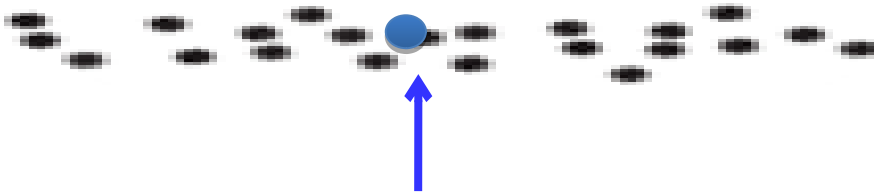
2. Summarizing of Data

2.1 Measures of Central Tendency

- A *measure of central tendency* is a descriptive statistic that describes the average, or typical value of a set of scores.
- It is also defined as a single value that is used to describe “center” of the data



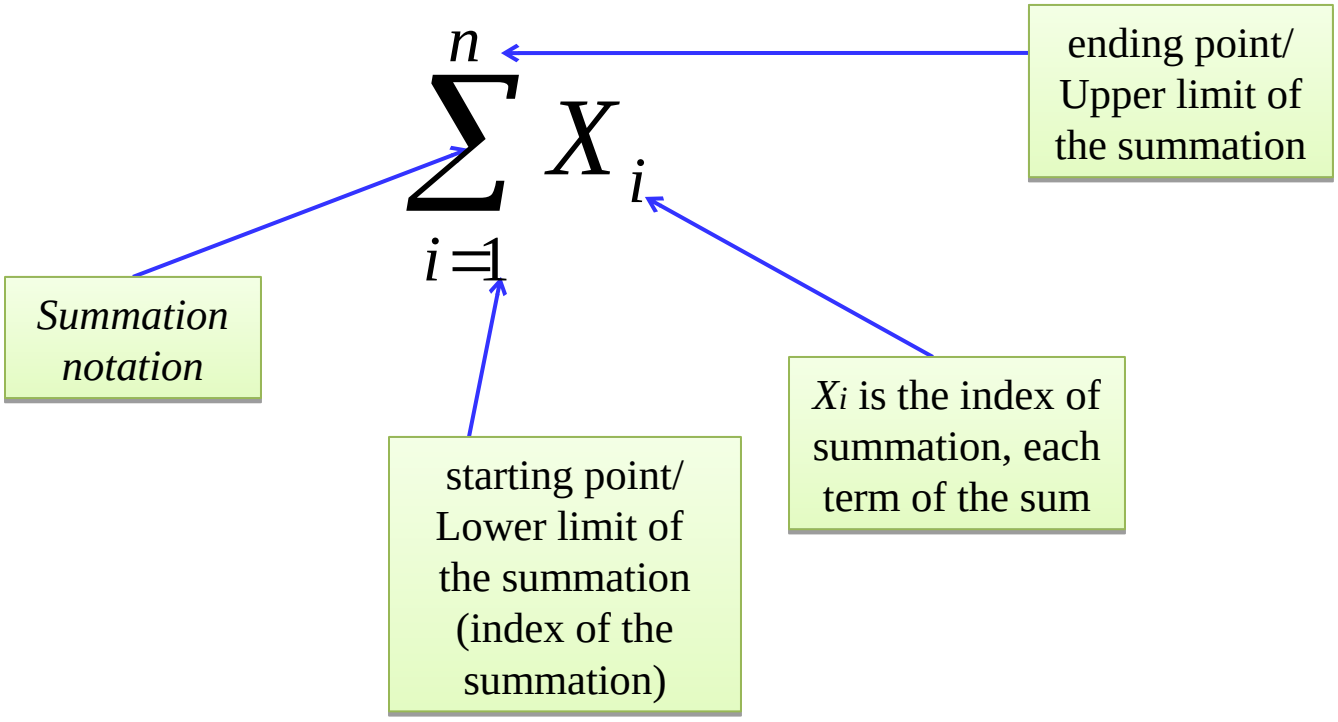
Typical value
(Center of data)

2.2 Types of measures of central tendency

- Good properties of typical average
  - Computation should be based on all the observed values.
  - It should be simple to understand and easy to interpret.
  - As little as affected by fluctuations of sampling.
  - should not unduly be influenced by extreme values.
  - it should be defined rigidly which means that it should have a definite value
- There are three common measures of central tendency
  - Mean
  - Median
  - Mode

The Summation Notation

- Also called Sigma notation
- Sigma is a Greek letter $\sum$ meaning "sum"
- Let X is a variable



The Summation Notation..

- *Properties of summation notation*

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

$$\sum_{i=1}^{n} X_i Y_i = X_1 Y_1 + X_2 Y_2 + \cdots + X_n Y_n$$

$$\sum_{i=1}^{n} X_i^2 = X_1^2 + X_2^2 + \cdots X_n^2$$

$$\sum_{i=1}^{n} CX_i = C\sum_{i=1}^{n} X_i = CX_1 + CX_2 + \cdots + CX_n$$

The Mean

- Mean is the most commonly used measure of central tendency. There are different types of mean
  - Arithmetic mean,
  - Weighted mean,
  - Geometric mean (GM) and
  - Harmonic mean (HM)
- If mentioned without an adjective (as mean), it generally refers to the arithmetic mean.

