University of Science and Technology in Zewail City

A Study of Income Inequality in Egypt

MATH 301 - Report

Made by:

Ahmed Elkot 202000418

Ahmed Nasser 202000466

Anas Nagaty 202000516

Mohamed Abdelsalam 202001631

Mahmoud Marzouk 202000258

Phelopater Ramsis 202001171

Supervised by:

Dr. Walaa El-Sharkawy

Table of Contents

Abstract

This research project aims to contribute to the understanding of income inequality and its underlying determinants. This investigation of income inequality in Egypt, which is a subset of the more general social inequality, are done by examining the impact of gender, household number, and education levels on annual income, we seek to uncover the mechanisms through which social inequality influences income outcomes. Through this knowledge, we aspire to inform evidence-based policies and interventions that promote greater income equality and foster inclusive economic growth, ultimately striving for a more equitable and just society for all individuals, regardless of their social backgrounds. The examination process will be mainly occurred by hypothesis testing (Point-biserial correlation, and Pearson correlation). Results in this paper show the relationship between income on one hand and on the other hand education, gender, and household number. Our findings included the big gender pay gap against women, correlation between education level and salary, and an unexpected increase of household members with increasing income.

1. Introduction

Social inequality is the unequal distribution of resources in different areas such as health, education, and employment, leading to disparities based on factors like gender, race/ethnicity, and social class [1]. Social inequality is a critical issue that permeates societies worldwide, encompassing disparities in access to resources, opportunities, and power among different social groups. Its importance lies in the profound impact it has on individuals, communities, and societies at large. In this research project, we aim to delve into the extent to which social inequality matters, particularly focusing on its relationship with annual income. Furthermore, we will investigate the

influence of gender, household number, and education levels on income disparities. Recognizing the importance of social inequality is crucial due to its wide-ranging consequences for individuals and society. Income inequality, as a manifestation of social inequality, can perpetuate social divisions, hinder social mobility, and undermine social cohesion. Moreover, it can have significant implications for economic growth and overall societal well-being. Understanding the magnitude and implications of social inequality on annual income is essential for developing effective policies and interventions aimed at reducing disparities and fostering more equitable and inclusive societies.

To achieve our research objectives, we will examine three key factors: gender, household number, and education levels, in relation to annual income. These factors have been widely acknowledged as influential determinants of income disparities and can provide valuable insights into the dynamics of social inequality. Gender plays a significant role in income inequality. Historical biases and gender-based discrimination have contributed to persistent wage gaps and occupational segregation. By analyzing the relationship between gender and annual income, we aim to explore how gender-based disparities in opportunities and wages contribute to income differences [2]. Household numbers, as a determinant of income inequality, examine the impact of family structure on individuals' annual income. Different household sizes and compositions can experience varied access to resources and opportunities, influencing overall income levels. By investigating the relationship between household numbers and annual income, we can gain insights into how family dynamics and support systems shape income outcomes [3]. Education levels are crucial in determining annual income and social mobility. Disparities in educational opportunities and attainment can perpetuate income inequality by limiting access to well-paid employment. By examining the relationship between education levels and annual income, we seek to understand how disparities in educational opportunities contribute to income disparities across different segments of society [4].

To investigate these factors and their relationships with annual income, we will employ a methodology that includes data collection, statistical analysis, and hypothesis testing. By utilizing relevant datasets and employing appropriate statistical techniques, we aim to generate empirical

evidence that sheds light on the complex interplay between social inequality and the variables under investigation, specifically in relation to annual income disparities.

Hypothesis testing is a statistical method used to draw conclusions about a population based on sample data. It involves comparing a null hypothesis, which assumes no significant difference or relationship, with an alternative hypothesis that suggests the presence of a significant difference or relationship. There are two types of hypothesis tests: one-tailed and two-tailed tests.

In a one-tailed test, the alternative hypothesis is directional and specifies whether the parameter being tested is greater than or less than a certain value. The critical region is defined on one side of the distribution, depending on the specified direction. The test statistic, such as a t-statistic or z-statistic, measures the difference between the sample data and the expected values under the null hypothesis. If the test statistic falls within the critical region, the null hypothesis is rejected in favor of the alternative hypothesis. While, in a two-tailed test, the alternative hypothesis is non-directional, indicating a significant difference or relationship without specifying the direction. The critical region is defined on both sides of the distribution. The test statistic is calculated similarly to the one-tailed test, and if it falls within either critical region, the null hypothesis is rejected. The figure (1) below illustrates one-tailed and two tailed tests schematically. The choice between one-tailed and two-tailed tests depends on the research question and hypothesis. One-tailed tests are used when a specific direction is expected, while two-tailed tests are suitable for detecting any significant difference or relationship, regardless of the direction.

