STEM & LEAF PLOTS

1.28 Preschool The ages (in months) at which 50 children were first enrolled in a preschool are listed below.

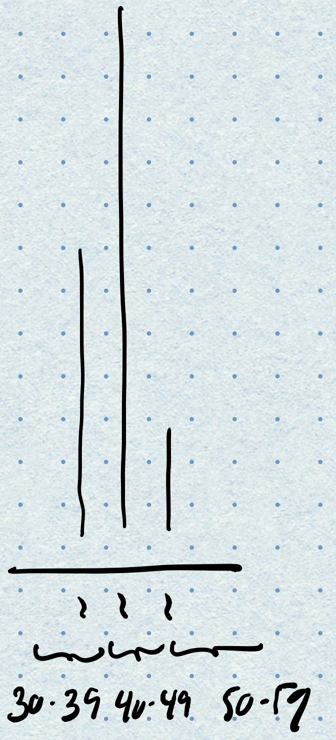
Data set EX0128

38	40	30	35	39	40	48	36	31	36
47	35	34	43	41	36	41	43	48	40
32	34	41	30	46	35	40	30	46	37
55	39	33	32	32	45	42	41	36	50
42	50	37	39	33	45	38	46	36	31

EACH # IS A LEAF ON A TREE...
THE FIRST DIGIT TELLS WHAT STEM ITS ON.

	STEMS	LEAVES
30'S	3	8 0 5 9 6 1 6 5 4 6
40'S	4	0 0 8 7 3 1 1 3 8 0
50'S	5	

ETC.

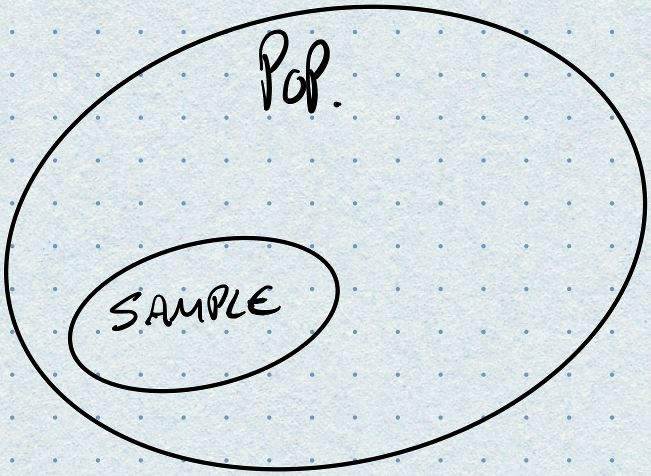


NOTE: THE SAME SET OF DATA CAN BE DISPLAYED WITH MANY DIFFERENT (REL. FREQ.) HISTOGRAMS DEPENDING ON CHOSEN CLASS WIDTH & # OF CLASSES.

CH 2. DESCRIBING DATA WITH NUMERICAL MEASURES.

DEF: PARAMETERS ARE MEASURES ASSOCIATED WITH A POPULATION.

STATISTICS ARE MEASURES ASSOCIATED WITH A SAMPLE.



§2.2 MEASURES OF CENTER

DATA: $\underbrace{1, 2, 2, 4\ 8}_{\text{SUMMARIZE WITH 1 \#}}$

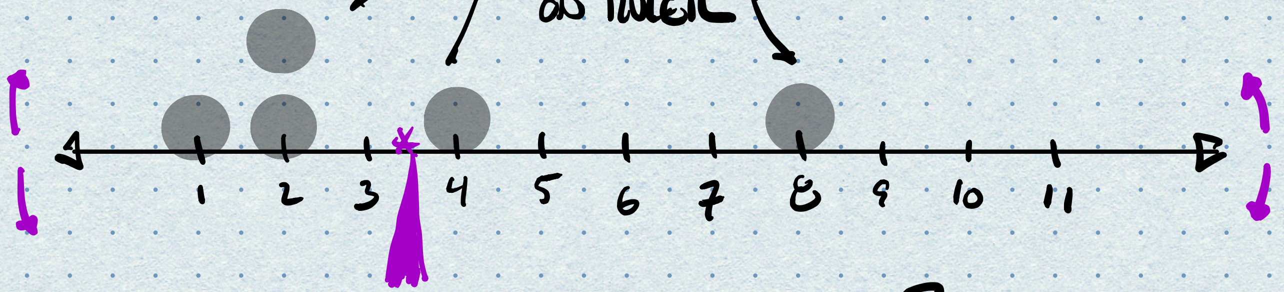
3 WAYS TO COMPUTE THE CENTER.

