Using chi-square categorical testing to analyse the survey data and find people's attitude towards inequalities

Quan Zhou
University of Bristol
Lu21887@bristol.ac.uk

Abstract. This study utilizes chi-square categorical testing to analyze survey data, aiming to delve into people's attitudes towards various forms of inequality. The research investigates individuals' perceptions concerning disparities in wealth, access to healthcare, and opportunities for personal and professional growth. Data was collected through a structured survey, presenting participants with relevant questions related to societal inequalities. By employing the chi-square test, the study examines the associations between respondents' demographic characteristics and their attitudes towards inequality. Additionally, the research explores potential correlations between attitudes towards inequality and factors such as education level, income, and social background. The findings shed valuable light on the extent to which individuals perceive and respond to societal disparities, providing crucial insights for policymakers and social scientists in their efforts to design effective strategies for addressing inequality and fostering a fairer and more cohesive society.

Keywords: chi-square; attitude; inequality.

1. Introduction

1.1. Background

While keeping an eye on the bigger picture and focusing on the global events, we need to pay attention to the domestic front, since a lot of people there experience severe inequity. Due to the growingly serious class entrenchment in the United Kingdom and the closed-loop resource circulation, it is even more challenging for them to ascend the social ladder and get access to greater resources and riches. All of this is closely connected to wellbeing and quality of life. Income and health can both have an impact on one another. Instead of raising and improving the level of care enjoyed by the higher income groups in our country, even if it costs them money, it would be more ethical if the scarce medical resources is distributed to the poor who cannot afford standard medical care.

Inequalities within societies have long been a subject of concern and debate. They encompass various dimensions, including wealth, healthcare, and opportunities, among others. These inequalities can have profound implications for individuals, communities, and the overall social fabric. Addressing and understanding these inequalities is crucial for fostering a just and equitable society.

One of the most visible and widely discussed forms of inequality is wealth disparity. Within societies, a significant portion of wealth often tends to be concentrated in the hands of a select few, while a large portion of the population faces financial challenges and limited economic opportunities. Such disparities can lead to social stratification, barriers to upward mobility, and the perpetuation of intergenerational inequality. Understanding the causes and consequences of wealth inequality is vital for designing effective policies that promote economic fairness and social cohesion.

Access to quality healthcare is a fundamental right, yet it remains unequal in many societies. Disparities in healthcare can manifest in various ways, such as differences in healthcare coverage, availability of medical facilities, and healthcare outcomes. Marginalized communities, socioeconomically disadvantaged individuals, and minority populations often face significant barriers in accessing timely and affordable healthcare services. Addressing healthcare inequalities requires a comprehensive approach that considers social determinants of health, promotes health equity, and ensures equitable distribution of resources and healthcare facilities.

Unequal access to opportunities is another critical aspect of societal inequality. Factors such as education, employment, and social networks play a pivotal role in determining individuals'

opportunities for personal and professional growth. Unfortunately, systemic barriers and discrimination can hinder equal access to these opportunities, limiting social mobility for certain groups. Examining the disparities in educational attainment, job prospects, and social capital sheds light on the structural challenges that impede equal opportunities. Efforts to reduce inequality should encompass policies that promote inclusive education, equal employment opportunities, and fair social networks.

Social inequalities related to wealth, healthcare, and opportunities persist in UK societies. These disparities pose significant challenges to achieving a fair and just society. Understanding the multifaceted nature of these inequalities is crucial for developing targeted interventions and policies that aim to reduce disparities and promote greater equality. By addressing these inequalities, societies can foster inclusive growth, enhance social cohesion, and create environments where individuals can reach their full potential regardless of their background or circumstances.

This essay explores the people's attitude towards the persistent disparities in wealth distribution, access to healthcare, and opportunities, and using chi-square testing to discover the potential factor.