The Arithmetic Mean

- It is computed by adding all the values in the data set divided by the number of observations in it.
- If we have the raw data, mean is given by the formula

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

- If we have frequency distribution (ungrouped) mean is given by the formula

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

- If we have frequency distribution (grouped) mean is given by the formula

LCB/UCB is lower/upper class boundary

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i m_i}{n}, \text{ where } m_i = \frac{LCB_i + UCB_i}{2}$$

The Arithmetic Mean …

- Example 1: The following data is the weight (in Kg) of eight youths: 32,37,41,39,36,43,48 and 36. Calculate the arithmetic mean of their weight. (Ans:312/8=39)
- Example 2: The ages of a random sample of patients in a given hospital in Ethiopia is given below: (Ans: 16.075)

The Arithmetic Mean …

- Example 3: Age in year of 20 women who attended health education at Jimma Health center in 1986 is summarized in the table. What is the mean age of these women. (Ans: 670/20=33.5)

Properties of Arithmetic Mean …

- It can be computed for any set of numerical data, it always exists, and unique.
- It depends on all observations.
- The sum of deviations of the observations about the mean is zero i.e.
- It is greatly affected by extreme values.
- It lends itself to further statistical treatment, for instance, combinations of means.
- It is relatively reliable, i.e. it is not greatly affected by fluctuations in sampling.
- The sum of squares of deviations of all observations about the mean is the minimum
- If a constant is added to all observations, the new mean is old mean plus constant
- If all observations are multiplied by a constant, the new mean is the multiple of the constant and old mean
- If wrong value is recorded and latter on it is discovered, the new corrected mean is

$$\overline{X}_{corr} = \overline{X}_{wrong} + \frac{\left(X_{corr} - X_{wrong}\right)}{n}$$

Weighted Mean

- Weighted mean is calculated when certain values in a data set are more important than the others.
- A weight w_i is attached to each of the values x_i to reflect this importance.
- The weighted mean is computed as

$$\overline{X}_w = \frac{\sum_{i=1}^{k} w_i x_i}{\sum_{i=1}^{k} w_i}$$

- Example: CGPA of a students (each result is weighted by credit of a course) [Ans: 2.88]

Course	Credit hour (w_i)	Grade
Math101	4	A=4
Bio101	3	C=2
Chem101	3	B=3
Phys101	4	B=3
Flen101	3	C=2

Geometric Mean

- It is defined as the arithmetic mean of the values taken on a *log* scale.
- It is also expressed as the n^{th} root of the product of an observation.

$$G.m = \begin{cases} \sqrt[n]{x_1 * x_2 * \dots * x_n} & \rightarrow for\ raw\ dtata \\ \sqrt[n]{{x_1}^{f_1} * {x_2}^{f_2} *, \dots, * {x_k}^{f_k}} & \rightarrow for\ ungrouped\ data\ \ where\ \ n = \sum f_i . \\ \sqrt[n]{{m_1}^{f_1} * {m_2}^{f_2} *, \dots, * {m_k}^{f_k}} & \rightarrow for\ grouped\ data \quad where\ n = \sum f_i . \end{cases}$$

- GM is an appropriate measure when values change exponentially and in case of skewed distribution that can be made symmetrical by a log transformation.
- **Note:** The geometric mean is useful in finding the average of *percentages, ratios*, *indexes*, or *growth rates*.
- One important disadvantage of GM is that it cannot be used if any of the values are zero or negative.

Geometric Mean…

Example 1:- The G.M of 4, 8 and 6 is**.**

Solution: $G.M = \sqrt[3]{4\,x\,8\,x\,6} = \sqrt[3]{192} = 5.77.$

Example 2: The man gets three annual raises in his salary. At the end of first year, he gets an increase of 4%, at the end of the second year, he gets an increase of 6% and at the end of the third year, he gets an increase of 9% of his salary. What is the average percentage increase in the three periods?

Solution:

$$G.M = \sqrt[3]{1.04 * 1.06 * 1.09} = 1.0631 \Rightarrow 1.0631 - 1 = 0.0631$$

$$Therefore, the\ average\ percentage\ increase\ is\ 6.31\%.$$

Properties of geometric mean

- Its calculations are not as such easy.
- It involves all observations during computation
- It may not be defined even it a single observation is negative.
- If the value of one observation is zero its values becomes zero.

Harmonic Mean

- Another important mean is the harmonic mean, which is suitable measure of central tendency when the data pertains to ***speed, rates and price.***
- It is the reciprocal of the arithmetic mean of the observations.
- Let $x_1, x_2, \dots, x_n$ be n variant values in a set of observations, then simple harmonic mean is given by:

$$S.H.M = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

- ***Note:*** *SHM is used for equal distances, equal costs and equal rates.*

Harmonic Mean

Example 1: A motorist travels for three days at a rate (speed) of 480 km/day. On the first day he travels 10 hours at a rate of 48 km/h, on the second day 12 hours at a rate of 40 km/h, on the third day 15 hours at a rate of 32 km/h. What is the average speed?

