MEASURES OF LOCATION

Using averages to describe data

FEATURES OF LOCATION

- Everyone has heard of averages
- Easy to understand and calculate
- Available on most packages

- Arithmetic average is one measure of location, but there are more.

- We will see three of them:
  - Arithmetic Mean
  - Median
  - Mode

A dataset can have several variables, but the location actually refers to the single variable.
For a multivariate dataset, we can compute a measure of location for each of them.

DEFINITIONS

- Arithmetic mean (also called average):
- *Sum of all numbers divided by size of dataset.*
- *Applies only to numeric data (discrete or continuous)*

- Median:
- *Middle value when data arranged in order*
- *Applies to ordinal and numeric data*

- Mode:
- *Most frequently occurring value*
- *Can be used with qualitative data, and numeric, but for continuous data, it must be applied to the distribution (or frequency table)*

RAW NUMERIC DATA – AN EXAMPLE

Let's start with a small data set (N=11): 1,2,4,5,5,5,6,6,8,9,10

The **mean** is: Sum/N = 61/11 = 5,545

The **median** is the middle one, that is, 5

1,2,4,5,5,5,6,6,8,9,10

The **mode** is the most frequent value, which is also 5

Let's start with a similar data set (N=12): 1,2,4,5,5,5,6,6,8,9,10,35

We just added 1 number at the end

The **mean** becomes 96/12 = 8

The **median** is between the numbers 5 and 6, so we decide 5.5

1,2,4,5,5,5,6,6,8,9,10,25

The mode remains at 5

TABULATED DATA – AN EXAMPLE

Data = {1, 1,1, …, 1, 1, 2, 2, …,2, 2, 3, 3, …,3, 3, 4, 4, …, 4, 4, 5, 5, … 5, 5 }, N =130

20 30 50 25 5

X	Freq	Freq x X
1	20	20
2	30	60
3	50	150
4	25	100
5	5	25
	130	355

Average: **355/130 = 2,73**

mean= (1+ 1 + …,+1 + 2 + 2 +,…+ 2+ 3+ 3+ … +3+ 4+4+ …+4+ 5+ 5+ … +5)/N

20 30 50 25 5

mean= (20 x 1 + 30 x 2 + 50 x 3 + 25 x 4 + 5 x 5)/N = Sum (Freq x X)/N = 355/130 = **2,73**

For the median, middle position of 130 is between position 65 and 66
Which is in the group of X=3
So the median is 3

For the mode, the group with the highest frequency is where X=3
So the mode is 3

Python practice: measures of location

- Let's compute some measure of location for the numeric data in file data_lecture3.xlsx