Chi-Square testing is a statistical method that plays a significant role in examining relationships and associations between categorical variables. It provides valuable insights into understanding the dependence or independence of variables based on observed frequencies within different categories. The importance of Chi-Square testing lies in its ability to analyze data that cannot be easily measured on a continuous scale. The primary purpose of a Chi-Square test is to compare observed frequencies (data collected from samples or populations) with expected frequencies (theoretical values based on assumed independence between variables). If the observed frequencies deviate significantly from the expected frequencies, it suggests a relationship or association between the variables under investigation. This allows researchers to draw conclusions about the presence or absence of associations between categorical variables based on statistical evidence. It helps answer research questions related to categorical data.

In this essay, Chi-Square testing is a powerful statistical tool which is used to measure the survey response and determine whether there is a relationship between age, gender and education level and their attitude towards inequalities.

1.2. Related Work

The Chi-square (X2) test is a widely used nonparametric statistical method primarily applied in experimental research dealing with frequency or count data. This type of data involves 'counts,' such as the number of boys and girls in a class who have undergone tonsillectomy, as opposed to continuous variables like temperature or height. The main objective of this test is to determine the probability of association or independence between different facts [1].

CHI-Squared Test of Independence, written by Zibran served as a comprehensive summary of the Chi-square test, starting from its basic principles and progressively delving into its practical application through descriptive examples. The ultimate goal of his work is to provide a quick and accessible overview of the Chi-square test, making it an ideal resource for individuals with limited knowledge in statistics who wish to use this test as a beginner's guide.

presents a literature review on the relationship between income inequality and health, with a focus on assessing whether wider income disparities play a causal role in contributing to worse health outcomes. Earlier reviews yielded differing interpretations, but a majority of studies reported worse health in more unequal societies. This review includes more recent studies and adopts an epidemiological causal framework to analyze the evidence comprehensively [2].

The body of evidence strongly supports the idea that income inequality has a significant impact on population health and wellbeing. The review finds that the causal criteria of temporality, biological plausibility, consistency, and lack of alternative explanations are well supported. While a small minority of studies did not find an association, the review explains that many of these cases can be attributed to measurement issues, inclusion of mediating variables, subjective health measures, or short follow-up periods.

Overall, the evidence strongly indicates that large income differences have detrimental effects on health and social outcomes. As inequality continues to increase in many countries, the paper emphasizes that narrowing the income gap can lead to improved health and wellbeing for populations.

2. Methodology

2.1. Data Selection

The data is collected from GESIS [3]. The ISSP Social Inequality module series comprises five surveys conducted in 1987, 1992, 1999, 2009 and 2019. Successive surveys are always partial replications of earlier surveys. ISSP Social Inequality modules mainly deal with issues, such as attitudes towards income inequality, views on earnings and incomes, legitimation of inequality, career advancement by means of family background and networks, social cleavages and conflict among groups, and current and past social position. This report takes a global survey in 2019 of attitudes to inequality and looks at the 1724 responses made from the UK. A relatively large dataset is taken for a more general conclusion. With a reasonably good amount of data, the result is unlikely to be affected by anomalies. The much earlier data is ignored since it is out of date and cannot represent the opinions of the current public. The excessive dataset will also alter the results.

Three typical questions are taken to represent people's attitude towards the inequalities, which can be seen as people's attitude towards the inequalities, the effect of the inequalities, and a purposed solution to the inequalities. Three potential factors which may affect people's decisions, age, gender, and education level are raised and selected to compare. Since some of the results are taken as an error due to multiple factors, "N/A" is filtered and removed before the data is extracted and processed so as not to affect the final assertion.

