IC 102: Data Analysis and Interpretation

Instructor: Guruprasad PJ
Dept. Aerospace Engineering
Indian Institute of Technology Bombay
Powai, Mumbai – 400076

Email: pjguru@aero.iitb.ac.in
Phone no.: 2576 7142

Topics covered until now

Chapter 1: The Role of Statistics
- Reasons to study statistics
- The nature and role of variability
- Statistics and data analysis
- Types of data and some simple graphical displays

Chapter 2: The data analysis process and collecting data sensibly
- Data collection
- The data analysis process
- Population and sampling
- Inferential statistics

Chapter 3: Graphical methods for describing and summarizing data
- Displaying categorical data: comparative bar charts and pie charts
- Displaying numerical data: stem-and-leaf displays, boxplots
- Displaying numerical data: Frequency distributions and histograms
- Displaying bivariate numerical data: scatterplot

Sample Mean, Median and Mode

- Sample Mean: Arithmatic average of the numerical values in a data set. Suppose the data set has the numerical values x_1, x_2, x_3,....x_n then the mean is defined as

$$\bar{x} = \sum_{i=1}^{n} x_i / n$$

- What about for the data set whose numerical values can represented by the below equation?

$$y_i = a\, x_i + b$$

Sample Mean, Median and Mode

- Suppose we want to determine the sample mean of a data set that is presented in a frequency table listing ***k*** distinct values v_1, v_2...v_k and having corresponding frequencies f_1, f_2...f_k, how do we do it?

$$\bar{v} = \sum_{i=1}^{k} f_i v_i / n$$

Sample Mean, Median and Mode

- Sample Median: Indicates the center of a data set.

- Definition: *Order the values of a data set of size n from smallest to largest. If n is odd, the sample median is the value in position (n+1)/2; if n is even, it is the average of the values in positions n/2 and n/2+1 .*

Sample Mean, Median and Mode

- Sample Mode: *Defined to be the value that occurs with the greatest frequency. If no single value occurs most frequently, then all the values that occur at the highest frequency are called modal values.*
- **Find the sample mean, median and mode for the below data set.**

Value	Frequency
1	9
2	8
3	5
4	5
5	6
6	7

Sample Variance and Standard Deviation

- Sample variance and standard deviation: Represents the spread or variability of the data values from the central tendency measure.

  The sample variance, denoted by s^2, of the data set x_1, x_2, x_3,...x_n, is defined by

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1)$$

- The positive square root of the sample variance is called the sampe standard deviation.

$$s = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1)}$$

Algebraic Identities

- Demonstrate the following:

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-n\bar{x}^2$$

- If the data values are represented by

$$y_i=a\,x_i+b$$

  then show that

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=a^2\sum_{i=1}^{n}(x_i-\bar{x})^2$$

Weekend Assignment

- Read Chapter 1 and upto section 2.4 in Chapter 2 from the textbook (Introduction to Probability and Statistics for Engineersand Scientists, Sheldon M. Ross)

- Note: I will upload class notes after I complete significant amount of topics in the course.