1. ARITHMETIC MEAN (MEAN) "AVERAGE"

CAN MEAN OTHER THINGS, DEPENDING ON SETTING.

PLACE DATA ON NUMBER LINE

1g WEIGHTS DISTRIBUTED ON RULER



BALANCED

WHERE WOULD THIS "RULER" BALANCE?

↳ ASSUME RULER IS MASSLESS (ZERO MASS)

* THE POINT ON # LINE WHERE THE DATA BALANCES IS THE MEAN (CENTER)

COMPUTE THE MEAN

$\bar{x}$ DENOTES THE MEAN OF A SAMPLE

μ DENOTES THE MEAN OF A POPULATION

"MU" GREEK LETTER FOR "m"

$q \quad \mu \quad p$

LET N = SIZE OF POPULATION

n = SIZE OF SAMPLE

SAMPLE DATA: $x_1, x_2, x_3, \ldots, x_n$

POPULATION DATA: $x_1, x_2, x_3, \ldots, x_N$

MEAN OF SAMPLE

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

SUM UP DATA FROM x_1 TO x_n

MEAN OF POPULATION

$$\mu = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N}$$

SUM FROM x_1 TO x_N

<u>ex.</u>

DATA: 1, 2, 2, 4, 8 (SAMPLE, $n = 5$)

$$\bar{x} = \frac{1 + 2 + 2 + 4 + 8}{5} = \frac{17}{5}$$

$$\bar{x} = 3.4$$

SIGMA (Σ) NOTATION

↑ GREEK CAPITAL "S"
STANDS FOR "SUM"
"SUMMATION"

END INDEX

INDEX

STARTING

n

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

SUM x_i's FROM $i=1$ TO $i=n$

$$= \frac{x_1 + x_2 + \ldots + x_n}{n}$$

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

SAME AS FORMULAS GIVEN ABOVE.

2. MEDIAN

<ins>Def</ins>: GIVEN A SET OF DATA, THE MEDIAN IS THE DATA VALUE IN THE MIDDLE POSITION WHEN THE DATA IS ARRANGED FROM LEAST TO GREATEST.

<ins>ex.</ins> DATA: 13, 5, 6, 5, 7, 11, 15

MEDIAN: $\underbrace{5, 5, 6}, 7, \underbrace{11, 13, 15}$

* EQUAL # OF DATA VALUES ABOVE/BELOW THE MEDIAN.
IMPORTANT CHARACTERISTIC.

IF # OF DATA VALUES IS EVEN, THEN THERE IS NO DATA VALUE IN THE MIDDLE POSITION.

INSTEAD, TWO #'S "STRADDLE" THE MIDDLE & THE MEDIAN IS THE MEAN OF THESE 2 #'S.

ex. 3, 4, 4, | 6, 9, 12 (n=6 EVEN)

$$\frac{4+6}{2} = \boxed{5 \text{ MEDIAN}}$$

NOTE: EQUAL # OF DATA VALUES (3) ABOVE / BELOW MEDIAN (5).

ex. SALARIES OF 7 EMPLOYEES AT TECH START UP.

24,000
60,000
60,000
60,000
60,000
120,000
350,000

SUMMARIZE WITH SINGLE # TO DESCRIBE THE CENTER

$$\mu = \frac{\sum \text{SALARIES}}{N}$$

$$= \frac{24{,}000 + \ldots + 350{,}000}{7}$$

$$= \frac{734{,}000}{7} = \$104{,}857$$

$\mu = \$104{,}857$ MEDIAN $= \$60{,}000$

ex. SAME COMP. NEXT YEAR

24,000
60,000
60,000
60,000
60,000
120,000
3,500,000

$$\mu = \frac{\sum_i x_i}{N} = \frac{3884000}{7}$$

$\mu \approx \$554,857$

MEDIAN = \$60,000

NOTE: MEAN $\mu, \bar{x}$ IS INFLUENCED BY EXTREME VALUES.

MEDIAN IS NOT.

