Employment Dynamics: Statistical Analysis & Tools

Vanshika Sharma
National College of Ireland
School of Computing
Dublin, Ireland
x23198389@student.ncirl.ie

I. Introduction

Employment plays a vital role in facilitating economic growth, minimises social differences, and shapes individual well-being. Creating jobs help the economy to grow by increasing the GDP. More the people are employed, the more demand increases which in a longer run, helps the country to grow and attain financial stability. This report studies the traditional 'age analysis' remains relevant in today's society, especially considering the factors like COVID-19. It also examines how age influences earnings, and how it affects the employment rate.

In this study, we aim to investigate the impact of earnings differentials within age groups and its broader implications for economic stability and growth. The distribution of earnings among various age groups become significant as long as societies undergo changes in age distribution, driven by factors such as aging populations and evolving labour market trends. Understanding the dynamics of earnings distribution across different age groups is a prominent step to comprehend the overall financial situation, particularly in the context of demographic shifts and workforce participation

Prior to the COVID-19 pandemic, global employment trends were characterized by relative stability, with existing disparities across different age groups in terms of their involvement in the workforce and income levels. The pandemic has brought about exceptional challenges across the globe, significantly impacting various aspects of society, including employment dynamics. In the middle of the challenges brought by the COVID-19 pandemic such as lockdowns and economic declination, the job market saw big changes. This affected how many people were employed and how much money they earned, especially across different age groups. In this study, we aim to investigate the multifaced effects of the COVID-19 pandemic on employment patterns, focusing specifically on its impact on the unemployment rate and mean earnings across different age groups.

The Consumer Price Index (CPI) is a fundamental tool in economic analysis, providing essential information on inflation, consumer behaviour, and economic equilibrium. A page by World Economic Forum[1] talks about the importance of CPI in determining various economic factors, such as finding inflation, adjusting interest rates and wage levels, etc. For Ireland, a country known for its dynamic economic landscape and rapid development, comprehending CPI data holds significant weight. This study aims to explore CPI's impact on growth, employment, and national development strategies in Ireland, while also investigating its role in decision-making processes like interest rate adjustments and wage levels. This research aims to provide insights into Ireland's economic condition and potential opportunities for immigrants.

In this study, various statistical methods and tests were employed to analyse interrelated datasets and draw conclusions which aligned with the research objectives. Basic methods were taken to check the status of data's normality, leading to its categorization into parametric and non-parametric datasets. Parametric tests such as t-tests were applied to analyse differences, while non-parametric tests like the Mann-Whitney U test and Kruskal-Wallis tests were utilised to handle non-normal distributions. Visualisation techniques such as clustered bar charts, Q-Q plots and line graphs were employed to depict trends and patterns effectively. Additionally, the selection of appropriate statistical tests was guided by careful consideration of the dataset's characteristics and research objectives, ensuring robust and accurate analyses.

In this study, various software tools were utilized for data analysis and implementation. Initially, Python scripting was employed to conduct the normality test on the datasets but it did not adhere to the CA description. Then, Real Statistics Resource Pack, which is compatible with Excel, was used for conducting statistical tests such as ANOVA. Python scripting was utilized to check for homogeneity in datasets. Visualization was carried out using Excel for clustered bar charts and line graphs. Additionally, Excel's Real Statistics add-in facilitated the execution of statistical analyses, ensuring robust and accurate results.

From the results of dataset which talked about the mean earnings per hour for different age, one could conclude that age difference can impact employment rates based on factors such as productivity and experience, leading to potential age discrimination as well sometimes. In research work by Broniatowska[2], the paper talks about there is a significant and negative relationship between wages and the proportion of older workers, the magnitude of this relationship differ depending on the specific occupational group. In occupational groups where skills and qualifications are crucial and require continuous updating, an increase in older workers leads to higher average wages. Mean earnings per hour can shape employment rates by affecting individuals' decisions to participate in or continue working.