Solution*:* Since the distance covered by the motorist is equal

($i.e.\ s_1 = 480, s_2 = 480,\ s_3 = 480$), so we use *SHM.*

$$S.H.M = \frac{3}{1/48 + 1/40 + 1/32} = 38.92$$

so the required average speed = 38.92 km/hr

We can check this, by using the known formula for average speed in elementary physics.

Check;

$$Average\ speed\ (V_{av}) = \frac{\text{total distance covered}}{\text{total time taken}} = \frac{S_T}{t_T} = \frac{480\text{km} + 480\text{km} + 480\text{km}}{10\text{hr} + 12\text{hr} + 15\text{hr}}$$

$$= \frac{1440\text{km}}{37\text{hr}} = 38.42\ km/$$

Weighted harmonic mean (WHM)

- *WHM* is used for different distance, different cost and different rate.

$$W.H.M = \frac{\sum w_i}{\sum \frac{w_i}{x_i}}$$

Example 1: A driver travel for 3 days. On the 1st day he drives for 10h at a speed of 48 km/h, on the 2nd day for 12h at 45 km/h and on the 3rd day for 15h at 40 km/h. What is the average speed?

Solution: since the distance covered by the driver is not equal, so we use WHM by taking the distance as weights (w_i).

$$v_{av} = w.h.m = \frac{\sum w_i}{\sum \frac{w_i}{x_i}} = \frac{(480 + 540 + 600)km}{\left(\frac{480}{40} + \frac{540}{45} + \frac{600}{40}\right)h_r} = 43.32\ km/hr.$$

Properties of harmonic mean

- It is based on all observation in a distribution.
- Used when a situations where small weight is give for larger observation and larger weight for smaller observation
- Difficult to calculate and understand
- Appropriate measure of central tendency in situations where data is in ratio, speed or rate.

Relation between AM, GM, and Hm

- If all the values in a data set are the **same**, then all the three means (arithmetic mean, GM and HM) will be **identical**.
- As the **variability** in the data **increases**, the **difference** among these means also **increases**.
- Arithmetic mean is always greater than the GM, which in turn is always greater than the HM.
  - **AM > GM > HM**

Median

- If the sample data are arranged in increasing order, the **median** is
  - if *n* is an odd number, median is middle value

$$\tilde{x} = the \left(\frac{n+1}{2}\right)^{th} positioned \text{ observation}$$

    - Example: **systolic blood pressure** of seven persons were given as 113, 124, 124, 132, 146, 151, and 170. what is the median systolic blood pressure? (Ans: 132)

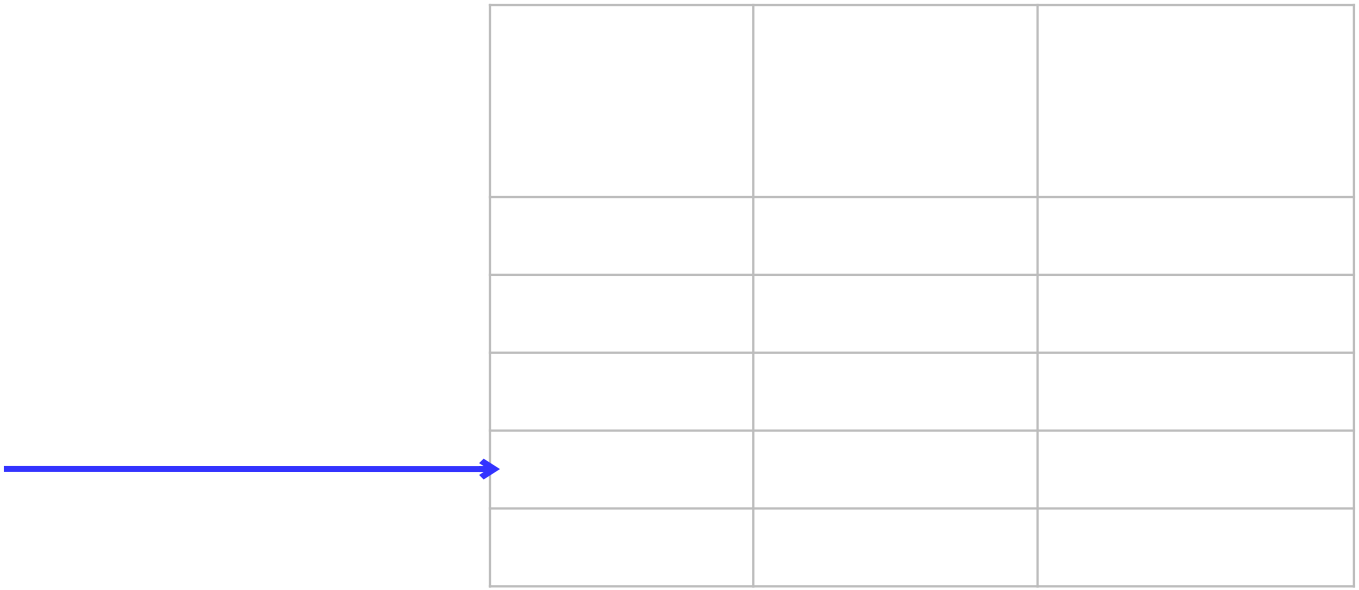
  - if *n* is an even number, midway between the two middle values

$$\tilde{x} = \frac{\left(\frac{n}{2}\right)\text{th observation} + \left(\frac{n}{2}+1\right)\text{th observation}}{2}$$

