UNIT – II

Measures of Central Tendency: Mean – Median-Mode –Geometric mean-Harmonic mean (only definition) - Measures of dispersion: Range-Quartile deviation-Mean deviation-Standard deviation-Lorenz curve

Measures of Central Tendency

Mean,*Median and Mode*
for Ungrouped Data
Basic Statistics

Measures of Central Tendency

Measures of central tendency are numerical descriptive measures which indicate or locate the center of a distribution or data set.

In layman's term, a measure of central tendency is an AVERAGE. It is a single number of value which can be considered typical in a set of data as a whole.

For example, in a class of 40 students, the average height would be the typical height of the members of this class as a whole.

"Add the numbers, divide by how many numbers you've added and there you have it-the average amount of minutes you sleep in class each day."

MEAN

The **MEAN** of a set of values or measurements is the sum of all the measurements divided by the number of measurements in the set.

Among the three measures of central tendency, the mean is the most popular and widely used. It is sometimes called the arithmetic mean.

If we compute the mean of the population, we call it the parametric or population mean, denoted by μ (read "mu").

If we get the mean of the sample, we call it the sample mean and it is denoted by $\bar{x}$ (read "x bar").

Mean for Ungrouped Data

For ungrouped or raw data, the mean has the following formula.

$$\bar{x} = \frac{\sum x}{n}$$

where
$\bar{x}$ = mean
$\sum x$ = sum of the measurements or values
n = number of measurements

Example 1:

Ms. Sulit collects the data on the ages of Mathematics teachers in Santa Rosa School, and her study yields the following:

38 35 28 36 35 33 40

Solution:

$$\bar{x} = \frac{38 + 35 + 28 + 36 + 35 + 33 + 40}{7}$$

$= \mathbf{35}$

Based on the computed mean, 38 is the average age of Mathematics teachers in SRS.

Your turn!

Mang John is a meat vendor. The following are his sales for the past six days. Compute his daily mean sales.

Tuesday	P 5 800
Wednesday	8 600
Thursday	6 500
Friday	4 300
Saturday	12 500
Sunday	13 400

Solution:

$$\bar{x} = \frac{5800 + 8600 + 6500 + 4300 + 12500 + 13400}{6}$$

$= 51,100$

The average daily sales of Mang John is P51,100.

Weighted Mean

Weighted mean is the mean of a set of values wherein each value or measurement has a different weight or degree of importance. The following is its formula:

$$\bar{x} = \frac{\sum xw}{\sum w}$$

where $\bar{x}$ = mean

x = measurement or value

w = number of measurements

Example

Below are Amaya's subjects and the corresponding number of units and grades she got for the previous grading period. Compute her grade point average.

Subject	Units	Grade
Filipino	.9	86
English	1.5	85
Mathematics	1.5	88
Science	1.8	87
Social Studies	.9	86
TLE	1.2	83
MAPEH	1.2	87

$$\bar{x} = \frac{86(.9) + 85(1.5) + 88(1.5) + 87(1.8) + 86(.9) + 83(1.2) + 87(1.2)}{.9 + 1.5 + 1.5 + 1.8 + .9 + 1.2 + 1.2}$$

$$= 86.1$$

Amaya's average grade is 86.1

Your turn!

James obtained the following grades in his five subjects for the second grading period. Compute his grade point average.

Subject	Units	Grade
Math	1.5	90
English	1.5	86
Science	1.8	88
Filipino	0.9	87
MAKABAYAN	1.5	87

Solution:

$$\bar{X} = \frac{90 * 1.5 + 86 * 1.5 + 88 * 1.8 + 87 * .9 + 87 * 1.5}{1.5 + 1.5 + 1.8 + .9 + 1.5}$$

$$= 87.67$$

James general average is 87.67

Properties of Mean

1. Mean can be calculated for any set of numerical data, so it always exists.
2. A set of numerical data has one and only one mean.
3. Mean is the most reliable measure of central tendency since it takes into account every item in the set of data.
4. It is greatly affected by extreme or deviant values (*outliers*)
5. It is used only if the data are interval or ratio.

MEDIAN

The **MEDIAN,** denoted Md, is the middle value of the sample when the data are ranked in order according to size.

16 17 18 19 20 21 22

↑ Median (19)

16 17 18 19 20 21 22 23

↑ Median (between 19 and 20)

$$\frac{19+20}{2} = 19.5$$

- **Median : (middle)**

- The "Median" of a data set is dependent on whether the number of elements in the data set is odd or even.
- First reorder the data set from the smallest to the largest
- Mark off high and low values until you reach the **middle**.
- If there 2 middles, add them and **divide** by 2.