From the result of one of the analysation which was based on the Consumer Price Index, one could conclude that with the growing years the prices of the goods also keep on increasing, leading to inflation. Moreover, it has an inverse relation with the GDP. According to The Irish Times[3], Ireland is expected to see an economic growth, with GDP to rise by 2.2% in 2024 and 3.8% in 2025. Employment is also expected to grow by 1.6% in 2024 and 1.8% in 2025 reflecting a robust labour market and high employment levels. Additionally, inflation is expected to be moderate, and interest rate changes are expected to improve economic prospects further.

Also, in an article by RTE[4], the paper talks about the positive impact of immigration on Ireland's workforce diversity and skill enhancement. Moreover, it emphasizes on the impact of immigrants by sharing the experience of Johnson Joseph who has contributed to sectors such as healthcare.

Moreover, after analyzation of a dataset, the result concluded was that there was a difference in the unemployment rates before and after COVID. After finding out the difference between unemployment rates between different age groups it was found that there was a declination in GDP. In the report by Ahmad et el[5], the report discusses about the impact of coronavirus on unemployment rates which concluded with the result indicating an increase in unemployment rates post-COVID. It was significantly due to lockdowns and business closures. Government interventions aimed to mitigate these effects, but recovery remains uncertain.

II. SAMPLE DATASETS

There were a variety of the datasets to choose from when it came for the selection of datasets. According to the description of the CA, the datasets could be collected from various official sites, including PISA[6], OECD.Stat[7], etc. On determining the objectives of the analysis, many datasets were picked from the site which talked about the Job Churn, Tax in Ireland, education, etc. Sticking to the motive, the datasets were narrowed down to three. These datasets gives information regarding the Consumer Price Index specifically in terms of food and non-alcoholic beverages, Mean earnings per hour between different age groups and Unemployment rates in different countries in Europe. These three datasets were closest to the research work and have been explained for the reasons in the specific paragraphs. Moreover, while working on these datasets, some unexpected findings were also noted. Due to personal preference and in order to adhere to the description of the CA, datasets were picked from The data was collected from the official site of Central Statistics Office[8]. The data was available to be downloaded in different formats, including JSON-stat, PX, XLSX, etc but CSV format was chosen because the analysis was to be conducted in excel. The data was available to be downloaded using filters which would adhere to the requirements of the analysis but the full dataset was chosen so that sampling could be done by hand as it is a good practice for future data scientist. Moreover, the datasets were available be viewed in different chart types like column, pyramid, polar area, radar, etc. Lastly, the data was available to be formatted in a map, with three different modes which were quantile, equidistant and k-means.

The first dataset selected was MIP11[9] which contained information of the Unemployment rates in Europe for countries like Finland, Germany, Hungary, Luxembourg, etc from years 2019 to 2022, the main aim to choose this dataset was to discover if there is a difference in the unemployment ratio between males and females, before and after COVID. This dataset was hard to interpret as it contained values for different European nations for different years. As the analysis was focused exclusively on Ireland and United Kingdom, data pertaining specifically to Ireland and United Kingdom as a country of interest was selected. The dataset contained information from year 2019 to 2022 but since the analysis was focused majorly on the impact of COVID-19, therefore, only 2019 and 2020 were chosen to be compared. Moreover, the initial dataset contained sex of male, female and others as well. The data overall was hard to be made into a sample data, since there were a lot of empty cells as well, therefore, countries which had values for all the specific domains that were being looked were observed. The sample dataset was created in such a way that the values for two years 2019 and 2020 were written with the values for each gender. The dataset was collected by Declan Smyth and was sourced from official page of Eurostat[10]