    - Six men with high cholesterol participated in a study to investigate the effects of diet on cholesterol level. At the beginning of the study, their **cholesterol levels** (mg/dL) were as follows:366, 327, 274, 292, 274 and 230. what is the median cholesterol level? (Ans:283)

Median …

- If the data is in ungrouped frequency distribution, median is the class with largest less than cumulative frequency smaller than or equal to half of the total observation
  - Example: Forty five students were taken to field and evaluated their **performance using 60m pure speed test**. The time is recorded in seconds, and the result is summarized in the table. What is the median performance of these students. (Ans: 19 secs)

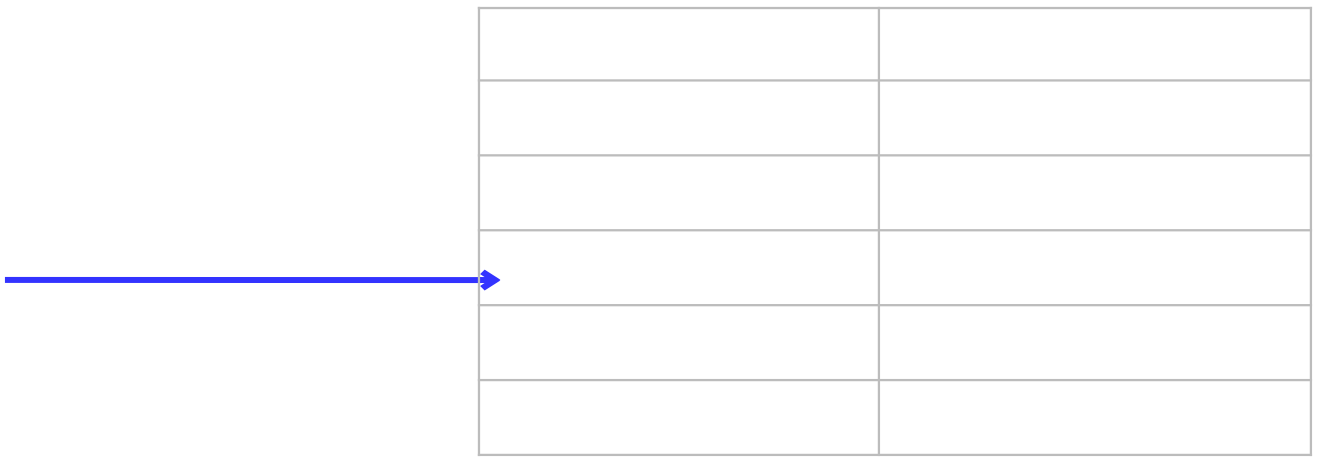


Median …

– If the data is in grouped frequency distribution, median is

$$\tilde{x} = L_m + \frac{\left(\frac{n}{2} - cf\right)}{f_m} w$$

- Example: fifty students were taken to field and evaluated their performance using **100 m pure speed test**. The time is recorded in seconds, and the result is summarized in the table. What is the median performance of these students. (Ans: 20.81 secs)



Mode

- The most frequent observation (value) in a data
- An observation with the largest frequency
  - There can be no mode Eg: 25, 27, 22, 18
  - There can be only one mode-unimodal Eg: 25, 27, 22, 25,18
  - There can be two mode-bimodal Eg: 25, 27, 22, 27, 25, 18, 20
  - There can be more than two mode-multimodal Eg: 25, 27, 22, 27, 25, 18, 20, 19, 22, 17
- Mode grouped frequency distribution

$$\hat{x} = L_1 + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right)w$$

  - f_1 = frequency of the modal class
  - f_0 = frequency of the class preceding the modal class
  - f_2 = frequency of the class next to the modal class

Mode…

- The most frequent observation (value) in a data
  - Example: Twenty five **amateur cyclists** were taken to field and their **time** is **recorded to complete a given distance**. The time is recorded in seconds, and the result is summarized in the table. What is the modal time to complete the distance. (Ans: 29.5 secs)

2.3 Quantiles

- *Quartiles* are three points which divide an array into four parts in such a way that each portion contains an equal number of elements.
  - First quartile (Q_1) 25% of the observations lies below or equal to it
  - Second quartile (Q_2) 50 % of the observations lies below or equal to it and
  - Third quartile (Q_3) 75% of the observations lies below or equal to it
- The i^{th} quartile for raw data is

$$Q_i = \frac{i(n+1)}{4}$$

- If there is an even number of data items, then we need to get the average of the middle numbers.