NEWS: "MEDIAN HOUSEHOLD INCOME"

3. MODE

PIE A LÁ MODE — FASHION (FRENCH)

POPULAR

Def: MODE FOR A DATA SET IS THE VALUE(S) THAT APPEAR(S) MOST OFTEN.

$$\left.\begin{array}{c} 24{,}000 \\ \boxed{\begin{array}{c} 60{,}000 \\ 60{,}000 \\ 60{,}000 \\ 60{,}000 \end{array}} \\ 120{,}000 \\ 3{,}500{,}000 \end{array}\right\}$$

MODE = 60,000

ex. $\underbrace{1, 1, 1, 2}_{4 \text{ BELOW}}, 3, \underbrace{4, 9, 9, 9}_{4 \text{ ABOVE}}$

MODES: 1 & 9 (APPEAR 3 TIMES, MORE THAN ANY OTHER VALUE)

MEDIAN: 3

MEAN: $\frac{\sum x_i}{n} = \frac{1+1+\ldots+9}{9} = \frac{39}{9} \approx$

≈ 4.3333

TIP WHEN CALCULATING MEDIAN:

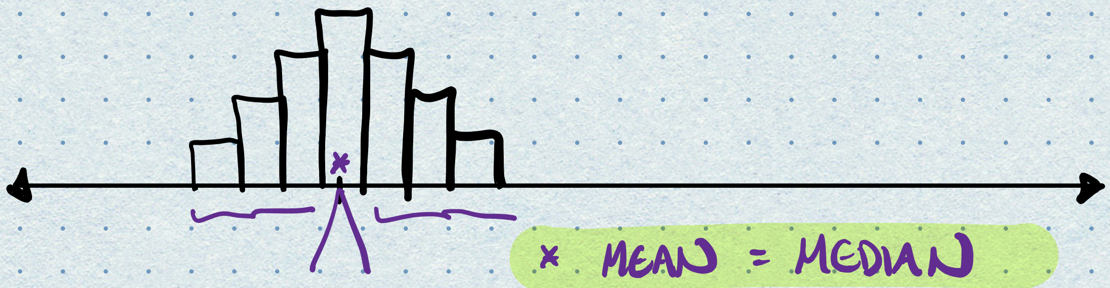
·) IF n IS ODD THEN THE MEDIAN IS IN THE

$\left(\frac{n+1}{2}\right)^{TH}$ POSITION

·) IF n IS EVEN THEN THE MEDIAN IS IN THE MEAN OF THE #'S IN POSITIONS

$$\left(\frac{n}{2}\right)^{TH} \;\&\; \left(\frac{n}{2}+1\right)^{TH} \text{ POSITION.}$$

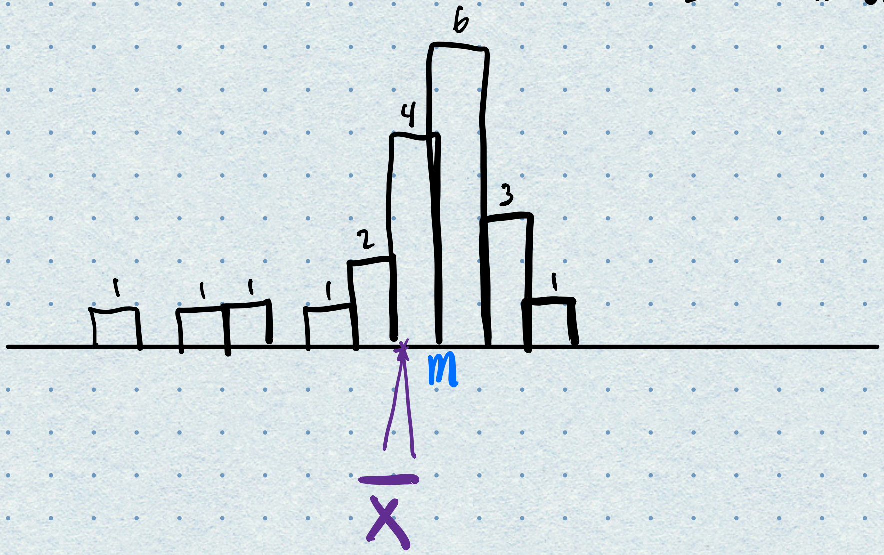
SYMMETRIC DISTRIBUTION



× MEAN = MEDIAN

SKEWED - LEFT

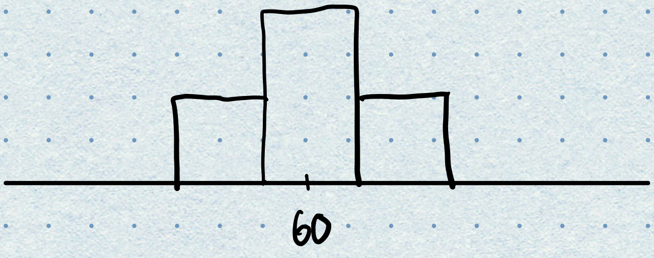
20 DATA VALUES



SKEWED LEFT : $\bar{x} < m$ (DEF)

