Chapter 1 - Lecture 3
Measures of Location

Andreas Artemiou

August 31st, 2009
Measures

General

Types of measures

Measures of location

Definition

Mean

Median

Data Characteristics

Skewness

Trimmed mean

Definition

Percentiles

Exercises
What might be interesting features of the data that we might want to know about?
In this class we will learn two categories of measures:
  - Measures of location:
    - Mean
    - Median
    - Trimmed mean
    - Percentiles
  - Measures of variability:
    - Range
    - Variance
    - Standard deviation
Measures of location are numbers that will give us a number of the data that might be of interest. Usually that point is the middle point of the dataset (like the average). But there are other interesting points, like the point that is above 75% of the rest of the points.
The most important location measure is the sample mean or the average of the dataset.

The formula to calculate this is:

\[ \bar{x} = \frac{x_1 + \ldots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i \]
Example

- I give a test in a class and I take a sample of 6 students. The grades are: 50, 70, 64, 94, 78, 88.
- Find the mean of the sample.
Median

- Median is the middle value when the sample values are ordered from the smallest to the largest.
- Formula:
  - If \( n \) is odd:
    \[
    \tilde{x} = \frac{n + 1}{2} \text{ ordered}
    \]
  - If \( n \) is even:
    \[
    \frac{n}{2} \text{ ordered} + \left(\frac{n}{2} + 1\right) \text{ ordered} \quad \frac{2}{2}
    \]
Example

- I give a test in a class and I take a sample of 5 students. The grades are: 50, 70, 64, 78, 88.
- Find the median of the sample.
Example

- I give a test in a class and I take a sample of 6 students. The grades are: 50, 70, 64, 94, 78, 88.
- Find the median of the sample.
How skewness determines the relationship between mean and median

- If a dataset is symmetric then the mean is equal to the median
- If a dataset is skewed to the left median is larger than the mean
- If a dataset is skewed to the right the mean is larger than the median
Example

- Suppose in a test I give in this class all the students take a grade between 80-100 and there is only one who takes a grade of 20. I take a sample of 10 students and I want to see the mean. The grades of the 10 people are: 80, 80, 85, 85, 20, 90, 90, 95, 95, 100.

- What is the mean of the above sample?
Because extreme values can occur in our sample and we want to avoid affecting our mean we use the idea of a trimmed mean.

- 10% trimmed mean is the mean calculated if we exclude the lower 10% and the upper 10% of the observations.

- 20% trimmed mean is the mean calculated when we exclude the lower 20% and the upper 20% of the observations.
Percentiles

- 90th percentile is the value in the dataset that is above 90% of the dataset and below 10% of the dataset.
- We will see later how to calculate this.
- The most important percentiles are the three quartiles:
  - 25th percentile = first quartile = Q1
  - 50th percentile = second quartile = median = Q2
  - 75th percentile = third quartile = Q3
Exercises

- Section 1.3 page 31
  - Exercises 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40