Firstly, the needed columns are extracted from the dataset, ensuring that any missing values are removed to ensure data integrity. Next, the factors are categorized into two groups based on gender (male/female), educational level (degree/no degree), and age (35-/35+). This categorization enables a comprehensive examination of how these variables contribute to agreement levels. To simplify the analysis, the levels of agreement are transformed into agree/disagree categories. Any agreement level greater than a specified threshold is considered as agree. This simplification helps in comparing the agreement levels across different factors. For each category, the total counts of agree and disagree are calculated to understand the overall distribution. This provides insights into the dominant agreement/disagreement patterns within each category. Visual representation is crucial in understanding the data effectively. A contingency table is created to present the aggregated data, and expected frequencies are calculated. The bar charts are utilized to visualize the agreement percentages. Comparisons between different categories can be made based on the percentage distributions. To further explore the relationships between factors, a ratio analysis is conducted, and an expected value table is constructed. This allows for a deeper understanding of the expected agreement levels based on the observed data. Statistical analysis is performed using the Chi-square test, comparing the observed values with the expected values. The test is conducted at a significance level of 5% to determine the presence of any significant associations between the factors and agreement levels. Based on the results of the analysis, a conclusion is drawn, summarizing the findings and highlighting the key factors that influence agreement levels. The essay provides insights into the relationship between the identified factors and agreement, contributing to a better understanding of the topic at hand.

2.2. Chi-Square Testing

The chi-square test is a statistical method used to determine the association between categorical variables. This statistical method allows for the evaluation of the association between categorical variables and provides valuable insights in scientific research. The chi-square test employs a mathematical process involving the construction of contingency tables, calculation of expected frequencies, and determination of the chi-square statistic [4].

Firstly, a contingency table is constructed, which displays the observed frequencies of the variables being analyzed. Each cell in the table represents the count or frequency of a particular combination of categories from the variables.

Next, the expected frequencies are computed under the assumption of independence between the variables. This is done by applying the principle of probability to determine the expected counts for each cell if the variables were independent.

$$x^2 = \sum \left[\frac{(\mathrm{O} - \mathrm{E})^2}{\mathrm{E}} \right] \tag{1}$$

The chi-square statistic is calculated by taking the sum of the squared differences between the observed and expected frequencies, divided by the expected frequencies, as shown in the equation above, where O represents the observed frequencies in each cell of the contingency table and E represents the expected frequencies under the assumption of independence between the variables. The squared differences are summed across all cells in the contingency table to obtain the chi-square statistic.

This statistic follows a chi-square distribution, which has degrees of freedom equal to the number of categories minus one for each variable. The chi-square statistic follows a chi-square distribution with degrees of freedom equal to $(\mathrm{r}-1)\times(\mathrm{c}-1)$ where r is the number of rows and c is the number of columns in the contingency table. The degrees of freedom represent the number of independent pieces of information available to estimate the expected frequencies.

Additionally, p-values can be calculated to quantify the strength of the evidence against the null hypothesis. The p-value represents the probability of obtaining a chi-square statistic as extreme as the observed one, assuming the null hypothesis is true. If the p-value is below a predetermined significance level (0.05 in this research), the null hypothesis is rejected, indicating a significant association between the variables.

In this essay, we use the function CHISQ.TEST to proceed the chi-square testing [5].

The CHISQ.TEST function is a statistical tool in spreadsheet software that requires two essential arguments: Actual_range and Expected_range.

The Actual_range is a required input and refers to the range of data containing the observed observations that need to be tested against expected values. This range typically includes the frequency or count data obtained from real-world observations or experiments.

The Expected_range is another required input, and it represents the range of data that contains the expected values derived from the product of row totals and column totals divided by the grand total. In statistical terms, this represents the expected frequency or count for each cell in a contingency table under the assumption of independence.

When the Actual_range and Expected_range have unequal data points, the CHISQ.TEST function returns an #N/A error value.

$$x^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(A_{ij} - E_{ij})^2}{Eij} \tag{2}$$

The chi-square statistic is calculated using the formula shown, where A_{ij} represents the actual frequency in the i_{th} row and j_{th} column, and E_{ij} represents the expected frequency in the same cell. The values ‘r’ and ‘c’ correspond to the number of rows and columns, respectively. A lower chi-square value indicates a stronger indication of independence between the variables. The chi-square statistic is always non-negative, being exactly zero when each A_{ij} equals E_{ij} for all i and j.

