2023) https://www.desmos.com

International Labour Organization, (2022) Women Work More, But are Still Paid Less, Accessed January 12 [online]. https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_008091/lang--en/index.htm

Official Statistics of Finland (OSF): Structure of Earnings (2020), Accessed December 15 [online].
https://www.stat.fi/til/pra/2020/pra_2020_2021-10-11_tau_001_en.html

Oxford IB Diploma Programme IB Mathematics analysis and approaches, Standard Level, by Paul La Rondie, Jill Stevens

Rodgers, Joseph Lee, and W. Alan Nicewander. “Thirteen Ways to Look at Correlation Coefficient.” The American Statistician (1988), Accessed December 15 [online].
https://www.jstor.org/stable/2685263

Summary of the Global Wage Report 2016/17: Wage inequality in the workplace (2016), Accessed December 15 [online].

https://www.ilo.org/global/research/global-reports/global-wage-report/2016/WCMS_537982/lang--en/index.htm