De La Salle University

Presented to the Department of

De La Salle University-Manila

Term 3, A.Y. 2023-2024

In partial fulfillment of the requirements in the course

STT101A

Relationship Between Unemployment and Mental Illness

Submitted by:

Malcolm Joaquin Cuady

John Kirbie Mendoza

Rashde Sy

Pulvert Valdez

Ericson Adler Tan

Submitted on July 28, 2024

Declaration of Generative AI and AI- assisted Technologies in the Writing Process

During [Chapter 2] [ChatGPT, Grammarly] was used to help check for grammar errors in the sentences for professionalism and for some paraphrasing.

During the preparation of [Chapter 4] of this work, the author(s) used [ChatGPT, Grammarly]. [ChatGPT is used to help in cleaning the data and for Data Preparation with the help of using python, Grammarly is used for paraphrasing and summarizing].

After using this/these tool/s/service/s, the
author(s) reviewed and edited the content as needed and take(s) full
responsibility for the content of the publication."

Signature over printed name:

Malcolm Joaquin L. Cuady

John Kirbie G. Mendoza

Rashde W. Sy

Pulvert Valdez

Ericson Adler Tan

Introduction

A. Background of the Study

The study examines the relationship between mental health and employment using material that includes demographic information, mental health conditions, and employment status. Mental illness often leads to higher unemployment rates and employment instability, exacerbated by employment gaps and legal disability (Luciano & Meara, 2014). The purpose of the research is to find the relationship between unemployment and mental health illness, to investigate the extent to which mental health conditions are associated with gaps in employment history, and to examine whether having regular access to a computer and the internet is associated with better employment outcomes for individuals with mental health conditions. These goals will help understand how mental health affects the employability of individuals. The results of this research will be able to give insight to individuals about the effects of mental health and other factors to employability.

B. Research Objectives

1. To describe the demographic characteristics of individuals within the sample.
2. To present statistics that show the association between unemployment and mental illness
3. To determine the association between unemployment and mental illness
4. To investigate the extent to which mental health conditions are associated with gaps in employment history.

C. Scope and Limitation

The dataset provides a comprehensive view of various demographic attributes, mental health indicators, employment status, technology access, and living situations. It includes information on age, gender, education level, household income, and regional distribution, as well as self-reported mental health conditions, hospitalization history, and legal disability status. Employment status and gaps in resumes are also covered, alongside data on access to computers, the internet, and the type of device used. Additionally, it notes whether individuals live with their parents or independently.

However, there are certain limitations for this study such that this dataset is collected through a survey, thus subjected to human biases. The geographical representation may be uneven, limiting the generalizability of findings across different regions. In addition to that, the dataset focuses on the presence of mental health issues but lacks detailed information on their severity and duration.

D. Significance of the Study

Studying and analyzing this data is essential for understanding how demographic factors affect mental health and resource availability, which can inform policy and lead to targeted mental health services and interventions. Insights gained can help organizations allocate resources more effectively, prioritize those most in need, and develop workplace mental health programs. Additionally, understanding the importance of technology use in managing mental health conditions can guide the design of digital mental health interventions, advancing public health research by revealing patterns and predictors of mental health problems and contributing to more informed decisions and intervention strategies.

Despite being based on a foreign demographic, this research is highly relevant to the Philippine context. By comparing demographic and mental health patterns, researchers can identify unique factors influencing mental health in the Philippines, aiding in targeted interventions and resource allocation. The data can guide the development of digital mental health interventions suited to the local technological landscape and inform workplace mental health programs tailored to Filipino workers. Adapting these insights to the Philippine setting can bridge gaps in local research and practice, significantly improving mental health outcomes across the country.

II. Review of Related Literature

Unemployment and Mental Health

- **Correlation between Unemployment and Mental Health Issues**

  Numerous studies have established a significant correlation between unemployment and mental health issues. Unemployment is frequently cited as a risk factor for developing or exacerbating mental health conditions, such as depression and anxiety. Conversely, individuals with mental health conditions often experience higher rates of unemployment due to various barriers.

- **Impact of Mental Health on Employment Opportunities**

  Research indicates that mental health conditions can negatively affect an individual's ability to secure and maintain employment. Factors contributing to this include the severity of the mental health condition, gaps in employment history, and legal disability status. For example, Luciano and Meara (2014) found that individuals with mental health conditions were significantly less likely to be employed compared to those without such conditions, highlighting the need for targeted interventions.

Barriers to Employment for Individuals with Mental Health Conditions

- **Social Stigma and Discrimination**

  Social stigma surrounding mental health conditions is a critical barrier to employment. Brouwers (2020) emphasized that stigma leads to discrimination in hiring practices and workplace exclusion, limiting career advancement opportunities for individuals with mental health issues. This stigma can discourage individuals from seeking employment or disclosing their mental health status, further complicating their job prospects.