CHISQ.TEST estimates the probability that a chi-square statistic value as extreme as the one computed by the formula occurred by chance, assuming independence between the variables. This probability is calculated using the chi-square distribution with an appropriate number of degrees of freedom. If both r and c are greater than 1, the degree of freedom equals $(\mathrm{r}-1)\times(\mathrm{c}-1)$. For the cases where r = 1 and c > 1, the degree of freedom becomes c - 1, and for r > 1 and c = 1, the degree

of freedom becomes r - 1. It is essential to note that the use of CHISQ.TEST is most appropriate when the expected frequencies (E_{ij}) are not excessively small, and some statisticians recommend a minimum threshold of 5 for each E_{ij}.

The chi-square test and the associated CHISQ.TEST function are valuable tools for determining the independence of categorical variables. Proper application and interpretation of the results are crucial, taking into consideration the number of data points, degrees of freedom, and the significance of the chi-square statistic for drawing meaningful conclusions about the relationship between the variables under study.

3. Conclusion

3.1. Data Analysis

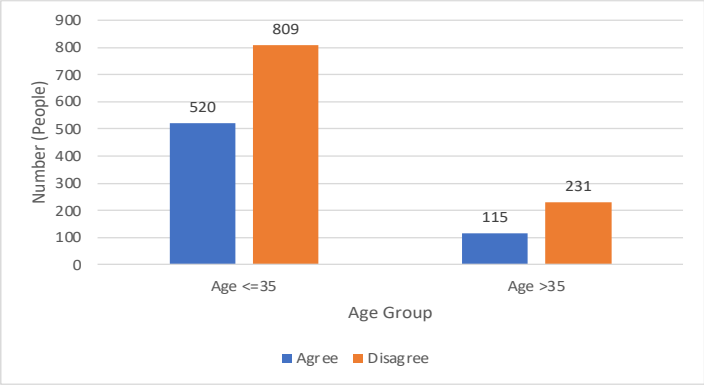


Fig. 1 Different Age Groups Attitude towards higher income better health care

When we compare the percentage of people in each group who have different opinions about whether it is fair for people with higher incomes to be able to afford better healthcare, we find that 39% of people in the age group 35 and under believe that people with higher incomes should be able to afford better healthcare, while 61% disagree, as shown in figure 1. Only 33% of people believe it is just that those with higher incomes can afford better health care, and this is because it is unjust for them to have this option. The majority of individuals concur that it is injustice to spend more money on better healthcare, however more persons over 35 share this opinion.

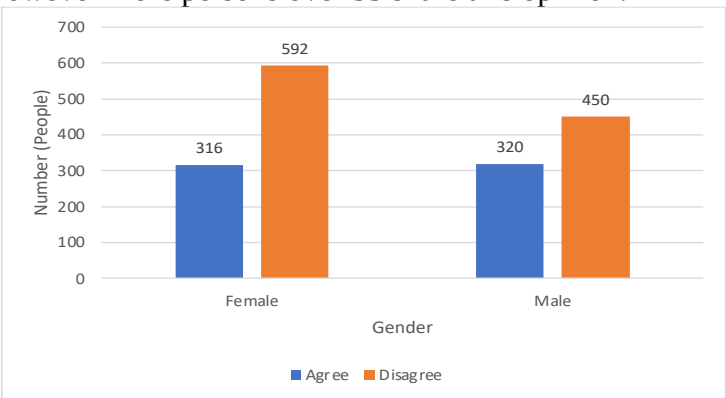


Fig. 2 Different Genders Attitude towards higher income better health care

In terms of gender, as illustrated in figure 2, 65% of women do not believe it is fair to purchase better healthcare with a higher salary, while 35% believe it is just to buy better healthcare with a higher income. Meanwhile, 58% of men believe it is unfair that a high income can afford better healthcare, while 42% believe it is just that a high income can afford good healthcare. As a result, women are more likely than men to believe it is injustice that a high salary cannot afford better healthcare. This might because people who are more likely to give a negative response when they have poor medical treatment experience. Women are less likely than males to receive CPR, but it is not the only way they are treated unfairly.