- **Examples : Odd Number of Elements**
- Data Set = 2, 5, 9, 3, 5, 4, 7
- Reordered = 2, 3, 4, 5, 5, 7, 9

  ^
- Median = 5

- **Examples : Even Number of Elements**
- Data Set = 2, 5, 9, 3, 5, 4
- Reordered = 2, 3, 4, 5, 5, 9

  ^ ^

  Median = (4 + 5) / 2 = 4.5

Properties of Median

1. Median is the score or class in the distribution wherein 50% of the score fall below it and another 50% lie.
2. Median is not affected by extreme or deviant values.
3. Median is appropriate to use when there are extreme or deviant values.
4. Median is used when the data are ordinal.
5. Median exists in both quantitative or qualitative data.

MODE

The **MODE,** denoted Mo, is the value which occurs most frequently in a set of measurements or values. In other words, it is the most popular value in a given set.

Examples:

Find the Mode.

1. The ages of five students are: 17, 18, 23, 20, and 19
2. The following are the descriptive evaluations of 5 teachers: VS, S, VS, VS, O
3. The grades of five students are : 4.0, 3.5, 4.0, 3.5, and 1.0
4. The weights of five boys in pounds are: 117, 218, 233, 120, and 117

- **Mode :** (most often)

- The "Mode" for a data set is the element that occurs the most often.

- It is not uncommon for a data set to have more than one mode.

- This happens when two or more elements occur with equal frequency in the data set.

- **Example :**
- Data Set = 2, 5, 9, 3, 5, 4, 7
- Mode = 5

- **Example:**
- Data Set = 2, 5, 2, 3, 5, 4, 7
- Modes = 2 and 5

Properties

1. It is used when you want to find the value which occurs most often.
2. It is a quick approximation of the average.
3. It is an inspection average.
4. It is the most unreliable among the three measures of central tendency because its value is undefined in some observations.

- **Range :**

- The "Range" for a data set is the difference between the largest value and smallest value contained in the data set.

- First **reorder** the data set from smallest to largest then **subtract** the first element from the last element.

Of all the measures of variability, the **range** is the **easiest and quickest way** to determine. It is simply the **difference of the highest (H) and the lowest (L) scores** in a set of data under consideration.

Formula for the Range:

$$\mathbf{r = H - L}$$

where: r = range
H = highest score
L = lowest score

Example:
The scores of Maria in her math quizzes are as follows: 12, 25, 27, 29, 36, 38, 40, 43, 50, and 62. Find its range.

Solution:
Highest score (H) = 62   Lowest score (L) = 12

$$r = H - L$$
$$= 62 - 12$$
$$= 50$$

Therefore, the range is 50.

The range

Here are the high jump scores for two girls in metres.

Joanna	1.62	1.41	1.35	1.20	1.15
Kirsty	1.59	1.45	1.41	1.30	1.30

Find the range for each girl's results and use this to find out who is consistently better.

Joanna's range = 1.62 – 1.15 = **0.47**

Kirsty's range = 1.59 – 1.30 = **0.29**

Joanna	1.62	1.41	1.35	1.20	1.15
Kirsty	1.59	1.45	1.41	1.30	1.30

Calculate the mean and the range for each girl.

	Joanna	Kirsty
Mean	1.35 m	1.41 m
Range	0.47 m	0.29 m

Use these results to decide which one you would enter into the athletics competition and why.

Calculating the mean, median and range

9, 7, 9, 3, 7, 7, 7, 9, 5

mean: 7

median: 7

range: 6

Advantages

- Gives **quick approximation** of the variability of data.
- It is **not** very sophisticated/complicated.
- Used when the **mode** is preferred measure of central tendency. (i.e./ when you have **nominal/titular** level data.)
- It is the simplest measure of variability/dispersion.

Disadvantages

- It is **limited/partial**, if extreme scores are not representative of the sample, but are included among the scores.
- It is **not** very informative, because it is based only on the most extreme scores.
- It is severely affected by extreme scores in your data distribution.

Geometric mean

GEOMETRIC MEAN

In Mathematics, the geometric mean is a type of mean or average, which indicates the central tendency or typical value of a set of numbers by using the product of their values(as opposed to the arithmetic mean which uses their sum).