In second dataset was CPA01[11], the dataset was about Consumer Price Index specifically in the domain of food and no-alcoholic beverages. The dataset was collected by Anthony Dawson while doing a survey on finding the change in the average prices of consumer goods and services. The dataset was huge, therefore, filter had to be applied as the initial dataset contained had a lot of commodity group like clothing, footwear, housing, water, electricity, etc. Since food is the most basic need of a human, therefore, food & non-alcoholic beverages were chosen to be the core commodity group for analysation. Moreover, even after filtering out the data the dataset was hard to read as it contained irrelevant information for the years which were not used and there were many empty cells which weren’t useful and it included data of years from 1971 to 2023, therefore, the dataset was selected in a way that only the values from the years 2021 to 2023 were selected to give a better analysis of the recent situation. Since the data was already filtered out to one group of commodity, therefore, there wasn’t a need to get a sample dataset for this dataset. The intention behind the selection of data was to analyse if there is a difference in the consumer prices for food and non-alcoholic beverages among different years that is 2021, 2022 and 2023.

TABLE I SUMMARY OF DATASETS

Criteria	MIP11	CPA01	SES08
File format	csv	csv	csv
Size	48 KB	168 KB	80 KB
Number of attributes	444	245	28
Number of instances	6	5	5
Examples of instances	Statistic Label, Year, Sex, Countries, UNIT, VALUE	Statistic Label, Year, Commodity Group, UNIT, VALUE	Statistic Label, Year, Age Group, UNIT, VALUE
Complexity	Hard to understand	Moderate to understand	Easy to understand
Year	2024	2024	2024
Survey size	50,000	N/A	40,000
Principal Variables	Consumer Price Index and Harmonised Index of Consumer Prices	N/A	Gender, nationality, place of residence, place of work, education, etc.
Collected by	Declan Smyth	Anthony Dawson	Darragh Turner
Normality Test	Shapiro-Wilk test	Shapiro-Wilk Test	Shapiro-Wilk Test
Normality Test Result	Not Normally Distributed	Not Normally Distributed	Normally Distributed
Homogeneity	Homogeneous	Homogeneous	Homogeneous
Skewness	Not skewed	Negatively Skewed	Negatively Skewed
Hypothesis Test	Mann-Whitney U Test	Kruskal- Wallis Test	T-test
Nature of data	Non-Parametric	Non-Parametric	Parametric
Data Type	Integer, Decimal, String	Integer, Decimal, String	Integer, Decimal, String
Periodicity	Annually	Monthly	Annually

The cells with the non-relevant information were cut down and the rows were deleted. It included information of solely the food and non-alcoholic beverages as a commodity. Initial dataset had 5 columns and 245 rows. After selecting the sample dataset, the number of columns were the same but the rows were 15.

The last dataset selected was of mean earnings per hour and was named as SES08[12] as it was the reference on the official site from where it was downloaded. This dataset was collected by Darragh Turner while doing a survey with a purpose to produce data on the structure and distribution of earnings. This dataset contained the mean and median of the earnings paid per hour or weekly. The dataset was made to be structured in a way that it was easy to understand. Therefore, the data was filtered out to the means earnings per hour for different age groups and the analysation was taken on that basis. This dataset was selected and a sample dataset was also picked based on in accordance for the analysation. This dataset contained information regarding the mean earnings per hour between different age groups in the year 2018 and 2022. The primary idea to use this dataset was to analyse if there was a difference between the mean earnings among different age groups in two different years. Before conducting the analysis, the dataset was sorted out by year and the values were based on the age gaps. The values for the different age gaps were collected independently.

III. Analysis

The analysis of the research objectives involved a comprehensive interpretation of interrelated datasets, leading to the utilization and application of various statistical tests to draw conclusions. Fundamental steps were taken across all selected datasets, aligning with research objectives, along with consultation of relevant references. Subsequently, the normality of the sampled data was checked in order to find out the normality of the data and its suitability to conduct different tests. Based on the analysation of the normality, the datasets were divided into two categories, namely parametric and non-parametric. Parametric tests are the statistical tests that make certain assumptions about the population distribution from which the data is drawn, typically assuming the data to be normally distributed and showing homogeneity of variance using the Real Statistics tool , some of the examples are, t-test, Linear Regression, etc. Non-parametric tests, on the other hand, are distribution free tests that do not rely on any assumptions about the population distribution, some of the examples are, Mann-Whitney U test, Spearman's rank correlation coefficient, etc. Careful considerations were given for the selection of tests that would yield accurate results in accordance with the level of significance, enhancing the overall value of the research. This section of report provides a summary of the statistical tests employed and the challenges encountered during their execution.