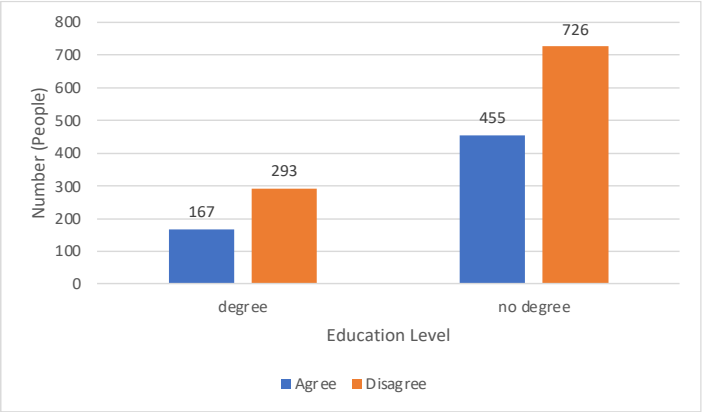


Fig. 3 Different Educational Level Attitude towards higher income better health care

Similarly, in figure 3, 64% of those with an education believe it is unfair that a high income will purchase better healthcare, while 37% believe it is fair. 61% of persons without a degree believe it is unfair that a higher income can purchase better health care, while 39% believe it is fair that a higher income can buy better health care. As a result, from figure 3, more educated people believe it is unfair to purchase better healthcare on a greater salary than uneducated people.

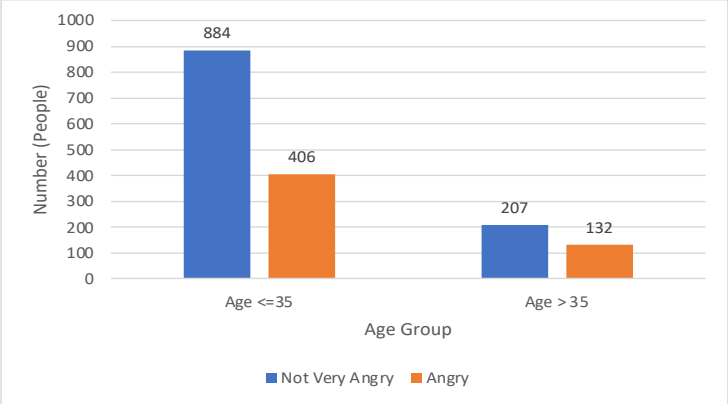


Fig. 4 Different Age Groups Attitudes towards Poor Countries People Allowed to Work in Wealthy Countries

We can see that 55% of people in the age group of 35 and under have a positive opinion on this issue, while 45% have the opposite opinion, when we compare the proportion of people in each group who have different attitudes to the question of whether it is acceptable for people from poor and backward countries to work in rich and advanced countries, as shown in figure 4. The remaining 38% regard this as acceptable. Most individuals concur that people from poorer countries migrate to work in wealthier nations, and this opinion is especially prevalent among those over 35.

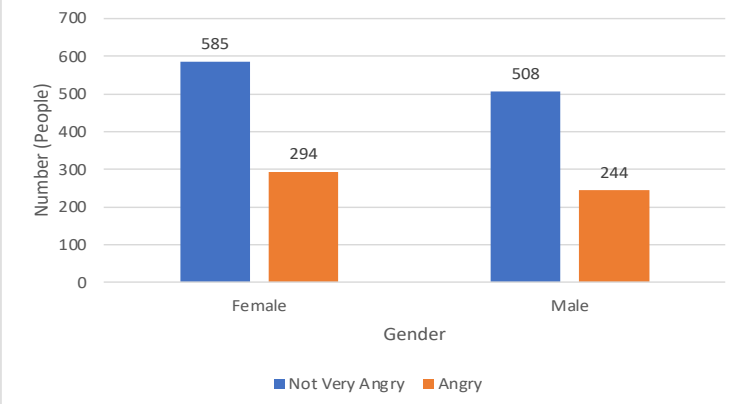


Fig. 5 Different Gender Attitudes towards Poor Countries People Allowed to Work in Wealthy Countries

