DATA: LIST OF n NUMBERS/MEASUREMENTS

$$x_1, x_2, x_3, \ldots, x_n.$$

EXAMPLE:

Table 3 Entrance Examination Scores of 100 Entering Freshmen

762	451	602	440	570	553	367	520	454	653
433	508	520	603	532	673	480	592	565	662
712	415	595	580	643	542	470	743	608	503
566	493	635	780	537	622	463	613	502	577
618	581	644	605	588	695	517	537	552	682
340	537	370	745	605	673	487	412	613	470
548	627	576	637	787	507	566	628	676	750
442	591	735	523	518	612	589	648	662	512
663	588	627	584	672	533	738	455	512	622
544	462	730	576	588	705	695	541	537	563

OUR GOAL IS TO EXTRACT USEFUL INFORMATION FROM THE DATA.

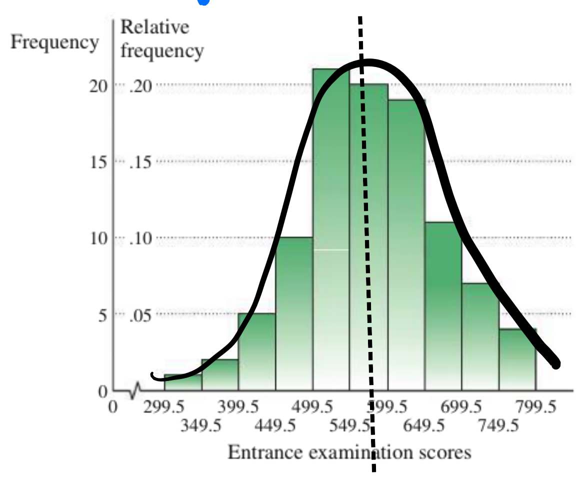
(DESCRIBE / SUMMARIZE)

TWO METHODS: I VISUALLY

II NUMERICALLY

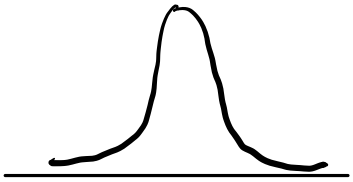
I. VISUALLY: FREQUENCY / RELATIVE FREQUENCY HISTOGRAM

DIVIDE FREQ. BY n



NOW WE CAN SEE THE DISTRIBUTION OF MEASUREMENTS.

WORDS TO DESCRIBE DISTRIBUTIONS:

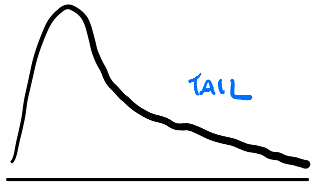


SYMMETRIC

e.g. SAT SCORES

HEIGHTS OF ADULT MALES

WEIGHTS OF APPLES PRODUCED BY 1 TREE

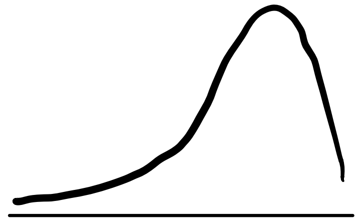


RIGHT-SKEWED

e.g. HOUSEHOLD INCOME

NUMBER OF SIBLINGS

SIZE OF NYC APTS



LEFT-SKEWED

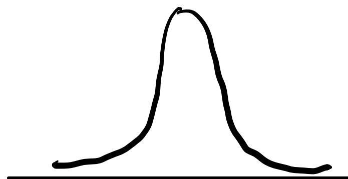
e.g. LIFETIME OF HUMANS

UNIFORM

e.g. # TIMES EACH FACE APPEARS ON 1000 ROLLS OF A DIE.