In the first dataset MIP11, the dataset was about information of the unemployment rates in Europe for countries like Finland, Germany, Hungary, Luxembourg, etc from years 2019 to 2022. In the dataset having years 2019 and 2020, the mean values for the numeric attributes differed slightly, with 2019 having mean value as 6.72 and 2020 slightly higher as 6.85. Additionally, the standard deviation, also known as a measure of dispersion of data around the mean, showed a similar pattern, where for the year 2019, the standard deviation was reportedly 1.93 whereas it was slightly lower with a proximity of 1.30 for the year 2020. Skewness, which gives the measure of symmetry of the data distribution, was observed to be negative in the year 2019 and close to zero in 2020, indicating a nearly symmetrical distribution in the latter year. For both the years, the distributions were platykurtic. These summarizations offered insights into the central tendency, variability, symmetry, and shape of the dataset across the two years.

In visualizing the sample dataset, a line graph, Q-Q plot, and box plot were chosen to depict the unemployment rate trends between the years 2019 and 2020. The line graph offered a straightforward comparison of unemployment rates over time for the values, with both years' data lines plotted on the same graph, showcasing any observable trends between the two years. While Q-Q plot was considered, its separate graphs for each year made comparing and interpretation challenging as compared to a line graph. Talking about the line graph, it revealed a general pattern of higher unemployment rates in 2020 compared to 2019. Additionally, the box plot provided detailed insights into the distribution characteristics for each year, allowing comparisons of central tendency, spread, and presence of outliers.

After conducting the Shapiro-Wilk test for normality using the Real Statistics add-in in Excel, it was determined that the data for both years did not follow a normal distribution. This finding was beneficial as it indicated that non-parametric tests had to be applied in order to conduct analysis. Given that the dataset comprised only two independent groups, an alternative to the t-test was sought. The Mann-Whitney U test seemed to be the most apt choice for the test. The Mann-Whitney U test is a non-parametric test used to compare two independent samples when the assumption test of normality is violated. Since the data didn't follow a normal pattern for both years, the Mann-Whitney U test allowed to ensure robustness against deviations from normality while effectively evaluating potential differences between the two groups. In conclusion to the normality test conducted, the Mann-Whitney U test was selected as an appropriate statistical approach for analyzing the dataset.

The Mann-Whitney U test was chosen as the main statistical test for analyzing the dataset due to several factors. First, the Shapiro-Wilk test revealed that the data did not adhere to the assumption of being normally distributed, making traditional tests like t-test, ANOVA, etc. inappropriate. The Mann-Whitney U test, being a non-parametric test, did not rely on the normality assumption, making it perfect to face deviations from normality. Moreover, the dataset consisted of only two groups which were independent of each other. By choosing this test, it was easy to ensure reliability and accuracy in evaluation of potential differences between the two groups, therefore, based on the normality test and characteristics of the dataset, the Mann-Whitney U test was deemed the most appropriate statistical approach for analyzing the data.

Hypothesis testing is a fundamental concept in statistical analysis, providing a systematic framework for evaluating conjectures about population parameters based on sample data. At the core of hypothesis testing are the null and alternative hypothesis, which help to play askewness crucial role in guiding statistical analyses. The null hypothesis stated for this analysis was that there was no significant difference between the unemployment rates for different sexes before and after COVID. Symbolically, it could be represented as

$$Ho: \mu1 = \mu2$$

On the other hand, the alternative hypothesis poised that there was a difference between the unemployment rates for different sexes, indicating that the mean unemployment rate for one sex was not equal to the mean unemployment rate for the other sex, before and after COVID. Symbolically, it was represented as

