Stat 231 - Statistics **Spring 2017**

Lecture 3: May 5th, 2017

Lecturer: Suryapratim Banerjee *Notes By: Harsh Mistry*

Objective : We want to extract and summarize important properties of data sets because

- A lot of information about the data can be captured by a few values
- We want to identify the "likely" distribution from which the data is drawn

By looking at the properties we can apply the right model, and use properties learned in stat 230.

Properties We Are Interested In :

- Measures of Location (Center)
- Measures of Dispersion (Variability)
- Measures of Skewness (Is Data Symmetric)
- Measures of Kurtosis (How common are extreme observations)

3.1 Measures of Location

Consider the data set : $\{y_1, \ldots, y_n\}$, where n is the sample size

- Sample mean : $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
- Geometric Mean : $G.M = (y_1 \cdot y_2 \cdot \ldots \cdot y_n)^{\frac{1}{n}}$
- Median (Less affected by extreme observations) : $\hat{m}$ = middle most observation. If there is no middle, the median is the sum of the two conflicting values divided by 2.
- Mode : The measures that occur with the maximum frequency

3.2 Measures of Dispersion

Consider : The Range = Maximum - Minimum Value

3.2.1 Sample Variance

$$s^2 = \frac{1}{n-1}\sum(y_i - \bar{y})^2$$