Quantiles

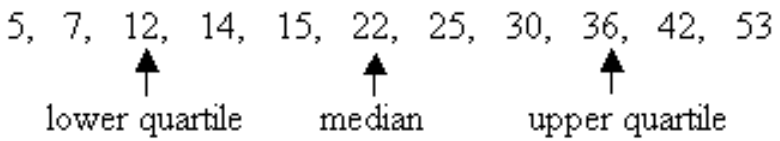
- *Example: Find* the median, lower quartile and upper quartile of the following numbers.

  a) 12, 5, 22, 30, 7, 36, 14, 42, 15, 53, 25

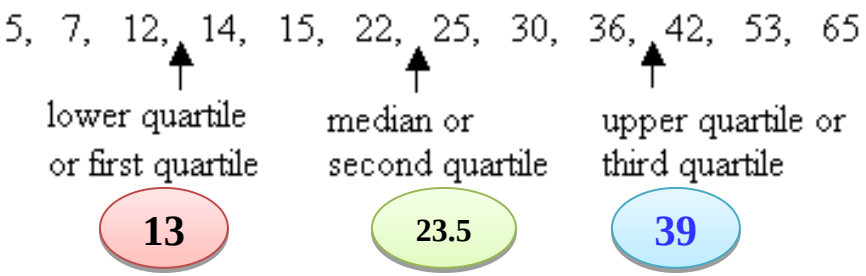
  b) 12, 5, 22, 30, 7, 36, 14, 42, 15, 53, 25, 65

- **Solution**: first arrange data from smallest to largest

a)



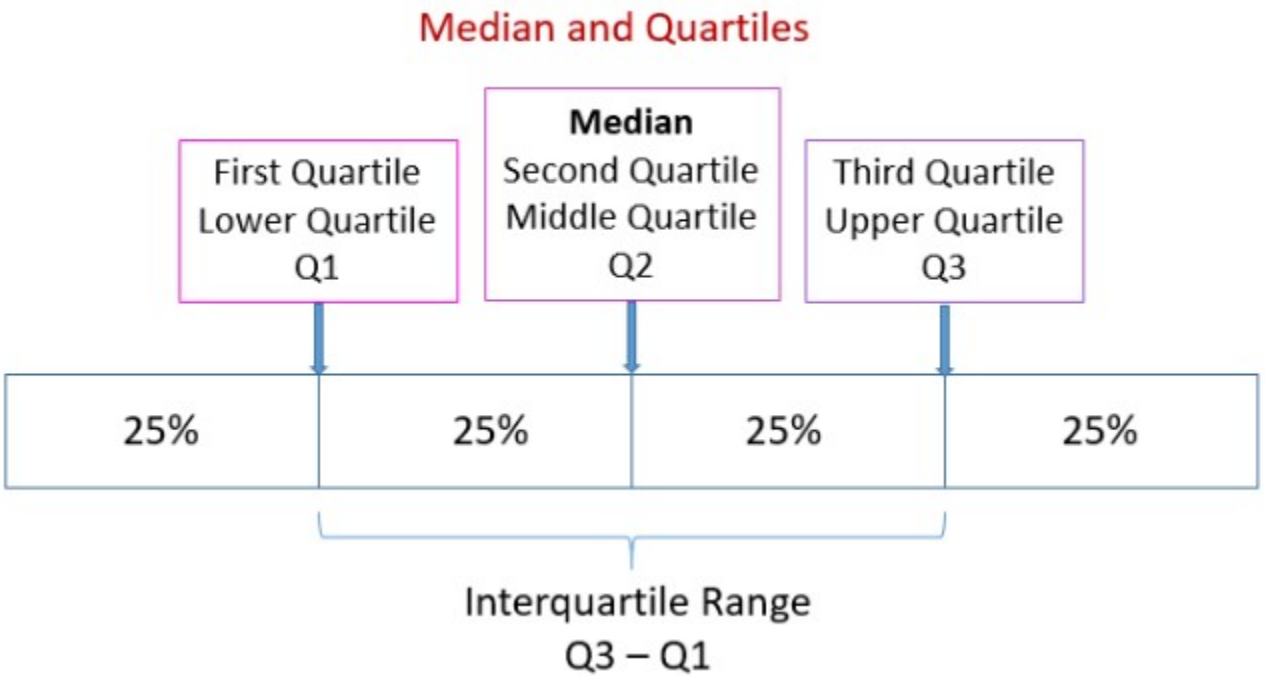
b)



Quantiles

- The i^{th} quartile for grouped frequency distribution is

$$\mathbf{Q_i} = \mathbf{lcb}_{q_i} + \frac{\left(\frac{i\,x\,n}{4} - lcf_p\right) x\, w}{f_{q_i}} \quad , \quad for \;\; i = 1, 2, 3.$$