- **Legal and Institutional Barriers**

  Legal disability status and institutional barriers, such as inadequate mental health support in the workplace, also hinder employment for those with mental health conditions. These barriers can manifest in various forms, including lack of reasonable accommodations and support services, which are essential for enabling individuals with mental health conditions to work effectively.

Role of Technology in Employment Outcomes

- **Access to Digital Resources and Employment**

  Access to computers and the internet has been shown to positively influence employment outcomes for individuals with mental health conditions. The study by Too, Leach, and Butterworth (2020) suggested that regular internet access can provide valuable resources for job searching, skill development, and social support, which are critical for enhancing employability.

- **Digital Inclusion and Mental Health Management**

  The integration of digital tools and resources can also support mental health management, thereby indirectly improving employment outcomes. Digital inclusion efforts, such as providing affordable internet access and digital literacy training, can help mitigate some of the employment challenges faced by individuals with mental health issues.

Integrated Approaches to Mental Health and Employment Support

- **Comprehensive Support Systems**

  Integrated approaches that combine employment support with mental health services are essential for addressing the dual challenges of unemployment and mental health issues. The Health Foundation (n.d.) highlighted the need for comprehensive support systems that address both mental health treatment and employment assistance, promoting overall well-being and helping individuals break the cycle of unemployment and mental health problems.

- **Policy Implications and Interventions**

  The findings show the importance of policy interventions that prioritize mental health in employment strategies. This includes enhancing workplace mental health programs, improving access to mental health services, and reducing stigma through education and awareness campaigns.

Summary

The reviewed literature highlights the complex relationship between unemployment and mental health, identifying several key themes. Firstly, individuals with mental health conditions are more prone to experiencing unemployment and employment instability. The availability of regular access to technology, such as the Internet, can have a positive impact on employment outcomes for these individuals. However, social stigma remains a significant barrier to employment, underscoring the need for targeted interventions to reduce discrimination. The literature emphasizes the importance of comprehensive support systems that integrate employment services with mental health care to enhance the employability of those with mental health conditions.

III. Methodology

Data Set Acquisition

The data set that will be used in this paper is a secondary data that was acquired from Kaggle, a major online hub for data scientists and machine learning enthusiasts. Kaggle offers a vast repository of datasets uploaded by users and organizations, covering broad topics and industries. Each dataset comes with metadata, such as the source, description, and instructions for use. Transparency from the metadata helps users understand the context, reliability, future use, and limitations of the data. Many datasets on Kaggle are provided by reputable organizations, including corporations, research institutions, and government agencies. Besides, it maintains an active community of different fields which helps in flagging any issues with published datasets, ensuring a significant and true analysis.

Details on Study

The acquired data set was published by Michael Corley, a data scientist at Sigma Data Science. Based on the dataset’s description, it is a survey exploring the relation between unemployment and mental illness. Although according to the author, similar surveys was conducted by National Alliance on Mental Illness (NAMI) that caught its positive significance, Corley's survey expounds more on its causation. The researcher took stratified sampling from a general population which did not target a specific demographic in order to avoid skewness of results. It has been stratified to characteristics like income and location. The balanced proportions of the number of respondents with mental illnesses and respondents without mental illnesses stayed consistent among different sized samples.

Data Cleaning

The data set is cleaned through the use of functions in the pandas library of python. The missing values of the data set are filled through the use of mean median mode imputation because of the nature of the data collected. In addition to that, we disregarded samples that had more than 50% missing values.

Data Analysis

A program was coded using the Python programming language to be used for data analysis. Different libraries were used for the implementation; Pandas to manipulate statistical data, NumPy for mathematical operations, Matplotlib and Seaborn for the graphs and visualization. Additionally, Microsoft Excel and PHStat functions were used in double-checking results that were acquired from the Python program.

IV. Results and Discussion

Descriptive Statistics

This section presents the descriptive statistics of individuals with mental health conditions, including mean age, gender distribution, and educational attainment. By summarizing these variables, the analysis aims to highlight patterns and trends that influence employability, helping to achieve the research objective.

