Measures of Central Tendency and Location

Measures of Central Tendency

The Summation Notation

Summation Notation

- Necessary for computing summary statistics
- Sums of numerical values
- Denoted by the capital Greek letter sigma (Σ)

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + X_3 + \cdots + X_n$$

Properties of Summation

- The summation of the sum (or difference) of two or more terms equals the sum (or difference) of the individual summations
- The summation of a constant, c, times a variable, X, equals the constant times the summation of the variable
- The summation of a constant, c, from i = 1 to n, equals the product of n and c.

Some Notes on Summation

- The subscript may be any letter, but the most common are i, j, and k.
- The lower limit of the summation may start with any number
- The lower limit of the summation is not necessarily a subscript
- The sum of the squared values of X is NOT equal to the square of the sum of X
- The summation of the square of (X+Y) is NOT equal to the summation of the sum of X^2 and the sum of Y^2
- The sum of the product of X and Y is NOT equal to the product of the sum of values of X and the sum of values of Y
- The sum of the quotient of X and Y is NOT equal to the quotient of the sum of X and the sum of Y
- The sum of the square root of X is not equal to the square root of the sum of X

The Mean

The Arithmetic Mean

- Most common average
- Sum of all observed values divided by the number of observations
- For ungrouped and grouped data
  - Population mean (μ)
  - Sample mean ($\bar{X}$)
  - Different formulas for grouped and ungrouped data

Some Remarks on the Mean

- Most common measure of central tendency
- Uses all observed values
- May or may not be an actual value in the data set
- May be computed for both grouped and ungrouped data sets
- Extreme observations affect the value of the mean

Two Mathematical Properties

- The sum of the deviations of the observed values from the mean is zero.

$$\sum_{i=1}^{n}(X_i - \bar{X}) = 0$$

- The sum of the squared deviations of the observed values from the mean is smallest.

Modifications of the Mean

- The Weighted Mean
  - There are weights
  - Values are not of equal importance
  - e.g. GWA
- The Combined Mean
  - Mean of several data sets
- The Trimmed Mean
  - Less affected by extreme observations
  - Order data
  - Remove a certain percentage of the lower and upper ends
  - Calculate arithmetic mean

The Median

The Median

- Middle value in an ordered set of observations
- Divides ordered set of observations into two equal parts
- Positional middle of an array
- Grouped and ungrouped data

Characteristics of the Median

- Extreme values affect the median less than the mean
- The median is used when
  - We want the exact middle value of the distribution
  - There are extreme observed values
  - FDT has open-ended intervals

The Mode

The Mode

- Most frequent observed value in the data set
- Small data sets: inspection
- Large data sets: array or FDT
- Less popular

Characteristics of the Mode

- Gives the most typical value of a set of observations
- Few low or high values do not easily affect the mode
- May not be unique and may not exist
- Several modes can exist
- Value of the mode is always in the data set
- May be used for both quantitative and qualitative data sets

Remarks

- Data is symmetric and unimodal → all three measures may be used
- Data is asymmetric → use median or mode (if unique); trimmed mean
- Describe shape of data → use all three

FORMULAS: MEAN

Ungrouped

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Grouped

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i}$$

FORMULAS: MEDIAN

Ungrouped

- n is odd

$$Md = X_{\left(\frac{n+1}{2}\right)}$$

- n is even

$$Md = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2}$$

Grouped

$$Md = LCB_{Md} + C\left(\frac{\frac{n}{2} - < CF_{Md-1}}{f_{Md}}\right)$$

FORMULA: MODE

- **Grouped**

$$Mo = LCB_{Mo} + C\left(\frac{f_{Mo} - f_1}{2f_{Mo} - f_1 - f_2}\right)$$

Examples

Find the MEAN, MEDIAN, and MODE of the ff:

1. A sample survey in a certain province showed the number of underweight children under five years of age in each barangay: 3 5 6 4 7 8 6 9 10 4 6 7 5 8 9 8 3 4 5 5
2. Given the frequency distribution table of scores

LCL	UCL	LCB	UCB	Frequency	RF	RF %	<CF	>CF
3.5	5.4	3.45	5.45	4	0.057971	5.797101	4	69
5.5	7.4	5.45	7.45	4	0.057971	5.797101	8	65
7.5	9.4	7.45	9.45	15	0.217391	21.73913	23	61
9.5	11.4	9.45	11.45	11	0.15942	15.94203	34	46
11.5	13.4	11.45	13.45	13	0.188406	18.84058	47	35
13.5	15.4	13.45	15.45	8	0.115942	11.5942	55	22
15.5	17.4	15.45	17.45	14	0.202899	20.28986	69	14
				69	1	100		

