Introduction to Statistics

Friday, October 7, 2022   8:01 PM

Intro to Statistics:
1. Descriptive stats
2. Inferential Stats

Descriptive Stats:
1. Measure of Central Tendency
2. Measure of Dispersion
3. Histograms, Box-Plot and whiskers
4. Summarizing the data(Pdf, cdf, probability, permutation, mean, median, mode, variance, standard dev., gaussian distribution, Lognormal Distribution, Bernoulli's Distribution, Pareto D. (Power Law Distribution), Standard Normal Distribution, Transformation and standardization, Q-Q plot).

Inferential Stats:
1. Z test
2. T test
3. Anova test (F test)
4. Chisq.
5. Hypothesis Testing (O values)
6. Confidence Intervals
7. Z table, T tables.

****

What is Stats??

    Stats is the sc. Of collecting, organizing and analyzing data.
    Done for better decision making.

Data:

    Facts or pieces of info, that can be measured.
    Example:
        IQ of a class: {98, 97, 60, 55, 75, 65}
        Age of students in a class: {30, 25, 24, 23, 27, 28}

Types of Stats:
1. Descriptive Stats:
    a. Consists of organizing and summarizing of data.
2. Inferential Stats:
    a. Technique where in we the data that we have measured to form conclusions.

Classroom of Maths Student (20)
    Marks of 1st Sem:
        {84, 86, 78, 72, 75, 65, 80, 81, 92, 95, 96, 97, …}
    Q1: Average marks of the whole class?? *(Descriptive Stats)*
    Q2: Are the marks of the students of this classroom similar to the age of the Maths classroom in the college?? *(Inferential Stats)*

Population and Sample:

[Live Day 1- Introduction To statistics In Data Science](#)

Sampling Techniques: ([Live Day 1- Introduction To statistics In Data Science](#)*)*

1. **Simple Random Sampling:** Every member of the population has an equal chance for ur sample(n).
2. **Stratified Sampling:** Technique where the population is split into non-overlapping grps. (Sampling done by dividing into grps)
3. **Systematic Sampling:** From N pick up every nth individual.
    a. Example, While doing a survey -> Catch every 8th person (for survey)
4. **Convenience Sampling:** Doing survey or sampling on those who know abt the topic that is surveyed on. (Sampling done based on a specific domain)

Variables:

A variable is a property that can take any value.

2 kinds of variables:

1. Quantitative - Measured Numerically, can perform (+, -, /, *, etc…)
2. Qualitative/categorical variables: Based on some characteristics we can derive categorical variables; eg: gender, IQ, T-shirt, Blood Grp, etc…

Quantitative:

1. **Discrete Variables (Whole Number)** eg, Number of Bank a/c, no. of children in a family.
2. **Continuous Variables (Any values)** eg, height, weight, rainfall(measured in mm, cm, inches)

Examples for variables: [Live Day 1- Introduction To statistics In Data Science](#)

Variables Measurement Scales:

1. Nominal: Categorical/Qualitative Data; eg: Color, gender, type of flavour
2. Ordinal: Order of the Data matters, but not the values.([Live Day 1- Introduction To statistics In Data Science](#))
3. Interval: Order and values matters, natural zero not present([Live Day 1- Introduction To statistics In Data Science](#))
4. Ratio:

Frequency Distribution:

[Live Day 1- Introduction To statistics In Data Science](#)

Bar Graphs:
[Live Day 1- Introduction To statistics In Data Science](#)

Histograms:
[Live Day 1- Introduction To statistics In Data Science](#)

Bar vs Histograms:
      Bar: for Discrete
      Histograms: For continuous

PDF: Probability Density Function