? # TIMES EACH DIGIT APPEARS IN FIRST 10 MILLION DIGITS OF π

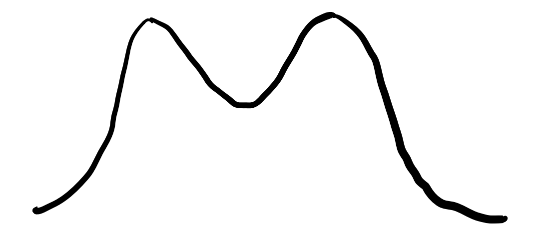
e



UNIMODAL

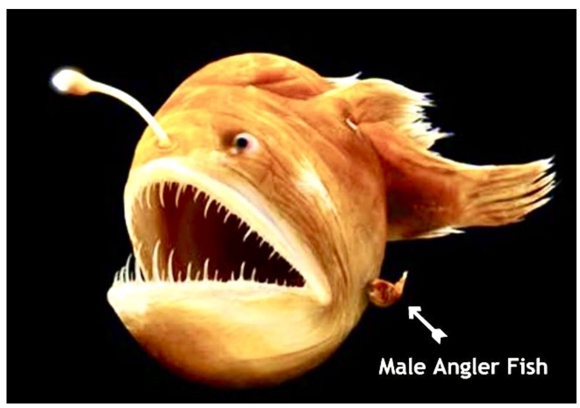
e.g. RESTING HEART RATE FOR HUMANS

PSI AT WHICH A PARTICULAR TYPE OF BICYCLE INNER TUBE POPS



BIMODAL

e.g. SIZE OF ANGLER FISH

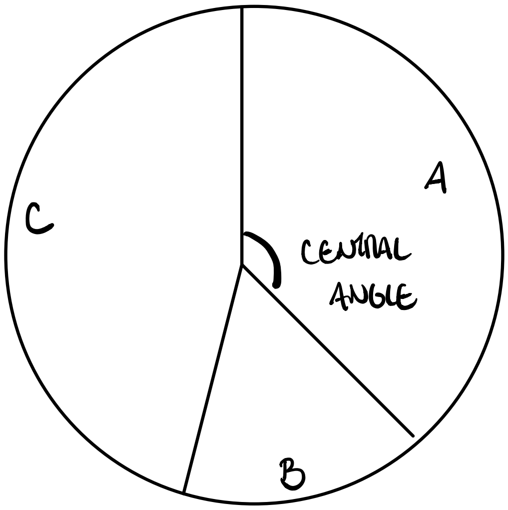


Pie Charts

$\frac{35}{100}$ $\rightarrow$ $\frac{\text{ANGLE}}{360}$ $\left(\frac{\text{ANGLE}}{2\pi}\right)$

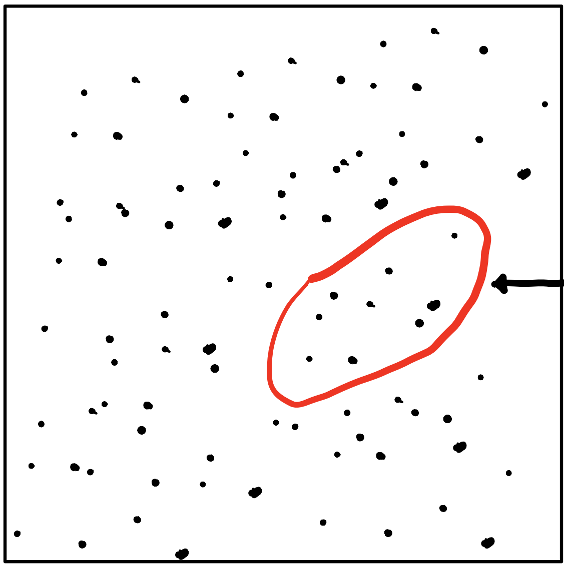
Proportion of 360°

	Percents		Central Angle
Type A	35%	$.35 \times 360 =$	126°
Type B	18%	$.18 \times 360 =$	64.8°
Type C	47%	$.47 \times 360 =$	169.2°



