§1.5 DISCRETE VARIABLE ONLY WHOLE #'S.

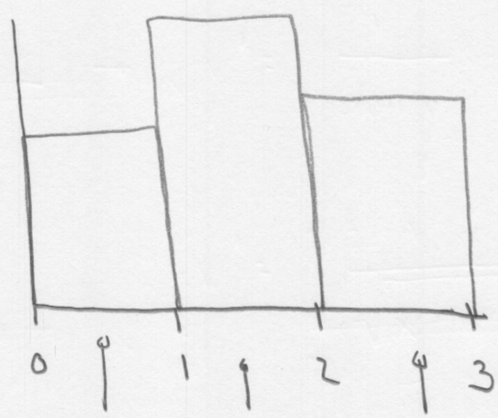
0, 1, 2, ...

IF YOUR DATA IS DISCRETE & ONLY TAKES ON A FEW DIFFERENT VALUES THEN

CLASS 1: $[0,1) \longrightarrow 0$

CLASS 2: $[1,2) \longrightarrow 1$

IN THIS CASE, OUR CLASSES ARE 0, 1, 2, ...



$\underbrace{0 \le \# < 1}_{0}$ $\underbrace{1 \le \# < 2}_{1}$ $\underbrace{2 \le \# < 3}_{2}$

§2.2 MEASURES OF CENTER

GIVEN DATA SET (LIST OF NUMBERS)

SUMMARIZE IT WITH A SINGLE #.

3 WAYS:

1. ARITHMETIC MEAN (AVERAGE)

TWO SETTINGS FOR STATISTICS:

POPULATION: DATA IS COLLECTED FROM ENTIRE SET OF INDIVIDUALS BEING STUDIED HUMAN

SAMPLE: DATA IS COLLECTED FROM A SUBSET OF THE SET OF INDIVIDUALS BEING STUDIED RANDOM SAMPLE OF PEOPLE.

THE AVERAGE OF A POPULATION IS DENOTED μ "MU"

THE AVERAGE OF A SAMPLE IS DENOTED $\overline{X}$. "EX BAR"

$$(\mu \text{ or}) \ \overline{x} = \frac{x_1 + x_2 + \dots + x_n}{n},$$ n IS # OF NUMBERS.

$$= \frac{\sum x_i}{n}$$ $\sum$ "SIGMA" S = SUM

SIGMA NOTATION:

$$\sum_{i=\boxed{2}}^{7} i = 2 + 3 + 4 + 5 + 6 + 7 = 27$$

i = INDEX

INTEGERS

STARTING AT $\boxed{\text{BOTTOM NUMBER}}$

ENDING AT TOP NUMBER

$$2i \quad 2i \quad 2i \quad 2i$$

·) $$\sum_{i=1}^{4} 2i = \underset{i=1}{2} + \underset{i=2}{4} + \underset{i=3}{6} + \underset{i=4}{8} = 20$$

·) $$1 + 2 + 3 + 4 + \dots + 5{,}000 = \sum_{i=1}^{5{,}000} i$$

$$2^i \quad 2^i \quad 2^i \quad 2^i$$

·) $$\sum_{i=0}^{3} 2^i = \underset{i=0}{2^0} + \underset{i=1}{2^1} + \underset{i=2}{2^2} + \underset{i=3}{2^3} = 15$$

ex. DATA: 20 #'S : $x_1, x_2, x_3, x_4, \dots, x_{20}$

1ST PIECE OF DATA 2nd 20TH PIECE OF DATA

AVERAGE $$\bar{x} = \boxed{\frac{\sum_{i=1}^{20} x_i}{20}} = \frac{x_1 + x_2 + x_3 + \dots + x_{20}}{20}$$

= ABBREVIATED $$\frac{\sum x_i}{20}$$

ex. SAMPLE CONTAINS FOLLOWING DATA:

2, 23, 48, 17, 20

MEAN $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{2+23+48+17+20}{5} = \boxed{22}$

