Week 5 Handout 2: Data Analysis Power Hour

How to represent data

List of numbers

See a raw list of numbers stemming from a survey on p. 286.
Ordered list: This is the ***conceptual basis*** for the other representations of data.

0, 0, 0, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4, 4, 4, 5

Frequency distribution

Shows all the ***different*** values in the list, x.
Also shows the ***frequency***, the number of times each number is repeated, f(x).

x	*f(x)*
0	3
1	5
2	7
3	6
4	3
5	1

Note that the frequencies can also be written as FDP's.

x	*f(x)*	P(*x*)	=P(x)	=P(x)
0	3	3/25	0.12	12%
1	5	5/25	0.2	20%
2	7	7/25	0.28	28%
3	6	6/25	0.24	24%
4	3	3/25	0.12	12%
5	1	1/25	0.04	4%

All are equivalent! (How would you turn a column of FDP's into a column of whole numbers?)

x	*P(x)*	*f(x)*
5	15%	
10	35%	
15	20%	
20	25%	
25	5%	

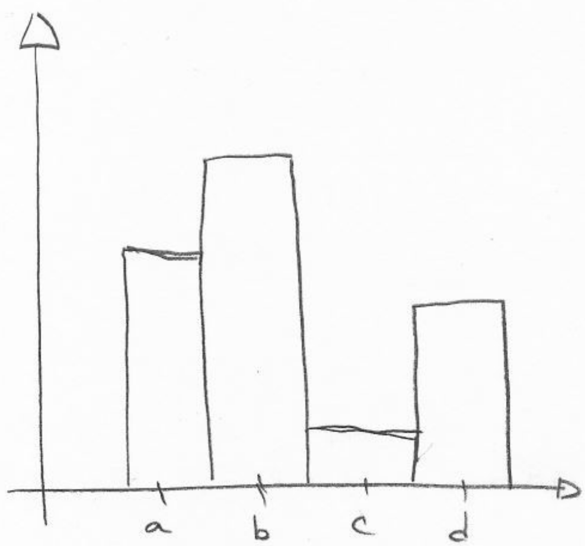
Bar Graph (p. 290)

2931 S. Sepulveda Blvd. Upstairs Ste. I (310) 499-3285
Los Angeles, CA, 90064 info@NthDegreeTutoring.com

To answer questions about a frequency distribution or graph, we almost always need to mentally ***convert it to an ordered list***. If this frequency distribution were converted into an ordered list of numbers:

x	$f(x)$
3	4
8	10
9	20
12	10
15	6

1) What's the largest number in the list?
2) What's the most common number in the list?
3) What's the 5th number in the list?
4) How many numbers are in the list?



5 Smallest

6 Least common

7 Largest

8 Most common

Summarizing Data

Measures of Center

"What is the mediumest number in the list?" Let's demonstrate with this short list:

3, 3, 6, 9, 30

Measure	aka	Calculation	Visual interpretation
Mean	Average (arithmetic mean)	Mean = Sum / Count.	Balancing point
Median	50^{th} percentile	Middle item in the list	Cuts area in half
Midrange		Mean of min & max	Midpoint of min & max
Mode		Most frequent	Highest point

Mean of 3, 3, 6, 9, 30 =

Median of 3, 3, 6, 9, 30 =

Midrange of 3, 3, 6, 9, 30 =

Mode of 3, 3, 6, 9, 30 =

Pointers

- Remember that the measures of center are always on the *x*-axis!
- For any two numbers, mean = median = midrange

  What's the average of 1,248,108 and 1,248,110?
  What point is exactly halfway between 3.6 and 9.1?
  What's the median of the list 4, 5 ?

Example: 6, 8, 8, 8, 10, 10, 12, 16, 30

Mean: Median: Midrange: Mode:

More about the Mean

1. Many problems *give* you the mean instead of asking for it. Interpretation: "Pretend they're all the same"

   Ex: 12 children have a mean weight of 80 pounds. Can they ride an elevator with a capacity of 1,000 pounds?

   Solution: <u>Pretend every child</u> weighs 80 pounds. Then total weight = ____________

2. The mean is a fraction, just like a rate. Use the units in the rate (given or implied) to decide what to add and divide.

   John worked for 4 hours at an average of $20 / hr and 8 hrs at an average of $30 / hr. What is his average overall wage? (It's NOT $25 / hr!)

   Solution: The units in the rate are $ / hr. This tells us to tally all the dollars, and all the hours, and divide. The denominator is usually easier.