$$Ha: \mu1 \neq \mu2$$

The significance level, denoted by α, represents the probability of rejecting the null hypothesis when it is actually true. In statistical hypothesis testing, α is typically set before conducting the test and serves as a threshold for determining whether results are statistically significant. In this case, choosing the significance level α involves balancing the risks of Type I and Type II errors. A Type I error occurs when the null hypothesis is rejected when it is actually true, implying that there is a significant difference between the unemployment rates for different sexes before and after COVID when, in fact, no such difference exists. On the other hand, a Type I error occurs when the null hypothesis is accepted when it actually wrong, suggesting no difference when there actually is one. Therefore, in order to adhere to the traditional practice and balance out Type I and Type II error, the level of significance was chosen to be 0.05.

The Mann-Whitney U test results revealed a statistically significant difference between the unemployment rates for different sexes before and after COVID. The Mann-Whitney U test results revealed a statistically significant difference between the unemployment rates for different sexes before and after COVID, U (18) = 18, $p < 0.05$. As U-statistic (U = 18) exceeded the critical value (U-critical = 5), the null hypothesis was rejected. Thus, it can be conducted that there was indeed a disparity in the unemployment rates for the different sexes. Therefore, the null hypothesis was rejected in favor of the alternative hypothesis, indicating significant changes in unemployment rates between the two time periods which were analyzed.

The decision to accept or reject the null hypothesis (Ho) was based on comparison of the obtained test statistic and the critical value or the p-value. In the Mann-Whitney U test, the obtained U-statistic (U= 18) was compared t the critical value (U-critical =5) for a significance level (α) of 0.05. If the obtained U-statistic exceeded the critical value, it suggested that the observed difference between the groups was statistically significant. In this case, the obtained U-statistic of 18 indicated a significant difference between the unemployment rates for different sexes before and after COVID as it was greater than the critical value of 5.

A journal by McCann et el[13] indicates that the COVID-19 pandemic has had significant impact on employment in Ireland, particularly for small and medium-sized enterprises (SMEs). It talks about the declination of 70% SMEs with a median decrease of 25%. Expenditure also decreased by 8.5% on average, with 40% of firms reducing spending. Additionally, it suggests that the uptake of policy support is more common among businesses most affected by the economic downturn, and smaller firms are less likely to utilize such support compared to larger firms.

In the second dataset CPA01 which talks about the Consumer Price Index, descriptive statistics were calculated to analyse the central tendency and variability of the data. The mean values for the years 2021, 2022, and 2023 were 74.28, 80.8, and 89.18, respectively. The measure of data dispersion around the mean, was approximately, 41.70, 41.44, and 44.46 for the years 2021, 2022, and 2023, respectively. For the values for three years, the skewness indicated asymmetrical distribution with negative values. The value for the peakiness of the distribution, was approximately 4.94 for three years. These statistics provide insights into the central tendency and variability of Consumer Price Index over the three-year period. The class frequency for each year's dataset reveals the occurrence count of individual values within the respective year. It turned out to be either one or zero for different values within the dataset.

For visualization, stacked column was chosen as it allows to compare multiple values for different groups (years in this case). Each column represents the total value for a specific year, and the total height of each column represents the total value for that specific year. Also, stacked column charts utilize more space effectively when it comes to having multiples categories(years) and several components (values) within each category. This representation emphasized to make the representation more clear and easy to read and interpret. Initially, clustered bar chart was chosen but then for the dataset that was being tested, it had numeric values across different years which wasn't suitable choice because clustered bar charts are a good choice for categorical data. Moreover, the dataset was already small, therefore making categorical groups was not a good option and was time consuming. The values of the Consumer Price Index peaked in 2023 indicating a gradual rather than sudden change. According to the official site of Central Statistics Office, the CPI[14] rose by 4.6% between December 2022 and December 2023.