II. NUMERICALLY : MEASURES OF CENTER & VARIATION "SPREAD"

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



POPULATION DATA
DESIRABLE
HARD TO ACQUIRE

SAMPLE DATA
NOT IDEAL
EASIER TO ACQUIRE

WE CAN USE SAMPLE DATA TO MAKE ESTIMATES FOR THE POPULATION (INFERENTIAL STATISTICS)

3 MEASURES OF CENTER

1. SAMPLE MEAN $\bar{x}$

POPULATION MEAN μ

$$\left.\begin{matrix}\bar{x} \\ \mu\end{matrix}\right\} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

SIGMA NOTATION

SUM OF SIMILAR OBJECTS

SUM

END INDEX

e.g. $\sum_{i=1}^{5} \frac{1}{2^i} = \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \frac{1}{2^5} = \frac{63}{64}$

INDEX

START INDEX

"TEMPLATE"

INDEX TAKES SUCCESSIVE INTEGER VALUES FROM START # TO END #

The mean of 4 numbers is 90. If the mean of the first three numbers is 88, find the fourth number.

$$\frac{x_1 + x_2 + x_3 + x_4}{4} = 90 \qquad \frac{x_1 + x_2 + x_3}{3} = 88$$

$$x_1 + x_2 + x_3 + x_4 = 360 \qquad x_1 + x_2 + x_3 = 264$$

$$\underbrace{x_1 + x_2 + x_3}_{264} + x_4 = 360$$

$$264 + x_4 = 360$$

$$x_4 = 96$$

Example. Suppose I buy 20 gallons of gas at an average price of \$2.40/gallon, and you buy 10 gallons of gas at an average price of \$2.10/gallon. Together, what is the average price per gallon that we've paid for gas?

WRONG: $\frac{2.40 + 2.10}{2} = \$2.25$

CORRECT. 30 GALLONS: TOTAL COST: $20(2.40) + 10(2.10) = \$69$

TOTAL GAS: $20 + 10 = 30$

AVERAGE COST $= \dfrac{20(2.40) + 10(2.10)}{20 + 10} = \$2.30$

WEIGHTED AVERAGES:

GIVEN n #'S THE AVERAGE (MEAN) IS

$$\frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{x_1}{n} + \frac{x_2}{n} + \frac{x_3}{n} + \dots + \frac{x_n}{n}$$

$$= \frac{1}{n}x_1 + \frac{1}{n}x_2 + \frac{1}{n}x_3 + \dots + \frac{1}{n}x_n$$

ADD UP TO 1. ALL THE SAME.

MORE GENERALLY, A WEIGHTED AVERAGE OF n #'S IS

$$\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \dots + \alpha_n x_n$$

ADD UP TO 1. ALL POSITIVE.

Example. Suppose you have a homework grade of 90, and quiz grade of 85, and an exam grade of 80. If homework counts for 20% of your grade, the quiz counts for 35% of your grade, and the exam counts for 45% of your grade, calculate your average for the class.

Example. It costs a shipping company \$8.75 to ship a small package overnight. Suppose the shipping company charges a flat rate of \$19 to ship a small package overnight, and if the package is late they refund the full amount. If 93% of all packages are delivered on time, what is the average profit per package that the shipping company earns?

2. MEDIAN

Arrange the data in order from least to greatest.

If the number of measurements is odd, then the median is the number in the middle position.

e.g. 1 1 3 3 4

$$1 \quad 1 \quad \underset{\uparrow \text{ MEDIAN}}{3} \quad 3 \quad 4$$

If the number of measurements is even, then the median is the mean of the two numbers that share/straddle the middle position.

e.g. 1 1 3 5 5 6

$$1 \quad 1 \quad \underbrace{3 \quad 5} \quad 5 \quad 6$$

$$\frac{3+5}{2} = 4$$

MEDIAN

NOTE: AN EQUAL NUMBER OF MEASUREMENTS LIE TO THE RIGHT/LEFT OF THE MEDIAN WHEN ARRANGED IN ORDER.