   Count (hours): John worked for _________ hours altogether

   Sum (dollars):
   When John worked for 4 hours at an average of $20 / hr, he earned $ _______

   When he worked 8 hours at an average of $30 / hr, he earned $ _________

   Total = $ ____________

   So what was John's average wage? ___________________

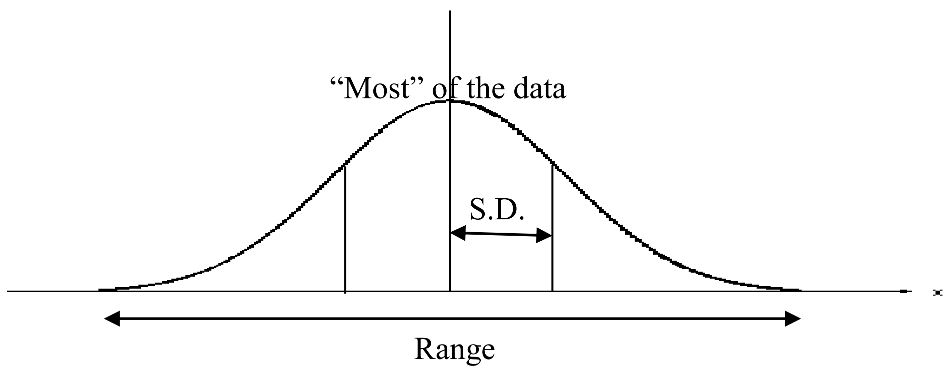
3. The mean is a fraction, just like the slope of a line.

On the graph on p. 322, what was the average annual increase in total expenditures?

Measures of Spread

"How spread out are the numbers in the list?"

Measure	Meaning	Calculation	Visual interpretation
Range	Difference between largest and smallest values	Max – Min	Width of entire graph
Standard deviation	The average "plus-or-minus" from the center	Not on the GRE	Half-width of the "heart" of the graph



If most GRE students are 25 ± 4 years old, then 25 is the approximate __________ and 4 is the approximate ___________________________ .

If all students in a class are between 19 and 32, then the range is _______________

Which one probably has a larger standard deviation?

A

Ages of people at Disneyland

B

Ages of people at the trendiest nightclub

Distribution

"How are the numbers repeated and spaced apart?" or
"What is the shape of the graph?"

Distribution affects almost all the measures above.

<u>The unknown distribution problem</u>

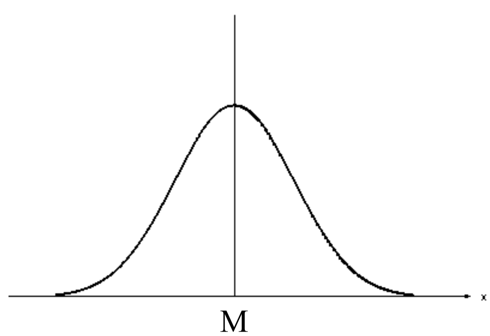
If we do not know how a set of numbers is distributed, then we usually can't draw conclusions about its mean, median, or percentiles.

The boys' mean test score was 65, and the girls' mean was 75

<u>A</u>	<u>B</u>
The boys' median score	The girls' median score

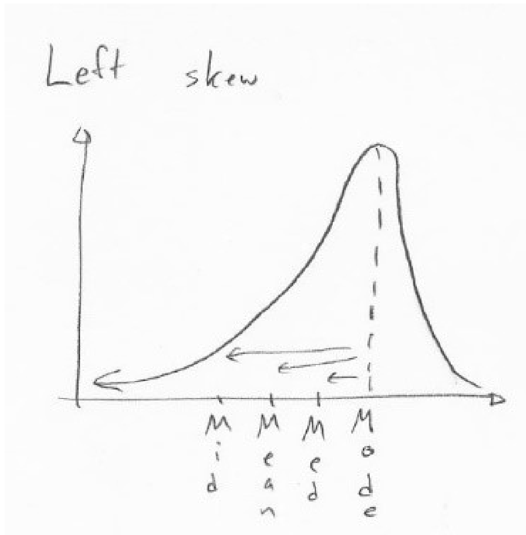
<u>Symmetry and skew</u>

If a graph is symmetrically bell-shaped, then mean = median = midrange = mode



If a graph is asymmetric (skewed):

1. The mean, median, and midrange are pulled away from the mode toward the tail.
2. The median stays closest to the mode.
3. The midrange gets pulled the most.



(Not drawn to scale)