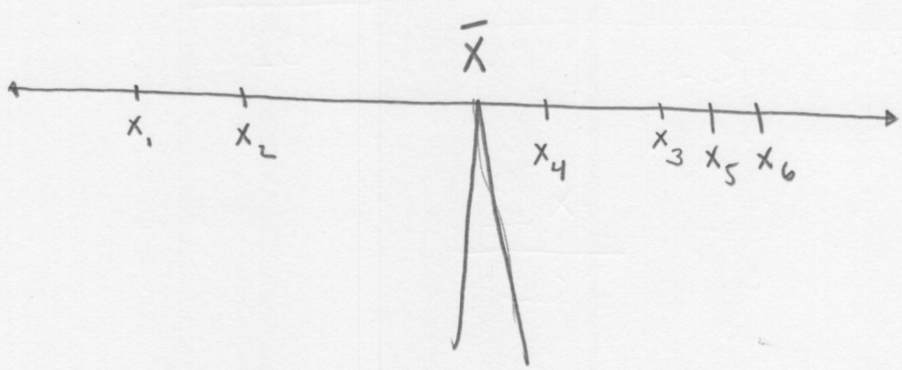
n IS NUMBER OF OBSERVATIONS IN DATASET

$(n = 5)$

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

~~35 + 50 + 65~~

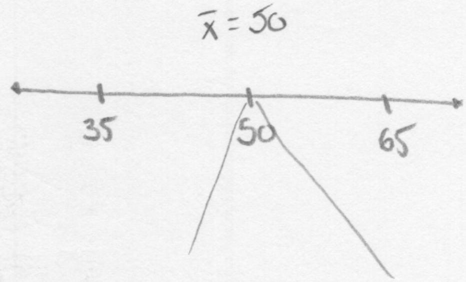
SUPPOSE YOU PLACE OBJECTS WITH EQUAL MASS ON A NUMBER LINE AT POSITIONS $x_1, x_2, \ldots, x_n$



THE MEAN $\bar{x} = \dfrac{\sum x_i}{n}$ IS THE POINT ON THE # LINE WHERE THE NUMBERS ARE BALANCED.

ex. $35 + 50 + 65$

$= 3(50) = 150$



$$\left(\sum x_i = n\bar{x} \right)$$

ex. POPULATION CONSISTS OF THE FOLLOWING DATA :

1.2 , 3.5 , 4.1 , 6.3 .

MEAN $\mu = \dfrac{\sum x_i}{n} = \dfrac{1.2 + 3.5 + \ldots + 6.3}{4} = \dfrac{15.1}{4}$

$= \boxed{3.775}$

2. MEDIAN : THE MEDIAN OF OF A SET OF n MEASUREMENTS IS THE VALUE X THAT FALLS IN THE MIDDLE POSITION WHEN THE MEASUREMENTS ARE ORDERED FROM LEAST TO GREATEST.

ex. 7, 5, -3, 12, 14

MEDIAN : $\underbrace{-3, 5}, 7, \underbrace{12, 14}$

MEDIAN = 7

2 MEAS. < 7

2 MEAS. > 7

NOTE: IF NUMBER OF MEASUREMENTS n IS EVEN

$$1,\ 3,\ 11,\ \underbrace{12,\ 13.5,\ 15}_{}\ ,\ 15,\ 16,\ 16$$

"MIDDLE" IS STRADDLED BY 2 NUMBERS

IN THIS CASE, MEDIAN IS MEAN OF THE 2 NUMBERS SHARING THE MIDDLE.

$$\text{MEDIAN} = \frac{12+15}{2} = \boxed{13.5}$$

SAME # OF MEASUREMENTS ABOVE/BELOW MEDIAN.

REMARK:

DATA SET #1 $\quad 1,\ 2,\ 3,\ 4,\ \boxed{5}$

MEAN $\bar{x} = \frac{15}{5} = 3$

MEDIAN $= 3$

CHANGE 1 MEASUREMENT.

DATA SET #2 $\quad 1,\ 2,\ 3,\ 4,\ \boxed{5,000,000}$

MEAN $\bar{x} = 1,000,002$

MEDIAN $= 3$

AVERAGE (MEAN) IS HIGHLY INFLUENCED BY EXTREME VALUES.

ex. HOUSEHOLD INCOMES ~~65,000~~ ~~48,000~~ ~~85,000~~ ~~36,000~~ 52,00

28,000 38,000 42,000 53,000 68,000,000

MEDIAN 42,000 ⟶ $42,000

MEAN 45,800 ⟶ $13,632,200

THE TYPICAL HOUSEHOLD INCOME IS $42,000 / $45,800

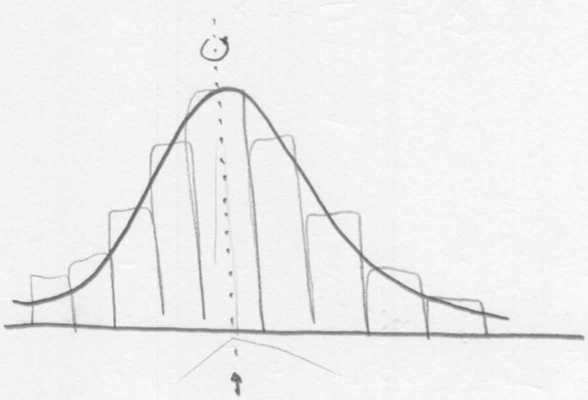
(EITHER ONE)