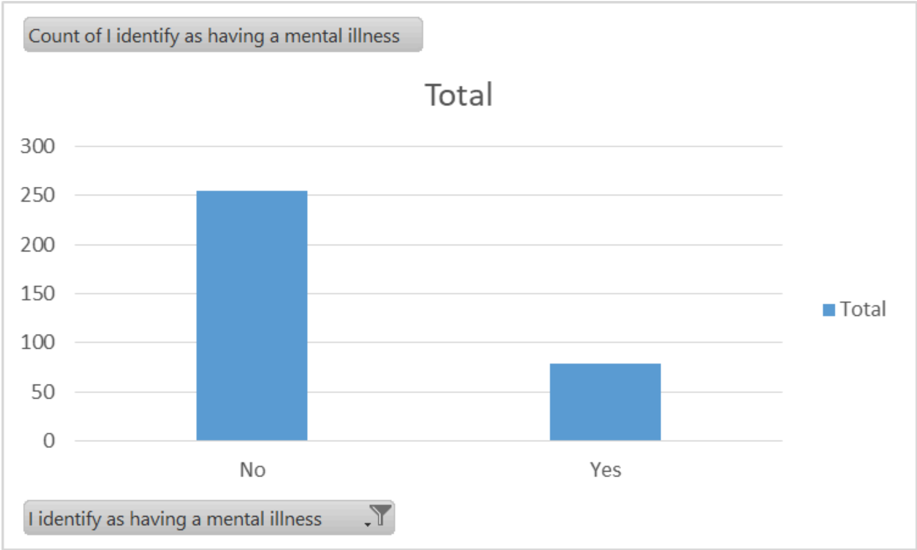

Fig. 1.0: Count of sample that has mental illness

This figure illustrates the distribution of responses to the question "Do you identify as having a mental illness?" The data reveals that out of a total of 333 respondents, 254 indicated "No," whereas 79 indicated "Yes." This highlights that a significantly larger portion of the surveyed population does not identify as having a mental illness compared to those who do.

	Mean Age
Mean	45.77177
Median	53
Mode	37
Minimum	24
Maximum	62
Range	38
Variance	175.7249
Standard Deviation	13.2561
Coeff. of Variation	28.96%
Skewness	-0.2817
Kurtosis	-1.2507
Count	333
Standard Error	0.7264

Table 1: Descriptive statistics of Mean Age

The table is generated through PHStat and it provides descriptive statistical measures for the ages of 333 individuals in a sample. The mean age is approximately 45.77 years old, with a median age of 53 years old. The mode, or the most frequently occurring age, is 37 years. Ages in the sample range from a minimum of 24 to a maximum of 62, with a range of 38 years. The variance is 175.7249, and the standard deviation is 13.2561, showing a moderate spread around the mean. The coefficient of variation is 28.96%, highlighting the relative variability of the ages. The skewness is -0.2817, indicating a slight left skew in the age distribution. Kurtosis is -1.2507, suggesting a relatively flat distribution compared to a normal distribution because of a negative kurtosis value. The standard error of the mean is 0.7264, reflecting the precision of the sample mean estimate.

Gender Distribution

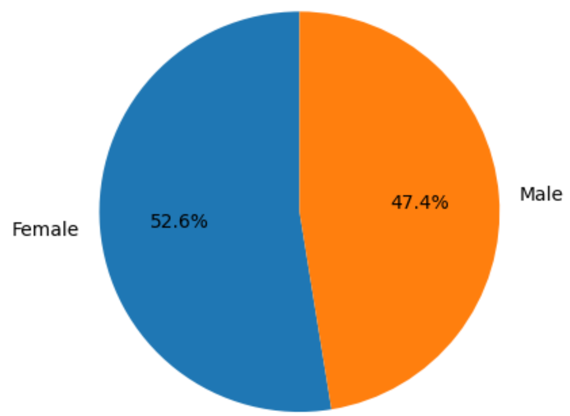

Fig. 1.1: Gender distribution of sample

This figure is a pie chart that represents the gender distribution in a sample. Here's the interpretation:

- **Female**: 52.6% of the 333 sample is female, represented by the blue section of the pie chart.
- **Male**: 47.4% of the 333 sample is male, represented by the orange section of the pie chart.

The figure indicates that there are slightly more females than males in the sample, with females constituting a majority.