In terms of gender, 41% of women, in Figure 5, dispute that individuals from poorer countries work in wealthier nations, compared to 59% of women who believe this to be true. In contrast, 53% of men believe that individuals from less affluent nations work in those from wealthier nations, while 47% of men do not. Therefore, women are more likely than men to concur that better health care cannot be purchased with higher incomes.

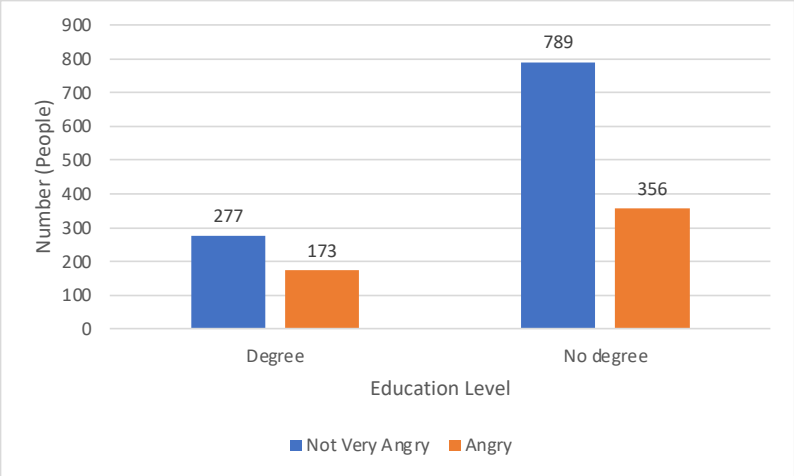


Fig. 6 Different Education Level Attitudes towards Poor Countries People Allowed to Work in Wealthy Countries

In the educated group, as shown in Figure 6, 35% of those with education do not believe that people from poor countries work in richer nations, while 65% of those with education believe that people from poor nations do. 53% of those without a degree believe that individuals from less wealthy nations may find employment in wealthy nations, while 47% of those without a degree disagree. Therefore, more people with credentials agree than those without credentials that people from less wealthy nations can work in wealthier nations.

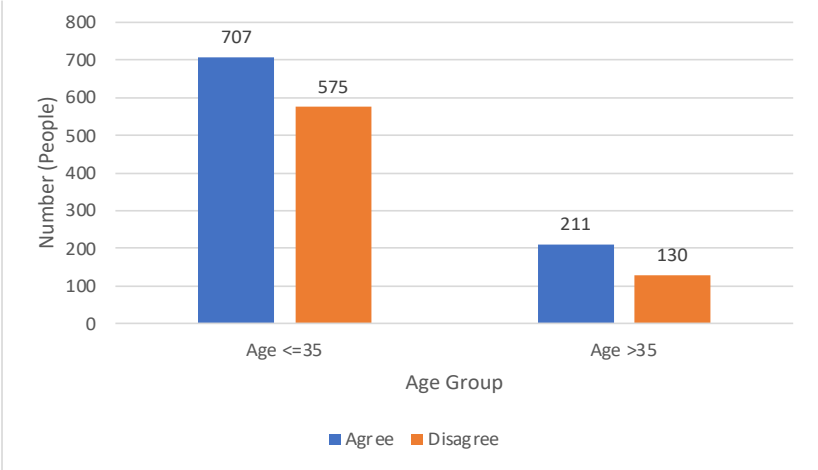


Fig. 7 Different Age Group Attitudes towards Difference in Wealth

When we compare the percentage of people in each group who have different attitudes towards the question of whether people feel angry about the difference between the rich and the poor, we can see that 69% of people in the age group of 35 and under accept the gap without anger on this issue, while 31% have the opposite opinion. In contrast, 39% of those aged 35 and older are angry about this issue, compared to 61% of those over 35. In Figure 7, the majority of people agree that they are not upset by the wealth gap's effects on society, and this viewpoint is particularly common among people under the age of 35.