Formulas

- For grouped data:

$$G = \text{Anti}\left(\sum \frac{log x}{n}\right)$$

- For Grouped data:

$$G = \text{anti}\left(\sum \frac{f \log x}{n}\right)$$

Question 1: Find the geometric mean of the following values:
15, 12, 13, 19, 10

Solution:

Given data, 15, 12, 13, 19, 10

$n=5, \sum \log x = 5.648$

As $G = \text{Anti}\left(\frac{\sum \log x}{n}\right)$

$G = \text{Anti}\left(\frac{5.648}{5}\right)$

$G = \text{Anti}(1.129)$

$G = 13.48$

x	*Log x*
15	1.1761
12	1.0792
13	1.1139
19	1.2788
10	1.0000
Total	5.648

The geometric mean is a type of average , usually used for growth rates, like population growth or interest rates. While the arithmetic mean **adds** items, the geometric mean **multiplies** items. Also, you can only get the geometric mean for positive numbers.

Merits of Geometric Mean:

- It is based on all the observations
- It is rigidly defined
- It is capable of further algebraic treatment
- It is less affected by the extreme values
- It is suitable for averaging ratios, percentages and rates.

Limitations of Geometric Mean:

- It is difficult to understand
- The geometric mean cannot be computed if any item in the series is negative or zero.
- The GM may not be the actual value of the series
- It brings out the property of the ratio of the change and not the absolute difference of change as the case in arithmetic mean.

CALCULATION

Formula:

G.M.= $\sqrt[n]{X1 * X2 * X3 * \ldots\ldots * X4} = (X1*X2*X_3*\ldots\ldots X_n)^{1/n}$

For example,

GM of two numbers 4 and 9 is $\sqrt{4 * 9} = \sqrt{36} = 6$

GM of three numbers 1, 4 and 128 is $\sqrt[3]{1 * 4 * 128} = \sqrt{512} = 8$

- The above method can be applied if there are two or three items. But if n is a very large number, the problem of finding the nth root is a tedious work. Therefore, we make use of logarithms. The above formula reduced to its logarithmic form is
- Geometric mean = Antilog of logXi + logX2 + logX3…..logXn

$$\frac{\log X_i + \log X_2 + \log X_3 \ldots \log X_n}{N}$$

[or] Geometric Mean = Antilog of log/N

CALCULATION OF GEOMETRIC MEAN – INDIVIDUAL SERIES

STEP: 1. FIND OUT THE LOGARITHM OF EACH VALUE OR THE SIZE OF ITEM FROM THE LOG TABLE : LOG X

2. ADD ALL THE VALUES OF LOG - ∑LOGX

3. THE SUM OF (∑LOG) IS DIVIDED BY THE NUMBER OF ITEMS. ∑LOGX/N

4. FIND OUT THE ANTILOG OF THE QUOTIENT (FROM STEP 3). THIS IS THE GEOMETRIC MEAN OF THE DATA

HARMONIC MEAN

- **Harmonic mean** (formerly sometimes called the **subcontrary mean**) is one of several kinds of average.

- The harmonic mean is a very specific type of average.
- It's generally used when dealing with averages of units, like speed or other rates and ratios.

What is the harmonic mean of 1,5,8,10?

Here,
N=4

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

H = 4 / (1/1) +(1/5) + (1/8) + (1/10)
H = 4/ 1.425

H = 2.80

Rahul drives a car at 20 mph for the first hour and 30 mph for the second. What’s his average speed?

We need the harmonic mean:

$= 2/(1/20 + 1/30)$

$= 2(0.05 + 0.033)$

$= 2 / 0.083$

$= 24.09624$ mph.

Harmonic Mean

- Harmonic mean is quotient of "number of the given values" and "sum of the reciprocals of the given values".
- For Ungrouped Data

$$H.M \text{ of } X = \bar{X} = \frac{n}{\Sigma\left(\frac{1}{x}\right)}$$

- For grouped Data

$$H.M \text{ of } X = \bar{X} = \frac{\Sigma f}{\Sigma\left(\frac{f}{x}\right)}$$

Calculate the harmonic mean of the numbers: 13.2, 14.2, 14.8, 15.2 and 16.1

Solution:

The harmonic mean is calculated as below:

AS

$$H.M \text{ of } X = \overline{X} = \frac{n}{\sum\left(\frac{1}{x}\right)}$$

$$H.M \text{ of } X = \overline{X} = \frac{5}{0.3417} = 14.63$$

X	$\frac{1}{X}$
13.2	0.0758
14.2	0.0704
14.8	0.0676
15.2	0.0658
16.1	0.0621
Total	$\sum\frac{1}{X}$ =0.3147