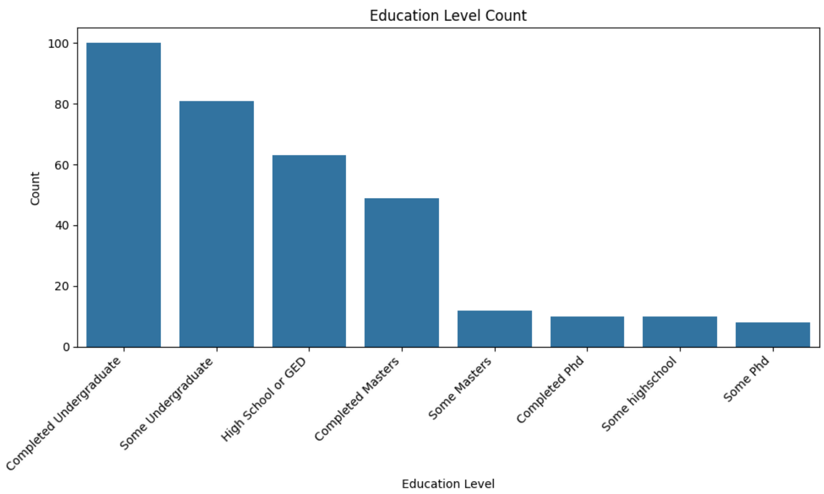

Fig. 1.2: Distribution of Education level of Sample

This figure presents the educational attainment levels of a sample population. The largest group comprises individuals who have completed their undergraduate degrees, totaling 100. This is followed by those with some undergraduate education, numbering 81, and those who have completed high school or obtained a GED, at 63. The sample also includes 49 individuals who have completed their master's degrees and 12 who have undertaken some master's coursework. Additionally, there are 10 individuals each who have completed their PhD or some high school education. The smallest group consists of those who have completed some PhD coursework, with 8 individuals. This distribution highlights a diverse range of educational backgrounds within the sample, with a notable concentration at the undergraduate level.

Hypothesis Testing

A z-test for differences in two proportions can be used to compare the difference in unemployment between two distinct groups: those who identify as having a mental illness and those who do not.

H0: There is no significant difference in unemployment between those who identify as having a mental illness, and those who do not.

Ha: There is a significant difference in unemployment between those who identify as having a mental illness, and those who do not.

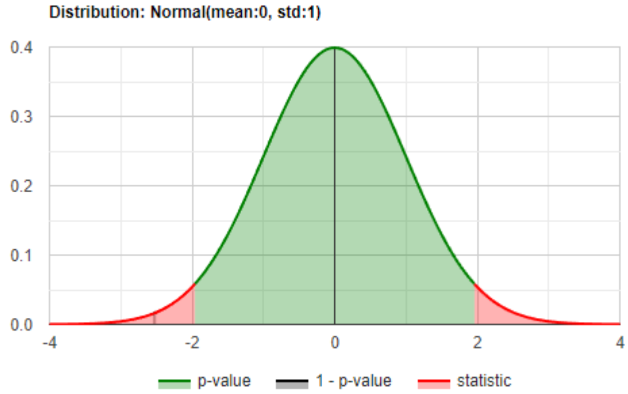

Fig. 2.0: Normal Distribution of people with mental illness

Using a two-tail normal distribution**,** it is assessed if there is a significant difference in unemployment between those who identify as having a mental illness, and those who do not. *(z-Test = -2.5306, Critical Value = ±1.9600, p-Value = 0.0114)*

Since the z-Test = -2.5306 < Lower Critical Value = -1.9600, *H0* is rejected.

Interpretation: There is enough evidence of significant difference in unemployment between those who identify as having a mental illness, and those who do not.

Moreover, a t-test for pooled variances can be used to compare the difference in total length of any resume gaps (in months) between two distinct groups: those who identify as having a mental illness and those who do not.

H0: There is no significant difference in total length of any resume gaps (in months) between those who identify as having a mental illness, and those who do not.

Ha: There is a significant difference in total length of any resume gaps (in months) between those who identify as having a mental illness, and those who do not.

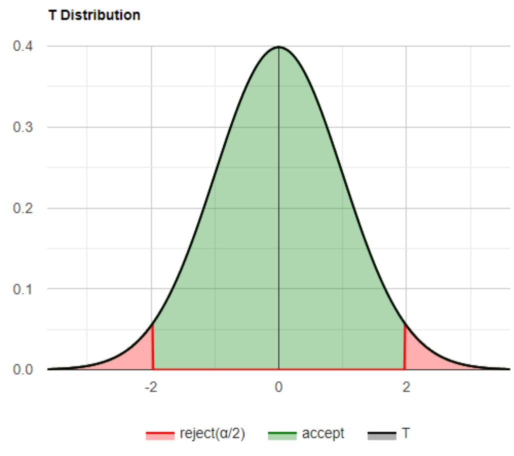

Fig. 2.1: T-distribution of resume gap

Using two-tail t-distribution, it is assessed if there is a significant difference in total length of any resume gaps (in months) between those who identify as having a mental illness and those who do not. *(t-Test = -4.1571, Critical Value = ±1.9672, p-Value = 0.000041)*

Since the t-Test = -4.1571 < Lower Critical Value = -1.9672, *H0* is rejected.

Through this analysis there is enough evidence of significant difference in total length of any resume gaps (in months) between those who identify as having a mental illness, and those who do not.

