**Course – Data Analytics Open Elective**

|  |  |
| --- | --- |
| **UID** | 2021300126 |
| **Name** | Pranay Singhvi |
| **Class and Batch** | TE Computer Engineering - Batch A |
| **Date** | 13-01-2024 |
| **Lab #** | 3 |
| **Aim** | To perform Hypothesis testing t test, z test, p value /ANOVA test |
| **Data Set** | <https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009> |
| **Purpose** | To test different hypothesis on the given dataset |
| **Theory** | **What is Hypothesis Testing?**   * Hypothesis testing is a systematic process for evaluating claims or assumptions about a population based on evidence from a sample. * It involves testing the plausibility of a statement (hypothesis) about a population parameter, such as the population mean or proportion.   **Key Components**   1. **Null Hypothesis (H₀):** This is the initial assumption you are trying to challenge or examine. It usually represents a statement of "no effect" or "no difference." For example:    * "The average weight of apples from a new orchard is the same as the national average."    * "There is no relationship between exercise frequency and stress levels." 2. **Alternative Hypothesis (H₁ or Ha):** This is the statement you will consider supporting if you find enough evidence to reject the null hypothesis. It's often what you aim to demonstrate through your research. For example:    * "The average weight of apples from the new orchard is greater than the national average."    * "There is a negative relationship between exercise frequency and stress levels." 3. **Test Statistic:** A numerical value calculated from your sample data, used to compare against a theoretical distribution. Some common test statistics include:    * **Z-score:** For testing hypotheses about means (when population standard deviation is known)    * **T-score:** For testing hypotheses about means (when population standard deviation is unknown)    * **Chi-square statistic:** For testing relationships between categorical variables 4. **Significance Level (α):** A pre-determined threshold of error you are willing to accept in rejecting the null hypothesis when it might be true (called a Type I error). Typical values for α are 0.05 (5%) or 0.01 (1%). 5. **P-value:** The probability of getting a test statistic as extreme or more extreme than the one you observed from your sample data, assuming the null hypothesis is true.   **Steps in Hypothesis Testing**   1. **State Hypotheses:** Formulate your null and alternative hypotheses clearly. 2. **Set the Significance Level:** Choose your α (usually 0.05). 3. **Select a Test Statistic:** Determine the appropriate test statistic based on your data and question. 4. **Calculate Test Statistic and P-value:** Calculate these values from your sample data. 5. **Decision:** Compare p-value to α.    * If p-value ≤ α: Reject the null hypothesis (statistically significant result).    * If p-value > α: Fail to reject the null hypothesis (not enough evidence to conclude the alternative hypothesis is true).   **Types of Hypothesis Tests**   * **One-tailed vs. Two-tailed Tests:**   + One-tailed tests specify a direction (greater than or less than) for the potential difference.   + Two-tailed tests simply look for any difference, regardless of direction. * **Parametric vs. Non-parametric Tests:**   + Parametric tests assume the sample data follow a specific distribution (e.g., normal distribution).   + Non-parametric tests make fewer assumptions about the distribution, useful for non-normal data. |
| **Code** | T Test  Null Hypothesis (H₀): The mean alcohol content is the same for high-quality and low-quality wines.  Alternative Hypothesis (H₁): There is a significant difference in the mean alcohol content between high-quality and low-quality wines.    Z test:  Null Hypothesis (H₀): The mean pH of the wines is equal to a standard pH value of 3.  Alternative Hypothesis (H₁): The mean pH of the wines is significantly different from the standard pH value.    P Test  Null Hypothesis (H₀): There is no association between chlorides and wine quality.  Alternative Hypothesis (H₁): There is a significant association between chlorides and wine quality.    ANOVA:  Null Hypothesis (H₀): The mean alcohol content is the same across all wine quality ratings.  Alternative Hypothesis (H₁): At least one wine quality rating has a different mean alcohol content. |
| **Conclusion** | In conclusion, I have learnt to test a hypothesis using different method like t Test, z test and ANOVA |