MEDIAN IS NOT INFLUENCED BY EXTREME VALUES.

ex. MEDIAN INCOME IN US IS $49,000

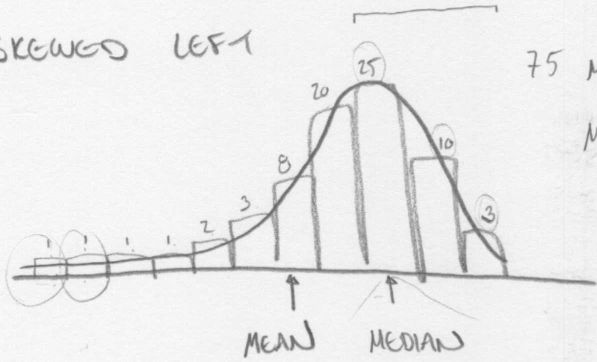
SHAPES OF DISTRIBUTIONS:

SYMMETRIC



MEAN = MEDIAN

SKEWED LEFT



75 MEASUREMENTS

MEDIAN = 38^{TH} MEASUREMENT

$$\underbrace{x_1, \dots, x_{37}}_{37\#'s}, x_{38}, \underbrace{x_{39}, \dots, x_{75}}_{37\#'s}$$

MEAN PULLED TOWARD TAIL

Def: SKEWED LEFT CAN BE DEFINED AS

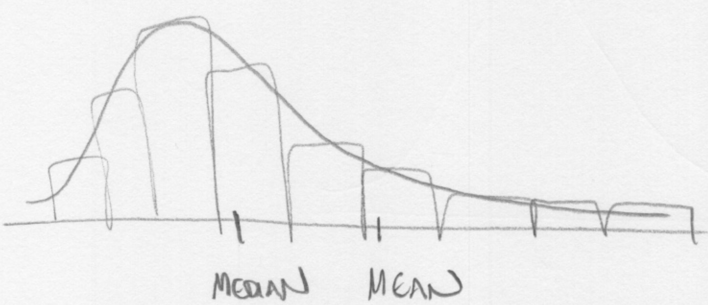
MEAN < MEDIAN

SYMMETRIC CAN BE DEFINED AS

MEAN = MEDIAN

SKEWED RIGHT CAN BE DEFINED AS

MEAN > MEDIAN



MEAN PULLED TOWARD TAIL.

3. MODE

THE MODE IS THE MOST FREQUANTLY OCCURRING MEASUREMENT.

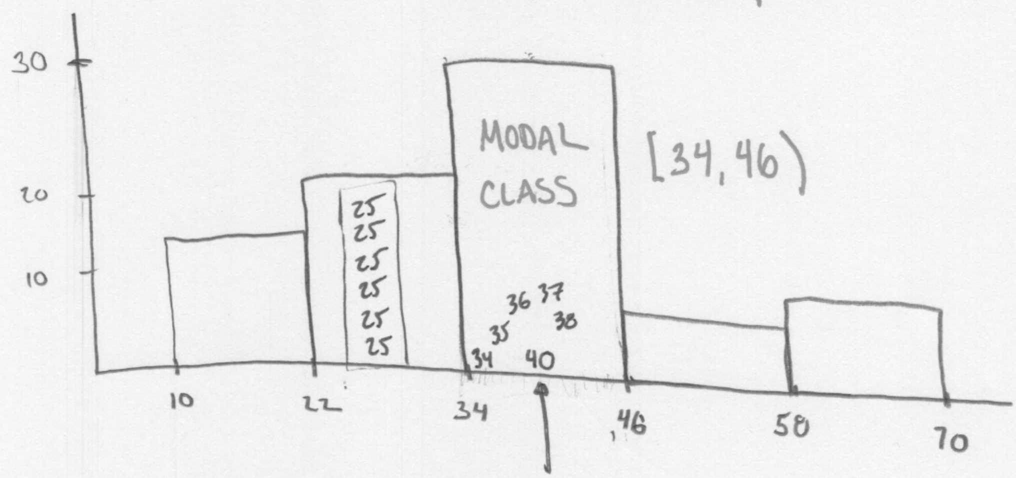
WHEN MEASUREMENTS ARE GROUPED INTO CLASSES, THEN THE CLASS WITH THE GREATEST FREQUENCY IS CALLED THE MODAL CLASS.