Chi-Square Test

Chi-square tests are used to fulfill our research objectives by allowing us to explore the associations between various categorical variables. For instance, testing the independence between 'gender' and 'I identify as having a mental illness' helps in understanding if mental illness prevalence varies across genders. Examining the association between 'education level' and 'employment status' provides insights into how educational attainment influences employment outcomes, which is crucial for understanding the broader relationship between mental health and employment. These analyses collectively help us understand the complex interplay between mental health and employment thereby addressing our research objectives comprehensively. The following Chi-square test uses the following hypotheses:

H0: Y and X is independent of each other

Ha: Y and X is dependent of each other

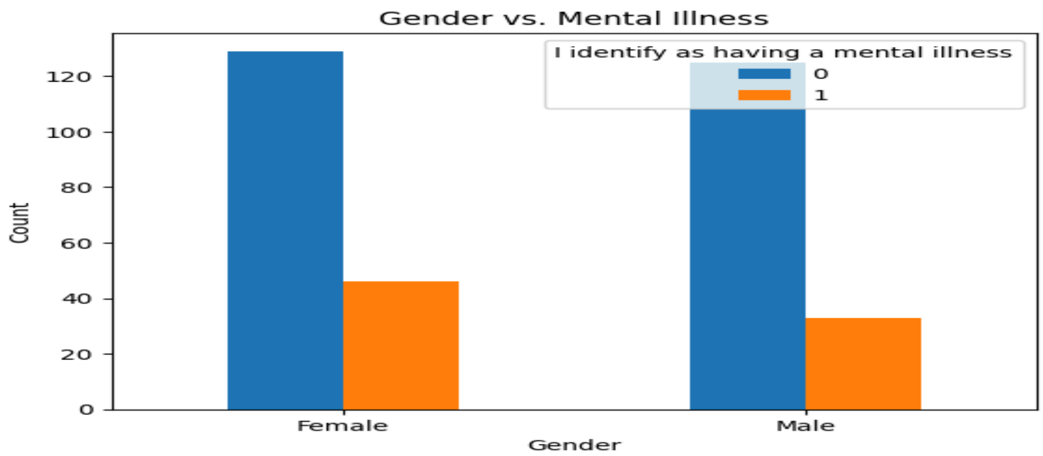

Fig. 3.0: Proportion of Genders that has mental illness

This figure shows the results of a chi-square test, with a chi-square statistic of 1.3379 and a p-value of 0.2474, and a critical value of 3.8415. These results indicate that there is no statistically significant association between gender and mental illness. The p-value is greater than the significance level (0.05), meaning that the correct action would not reject the null hypothesis. Therefore, the data do not provide sufficient evidence that gender and mental illness are related; rather, they appear to be independent variables.

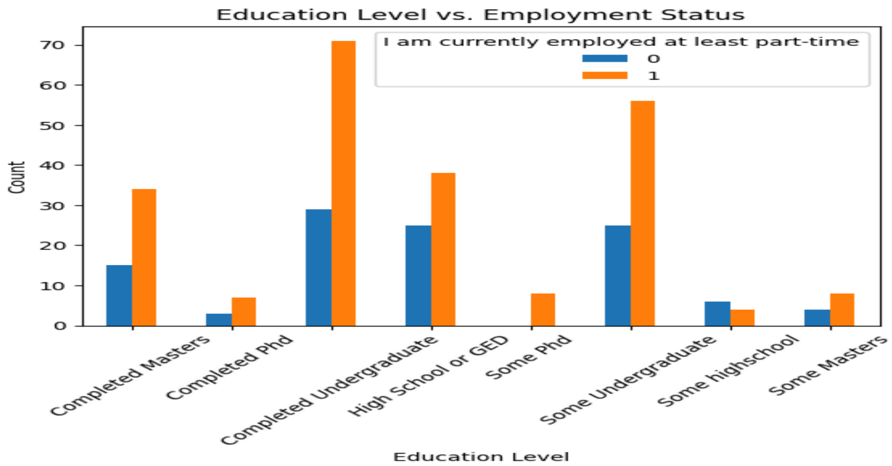

Fig. 3.1 Proportion of different education level that are employed

The results of the chi-square test, with a chi-square statistic of 9.5860 and a p-value of 0.213, and a critical value of 14.06714 indicating that there is no statistically significant association between education and employment status. Due to the p-value being greater than the level of significance (0.05), resulting in not rejecting the null hypothesis. Therefore, the data do not provide sufficient evidence that education and employment status are related; rather, they appear to be independent variables.