Example: Calculate the harmonic mean for the given below:

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
F	2	3	11	20	32	25	7

Solution: Now We'll find H.M as:

$$\bar{X} = \frac{\Sigma f}{\Sigma\left(\frac{f}{x}\right)} = \frac{100}{1.4368} = 69.60$$

Marks	x	f	$\frac{f}{x}$
30-39	34.5	2	0.0580
40-49	44.5	3	0.0674
50-59	54.5	11	0.2018
60-69	64.5	20	0.3101
70-79	74.5	32	0.4295
80-89	84.5	25	0.2959
90-99	94.5	7	0.0741
Total		4.5	0.4295

MEASURE OF DISPERSION:

(RANGE, QUARTILE DEVIATION, MEAN DEVIATION, STANDARD DEVIATION)

QUARTILE DEVIATION

The measure of dispersion is provided by quartile-deviation or semi - inter –quartile range. The formula to find quartile-deviation is given by

$$Q_d = \frac{Q_3 - Q_1}{2}$$

Coefficient of Quartile Deviation

A relative measure of dispersion using quartiles is given by coefficient of quartile-deviation which is

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

Properties of Quartile Deviation

1) Quartile-deviation provides the best measure of dispersion for open-end classification.

2) It is less affected due to sampling fluctuations.

3) Like other measures of dispersion, quartile-deviation remains unaffected due to a change of origin but is affected in the same ratio due to change in scale.

i.e. if y = a + be, a and b being constants,

$$\text{then QD of } y = |b| \times \text{QD of } x$$

Solved Problems

Problem 1 :

Following are the marks of the 10 students :
56, 48, 65, 35, 42, 75, 82, 60, 55, 50
Find QD and also its coefficient.

Solution :

After arranging the marks in an ascending order, we get

35, 42, 48, 50, 55, 56, 60, 65, 75, 82

First quartile (Q_1) :

= (n+1)/4 th observation

= (10+1)/4 th observation

= 2.75th observation

= 2nd observation + 0.75 × difference between the 3rd and the 2nd observation

= 42 + 0.75 (48 - 42)

= 42 + 0.75 x 6

= 42 + 4.5

Q_1 = 46.5

Third quartile (Q_3) :

$= 3(n+1)/4$ th observation
$= 3(10+1)/4$ th observation
$= 8.25^{th}$ observation

$= 8^{th}$ observation + 0.25 × difference between the 9^{th} and the 8^{th} observation

$= 65 + 0.25\ (75 - 65)$

$= 65 + 0.25 \times 10$
$= 65 + 2.5$

$Q_3 = 67.5$

The formula to find QD is given by

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

$$QD = (67.50 - 46.50) / 2$$

$$QD = 21 / 2$$

$$QD = 10.5$$

The formula to find coefficient of QD is given by

$$= [(Q_3 - Q_1) / (Q_3 + Q_1)] \times 100$$

$$= [(67.50 - 46.50) / (67.50 + 46.50)] \times 100$$

$$= [21 / 114] \times 100$$

Coefficient of QD = 18.42

Problem 2 :

If the QD of x is 6 and 3x + 6y = 20, what is the QD of y ?

Solution :

Let us write the given equation 3x + 6y = 20 as y = ax + b,
So, we get

$$y = (20/6) + (-3/6)x$$

When 'x' and 'y' are related as y = a + bx, then

$$\text{QD of 'y'} = |b| \times \text{QD of 'x'}$$
$$\text{QD of 'y'} = |-3/6| \times 6$$
$$\text{QD of 'y'} = 3$$

MEAN DEVIATION

WHAT IS MEAN DEVIATION?

THE MEAN DEVIATION IS THE FIRST MEASURE OF DISPERSION THAT WE WILL USE THAT ACTUALLY USES EACH DATA VALUE IN ITS COMPUTATION. IT IS THE MEAN OF THE DISTANCES BETWEEN EACH VALUE AND THE MEAN. IT GIVES US AN IDEA OF HOW SPREAD OUT FROM THE CENTER THE SET OF VALUES IS.

FORMULA FOR MEAN DEVIATION: (UNGROUPED DATA)

$$MD = \frac{\sum |x - \mu|}{n}$$

WHERE,

μ = MEAN

X = EACH VALUE

N = NUMBER OF VALUES

EACH VALUE FROM THEIR MEAN.

THREE STEPS ON FINDING THE MEAN:

1) FIND THE MEAN OF ALL VALUES.