Measures of Location

RECALL: MEDIAN

- Measure of central tendency
- Measure of location
- Positional middle

Measures of Location

- Percentiles
- Deciles
- Quartiles

Percentiles

- Divide ordered observations into 100 *equal* parts
- 99 percentiles; roughly 1 percent of observations in each group
- Interpretation:

  P_1, the first percentile, is the value below which 1 percent of the ordered values fall. (ETC.)
- Ungrouped data
  - Empirical Distribution Number with Averaging
  - Weighted Average Estimate
- Grouped data

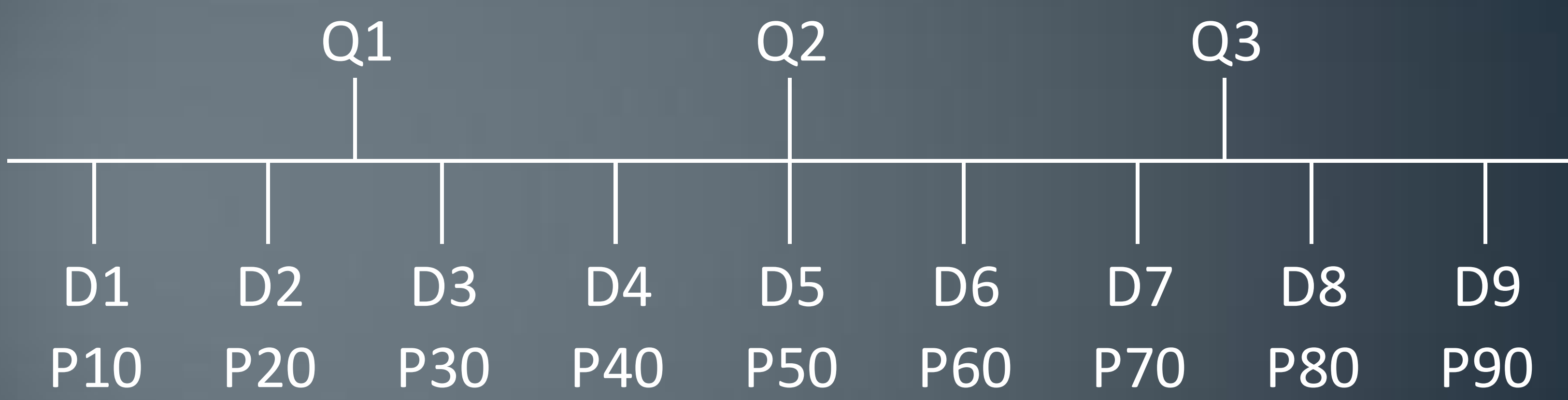
Deciles

- Divide the ordered observations into 10 equal parts
- Each part has ten percent of the observations

Quartiles

- Divides observations into 4 equal parts
- Each part has 25 percent of the observations

Relationship Among the Three Measures



Formulas: Percentiles (Ungrouped)

Empirical

- nk/100 is an integer

$$P_k = \frac{X_{\left(\frac{nk}{100}\right)} + X_{\left(\frac{nk}{100}+1\right)}}{2}$$

- nk/100 is not an integer

$$P_k = X_{\left[\!\left[\frac{nk}{100}+1\right]\!\right]}$$

Weighted Average

$$\frac{(n+1)k}{100} = j + g$$

$$P_k = (1-g)X_{(j)} + gX_{(j+1)}$$

Formulas: Percentiles (Grouped)

$$P_k = LCB_{P_k} + C\left(\frac{\frac{nk}{100} - < CF_{P_{k-1}}}{f_{P_k}}\right)$$

Example

- The number of incorrect answers on a true-false exam for a random sample of 20 students was recorded as follows: 2
- and 2 find the 90^{th} percentile, 7^{th} decile, and 1^{st} quartile
- Given the frequency distribution of scores of 200 students in an entrance exam in college, find the 95^{th} percentile, 4^{th} decile, and 3^{rd} quartile.

Scores	Freq.	<CFD	LCB	UCB
59 – 62	2	2	58.5	62.5
63 – 66	12	14	62.5	66.5
67 – 70	24	38	66.5	70.5
71 – 74	46	84	70.5	74.5
75 – 78	62	146	74.5	78.5
79 – 82	36	182	78.5	82.5
83 – 86	16	198	82.5	86.5
87 – 90	2	200	86.5	90.5