Based on the chi-square tests conducted, there was no statistically significant evidence to support a relationship of dependence between gender and mental illness ($\chi^2 = 1.3379$, $p = 0.2474$). Similarly, education level and employment status ($\chi^2 = 9.586$, $p = 0.213$) was found to be independent of one another. Therefore, based on these analyses, both gender and education level appear to be independent of mental illness prevalence and employment status, respectively.

ANOVA Results Interpretation

The study used Analysis of Variance (ANOVA) to investigate whether there were statistically significant differences in the identification of mental illness based on employment status. Participants were divided into two groups: those employed at least part-time and those not employed. This statistical method is advantageous for comparing means across multiple groups while mitigating the risk of inflated Type I errors from multiple t-tests (Field, 2013; Howell, 2010). The primary hypothesis was to determine if there is a significant difference in mental illness identification between the currently employed (at least part-time) and the unemployed. The sample consisted of 666 participants, with 333 in each group, and the analysis was conducted at a 0.05 significance level.

Statistic	Value
Between Groups Variance (SS)	32.445
Degrees of Freedom (Between Groups)	1
Within Groups Variance (SS)	132.878
Degrees of Freedom (Within Groups)	664
Total Variance (SS)	165.323
Degrees of Freedom (Total)	665
Mean Square Between (MSB)	32.445
Mean Square Within (MSW)	0.2001
F-Statistic	162.1359
P-Value	< 0.0001 (2.19E-33)
F critical	3.85501

Table 2: Results of ANOVA

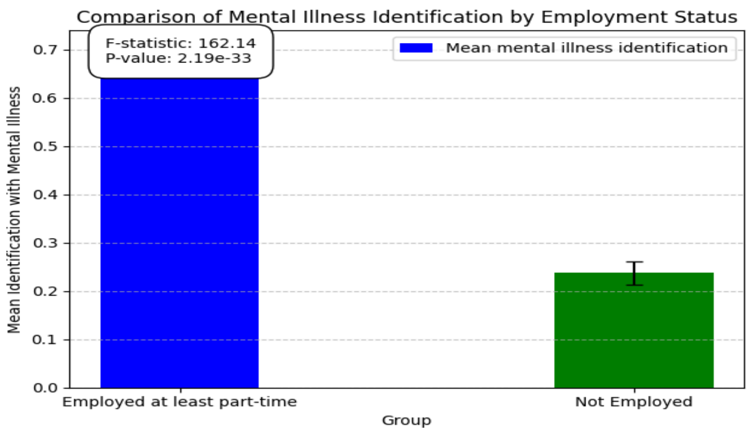

Fig. 4.0: Comparison of Mental Illness and Employment

F-statistic conducted ANOVA was highly significant with a value of 162. 1359 and an assisted p-value of < 0. Thus, the value of F equals 0001, which supports the rejection of the null hypothesis suggesting the absence of a significant difference in the overall accuracy of mental illness identification between the employment status groups. The obtained p-value of 2.19E-33 means that there is a statistically significant effect of the employment status on the identification of mental illnesses among the participants.

The F-statistic directly stands for the magnitude of the effect and, when compared with the critical F-value of 3. 85501, it is evident that the difference in means between the groups is unlikely to have occurred by chance and is most probably due to the employment status. The findings of this analysis based on the ANOVA indicate that there is a significant difference of employment status at least the part time employment in relation to the rates of self-identification of being mentally ill compared to not being employed at all. This result implies the psychological consequences of employment status and conceivably indicates that employment, owing to factors including stress or job demands, plays a role in how people see and manage mental health disorders.

Tukey-Kramer Multiple Comparisons Test(Post Hoc)

The Tukey-Kramer post hoc test was used to ascertain which of the groups was responsible for the significant differences after having an F statistic that suggested there was a significant difference in the means after the ANOVA had been conducted. This test is quite valuable because it eliminates the possibility of committing a Type I error especially when making several comparisons. For this comparison, we selected the Tukey-Kramer test to determine the probability of differences of the mean levels of mental illness identification between the groups of people with a current paid employment status at least part time, and those not in paid employment. This post hoc analysis enabled us to identify the specifics of the differences, which consequently showed that the means are indeed significantly different. This approach is considered important as it helps identify where the major discrepancies are, thus revealing more information concerning the employment status and the mental health of the population (Field, 2013; Howell, 2010)

Ho: $\mu A = \mu B$

Ha: there is a significant difference in the means of identifying as having a mental illness between those who are currently employed at least part-time and those who are not.

