Descriptive Statistic

- summarizes or describes the characteristics of a data set

Absolute Frequency

- data set having a relatively small number of distinct values can be conveniently presented in a frequency table.

Test of Hypothesis and Regression Models

Hypothesis - testable statement about the relationship between two or more variables

Hypothesis Testing - Process to determine if there is enough evidence to reject a null hypothesis

Null Hypothesis - No effect or difference

Alternative Hypothesis - There is an effect or difference.

What Makes a Good Hypothesis

1. Testable, 2. Specific, 3. Based on Existing Knowledge, 4. Predictive.

Importance of Assumptions in Hypothesis Testing

1. Validity of Results, 2. Appropriate Test Selection, 3. Statistical Power and Efficiency.

Common Assumptions for Parametric Tests

1. Normality, 2. Homegeneity of Variances, Independence, No Outliers.

When any of the Assumptions is violated, it is said to be Non-Parametric

Advantages of Non-Parametric Tests

1. Flexibility, 2. Robustness.

Disadvantages of Non-Parametric Tests

1. Less Power

Type I Error (False Positive)

- occurs when the null hypothesis is TRUE but is incorrectly rejected

- it is denoted by alpha α. To control, set at an acceptable significance level at $\alpha = 0.05$

Type II Error (False Negative)

- occurs when the null hypothesis is FALSE but is incorrectly accepted

- it is denoted by beta β. To control, increase sample size, improve study design, and increase effect size of the outliers.

One-Tailed Test - used wheen the research hypothesis specifies a direction of the effect.

Two-Tailed Test - used when the research hypothesis does not specify a direction of the effect but only that there is a difference.

Introduction to DOX - an experiment is a test or a series of tests.

The Basic Principles of DOX 1. Randomization, 2. Replication, 3. Blocking

Strategy of Experiment 1. "Best-Guess" experiment, 2. One-factor-at-a-time (OFAT) experiments, Statistically designed experiments **Statistical Methods**

1. Independent T-test (Parametric)
   (a) Normality, Homogeneity of Variances, Independence
   (b) Comparing means of two independent groups

   Process

   1. Calculate the means $\mu = \frac{\sum_{i=0}^{n} x_i}{n}$,

   2. Calculate the standard deviation $\sigma^2 = \sqrt{\frac{\sum_{i=0}^{n} x_i - \mu}{n-1}}$,

   3. Calculate the pooled standard deviation $s_p = \sqrt{\frac{(n_A-1)s_A^2 + (n_B-1)s_B^2}{n_A+n_B-2}}$,

   4. Calculate the t-Statistic $\frac{\bar{X_A} - \bar{X_B}}{s_p\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}}$,

   5. Determine the critical value and Compare $df = n_A + n_B - 2at\alpha = 0.05$.

   6. Conclude

2. Mann-Whitney U test (Non-Parametric)
   (a) Independence, Continuous/Ordinal data
   (b) Comparing distributions of two independent groups

   Process

   1. Combine and rank data,

   2. Assign individual ranks,

   3. Calculate the sum of the ranks, 4. Compute for the U statistic for the samples and choose the minimum,

   $U_A = n_A \cdot n_B + \frac{n_A(n_A+1)}{2}; U_B = n_A \cdot n_B + \frac{n_B(n_B+1)}{2}; U = \min(U_A, U_B)$,

   5. Determine the critical value and Compare n_A, n_B at $\alpha = 0.05$.

   6. Conclude

3. Paired t-test (Parametric)
   (a) Normality of differences, Independence of Pairs
   (b) Comparing means of paired observation (cause/effect)

   Process

   1. Calculate the differences,

   2. Calculate the mean difference,

   3. Calculate the standard deviation of the difference,

   4. Calculate the t-Statistic,

   5. Determine the critical value and Compare $df = n - 1at\alpha = 0.05$.

   6. Conclude

4. Wilcoxon Signed-Rank Test (Non-Parametric)
   (a) Paired data, Continuous/Ordinal data
   (b) Comparing distributions of paired observations (cause/effect)

   Process

   1. Calculate the differences and absolute values,

   2. Rank absolute differences,

   3. Assign ranks with signs,

   4. Calculate test statistic,

   5. Determine the critical value and Compare $n\alpha = 0.05$.

   6. Determnine critical value and Compare $W \leq / \geq W_{\text{critical}}$

5. RMANOVA (Parametric)
   (a) Normality, Sphericity, Independence of Variances
   (b) Comparing means of related groups measured under different conditions

Process

1. Calculate the group mean,
2. Calculate the overall mean,
3. Calculate the Total Sum of Square $SST = \sum (X_{ij} - \bar{X})^2$,
4. Calculate the Between-Groups Sum of Squares $SSB = \sum n_i(\bar{X}_i - \bar{X})^2$
5. Determine the SSW $SSW = SST - SSB$.
6. Calculate the Between-Groups Degrees of Freedom $df_B = k - 1$
7. Calculate the Within-Groups Degrees of Freedom $df_w = N - k$
8. Calculate the Mean Square Between $MSB = \dfrac{SSB}{df_B}$
9. Calculate the Mean Square Within $MSW = \dfrac{SSW}{df_W}$
10. Calculate the F statistic $F = \dfrac{MSB}{MSW}$
11. Determine the critical value and Compare
12. Conclude