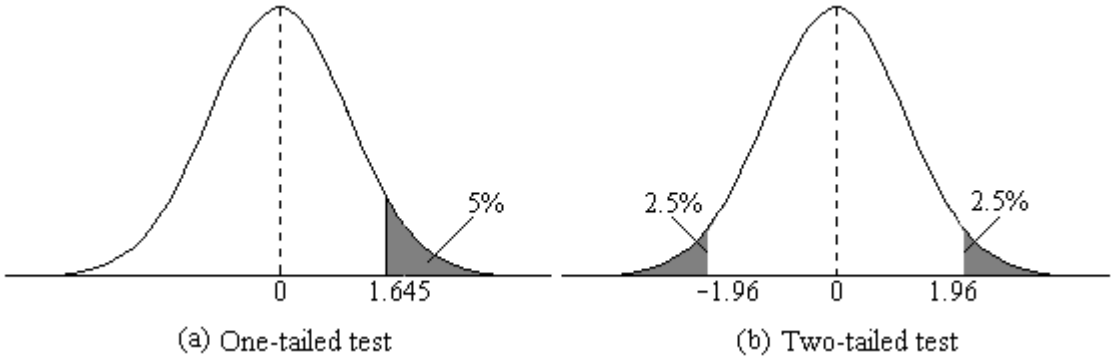

Figure 1: Rejection and Non-rejection regions for H0 in both one- tailed and two-tailed hypothesis testing.

Hypothesis testing enables researchers to make informed decisions based on statistical evidence from sample data. It plays a crucial role in scientific research by allowing for conclusions and generalizations from samples to larger populations. By rigorously testing hypotheses, researchers can validate or refute claims, contributing to the advancement of knowledge and decision-making in various fields. The type that will be used will be one tail.

Correlation refers to the statistical relationship between two variables. It measures the strength and direction of the linear association between the variables. Correlation does not imply causation.

The Pearson correlation coefficient, denoted as r, is a measure of the linear correlation between two variables. It quantifies the degree to which the variables are linearly related. The coefficient ranges from -1 to +1. A positive value indicates a positive linear relationship, meaning that as one variable increases, the other tends to increase as well. A negative value indicates a negative linear relationship, suggesting that as one variable increases, the other tends to decrease. A value of zero suggests no linear correlation between the variables.

The magnitude of the Pearson correlation coefficient reflects the strength of the relationship. A coefficient closer to -1 or +1 indicates a stronger linear relationship, while a coefficient closer to zero suggests a weaker relationship. The coefficient does not indicate the slope or the steepness of the relationship, but rather the consistency and direction of the relationship between the variables.

2. Methodology

Data was collected from Central Agency for Public Mobilization and Statistics (CAPMAS). Data on Household leader gender was obtained from various tables. Our findings are summarized in table 1.

Income	M HH number	F HH number	M HH percentage	F HH percentage		M HH percentage	F HH percentage
<10.000	7.245	17.869	0.032679296	0.08059991	5000	0.000399595	0.004432771
10.000-	116.009	327.753	0.523270185	1.478362652	15000	0.006398421	0.081305833
20.000-	240.2	302.762	1.083446098	1.36563825	22500	0.013248116	0.075106304
25.000-	526.323	347.583	2.374032476	1.567807848	27500	0.029029093	0.08622507
30.000-	912.482	326.723	4.115841227	1.473716734	32500	0.050327508	0.08105032
35.000-	1318.303	349.187	5.946337393	1.575042851	37500	0.07271037	0.086622975
40.000-	1562.277	331.861	7.046806495	1.496892197	42500	0.086166639	0.082324906
45.000-	1553.716	300.727	7.008191249	1.356459179	47500	0.085694462	0.074601481
50.000-	1617.264	248.427	7.294830853	1.120554804	52500	0.089199421	0.061627397
55.000-	1554.519	247.736	7.011813261	1.117437979	57500	0.085738751	0.06145598
60.000-	1234.401	176.476	5.567889039	0.79601263	62500	0.068082796	0.04377848
65.000-	1191.315	156.746	5.373545332	0.707018493	67500	0.065706408	0.03888405
70.000-	1701.02	246.575	7.672620659	1.112201173	75000	0.093818943	0.06116797
80.000-	1168.106	171.215	5.268858818	0.772282364	85000	0.064426327	0.042473381
90.000-	848.593	119.173	3.827663509	0.537541723	95000	0.04680374	0.029563299
100.000-	1052.501	146.372	4.747410916	0.66022553	110000	0.058050188	0.036310567
120.000-	699.581	108.403	3.155529995	0.488962562	130000	0.038585055	0.026891581
150.000-	445.82	48.069	2.010915652	0.216820027	175000	0.024588989	0.011924498
200.000+	381.205	57.456	1.719463239	0.259161028	300000	0.021025179	0.014253136
Total	18130.88	4031.113	81.78114569	18.18273793	1440000	1	1