Tukey-Kramer Multiple Comparisons

Group	Sample Mean	Sample Size	Comparison	Absolute Difference	Std. Error of Difference	Critical Range	Results
1: I am currently em	0.678679	333	Group 1 to Group 2	0.441441	0.02451425	0.0679	Means are different
2: I identify as havin	0.237237	333					

Other Data	
Level of significance	0.05
Numerator d.f.	2
Denominator d.f.	664
MSW	0.200116
Q Statistic	2.77

Fig. 5.0: Results of Post Hoc test

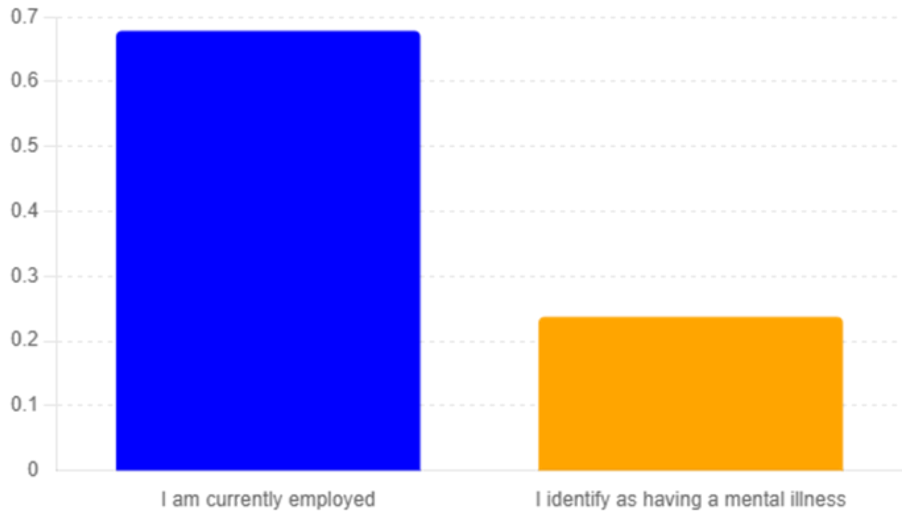

Fig. 5.0:Graph of Count of having a Mental Illness and employed

The test indicated that there is actually a difference in the means of these two groups. This is accented by the absolute difference of means = 0. 441441 which is greater than the critical range = 0. 0679. This raises the prominence of employment status as dependent on the population's rate of identifying as a mental illness. Which means that the null hypothesis is rejected. In addition to that, awareness of these particular distinctions is useful because it identifies clearly where the difference between the groups is, which is why it can contribute to achieving the objectives of the given study more effectively.

V. Conclusion and Recommendation

Unemployment and mental health have consistently shown negative effects on people's lives. This study has aimed to find the correlation that exists between these two variables. Under random sampling and data analysis, it is seen that these two variables have significant association with each other. This implies that they do have a correlation between them, which is important for mental health institutions to take into consideration, especially because of the social stigma with mental health ingrained into Filipino culture. However, the limitations for this study restrains it with ambiguous data, which may not be fully accurate. It is recommended for future research to utilize a more effective and detailed data collection method. By understanding different aspects of people, we also enhance our understanding of our society and the factors that affect it.

References:

Unemployment and mental illness survey. (2019, April 2). https://www.kaggle.com/datasets/michaelacorley/unemployment-and-mental-illness-survey?resource=download

Too, L. S., Leach, L., & Butterworth, P. (2020). Mental health problems and internet access: results from an Australian National Household survey. JMIR Mental Health, 7(5), e14825. https://doi.org/10.2196/14825

Brouwers, E. P. M. (2020). Social stigma is an underestimated contributing factor to unemployment in people with mental illness or mental health issues: position paper and future directions. BMC Psychology, 8(1), 36–36. https://doi.org/10.1186/s40359-020-00399-0

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). SAGE Publications. Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Wadsworth.

Luciano, A., & Meara, E. (2014). Employment status of people with mental Illness: National Survey data from 2009 and 2010. Psychiatric Services, 65(10), 1201–1209. https://doi.org/10.1176/appi.ps.201300335

The Health Foundation. (n.d.). Unemployment and mental health - The Health Foundation. https://www.health.org.uk/publications/long-reads/unemployment-and-mental-health

Wilson, H., & Finch, D. (2021, April 16). *Unemployment and mental health*. The Health Foundation. Retrieved July 15, 2024, from https://www.health.org.uk/publications/long-reads/unemployment-and-mental-health

VI. Appendix

Appendix A: Data Clean Summary

1. **Introduction**

   This appendix contains the summary of the cleaned data from the uploaded CSV file, including key variables and their descriptions.