"HALF THE DATA IS LESS THAN THE MEDIAN, & HALF THE DATA IS GREATER THAN THE MEDIAN."

GOOD WHEN YOU DON'T WANT THE "CENTER" TO BE INFLUENCED BY EXTREME VALUES

35,000 42,000 46,000 48,000 59,000 63,000

MEAN 48,833

MEDIAN 47,000 (HALF ABOVE, HALF BELOW)

35,000 42,000 46,000 48,000 59,000 63,000,000

MEAN 10,538,333

MEDIAN 47,000 (HALF ABOVE, HALF BELOW)

Blood cholesterol levels. Find the mean and median for the data in the following table.

Blood Cholesterol Levels (milligrams per deciliter)

MIDPOINT	Interval	Frequency	
159.5	149.5–169.5	4	1 - 4
179.5	169.5–189.5	11	5 - 15
199.5	189.5–209.5	15	16 - 30
219.5	209.5–229.5	25	31 - 55
239.5	229.5–249.5	13	
259.5	249.5–269.5	7	
279.5	269.5–289.5	3	
299.5	289.5–309.5	2	

MODE

(TOTAL 80)

MODAL CLASS

WE ASSUME ALL MEASUREMENTS IN EACH CLASS/BIN ARE EQUAL TO THE MIDPOINT OF THAT CLASS/BIN'S INTERVAL

$x_{40} = x_{41} = 219.5$

MIDPOINT OF THE INTERVAL

MEAN $\bar{x} = \dfrac{4(159.5) + 11(179.5) + 15(199.5) + 25(219.5) + 13(239.5) + 7(259.5) + 3(279.5) + 2(299.5)}{80}$

MEDIAN: $x_1, x_2, x_3, \ldots, x_{40}, x_{41}, \ldots, x_{79}, x_{80}$ IN ORDER

MEDIAN = MEAN OF 40^{th} & 41^{st} #

219.5

3. MODE: MOST FREQUENTLY OCCURRING MEASUREMENT(S).

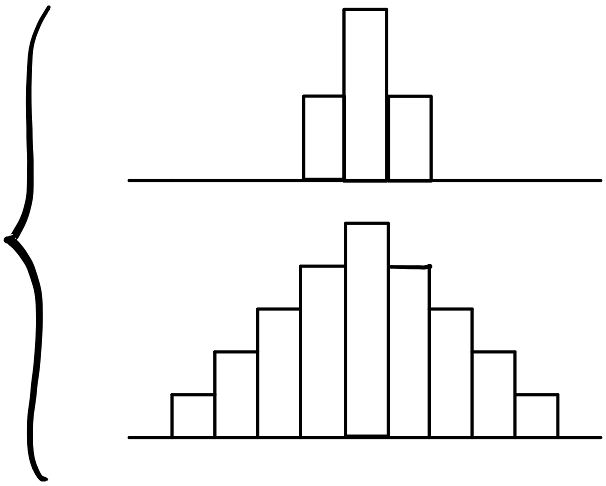
YES, THERE MAY BE MULTIPLE MODES (TIES ALLOWED).