Quantiles …

- *Deciles* are nine points which divide an array into 10 parts in such a way that each part contains equal number of elements.
  - The nine deciles are denoted by $D_1, D_2, \ldots, D_9$
  - First decile (D_1) 10% of the observations lies below or equal to it
  - Second decile (D_2) 20% of the observations lies below or equal to it etc
- The i^{th} decile for grouped frequency distribution is

$$\mathbf{D_i} = \mathbf{lcb}_{d_i} + \frac{\left(\frac{i\,x\,n}{10} - lcf_p\right) x\, w}{f_{d_i}} \quad , \quad for \quad i = 1, 2, \ldots, 9$$

Quantiles …

- *Percentiles* are 99 points which divide an array into 100 parts in such a way that each part consists of equal number of elements.
  - The ninty nine percentiles are denoted by P_1, P_2, …, P_{99}
  - First percentile (P_1) 1% of the observations lies below or equal to it
  - Second percentile (P_2) 2% of the observations lies below or equal to it etc
- The i^{th} percentile for grouped frequency distribution is

$$\mathbf{P_i} = \mathbf{lcb}_{p_i} + \frac{\left(\frac{i\ x\ n}{100} - lcf_p\right) x\ w}{f_{p_i}} \quad , \quad for \quad i = 1, 2, \ldots, 99$$

Quantiles …

- Example:- The following frequency distribution is the **score** of 25 students.

Compute the following quantities

- First quartile (Ans:44.92)
- Ninth decile (Ans:65.75)
- forty fifth percentile (Ans:51.38)

Remark:

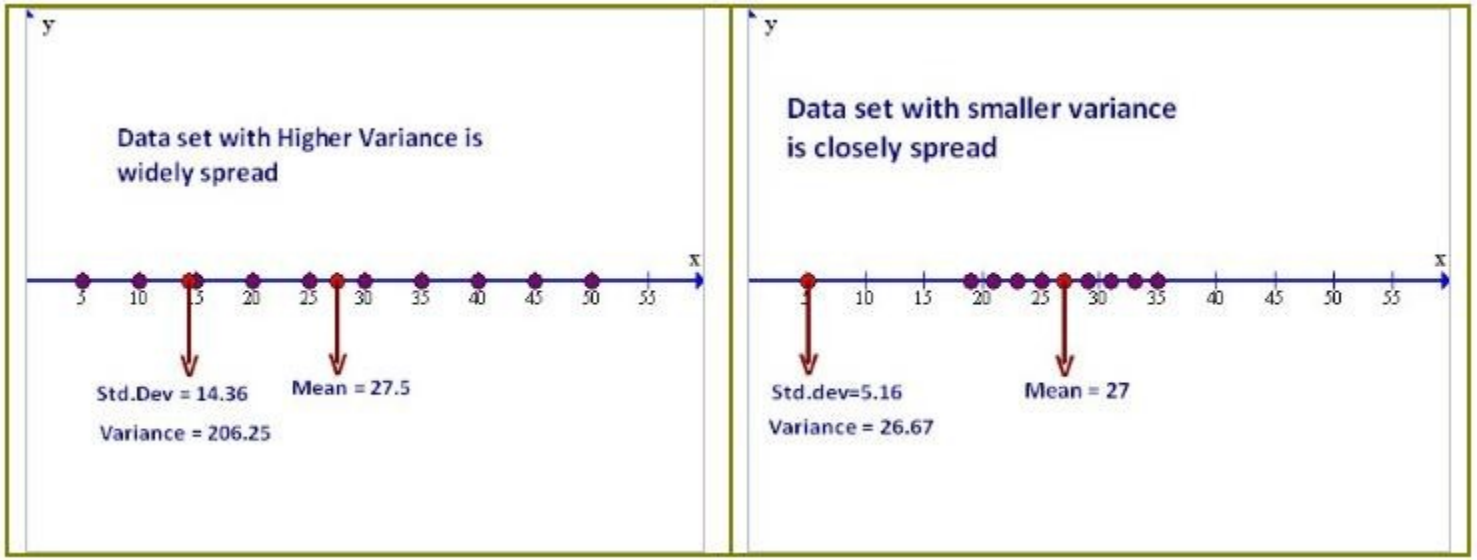
$$Q_1 = P_{25}$$

$$Q_2 = D_5 = P_{50} = Median$$

$$Q_3 = P_{75}$$

$$D_1 = P_{10}; D_2 = P_{20}; \cdots; D_9 = P_{90}$$

2.4 Measures of Dispersion



Introduction

- Central tendency measures do not reveal the variability present in the data.
- Dispersion is the scatteredness of the data series around it average.
- Dispersion is the extent to which values in a distribution differ from the average of the distribution
- A measure of statistical dispersion is a nonnegative real number that is zero if all the data are the same and increases as the data become more diverse.

- **Why we need measures of dispersion?**
  - Determine the reliability of an average
  - Serve as a basis for the control of the variability
  - To compare the variability of two or more series and
  - Facilitate the use of other statistical measures.