AND THE MIDPOINT OF THE MODAL CLASS IS THE MODE.

ex. DATA SET: 3.1, 4.2, 5.3, 4.2, 7.1, 7.1

MODE = 4.2, 7.1

ex.

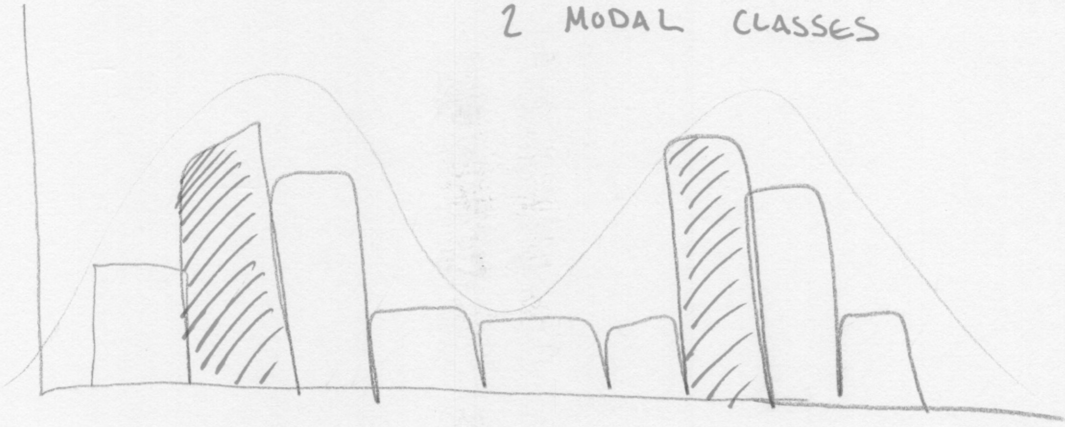


$$\text{MODE} = \frac{34 + 46}{2} = 40$$

(MIDPOINT)

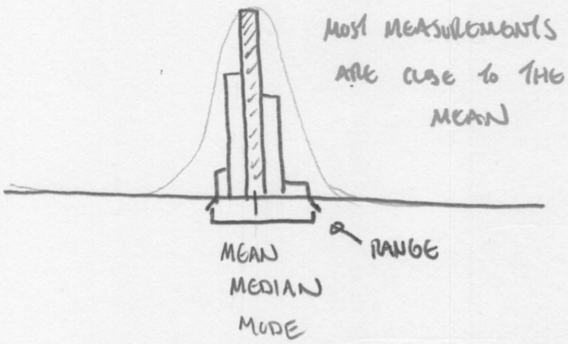
BIMODAL

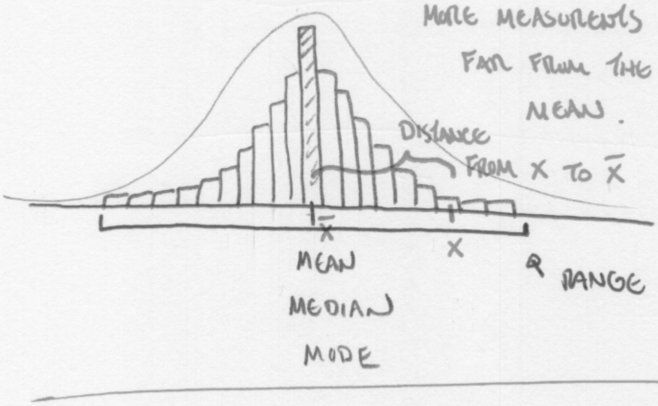
2 MODAL CLASSES



§2.3 MEASURES OF VARIABILITY (VARIATION)

CONSIDER TWO DISTRIBUTIONS OF MEASUREMENTS





SAME "CENTER"

1. RANGE = MAX - MIN
(LENGTH OF SMALLEST INTERVAL CONTAINING ALL MEASUREMENTS.)

2. VARIANCE MEASURES THE "AVERAGE" DISTANCE FROM A MEASUREMENT TO THE MEAN.

HOW DO YOU MEASURE THIS? (NEXT CLASS)