To check the normality of the dataset, Excel's add in was used called RealStats, which was downloaded from the official site of Real Statistics. To add in the extension, various steps were followed. With the guidance of professor and using the help of the ppts in class, the extension was inserted. For the analysis, the level of significance was chosen to be 0.10 to adhere to the conventional practice and reduce Type I and Type II errors. There were various options to choose the normality test from however due to personal preference, Shapiro-Wilk test was chosen as it was already practiced in the class. Moreover, with the help of the tool, a table was to be formulated which would give a clear-cut answer if the values were normally distributed. After the conduction of the test, the W-stat value came out to be 0.602, 0.605 and 0.607 for the years 2021, 2022 and 2023, nonetheless, to interpret the normality, the focus was to check the p-value for each year, it was 0.00068 for 2021, 0.00075 for 2022 and 0.00078 for 2023 which concluded that the data was not normally distributed, hence the next question was to think of a suitable non-parametric test to be applied. Therefore, it was concluded that the data didn't appear to be normally distributed. This indicated that non-parametric tests had to be applied for the analysis. These findings provided a confidence in the statistical assumptions underlying further analyses and interpretations of the dataset.

Later, the homogeneity was checked for the datasets using Python script and it turned out that the data was homogenous. Since, it wasn't normally distributed, therefore, having assumptions wouldn't be a great idea. Therefore, non-parametric tests had to be applied. Most common tests were Mann-Whitney U test, Wilcoxon Signed-Rank Test and

Kruskal-Wallis Test. Among these, since the dataset included three different groups, therefore Kruskal-Wallis Test was the best suited as it allows to compare three or more independent groups. Therefore, for conducting the test, the first thing which was done was that the null hypothesis and alternative hypothesis were clearly specified. In Kruskal-Wallis test, the null hypothesis stated that the values of the commodities for food and non-alcoholic beverages were same for the three years, symbolically, it could be stated as

$$Ho: \mu 1 = \mu 2 = \mu 3$$

and the alternative hypothesis stated that the values of the commodities for food and non-alcoholic beverages were not the same for the three years, symbolically, it could be stated as

$$Ha: \mu 1 \neq \mu 2 \neq \mu 3$$

This was done because the goal was to check if there were any differences in the consumer price index (for food and beverages specifically) in three alternative years.

The level of significance chosen would normally be 0.05 for conducting a research test but the chosen value of alpha was 0.10 since the research was related to Consumer Price Index which is a complex economic indicator affected by various factors, and small changes may not always be immediately apparent. A significance level of 0.10 allows for more sensitive analysis, keeping in mind the fluctuations in CPI that might have been overlooked with a stricter significance level. Moreover, the previous level of significance of 0.05 did not align with the findings and references.

In the Kruskal-Wallis test, firstly, the rank was determined for all the values within the three groups. These ranks were then split down into three groups and there summation was noted down to be applied on the formula. The number of groups, k, was determined to be three (k=3). With the application of the formula the value of H-statistic was determined to be 5.78, on checking the degree of freedom (df), which was 2, the value of H-statistic was determined to be 4.605, on comparing the values, H-statistic was greater than H-critical, therefore, the null hypothesis was rejected or one could say that there was difference between the values of the commodities for food and non-alcoholic beverages in the years 2021, 2022 and 2023.

In an article on Investopedia by Fernando[15], the article talks about the relation of CPI with employment. It says that there is an inversely related connection between CPI and unemployment rate.

For the last dataset SES08, the mean earnings per hour was 21.4 in 2018 and was 25.7 in 2022. For the years 2018 and 2022, the skewness was reported to be negatively skewed. Moreover, the values of the standard deviation was reported to be 4.62 in 2018 and was 4.73 in 2022. Nonetheless, talking about the categorical attributes, the class frequency was the same for both of the years. However, no mode was available for the data.