Introduction…

- **Properties of a good measures of dispersion**
  - It should be rigidly defined
  - It should be easy to understand and to calculate
  - It should be based on all observations of data
  - It should be easily subjected to further mathematical treatment
  - It should be least affected by sampling fluctuation
  - It shouldn’t be unduly affected by extreme values

Introduction…

- There are many types of dispersion measures
  - Range /Relative Range (Coefficient of range)
  - Inter Quartile Range/ coefficient of quartile deviation
  - Mean Absolute Deviation /Coefficient of mean deviation
  - Variance/Standard Deviation/ coefficient of variation
- Measures of dispersion cane be absolute or relative.
  - When measurements are observed with different units, or have different averages use relative measures of dispersion.

Range (R)

- Range is the difference between two extreme values in a data
- Denoted by R

$$R = \max - \min$$

- Only two values are used in its calculation.
- It is influenced by an extreme value (non-robust).
- It is easy to compute and understand.

Relative Range (RR)

- Relative range is the ratio of the difference and sum of the two extreme values in a data
- Denoted by RR/CR

$$RR = \frac{\max - \min}{\max + \min}$$

- Example: what is the range and relative range of the following data: 4, 8, 1, 6, 6,2, 9, 3, 6, 9. (Ans: R=8, RR=0.8)

Properties of range

- It is the simplest crude measure and can be easily understood
- It takes into account only two values which causes it to be a poor measure of dispersion
- Very sensitive to extreme observations
- The larger the sample size, the larger the range

Inter Quartile Range

- Measures the range of the middle 50% of the values only
- Is defined as the difference between the upper and lower quartiles
- Interquartile range = upper quartile - lower quartile

$$= Q_3 - Q_1$$

- The **semi-interquartile range** (or SIR) is defined as the difference of the first and third quartiles divided by two

$$SIR = (Q_3 - Q_1) / 2$$

- The SIR is often used with skewed data as it is insensitive to the extreme scores

Coefficient of Quartile Deviation

- The ratio of the difference to sum of the two extreme quartiles of a data. Denoted by CQD

$$CQD = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

- Example: The following data are recorded: 9, 7, 3, 7, 1, 2, 5, 4, 5, 10, 10, 2, 2, 2, 6, 7, 9, 8, 5, 6. What are the SIR and CQD for the recorded data?
- Solution: put in ascending order: 1, 2, 2, 2, 2, 3, 4, 5, 5, 5, 6, 6, 7, 7, 7, 8, 9, 9, 10, 10. (Ans: SIR=2.5, CQD=0.5)

Properties of IQR

- It is a simple and versatile measure
- It encloses the central 50% of the observations
- It is not based on all observations but only on two specific values
- Since it excludes the lowest and highest 25% values, it is not affected by extreme values
- Less sensitive to the size of the sample

Mean Absolute Deviation (MAD)

- Measures the ‘average’ distance of each observation away from the mean of the data
- Gives an equal weight to each observation
- Generally more sensitive than the range or interquartile range, since a change in any value will affect it
- The Mean Absolute Deviation of a set of n numbers is

$$MAD = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

  - All values are used in the calculation.
  - It is not unduly influenced by large or small values (robust)
  - The absolute values are difficult to manipulate.

Coefficient of Mean Deviation (CMD)

$$CMD = \frac{MAD}{\bar{x}}$$

- All values are used in the calculation.
- It is not unduly influenced by large or small values (robust)
- The absolute values are difficult to manipulate.

- **Example**: For the following data

  52.5, 46.8, 38.8, 37.6, 32.3.

- Compute MAD and CMD?
- Solution: (**Ans: MAD=6.44, CMD=0.16**)

Solution

Step 2

Step 3

Observation	x	$x - \bar{x}$	$\lvert x - \bar{x} \rvert$
1			
2			
3			
4			
5			
Total			
Mean	Step 1	0	Step 4 / Step 5 **6.44**

Properties of mean deviation

- MD removes one main objection of the earlier measures, that it involves each value
- It is not affected much by extreme values
- Its main drawback is that algebraic negative signs of the deviations are ignored which is mathematically unsound

Variance

- Variance is the mean of squared deviation of observations from their arithmetic mean

$$Population\ variance = \sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} \quad \rightarrow for\ population.$$

$$Sample\ variance = s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \quad \rightarrow for\ sample.$$

  - All values are used in the calculation.
  - It is not extremely influenced by outliers (non-robust).
  - The units of variance are awkward: the square of the original units.
    - Therefore standard deviation is more natural since it recovers the original units.

- In general, the sample variance is computed by:

$$s^2 = \begin{cases} \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \dfrac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}. & \rightarrow for\ raw\ data. \\ \dfrac{\sum_{i=1}^{k} f_i(x_i - \bar{x})^2}{\sum_{i=1}^{k} f_i - 1} = \dfrac{\sum_{i=1}^{k} f_i x_i^2 - n\bar{x}^2}{n-1}. & \rightarrow for\ ungrouped\ data. \\ \dfrac{\sum_{i=1}^{k} f_i(m_i - \bar{x})^2}{\sum_{i=1}^{k} f_i - 1} = \dfrac{\sum_{i=1}^{k} f_i m_i^2 - n\bar{x}^2}{n-1}. & \rightarrow for\ grouped\ data. \end{cases}$$

Standard Deviation

- One of the most useful measures of dispersion is the ***standard deviation*.**
- It is based on ***deviations from the mean*** of the data.
- The **sample standard deviation** is found by calculating the square root of the **variance**.

