I have worked on a project based on analyzing the social and psychological effect of covid-19 population in turkey. The dataset used for the project was taken from a research paper that is about the collection of dataset through a survey. The data of close to 3000 people were collected and compiled. It had various demographic, health, economic, behavior and security related questions. So firstly the data was preprocessed and the variables were converted into analysis friendly vars like continuous variable (range), categorical variable (male, femal), ordinal variables (low, medium, high) and they were mapped to integer values. Further we carried out 8 hypothesis tests like analyzing the change in stockpiling behavior pre and post covid, usage of online payment options pre-post covid, and the change in their spending and saving habits wtc. There were several tests used for determining the results of these hypotheses like chi square test, mann whitney u test, anova test etc.

Chi-Square Test:

- Used to determine if there is a significant association between two categorical variables.
- Compares the observed frequencies in each category with the frequencies expected if there were no association.
- Requires a contingency table of observed frequencies.
- Degrees of freedom = (number of rows - 1) * (number of columns - 1).
- Null hypothesis: No association between the variables.
-
    - **Formula:** $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$
    - **Variables:** O_i = observed frequency, E_i = expected frequency

Mann-Whitney U Test:

- A non-parametric test for comparing differences between two independent groups.
- Used when the assumptions of a t-test are not met, especially for non-normally distributed data.
- Ranks all the data points together, then compares the sum of ranks between groups.
- Null hypothesis: The distributions of the two groups are the same.
- Does not assume equal variances between the groups.
-
    - **Formula:** $U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$
    - **Variables:** n_1 = sample size of group 1, n_2 = sample size of group 2, R_1 = sum of ranks for group 1

One-Way ANOVA:

- Tests for significant differences between the means of three or more independent groups.

- Assumes normal distribution and equal variances within groups.
- Compares the variance between groups to the variance within groups.
- Null hypothesis: All group means are equal.
- Post-hoc tests are used to determine which groups differ if the null hypothesis is rejected.
-
  - **Formula**: $F = \frac{MS_{between}}{MS_{within}}$
  - **Variables:** $MS_{between}$ = mean square between groups, MS_{within} = mean square within groups

Kruskal-Wallis Test:

- A non-parametric alternative to one-way ANOVA for comparing more than two independent groups.
- Uses ranks of the data rather than the raw data itself.
- Tests whether samples originate from the same distribution.
- Null hypothesis: All groups have the same median.
- Suitable for ordinal data or when assumptions of ANOVA are violated.
-
  - **Formula**: $H = \frac{12}{N(N+1)} \sum \frac{R_i^2}{n_i} - 3(N+1)$
  - **Variables:** N = total number of observations, R_i = sum of ranks for group i, n_i = number of observations in group i

Spearman Correlation:

- Measures the strength and direction of association between two ranked variables.
- Non-parametric alternative to Pearson correlation, based on rank-order.
- Values range from -1 to 1, where 1 means a perfect positive rank correlation and -1 means a perfect negative rank correlation.
- Null hypothesis: No association between the ranks of the variables.
- Does not require the assumption of normality.
-
  - **Formula**: $\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$
  - **Variables**: d_i = difference between ranks of each pair of values, n = number of pairs

Multiple Linear Regression:

- Models the relationship between one dependent variable and two or more independent variables.

- Estimates the coefficients of the linear equation, involving the independent variables that best predict the dependent variable.
- Assumes linearity, independence, homoscedasticity, and normality of residuals.
- Allows for hypothesis testing of individual predictors and overall model fit.
- Useful for predicting outcomes and understanding relationships between variables.
-
    - **Formula:** $\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$
    - **Variables:** $\hat{Y}$ = predicted value, β_0 = intercept, $\beta_1, \beta_2, \ldots, \beta_k$ = regression coefficients, $X_1, X_2, \ldots, X_k$ = independent variables

Shapiro-Wilk Normality Test:

- Tests whether a sample comes from a normally distributed population.
- Suitable for small to moderate sample sizes.
- Produces a W statistic and a p-value; small p-values suggest deviation from normality.
- Null hypothesis: The data is normally distributed.
- Frequently used to check normality assumption before applying parametric tests.
-
    - **Formula**: $W = \frac{(\sum a_i x_{(i)})^2}{\sum (x_i - \bar{x})^2}$
    - **Variables**: a_i = constants derived from the sample size, $x_{(i)}$ = ordered sample values, x_i = sample values, $\bar{x}$ = sample mean