Table 1

Data on Education Level vs income was obtained. The data are summarized in table 2.

No. of H.H. Members	10806	1168	252	2377	1451	115	885
Income	University an	Above Intermediate & less	Intermediate weight	Below Intermediate w	Literacy Certificate wei	Read and write weigh	Illiterate weight
<10.000	0	0.000896473	0.000229758	0	0	0.000841629	0.003938019
10.000-	0.00233998	0	0.003378548	0.010421547	0.006211302	0.020085973	0.063906546
20.000-	0.00488667	0.007588168	0.009986854	0.021586909	0.01844226	0.031719457	0.058380189
25.000-	0.00801995	0.004103527	0.027757096	0.035694331	0.034299754	0.04860181	0.081238656
30.000-	0.01569992	0.024414107	0.048493427	0.061586909	0.070081081	0.072167421	0.088141533
35.000-	0.02664644	0.054006826	0.076378548	0.084595052	0.080176904	0.090058824	0.0992877
40.000-	0.03973982	0.047105802	0.090574395	0.093917319	0.105056511	0.103565611	0.102411662
45.000-	0.04988667	0.069588168	0.093033911	0.103983401	0.105761671	0.087524887	0.081451439
50.000-	0.05167997	0.076343572	0.095013146	0.099239587	0.086584767	0.101339367	0.076615177
55.000-	0.06167997	0.105759954	0.089736331	0.093834638	0.087142506	0.08399095	0.07104441
60.000-	0.06020671	0.070962457	0.070128025	0.064785155	0.076857494	0.076764706	0.049185943
65.000-	0.05632003	0.05592719	0.074161936	0.064537112	0.069169533	0.058733032	0.045349681
70.000-	0.11320671	0.114	0.091195847	0.08042969	0.086223587	0.085615385	0.066575401
80.000-	0.08728013	0.085825939	0.066425605	0.057512371	0.056115479	0.04399095	0.039123962
90.000-	0.07620671	0.067030717	0.042	0.043768556	0.018230958	0.033891403	0.026123962
100.000-	0.09894015	0.090687144	0.059506573	0.044066082	0.050523342	0.030366516	0.026898243
120.000-	0.08958659	0.071343572	0.034344637	0.020553711	0.034646192	0.018375566	0.013327476
150.000-	0.07594015	0.029207053	0.016195847	0.011595052	0.010823096	0.008316742	0.004265495
200.000+	0.08184677	0.022830489	0.010344637	0.00804134	0.004565111	0.005049774	0.003367252
Total	1.00	1.00	1.00	1.0	1	1.00	1.000632748

Table 2

Data on Number of people per household were obtained vs yearly income and it is summarized in table 3.

Income	8 or more persons	6-7 persons	5 Persons	4 Persons	3 Persons	2 Persons	1 Person
<10.000	0	0	0.001	0	0	0.002	0.008
10.000-	0	0.001	0.003	0.005	0.007	0.023	0.183
20.000-	0	0.005	0.006	0.008	0.018	0.044	0.154
25.000-	0.005	0.01	0.016	0.023	0.041	0.082	0.148
30.000-	0.009	0.023	0.03	0.051	0.062	0.106	0.126
35.000-	0.03	0.042	0.057	0.075	0.094	0.118	0.101
40.000-	0.044	0.056	0.075	0.091	0.114	0.111	0.069
45.000-	0.05	0.073	0.09	0.091	0.101	0.079	0.049
50.000-	0.073	0.088	0.09	0.095	0.094	0.07	0.036
55.000-	0.06	0.091	0.092	0.089	0.084	0.073	0.024
60.000-	0.064	0.08	0.075	0.072	0.063	0.037	0.02
65.000-	0.083	0.081	0.076	0.065	0.05	0.041	0.013
70.000-	0.122	0.114	0.11	0.094	0.073	0.061	0.018
80.000-	0.127	0.09	0.069	0.059	0.058	0.034	0.009
90.000-	0.073	0.067	0.055	0.044	0.031	0.026	0.009
100.000-	0.134	0.08	0.063	0.058	0.04	0.029	0.008
120.000-	0.08	0.047	0.043	0.034	0.032	0.031	0.007
150.000-	0.024	0.029	0.032	0.022	0.016	0.014	0.006
200.000+	0.019	0.022	0.019	0.024	0.018	0.019	0.01
Total	1.00	1.00	1.00	1	1.00	1	1.00