2) FIND THE DISTANCE OF EACH VALUE FROM THAT MEAN.

3) FIND THE MEAN OF THOSE DISTANCES.

Example :

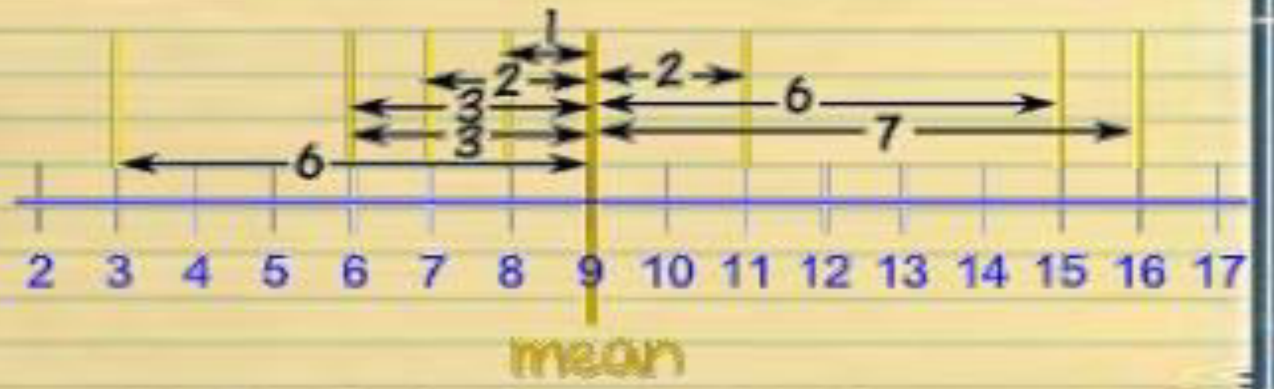
Find the mean deviation of 3, 6, 6, 7, 8, 11, 16 and 15.

Step 1 : Find the mean.

$$\text{Mean} = \frac{3+6+6+7+8+11+16+15}{8} = \frac{72}{8} = 9$$

STEP 2: FIND THE DISTANCE OF EACH VAUE FROM THE MEAN.

Value	Distance from 9
3	6
6	3
6	3
7	2
8	1
11	2
15	6
16	7



6+3+3+2+1 = 2+6+7

15 = 15

It tells us how far, on average, all values are from the middle.

Step 3: Find the mean of those distances.

$$MD = \frac{6+3+3+2+1+2+6+7}{8} = \frac{30}{8} = 3.75$$

So, the mean = 9 and the mean deviation is 3.75.

EXERCISES :

1) A BOOKLET HAS 12 PAGES WITH THE FOLLOWING NUMBERS OF WORDS:

271, 354, 296, 301, 333, 326, 285, 298, 327, 316, 287 AND 314

WHAT IS THE MEAN DEVIATION OF THE NUMBER OF WORDS PER PAGE?

STEP 1:

FIND THE MEAN.

RAW DATA:

271, 354, 296, 301, 333, 326, 285, 298, 327, 316, 287, 314

$$\mu = \frac{271 + 354 + 296 + 301 + 333 + 326 + 285 + 298 + 327 + 316 + 287 + 314}{12}$$

$$\frac{3708}{12} = 309$$

STEP 2: FIND THE ABSOLUTE DEVIATIONS.

x	$\lvert x-\mu \rvert$
271	38
354	45
296	13
301	8
333	24
326	17
285	24
298	11
327	18
316	7
287	22
314	5
	$\Sigma \lvert x-\mu \rvert = 232$

STEP 3: FIND THE MEAN DEVIATION

$$M.D. = \frac{\sum \left| x - \mu \right|}{N} = \frac{232}{12}$$

19.33

FORMULA FOR MEAN DEVIATION: (GROUPED DATA)

$$MD = \frac{\sum f\,|x - \mu|}{\sum f}$$

WHERE,

μ = mean

x = each value

f = frequency

THREE STEPS ON FINDING THE MEAN:

1) FIND THE MEAN BY USING THE FORMULA

$$\mu = \frac{\Sigma fx}{\Sigma f}$$

2) SOLVE FOR $|x - \mu|$ AND MULTIPLY IT TO THE FREQUENCY OF EACH CLASS. FIND THE

$$\Sigma f|x - \mu|$$

3) DIVIDE THE ANSWER OF $\Sigma f|x - \mu|$ TO THE Σf

STEP 1: FIND THE MEAN BY USING THE GIVEN FORMULA:

$$\mu = \frac{\sum fx}{\sum f}$$

x	f	fx
0	4	0
1	12	12
2	8	16
3	2	6
4	1	4
5	2	10
6	1	6
	$\sum f = 30$	$\sum fx = 54$