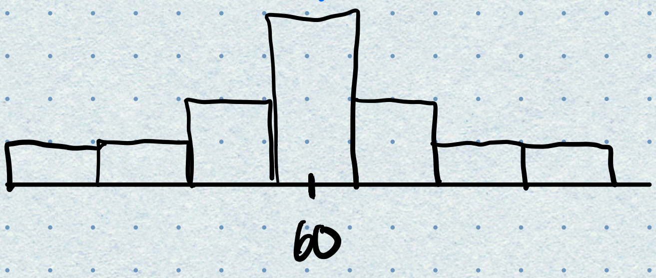
SIMILARLY, SKEWED RIGHT : $\bar{x} > m$ (DEF)

§2.3 MEASURES OF VARIABILITY (VARIATION)



SAME CENTER C
(MEAN, MEDIAN, MODE ALL = C)

GREATER VARIATION



← GREATER RANGE

IF THESE HISTOGRAMS REPRESENT GRADES ON A QUIZ FOR 2 CLASSES, WHAT'S THE DIFFERENCE

WIDER DISTRIBUTION

<u>Def</u>: RANGE OF A DATA SET IS

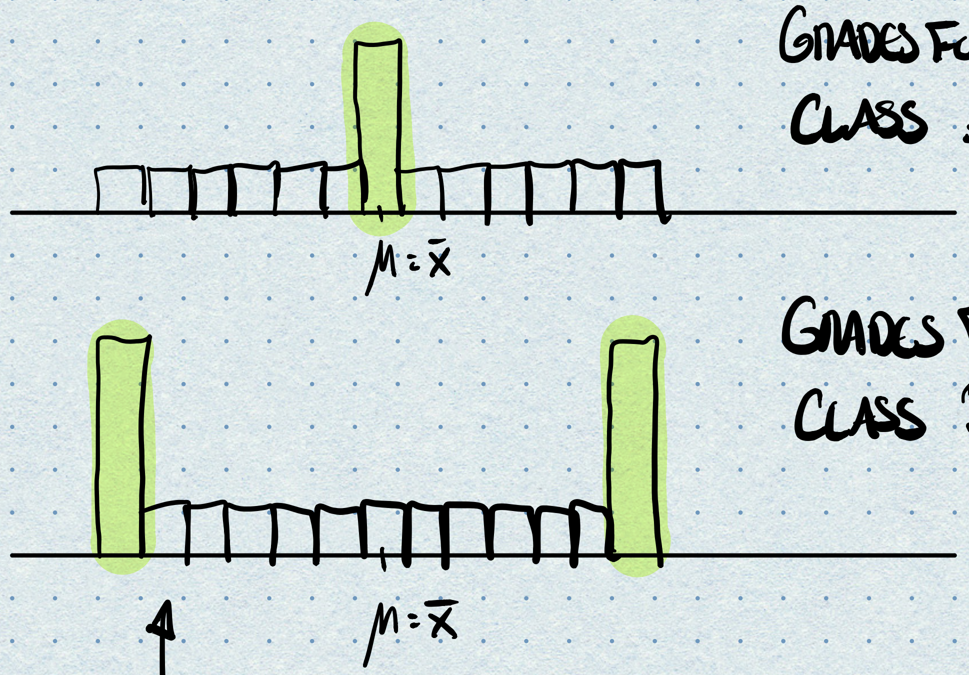
MAX VALUE − MIN VALUE

WE'RE MEASURING THE VARIATION (VARIABILITY) OF DATASETS ex

SAME RANGE

GRADES FOR CLASS A

GRADES FOR CLASS B



GREATER VARIATION

TO CAPTURE THE DIFFERENCE IN VARIATION THAT WE SEE HERE WE WILL NEED NEW TOOLS:

- VARIANCE
- STANDARD DEVIATION

(NEXT TIME)