Table 3

Data was collected using Excel and visualized (discussed in the results section). Then we did our analyses. Each frequency of salaries was represented in its middle point. Pearson Coefficient correlated values in 3 different areas: Education level (from illiterate to university level), number of people per household (from 1 to +8), and gender (men and women) vs annual salary. Pearson Correlation coefficient r can be calculated with the following formula:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[ln\sum x^2 - (\sum x)^2][ln\sum y^2 - (\sum y)^2]}}$$

A two-sample hypothesis test of means is performed between a sample of households whose head is male and those whose head is female to test whether the latter group's average income is lower than the former in the population. The one-tailed independent t-test is chosen, and since the sample sizes are so huge, they are treated as following a normal distribution and the test statistic z is used.

It can be calculated with the following formula, where S1 is standard deviation of sample S1 and S2 is the standard deviation of sample 2:

$$z = \frac{(\bar{x}_1) - (\bar{x}_2) - (\mu 1) - (\mu 2)}{\sqrt{\frac{(S1)\text{^}2}{n1} + \frac{(S2)\text{^}2}{n2}}}$$

Based in the feedback, we decided to add a step for Testing the Significance of the Correlation Coefficient. It is like the hypothesis test with 2 hypotheses. The null being the results are not significant and the alternate one being that it is significant. To get the best results, we will use 99% confidence level with alpha = 0.01. We should calculate p value that is calculated from the t-graph with t-score and freedom level of n-2 (it has a function in excel). and t-score is calculated from the following formula: $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

3. Results

A. Data Visualization

We first visualized the data and obtained the following graphs:

- The different incomes with householders heads male and female:

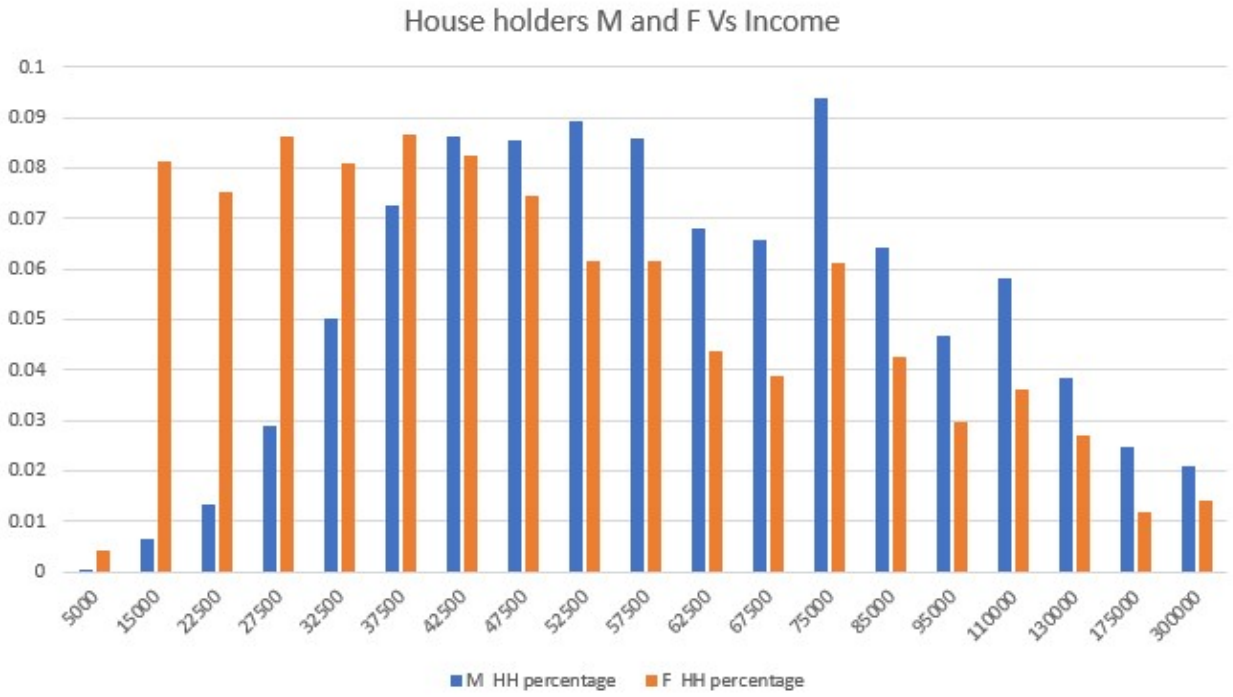

Figure 2: Par graph of the income distribution of the male and female householders

We can conclude that the women householders are most likely with lower income comparing to men

- The different incomes with different levels of education categories:

Level of education	Income
Illiterate avg	50794.15
Read and write avg	57571.42
Literacy Certificate avg	61490.77
Below Intermediate avg	61769.5
Intermediate avg	66784.08
Above Intermediate & less than University avg	80328.53

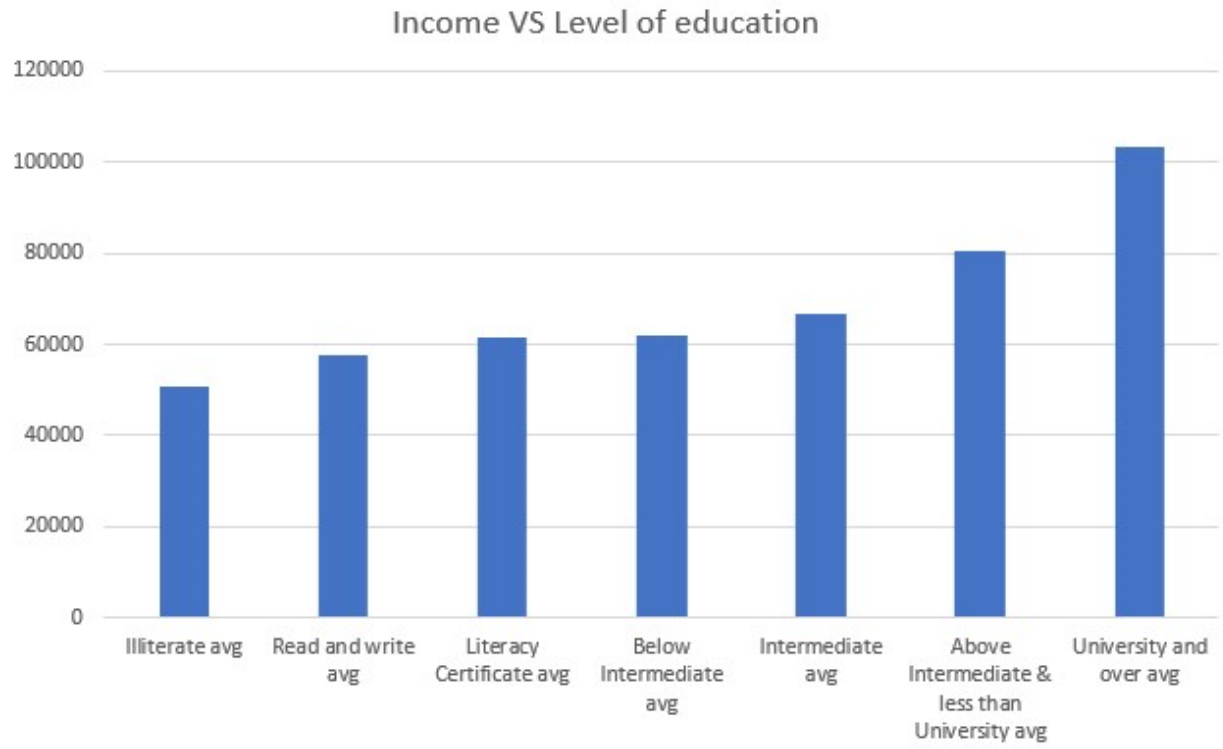

Figure 3 Par graph of the level of education

We can conclude that the level of education slightly on the income for different levels of education

- Box plot for different education categories:

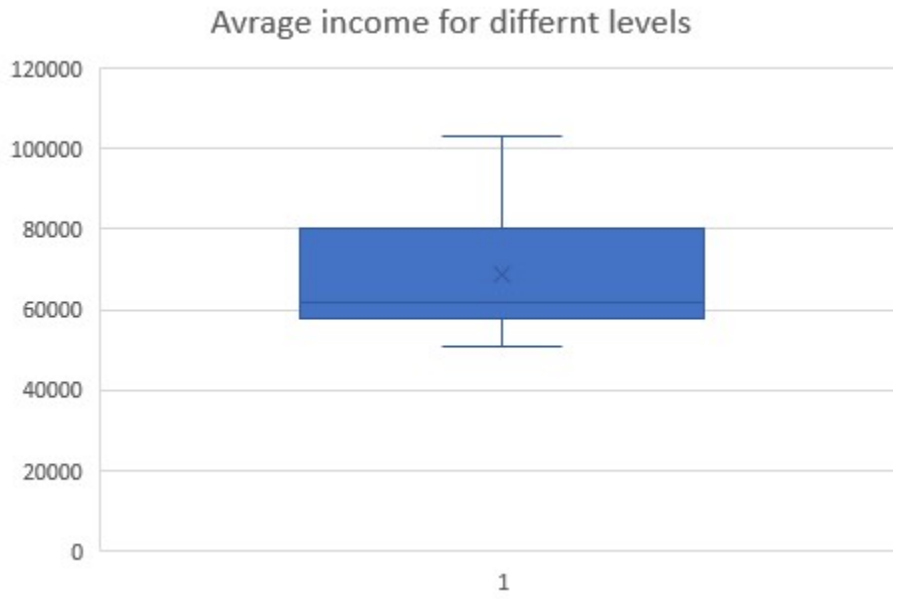

Figure 3 Par graph of the level of education Vs income

- The different incomes with different household size categories:

income	classification
37745.49	1 Person
58197.5	2 Persons
64264.56	3 Persons
70635	4 Persons
74044.41	5 Persons
78120.62	6-7 persons
84859.58	8 or more persons

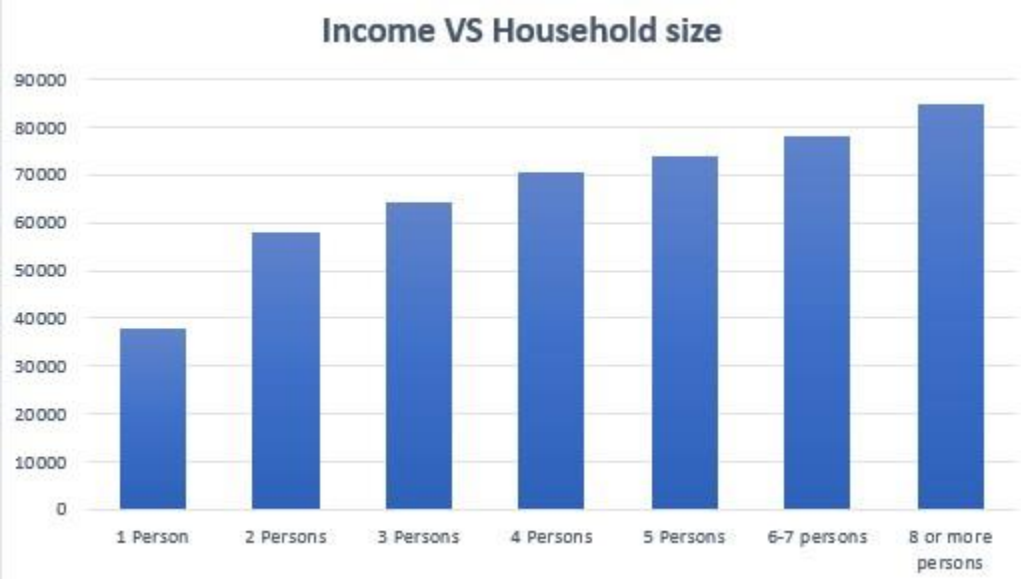

Figure4: Par graph of the household size Vs income

We can conclude that the household size also affects the income while the number increases the income increases.

B. Pearson & testing its significance

Results will be commented on in the discussion section. Zeroes indicate that the number is very small that it couldn't be displayed by Excel.

