Testing of Hypothesis using Python

Dr. Hakeem–Ur–Rehman

CSAS–PU

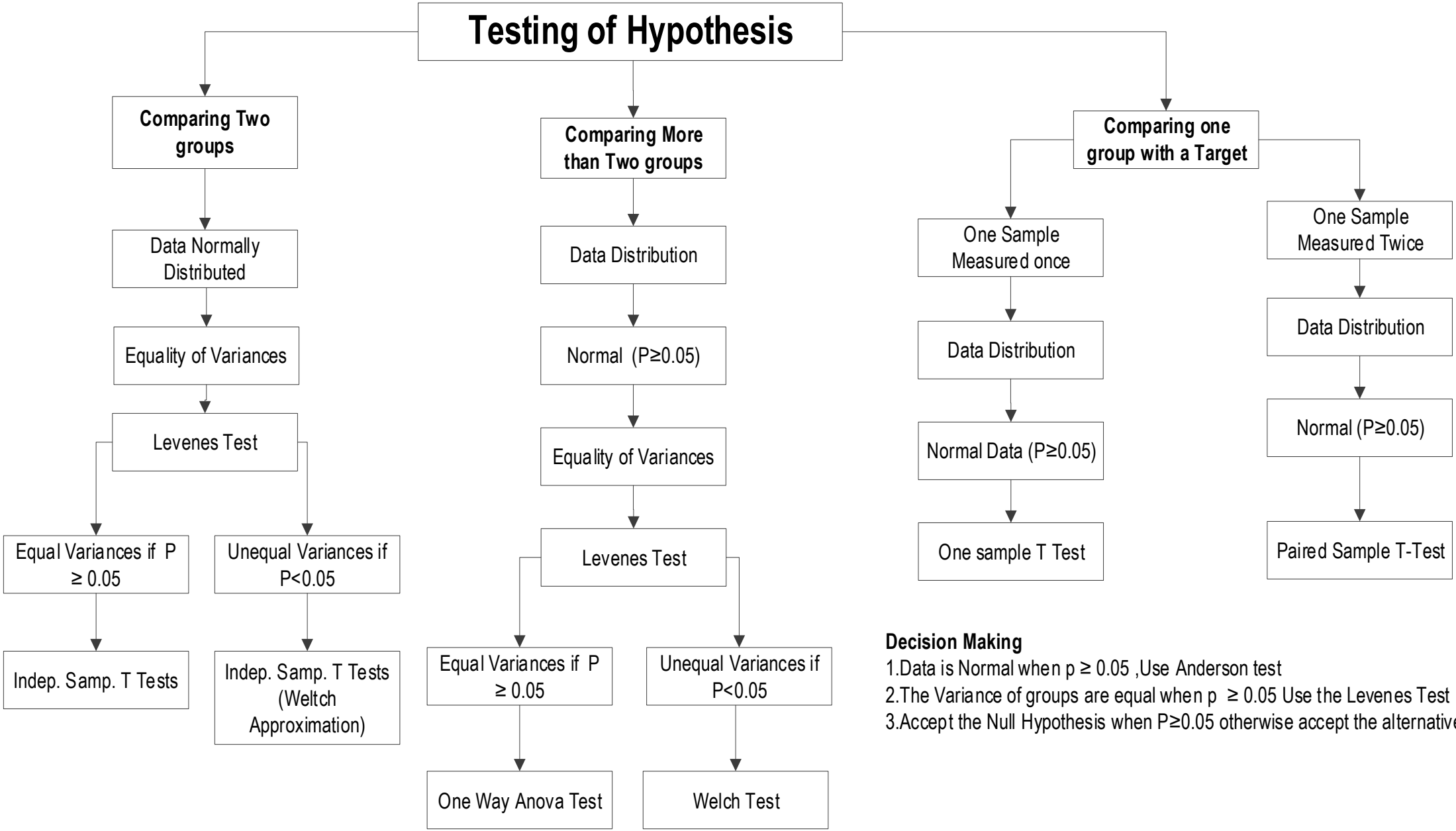
Bi-variate Analysis

		Dependent Variable (DV=1)	
		Quantitative (Metric)	**Qualitative (Non-Metric)**
Independent Variable (IV=1)	**Quantitative (Metric)**	▪ Correlation Analysis ▪ Simple Regression	▪ Binary Logistic Regression ▪ Linear Discriminant Analysis
	Qualitative (Non-Metric)	Testing of Hypothesis: Parametric & Non-Parametric (Compare Averages)	Cross-tab / Chi-Square Test

- If Independent Variables (IV) has two categories → t-test
- If Independent Variables (IV) has more than three categories → One Way ANOVA

- If Dependent Variables (DV) has two categories → Binary Logistic Regression
- If Dependent Variables (DV) has more than three categories → Multinomial Logistic Regression

PARAMETRIC TESTS



Decision Making

1.Data is Normal when p ≥ 0.05 ,Use Anderson test

2.The Variance of groups are equal when p ≥ 0.05 Use the Levenes Test

3.Accept the Null Hypothesis when P≥0.05 otherwise accept the alternative hypothesis

Parametric Vs Nonparametric Tests

Nonparametric test	parametric test
One Sample Wilcoxon Signed Rank test	1-sample t-test
Wilcoxon Signed Rank test	Paired t-test
Mann-Whitney U test	2-sample t-test
Kruskal-Wallis test	One-way ANOVA
Mood's Median test	One-way ANOVA

Testing of Hypothesis: Students Performance Data (Assignment)

- Data set: https://www.kaggle.com/spscientist/students-performance-in-exams.

1. **Hypothesis 1:** There is no difference in means of student performance in any of basic literacy skills i.e. reading, writing, math.

   Ho: There is no difference in performance of students between math, reading and writing skills.

   Ha: There is a difference in performance of students between math, reading and writing skills.

2. **Hypothesis 2:** An educational consultant claims that on average students receive a respectable score of 70 or more

   Ho: The overall performance of students is greater than or equal to a respectable score of 70.

   Ha: The overall performance of students is less than 70.

3. **Hypothesis 3**: There is no significant difference in the mean score of students who have taken test preparation and those students who have not taken any test preparation

   Ho: There is no difference in overall performance of students between those who have taken test preparation and those who have not.

   Ha: There is a difference in overall performance of students between those who have taken test preparation and those who have not.

4. **Hypothesis 4**: There is no relation between the gender of a student and their corresponding academic performance.

   Note: Here we will be creating two features called 'Verdict' and 'Overall Performance' (average of all three scores of a student) which will let us know whether a student's performance is acceptable or not. I have considered 40 as the threshold for overall performance to be considered acceptable.

   Ho: There is no relation between the gender of a student and their corresponding academic performance.
   Ha: There is a relation between the gender of a student and their corresponding academic performance.

QUESTIONS