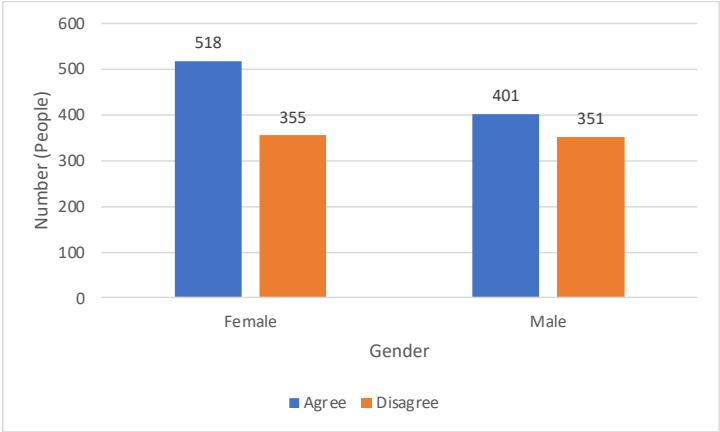


Fig. 8 Different Gender Attitudes towards Difference in Wealth

Comparing women by gender, 33% of them are furious about the wealth gap, while 67% of them are not. 32% of males are furious about the wealth gap, compared to 68% of men who are not. The majority of people agree that they are not unhappy about the income gap's effects, and men are particularly likely to hold this view, shown in Figure 8.

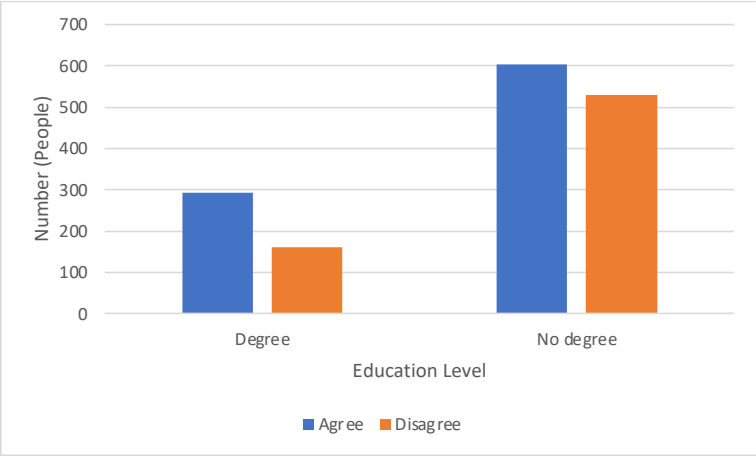


Fig. 9 Different Education Level Attitudes towards Difference in Wealth

In the highly educated group illustrated in Figure 9, 62% among those with education do not find the wealth disparity upsetting, but 38% of those with education find the wealth disparity upsetting. 31% of those without degrees do not accept the disparity in wealth, compared to 69% who do. As a result, despite the economic gap, persons with degrees are less angry than those without credentials.

3.2. Limitations

The chi-square test is a widely used statistical method for analyzing categorical data and assessing the association between variables. However, like any statistical tool, it has certain limitations that researchers need to be aware of when interpreting the results. Understanding these limitations is crucial for drawing accurate conclusions and avoiding misinterpretations [6].

An important assumption of the chi-square test is the assumption of independence. It assumes that the observations are independent of each other. Violation of this assumption, such as when the observations are correlated or clustered, can invalidate the results of the chi-square test. Therefore, researchers should ensure that their data meets the assumption of independence before using the chi-square test.