For the visualization, initially Q-Q plot was utilized to depict the data's characteristics and trends. However, despite its potential, the plot did not contribute significantly to the visualization or interpretation of the data as it was hard to understand the data through these plots. The complexity of the dataset and the nature of variables hindered its effectiveness in conveying meaningful insights. Later a pie chart was chosen but no information could be refrained from it as there were too many categories in the dataset, making the pie chart cluttered and difficult to interpret, hence, as an alternative, clustered bar was opted for the visualization purpose. It was able to display and compare multiple categories simultaneously in a clear and intuitive manner. The clustered bar displayed the mean earnings per hour for different age groups in 2018 and 2022. The x-axis represented the age groups, while the y-axis represented the mean earnings per hour in accordance to the age group. The chart clearly showed an increase in the mean age across all age groups from 2018 to 2022. This illustrated changes in mean earnings, upon examination, it was evident that there was a general upward trend in mean earnings per hour across different age groups. The greatest difference in the mean earnings per hour was observed in the age group of 30-39 years. On the other hand, the least difference was observed for the age groups 50-59 years and 60 above. Overall, the chosen plots provided valuable insights into the dataset's trends and patterns, facilitating a comprehensive understanding of the data's characteristics. The implications from the graph could conclude to highlight a positive trend indirectly, referring to the economic growth[16].

For the normality test, Shapiro-Wilk test was conducted on the dataset representing earnings per hour across different age groups in 2018 and 2022. The result indicated that the data for both years did not significantly deviate from a normal distribution. In 2018, the Sharp-Wilk test statistic was 0.9183 with a corresponding p-value of 0.4564 while in 2022, the test statistic was 0.8950 with a p-value of 0.3016. For both years, the p-value exceeded the commonly used significance level of 0.05, leading to the final outcome which suggested that the distribution earnings per hour across various age groups in year 2018 and 2022 could be considered approximately normal. The tools used to conduct the normality test was Real Statistics Resource Pack[17] which was compatible with excel.

Talking about the selection of the test, there were a variety of tests to pick from. In order to select the most suitable one, firstly all the parametric tests were listed down, including ANOVA, Linear Regression, ANCOVA, etc. The sole purpose for parametric tests was that the assumptions made were fulfilled and these assumptions included criteria as being normally distributed and having homogeneity of variances. Since the dataset contained means of two independent groups, therefore depending upon the nature and characteristics of dataset, t-test was the most suitable test for analysing of data. The independent t-test provided a clear comparison of the means between two groups, allowing to determine whether there would be a statistically significant difference in the mean earnings per hour between 2018 and 2022. Moreover, the results of the independent t-test were easy to interpret. This test provided a t-statistic and a p-value, which allowed to determine whether the difference in means between two groups was statistically significant at a given level of significance.

For the analysation of the t-test, two hypothesis had to be given in order to conduct the research. The null hypothesis stated that the mean earnings per hour for different age groups was more in the year 2022 than in 2018. In the generic equation it could be stated as

$$Ho: \mu 1 < \mu 2$$

whereas the alternative hypothesis stated that the difference in the mean earnings per hour for different age groups in the year 2018 was less than or equal to 2022. In generic equation, it could be stated as

$$Ha: \mu 1 \geq \mu 2$$

The level of significance, which is the probability of rejecting the null hypothesis when it is true, was chosen to be 0.05. The value of alpha was chosen to be 0.05 as it is a standard and a common practice in many fields of research. It is a convention that is commonly accepted in order to balance out Type I and Type II errors.

Based on the chosen significance level of $\alpha = 0.05$ and the one-tailed nature of the test, the calculated p-value ($p=0.056$) was compared to the significance level. Since the p-value exceeded the chosen significance level of 0.05, the null hypothesis couldn't be rejected. Therefore, there was insufficient evidence to conclude that the mean earnings for different age groups in 2022 were significantly less than or equal to those in 2018. Thus, in accordance with the data and the specified level of significance, we did not reject the null hypothesis and concluded that the mean earnings per hour for different age groups was more in the year 2022 than in 2018. In a research work by Klevmarken[18], the paper revealed the relationship between physical age, work experience, and income in context of human capital theory. It helped to conclude that age difference has a significant impact on the employment rate.