SO,

$$\mu = \frac{\sum fx}{\sum f} = \frac{54}{30} =$$

1.8

STEP 2: COMPLETE THE TABLE.

x	f	fx	$\lvert x-\mu \rvert$	$f\lvert x-\mu \rvert$
0	4	0	1.8	7.2
1	12	12	0.8	9.6
2	8	16	0.2	1.6
3	2	6	1.2	2.4
4	1	4	2.2	3.3
5	2	10	3.2	6.4
6	1	6	4.2	4.2
	$\Sigma f = 30$	$\Sigma fx = 54$		$\Sigma f\lvert x \quad \mu \rvert = 33.6$

STEP 3: DIVIDE THE ANSWER OF $\Sigma f|x-\mu|$ TO THE SUMMATION OF Σf

$$\frac{\Sigma f|x-\mu|}{\Sigma f} = \frac{33.6}{30} = 1.12$$

MEAN DEVIATION = 1.12

VARIANCE

Variance

Variance is the average squared deviation from the mean of a set of data. It is used to find the **standard deviation**.

Variance

1. Find the mean of the data.

   Hint – mean is the average so add up the values and divide by the number of items.

2. Subtract the mean from each value – the result is called the **deviation from the mean**.
3. Square each deviation of the mean.
4. Find the sum of the squares.
5. Divide the total by the number of items.

Variance Formula

The variance formula includes the Sigma Notation, $\sum$, which represents the sum of all the items to the right of Sigma.

$$\frac{\sum (x-\mu)^2}{n}$$

Mean is represented by μ and n is the number of items.

STANDARD DEVIATION

Standard Deviation

- The concept of standard deviation was first introduced by **Karl Pearson** in 1893.
- Karl Pearson after observing all these things has given us a more scientific formula for calculating or measuring dispersion. While calculating **SD we take deviations of individual observations from their AM and then each squares. The sum of the squares is divided by the Total number of observations. The square root of this sum is knows as standard deviation.**
- The standard deviation is the most useful and the most popular measure of dispersion.
- It is always calculated from the **arithmetic mean**, **median** and **mode** is not considered.

Standard Deviation

Standard Deviation shows the variation in data. If the data is close together, the standard deviation will be small. If the data is spread out, the standard deviation will be large.

Standard Deviation is often denoted by the lowercase Greek letter σ

sigma, .

Standard Deviation

Find the **variance**.

a) Find the **mean** of the data.

b) Subtract the mean from each value.

c) Square each deviation of the mean.

d) Find the sum of the squares.

e) Divide the total by the number of items.

Take the square root of the variance.

Standard Deviation Formula

The standard deviation formula can be represented using Sigma Notation:

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{n}}$$

Notice the standard deviation formula is the square root of the variance.

Find the variance and standard deviation

The math test scores of five students are: 92,88,80,68 and 52.

1) Find the **mean**: (92+88+80+68+52)/5 = 76.

2) Find the **deviation from the mean**:

92-76=16

88-76=12

80-76=4

68-76= -8

52-76= -24

Find the variance and standard deviation

The math test scores of five students are: 92,88,80,68 and 52.

3) Square the deviation from the mean:

$$(16)^2 = 256$$

$$(12)^2 = 144$$

$$(4)^2 = 16$$

$$(-8)^2 = 64$$

$$(-24)^2 = 576$$

Find the variance and standard deviation

The math test scores of five students are: 92,88,80,68 and 52.

4) Find the sum of the squares of the deviation from the mean:
   256+144+16+64+576= 1056

5) Divide by the number of data items to find the **variance**:
   1056/5 = 211.2

Find the variance and standard deviation

The math test scores of five students are: 92,88,80,68 and 52.

6) Find the square root of the variance: $\sqrt{211.2} = 14.53$

Thus the standard deviation of the test scores is 14.53.

Find The Variance And Standard Deviation

The math test scores of five students are:

92,88,80,68 and 52.

1) Find the **mean**:

(92+88+80+68+52)/5 = 76.