ex. PREVIOUS EXAMPLE

3 MEASURES OF VARIATION

1. RANGE = MAX - MIN

SAME MEAN, MEDIAN, MODE



← LESS VARIATION

SMALLER RANGE

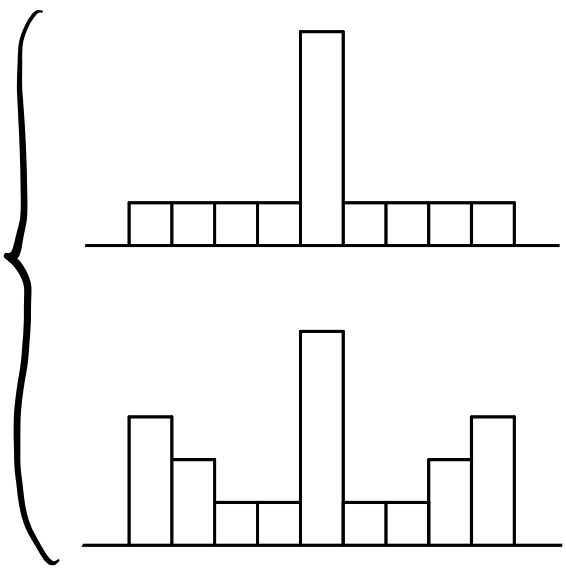
← GREATER VARIATION

LARGER RANGE

WHAT ABOUT THE FOLLOWING DISTRIBUTIONS:

SAME MEAN, MEDIAN, MODE

SAME RANGE.



← LESS VARIATION

← MORE VARIATION

WE CAN MEASURE THE DIFFERENCE IN VARIATION HERE WITH VARIANCE & STANDARD DEVIATION.

Variance σ^2 & Standard Deviation σ for Population

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

VARIANCE $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$ DEVIATION FROM MEAN, SQUARED

STANDARD DEVIATION $\sigma = \sqrt{\sigma^2}$ (SQRT OF VARIANCE)

Variance s^2 & Standard Deviation s for Sample

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

VARIANCE $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ DIVIDE BY $n-1$

STANDARD DEVIATION $s = \sqrt{s^2}$ (SQRT OF VARIANCE)

s^2 & s ARE GOOD ESTIMATES FOR σ^2 & σ

e.g. DATA : 13 14 17 25 26 (MEAN = 19)

CALCULATE THE VARIANCE & STANDARD DEVIATION

ASSUMING THE DATA COMES FROM A (a) POPULATION

(b) SAMPLE

(c) HOW MANY MEASUREMENTS LIE WITHIN 1 STND. DEV. OF MEAN?

(d) HOW MANY MEASUREMENTS LIE WITHIN 2 STND. DEV. OF MEAN?

(a) POPULATION VARIANCE $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$

x_i	$x_i - \overset{19}{\mu}$	$(x_i - \mu)^2$
13	-6	36
14	-5	25
17	-2	4
25	6	36
26	7	49
+		+

ADD THESE UP $\sum_{i=1}^{5}(x_i - \mu)^2$

VAR. $\sigma^2 = \frac{1}{5}(150) = 30$

STAND DEV. $\sigma = \sqrt{\sigma^2}$

$\sigma = \sqrt{30} = 5.4772$

0 150

(b) SAMPLE VARIANCE $s^2 = \frac{1}{n-1}\sum_{i=1}^{5}(x_i - \bar{x})^2 = \frac{1}{5-1}(150)$

$$s^2 = \frac{150}{4} = 37.5$$

SAMPLE STANDARD DEV. $s = \sqrt{s^2} = \sqrt{37.5} = 6.1237$

In Problems 11 and 12, find the standard deviation for each set of grouped sample data using formula (5) on page 525.

11.

	Interval	Frequency
2	0.5–3.5	2
5	3.5–6.5	5
8	6.5–9.5	7
11	9.5–12.5	1

TOTAL 15

VARIANCE $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$$\bar{x} = \frac{2(2) + 5(5) + 7(8) + 1(11)}{15} = \frac{96}{15} = 6.4$$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	FREQUENCY f_i
2	$2 - 6.4 = -4.4$	19.36	2
5	$5 - 6.4 = -1.4$	1.96	5
8	$8 - 6.4 = 1.6$	2.56	7
11	$11 - 6.4 = 4.6$	21.16	1

SUM OF THESE PRODUCTS

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{2(19.36) + 5(1.96) + 7(2.56) + 1(21.16)}{15 - 1}$$

$$s^2 = \frac{88}{14} = 6.2857$$

STANDARD DEVIATION $s = \sqrt{s^2} = \sqrt{6.2857} = 2.5071$