2. **Content/Variables**

   Employment and Mental Health

   - I am currently employed at least part-time: Indicator of current employment status (1: Yes, 0: No).
   - I identify as having a mental illness: Indicator of self-identified mental illness (1: Yes, 0: No).
   - I have been hospitalized before for my mental illness: Indicator of past hospitalization for mental illness (1: Yes, 0: No).
   - I am legally disabled: Indicator of legal disability status (1: Yes, 0: No).
   - I have a gap in my resume: Indicator of employment gap (1: Yes, 0: No).
   - Total length of any gaps in my resume in months: Duration of employment gaps in months.

   Demographics

   - Education: Highest level of education attained (e.g., High School or GED, Some Phd, Completed Undergraduate).
   - Gender: Gender of the respondent (e.g., Male).
   - Household Income: Income range of the respondent's household (e.g., $25,000-$49,999).
   - Region: Geographic region of the respondent (e.g., Mountain, Pacific).
   - Mean Age: Average age of the respondents.

Technology Access

- I have my own computer separate from a smartphone: Indicator of having a separate computer (1: Yes, 0: No).
- I have my regular access to the internet: Indicator of regular internet access (1: Yes, 0: No).
- Device Type: Type of device used (e.g., Android Phone / Tablet, MacOS Desktop / Laptop).

Symptoms and Mental Health Experiences

- Obsessive thinking: Indicator of obsessive thinking (1: Yes, 0: No).
- Mood swings: Indicator of experiencing mood swings (1: Yes, 0: No).
- Panic attacks: Indicator of experiencing panic attacks (1: Yes, 0: No).
- Compulsive behavior: Indicator of experiencing compulsive behavior (1: Yes, 0: No).
- Tiredness: Indicator of experiencing tiredness (1: Yes, 0: No).

3. **References**

The dataset used for this appendix can be accessed at the following link: https://drive.google.com/file/d/1lvMlYwrLL9dME60ceGm_3-EjGieUp1Sm/view?usp=drive_link

Appendix B: Raw Data Summary

1. **Introduction**

   This appendix contains the summary of the raw data from kaggle, including key variables and their descriptions.

2. **Content/Variables**

   Employment and Mental Health

   - I am currently employed at least part-time: Indicator of current employment status (1: Yes, 0: No).
   - I identify as having a mental illness: Indicator of self-identified mental illness (1: Yes, 0: No).
   - I have been hospitalized before for my mental illness: Indicator of past hospitalization for mental illness (1: Yes, 0: No).
   - I am legally disabled: Indicator of legal disability status (1: Yes, 0: No).
   - I have a gap in my resume: Indicator of employment gap (1: Yes, 0: No).
   - Total length of any gaps in my resume in months: Duration of employment gaps in months.

   Demographics

   - Education: Highest level of education attained (e.g., High School or GED, Some Phd, Completed Undergraduate).
   - Gender: Gender of the respondent (e.g., Male).
   - Household Income: Income range of the respondent's household (e.g., $25,000-$49,999).
   - Region: Geographic region of the respondent (e.g., Mountain, Pacific).
   - Age: The age range of the sample (e.g., 18 - 29)

Technology Access

- I have my own computer separate from a smartphone: Indicator of having a separate computer (1: Yes, 0: No).
- I have my regular access to the internet: Indicator of regular internet access (1: Yes, 0: No).
- Device Type: Type of device used (e.g., Android Phone / Tablet, MacOS Desktop / Laptop).

Symptoms and Mental Health Experiences

- Obsessive thinking: Indicator of obsessive thinking (1: Yes, 0: No).
- Mood swings: Indicator of experiencing mood swings (1: Yes, 0: No).
- Panic attacks: Indicator of experiencing panic attacks (1: Yes, 0: No).
- Compulsive behavior: Indicator of experiencing compulsive behavior (1: Yes, 0: No).
- Tiredness: Indicator of experiencing tiredness (1: Yes, 0: No).

3. References

Unemployment and mental illness survey. (2019, April 2). https://www.kaggle.com/datasets/michaelacorley/unemployment-and-mental-illness-survey?resource=download

Appendix C: Data Cleaning Algorithm Overview

1. **Introduction**

   This Python script is designed to clean a dataset by handling missing values, transforming specific columns, and preparing the data for further analysis. It employs various techniques to ensure the dataset is consistent and ready for statistical analysis.

2. **Content**

   Libraries Used

   - pandas is a python library that is commonly used to manipulate and handle data

   Process Taken

   - The missing values are filled in using mean median mode imputation and entries with more than 50% values are dropped
   - The Age column is converted to Mean age and each entry from the column are converted into the mean value from the age range

3. **References**

   The python script used for this appendix can be accessed at the following link: https://drive.google.com/file/d/1zVyj-UQS0r_VlEmVg2U8qvhB29qI7muE/view?usp=drive_link