2) Find the **deviation from the mean**:

92-76=16

88-76=12

80-76=4

68-76= -8

52-76= -24

3) Square the deviation from the mean:

$$(16)^2 = 256$$
$$(12)^2 = 144$$
$$(4)^2 = 16$$
$$(-8)^2 = 64$$
$$(-24)^2 = 576$$

4) Find the sum of the squares of the deviation from the mean:

256+144+16+64+576= 1056

5) Divide by the number of data items to find the **variance**:

1056/5 = 211.2

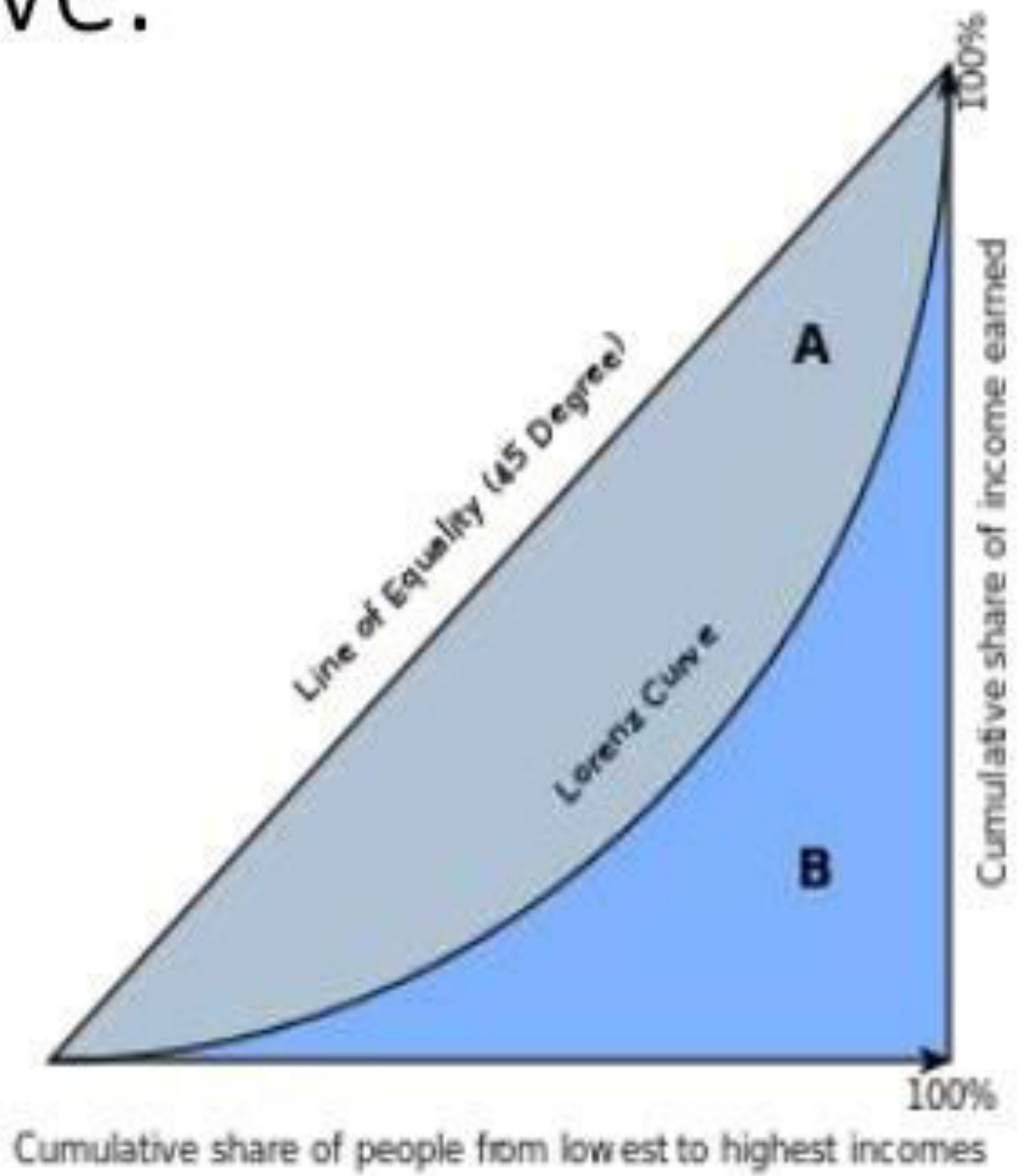
6) Find the square root of the variance:

$$\sqrt{211.2} = 14.53$$

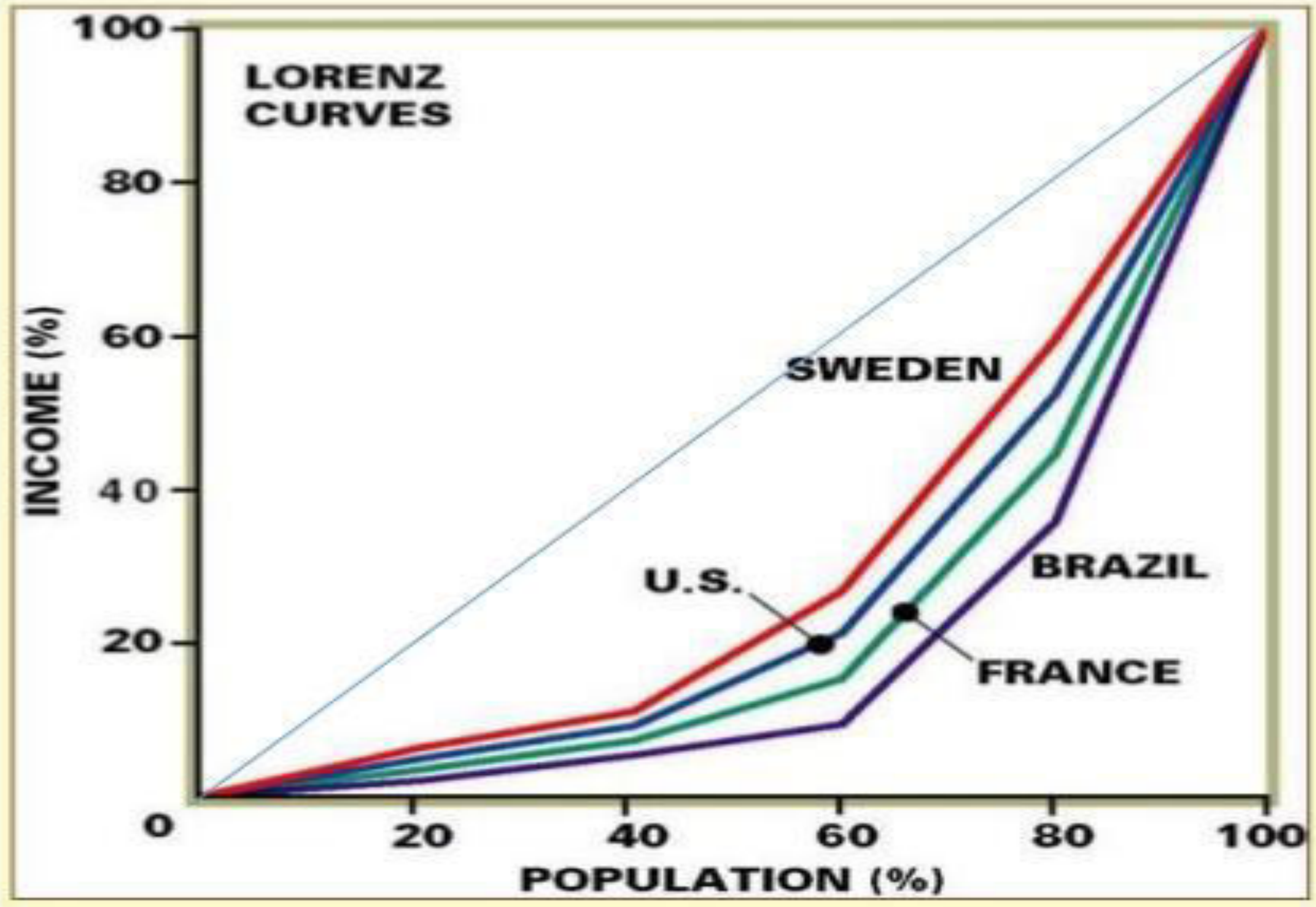
Thus the standard deviation of the test scores is 14.53.

The Lorenz curve:

The Lorenz curve represents income distribution. The line of equality shows what an equal distribution would look like, whereas he Lorenz curve shows the actual distribution.
The greater the degree of the bow the greater the …………..
………………………………



Annotate your diagram to show x and y axis.


LORENZ
CURVES
100
80
60
40
20
0
20
40
60
80
100
INCOME (%)
POPULATION (%)
SWEDEN
U.S.
BRAZIL
FRANCE

To draw a Lorenz Curve, follow these steps:

1. Gather the data (e.g. census data from two cities)
2. For each set of data, rank the categories and order them by rank in a table.
3. Convert each value in a % of the total.
4. Calculate the running totals (ie cumulative %, by adding the % of one line to the ones before)