Sample size requirements are another limitation of the chi-square test. In order to produce reliable results, the test often requires a minimum sample size. If the sample size is too small, the test may have low power or result in inaccurate conclusions. However, if the sample size is too big, the results

conducted might not be able to represent appropriately. Alongside ensuring the sample size is sufficient for conducting a chi-square test, the size should not be too big.

The sensitivity of the chi-square test to small expected cell frequencies is another concern. When the expected frequencies in some cells are very low, the chi-square test may yield inaccurate results or fail to detect true associations. This limitation is particularly relevant when dealing with sparse or imbalanced data. In such cases, alternative tests or techniques, such as Fisher's exact test or simulation-based approaches, may be more appropriate.

Multiple comparisons are another challenge when using the chi-square test. Conducting multiple chi-square tests on the same data increases the likelihood of obtaining a significant result by chance alone. Researchers should be cautious when performing multiple comparisons and consider adjustments, such as Bonferroni correction, to control the overall type I error rate.

It is essential to remember that statistical significance does not imply causation. A significant chi-square result indicates that there is evidence of an association, but it does not establish a causal relationship between variables. Researchers should exercise caution when interpreting significant results and avoid making causal claims solely based on statistical significance.

While the chi-square test is a valuable tool for analyzing categorical data, it has limitations that must be considered. These limitations include the assumption of independence, sample size requirements, sensitivity to small expected cell frequencies, issues with multiple comparisons, and the need to avoid drawing causal conclusions based solely on statistical significance.

4. Results

The Chi-square hypothesis testing has provided valuable insights into the relationship between age, gender, education level, and attitudes towards inequality. When considering people's opinions on whether it is fair for those with higher income to buy better healthcare, we found a significant relationship between age and gender and their attitudes. However, education level did not show a clear association in this context.

Similarly, when examining people's feelings towards differences in wealth between the rich and the poor, we observed a relationship with age and gender, but education level did not play a significant role in shaping their attitudes.

On the other hand, when asking about the possibility for individuals from poorer countries to seek employment in prosperous nations, all three variables—age, gender, and education level—showed a relationship with people's views.

These findings highlight the complexity of attitudes towards inequality and how they can vary based on different factors. Age and gender appear to have more consistent associations with attitudes, whereas education level may not always be a determining factor.

Overall, the Chi-square hypothesis testing has shed light on the nuanced relationship between demographic factors and attitudes towards inequality. Understanding these relationships can aid policymakers and researchers in devising targeted strategies to address societal disparities and foster more equitable perspectives.

References

[1] Zibran, M. F. CHI-Squared Test of Independence. Department of Computer Science, University of Calgary, Alberta, Canada.

[2] Pickett, K. E., & Wilkinson, R. G. (2015). Income inequality and health: A causal review. Social Science & Medicine, 128, 316-326. ISSN 0277-9536. https://doi.org/10.1016/j.socscimed.2014.12.031.

[3] GESIS - Leibniz Institute for the Social Sciences. (n.d.). ISSP - Modules by Topic: Social Inequality. GESIS. https://www.gesis.org/en/issp/modules/issp-modules-by-topic/social-inequality (Accessed July 17, 2023).

[4] Rana, R., & Singhal, R. (2015). Chi-square test and its application in hypothesis testing. Journal of the Practice of Cardiovascular Sciences, 1(1), 69-71. https://doi.org/10.4103/2395-5414.157577

[5] Microsoft. (n.d.). CHISQ.TEST function. Microsoft Office Support. https://support.microsoft.com/en-au/office/chisq-test-function-2e8a7861-b14a-4985-aa93-fb88de3f260f

[6] Lewis, D., & Burke, C. J. (1949). The use and misuse of the chi-square test. Psychological Bulletin, 46(6), 433-489. https://doi.org/10.1037/h0059088