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}.$$

- To calculate standard deviation follow this step

1. Calculate the mean of the numbers
2. Find the deviations from the mean.
3. Square each deviation
4. Sum the squared deviations
5. Divide the sum in Step 4 by $n - 1$
6. Take the square root of the quotient in Step 5

Example 1: Compute the variance for the sample: 5, 14, 2, 2 and 17.

Solution: $n = 5$, $\sum_{i=1}^{n} x_i = 40$, $\bar{x} = 8$, $\sum_{i=1}^{n} x_i^2 = 518$.

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} = \frac{518 - 5\,x\,8^2}{5-1} = 49.5. \quad , \qquad S = \sqrt{49.5} = 7.04.$$

Example 2: Suppose the data given below indicates time in minute required for a laboratory experiment to compute a certain laboratory test. Calculate the mean, variance and standard deviation for the following data.

x_i	32	36	40	44	48	***Total***
f_i	2	5	8	4	1	***20***
$f_i x_i$	64	180	320	176	48	***788***
$f_i x_i^2$	2048	6480	12800	7744	2304	***31376***

$\bar{x} = 39.4$, $s^2 = \frac{31376 - 20\,x\,39.4^2}{19} = 17.31.$, $S = \sqrt{17.31} = 4.16.$

Properties of Variance

- The variance is always non-negative ($s^2 \geq 0$).
- If every element of the data is multiplied by a constant "c", then the new variance

$$s^2_{new} = c^2 \, x \, s^2_{old} \, .$$

- When a constant is added to all elements of the data, then the variance does not change.
- The variance of a constant (c) measured in ***n times*** is zero. i.e. (var(c) = 0).

Coefficient of Variation

- The Coefficient of Variation (CV) for a data set defined as the ratio of the standard deviation to the mean
- It shows the extent of variability in relation to mean of the population.
- It is a normalized measure of dispersion of a probability distribution or frequency distribution.

$$CV = \frac{s}{\bar{x}} \times 100\%$$

  - All values are used in the calculation.
  - The actual value of the CV is independent of the unit in which the measurement has been taken, so it is a dimensionless number.
  - For comparison between data sets with different units or widely different means, one should use the coefficient of variation instead of the standard deviation.

Coefficient of Variation

Example: Last semester, the students of Biology and Chemistry Departments took *Stat 273* course. At the end of the semester, the following information was recorded.

Department	**Biology**	**Chemistry**
Mean score	79	64
Standard deviation	23	11

Compare the relative dispersions of the two departments' scores using the appropriate way.

Solution:

Chemistry Department	Biology Department
$CV = \frac{11}{64} \times 100 = 17.19\%$	$CV = \frac{23}{79} \times 100 = 29.11\%$

Since the CV of Biology Department students is greater than that of Chemistry Department students, we can say that there is more dispersion in the distribution of Biology students' scores compared with that of Chemistry students.

2.5 Standard Score

- If X is a measurement from a distribution with mean $\overline{X}$ and standard deviation S, then its value in standard units is

$$Z = \frac{X - \overline{X}}{S}$$

- Z gives the deviations from the mean in units of standard deviation
- Z gives the number of standard deviation a particular observation lie above or below the mean.
- It is used to compare two observations coming from different groups

Standard Score

- Example: Two groups of people were trained to perform a certain task and tested to find out which group is faster to learn the task. For the two groups the following information was given:

Value	Group one	Group two
Mean	10.4 min	11.9 min
Stan.dev.	1.2 min	1.3 min

- Relatively speaking:

a) Which group is more consistent in its performance? (Ans: Group 2)

b) Suppose a person A from group one take 9.2 minutes while person B from Group two take 9.3 minutes, who was faster in performing the task? Why? (Ans: person B is faster)

Solution

Coefficient of variation for group 1: $CV = \frac{S_1}{\bar{x}_1} \times 100\% = \frac{1.2}{10.4} \times 100\% = 11.54\%$

Coefficient of variation for group 2: $CV = \frac{S_2}{\bar{x}_2} \times 100\% = \frac{1.3}{11.9} \times 100\% = 10.92\%$

CV for group 2 < CV for group 1 ▭ group 2 is more consistent

Z-score of Person A: $Z = \frac{x_A - \bar{x}_1}{S_1} = \frac{9.2 - 10.4}{1.2} = -1.00$

Z-score of Person B: $Z = \frac{x_B - \bar{x}_2}{S_2} = \frac{9.3 - 11.9}{1.3} = -2.00$

Z-score of Person B < Z-score of Person A ▭ Person B is faster than person A