- Level of Education:

Level of Education	University and over	Above Intermediate	Intermediate
r coefficient	0.632	0.102	-0.216
T-score	48.9754553	3.047401472	-18.05399999
P value	0	0.002377716	3.52801E-71

Level of Education	Below intermediate	Literacy Certificate	Read and write	Illiterate
r coefficient	-0.332	-0.324	-0.430	-0.621
T-score	-19.85662839	-6.894698175	-22.39329791	-56.95739666
P value	7.957E-83	2.08119E-11	2.8792E-100	0

- Household Size:

HH size	8+ persons	6-7 persons	5 persons	4 persons	3 persons	2 persons	1 person
r coefficient	0.166640833	0.052850922	-0.042021863	-0.1284452	-0.2722096	-0.4119404	-0.5183855
T-score	3.993337282	3.157994916	-2.950981623	-9.1289574	-16.570849	-24.777956	-24.991028
P value	7.38688E-05	0.001601876	0.003182599	9.853E-20	2.1704E-59	1.748E-123	1.187E-117

- Gender:

HH head gender	Male	Female
r coefficient	-0.142	-0.614
T-score	-19.28227192	-49.36795188
P value	5.01397E-82	0.0000000

C. Hypothesis Test

- H0 : $\mu_{(F)} \geq \mu_M$, or $\mu_{(F)} - \mu_M \geq 0$. i.e., the average pay of both genders is at most equal.
- H1 : $\mu_{(F)} < \mu_M$, or $\mu_{(F)} - \mu_M < 0$. i.e., there indeed exists a pay gap between both genders.

From the Pearson coefficient of correlation between HH head gender and income, we are expecting that H0 be rejected. From CAPMAS census data, $\mu_{(F)}$= 49,237 EGP, and μ_F = 67,813 EGP. SF = 6,550 EGP, and SM = 4,320 EGP. We thus compute the test statistic $z \approx -172$. Such value is so high that it leads to rejecting H0 for practically any desired significance level.

4. Discussion

In the correlation part, some of the results were expected while others were surprising. For Level of education, we have indeed expected that higher levels of education get to be paid more, while the other end of the spectrum is illiterate people (from +ve 0.632 to -0.621). Surprisingly, almost all other categories had negative correlation. This shows that a part of the social inequality is having access to educational opportunities. This also shows the huge gap between highly paid highly educated class and low paid low educated class. Only the tip of the pyramid are likely to have a decent annual salary.

Another point discussed was the gender pay gap. We expected that both would be negative based on our previous study as most citizens are in the low class of the society. We also expected that women are on average paid less than men, backed by previous research. What was surprising was how wide was the gap. The correlation coefficient was 5 times more negative in women than in men. This shows the substantial difference in pay gap between the 2 genders. Our Z test further enforced that result by rejecting the non-gap hypothesis and accepting the gap hypothesis with very high confidence level.

Lastly, we studied the number of people in the household for the sake of confirming that poor people have more members in the household. Unexpectedly, the only +ve correlation, even though it was weak, was in the highest 2 categories only (+6 people). This didn't confirm our initial thought, but it revealed an eye-opening fact: Rich people are more likely to have more people per household. Significance test showed that all correlations were significant and less than alpha = 0.01 so the null hypothesis is rejected, and correlation is significant.

5. Conclusion

Wealth among the population due to, among other factors, social prejudices, and gender-based discrimination, which sheds light on more inherent socio-economic issue that our country faces. That is, it was shown, with very high confidence, that the mean income for households whose head is male is larger than those whose head is female. An unanticipated result was the positive correlation between household size and income, which can likely be attributed to the fact that it is easier for the more prosperous families to increase their household size, along with the aforementioned reasons. Overall, this data provides a gateway for further investigations of social classes in Egypt.

6. References

[1] Social inequality — department of psychology. (n.d.). Psu.edu. Retrieved May 21, 2023, from https://psych.la.psu.edu/research/social-inequality.

[2] Ortiz-Ospina, E., & Roser, M. (2018). Economic inequality by gender. Our World in Data. https://ourworldindata.org/economic-inequality-by-gender.

[3] US Census Bureau. (2022). Understanding the relationship between individual earnings and household income. https://www.census.gov/newsroom/blogs/random-samplings/2017/11/earnings-income.html.

[4] Stryzhak, O. (2020). The relationship between education, income, economic freedom and happiness. SHS Web of Conferences, 75, 03004. https://doi.org/10.1051/shsconf/20207503004.

[5] CAPMAS 2021 Income and Spending Comprehensive Study.