6. Friedman Test (Non-Parametric)
   (a) Paired/Repeated measures, Continuous/Ordinal data
   (b) Comparing distributions of more than two related groups

Process

1. Rank each set of conditions for each transmitter,
2. Calculate the sum of the ranks for each conditions,
3. Calculate the Friedman test statistic using the chi-squared aprroximation $\chi^2_F = \dfrac{12}{nk(k+1)} \sum R_j^2 - 3n(k+1)$,
4. Determine critical value,
5. Conclude

7. ANOVA (Parametric)
   (a) Normality, Homeogeneity of variances, Independence
   (b) Comparing the means of more than two groups

Process

Same as RMANOVA

8. Kruskal-Wallis Test (Non-Parametric)
   (a) Paired/Repeated measures, Continuous/Ordinal data
   (b) Comparing the distributions of more than two groups

Process

1. Combine and rank all data from all groups,
2. Sum the ranks for each group,
3. Calculate H statistic $H = \dfrac{12}{N(N+1)\sum_{R_i^2}^{n_i} - 3(N+1)}$,
4. Determine critical value,
5. Conclude

9. Pearson R
   (a) Interval/Ratio data, Linearity, Homogeneity of Variances
   (b) Measures the strength and direction of relationship between variables

Process

1. Compute the mean of distances and the mean of signal strengths $\bar{x} = \dfrac{\sum_{i=0}^{n} X_i}{n}; \bar{y} = \dfrac{\sum_{i=0}^{n} Y_i}{n}$,
2. Compute the mean of distances and the mean of signal strengths $\sum (x_i - \bar{x})(y_i - \bar{y})$,
3. Calculate the sum of squared deviations $\sum (x_i - \bar{x})^2; \sum (y_i - \bar{y})$
4. Compute for the Pearson R $r = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})(y_i - \bar{y})}}$,
5. Conclude

r value	Interpretation
$r/\rho = 1$	Perfect positive linear/monotonic corre
$1 > r/\rho \geq 0.8$	Strong positive linear/monotonic corre
$0.8 > r/\rho \geq 0.4$	Moderate positive linear/monotonic cor
$0.4 > r/\rho > 0$	Weak positive linear/monotonic correl
$r/\rho = 0$	No correlation
$0 > r/\rho \geq -0.4$	Weak negative linear/monotonic corre
$-0.4 > r/\rho \geq -0.8$	Moderate negative linear/monotonic cor
$-0.8 > r/\rho > -1$	Strong negative linear/monotonic corre
$r/\rho = -1$	Perfect negative linear/monotonic corr

10. Spearman Rho
    (a) Continuous/Ordinal data, Monotonic Relationship
    (b) Measures the strength and direction of relationship between variables

Process

1. Rank distances and strengths separately,
2. Calculate the difference in ranks and square them,
3. Compute the Spearman Rho $\rho = 1 - \dfrac{6\sum d_i^2}{n(n^2-1)}$
4. Conclude

11. Linear Regression Parametric
    (a) Linearity, Independence, Homegeneity of Variances, Normality
    (b) For modeling and predicting the relationship between a dependent variable (Y) and one or more independent variables (X).

Process

1. Rank distances and strengths separately,
2. Calculate the difference in ranks and square them,
3. Compute the Spearman Rho $\rho = 1 - \dfrac{6\sum d_i^2}{n(n^2-1)}$
4. Conclude

12. Multiple Regression Parametric
    (a) Linearity, Independence, Homeogeneity of Variances, No Multicollinearity
    (b) For modeling and predicting the relationship between a dependent variable (Y) and one or more independent variables (X).

Process

1. Calculate means $\bar{X} = \frac{\sum_{i=0}^{n} X_i}{n}; \bar{Y} = \frac{\sum_{i=0}^{n} Y_i}{n}$,
2. Calculate slope,
3. Calculate the y - intercept $a = \bar{Y} - b\bar{X}$,
4. Write the regression equation
5. Conclude

13. ANCOVA (Parametric)
    (a) Linearity, Homegeneity of Regression Slopes, Independence, Measurement of the Covariate
    (b) Compares a response variable by both a factor and a continuous independent variable

Process

1. Understand the given data
2. Compute the Overall Mean for Transmission Rate,
3. Compute the factor means,
4. Compute for the adjusted means, if needed.
5. Perform ANOVA on the adjusted means
6. Conclude

14. MANOVA (Parametric)
    (a) Independence of Observations, Linearity, Adequate Sample Size
    (b) ANOVA with two or more continuous response variables

15. MANCOVA (Parametric)
    (a) Independence of Observations, Linearity, Adequate Sample Size, Linearity and Homegeneity of Regression Slopes, Measurement of the Covariate
    (b) One or more covariates are added to the mix