Since the value for t-statistic after the overall test came out to be -1.73, on comparison with the t-critical for a one-tailed which was 1.78, the outcome was the failure to reject the null hypothesis. In other words, the mean earning for different age groups in the year 2022 was more than the mean earning for different age groups in 2018. Hence, the null hypothesis was accepted and alternative hypothesis was rejected.

REFERENCES

[1] "What is the Consumer Price Index and why is it important?," World Economic Forum. Accessed: Apr. 28, 2024. [Online]. Available: https://www.weforum.org/agenda/2022/05/what-is-the-consumer-price-index/

[2] P. Broniatowska, A. Majchrowska, and M. Nasiński, "Age Structure of Employment and Wages. An Analysis Across Occupational Groups," 2020.

[3] "Irish economy expected to resume strong growth path in 2024, predicts EY," The Irish Times. Accessed: Apr. 30, 2024. [Online]. Available: https://www.irishtimes.com/business/2024/02/08/irish-economy-expected-to-resume-strong-growth-path-in-2024-predicts-ey/

[4] W. Goodbody, "Immigration and its growing role in Irish labour force," Feb. 2024, Accessed: Apr. 30, 2024. [Online]. Available: https://www.rte.ie/news/business/2024/0202/1429998-immigration-employment/

[5] M. Ahmad, Y. A. Khan, C. Jiang, S. J. H. Kazmi, and S. Z. Abbas, "The impact of COVID-19 on unemployment rate: An intelligent based unemployment rate prediction in selected countries of Europe," *Int. J. Finance Econ.*, vol. 28, no. 1, pp. 528–543, 2023, doi: 10.1002/ijfe.2434.

[6] "Database - PISA." Accessed: Apr. 27, 2024. [Online]. Available: https://www.oecd.org/pisa/pisa-for-development/database/

[7] "OECD Statistics." Accessed: Apr. 27, 2024. [Online]. Available: https://stats.oecd.org/

[8] "Data." Accessed: Apr. 21, 2024. [Online]. Available: https://data.cso.ie/

[9] D. Smyth, "MIP11 - Unemployment rates." Accessed: Apr. 22, 2024. [Online]. Available: https://data.cso.ie/

[10] "Statistics | Eurostat." Accessed: Apr. 27, 2024. [Online]. Available: https://ec.europa.eu/eurostat/databrowser/view/lfsa_urgan/default/table?lang=en

[11] "CPA01 - Consumer Price Index." Accessed: Apr. 21, 2024. [Online]. Available: https://data.cso.ie/

[12] D. Turner, "SES08 - Mean and Median earnings per hour and paid weekly hours." Accessed: Apr. 15, 2024. [Online]. Available: https://data.cso.ie/

[13] F. McCann, C. O'Toole, M. Lawless, J. Kren, and J. McQuinn, "New Survey Evidence on COVID-19 and Irish SMEs: Measuring the Impact and Policy Response," *Econ. Soc. Rev.*, vol. 52, no. 2, Summer, Art. no. 2, Summer, Jun. 2021.

[14] "Consumer Price Index December 2023 - CSO - Central Statistics Office." Accessed: Apr. 24, 2024. [Online]. Available: https://www.cso.ie/en/releasesandpublications/ep/p-cpi/consumerpriceindexdecember2023/

[15] J. Fernando, "Consumer Price Index (CPI): What It Is and How It's Used," Investopedia. Accessed: Apr. 25, 2024. [Online]. Available: https://www.investopedia.com/terms/c/consumerpriceindex.asp

[16] "World Bank Open Data," World Bank Open Data. Accessed: Apr. 24, 2024. [Online]. Available: https://data.worldbank.org

[17] "Home Page (Welcome) | Real Statistics Using Excel." Accessed: Apr. 24, 2024. [Online]. Available: https://real-statistics.com/

[18] A. Klevmarken and J. M. Quigley, "Age, Experience, Earnings, and Investments in Human Capital," *J. Polit. Econ.*, vol. 84, no. 1, pp. 47–72, Feb. 1976, doi: 10.1086/260409.