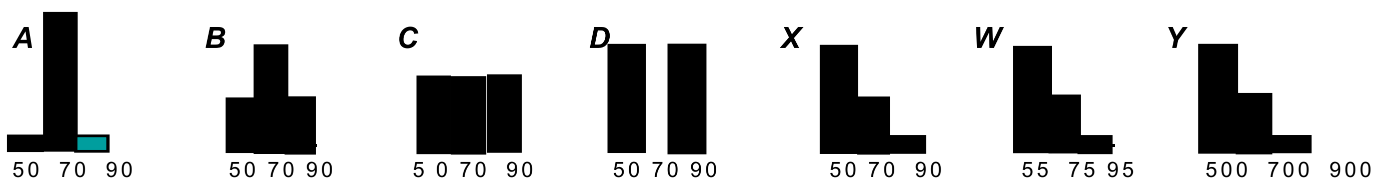
Suppose two students in your class earn 100 on a test, five earn 85, and thirteen earn 70.
Would the mean score for those students on that test be (100+ 85+ 70)/3 = 255/3 = 85?

12. For the following simplistic sets of grouped data—find the median & the mean:

a	f	b	f	c	f	d	f	x	f	w	f	y	f
50	1	50	3	50	4	50	6		50 6		55 6		500 6
70	8	70	6	70	4	70	0		70 3		75 3		700 3
90	1	90	3	90	4	90	6		90 1		95 1		900 1



mean:

median:

13. The first four distributions are symmetric.
When the distribution is symmetric, the mean & median are ...

The fifth (w) distribution is asymmetric, "**skewed to the right**".
Where is the mean, relative to the median?

14. The values in the ***W*** distribution are ***each 5 more*** *than* those in the ***X*** distribution ($w_1 = 5 + x_1$, etc).
How do the means for ***W*** & ***X*** compare? Does this make sense?
If every value in a set of data is increased by an amount "a" (i.e. amount *a* is ADDED), then the mean is

15. The values in the ***Y*** distribution are ***each 10 times*** those in the ***X*** distribution ($y_1 = 10x_1$, $y_2 = 10x_2$, etc).
How does the mean for *Y* compare with the mean for *X*? Does this make sense?
If every value in a sample or population is MULTIPLIED by a factor "r", then the mean is

16. Suppose the example ***X*** data above is doubled and then increased by 3. What is the new mean?

At right are the selling prices of the 30 single-family homes sold in Northridge in January, 1995.

$	f
125K	15
175K	11
250K	2
1000K	1
3000K	1

17. Find the median.

18. Is the mean higher or lower than the median?
Find it.

19. Find the mode. (OK, the "modal class")

20. If you are interested in the price of housing in a particular area, which of the above "average" statistics would you want to know, to estimate the price of houses in that area?

> Mode = most frequent value
> Median = the middle score [the *(n+ 1)/2th*]
> Mean = the evenly distributed total; also the balancing point of the distribution

Measures of Central Tendency: Among statistics commonly used to describe the "variation", or spread, in a set of data—are the **STANDARD DEVIATION**, the **RANGE** and the **INTERQUARTILE RANGE.**

The **RANGE** is the total span or spread of the data, ie the highest value – the lowest value.

Just as the median divides the ordered data into two equal groups, in order, the **QUARTILE MARKS** divide the data into four equal groups. These quartile marks are referred to as the **first** and **third quartile** marks, and the **second quartile mark**, which is also the median.

21. Compute the range and interquartile range of the scores in Mr. Jones' class.

range = maximum data value – minimum data value = 28

Interquartile range = $Q_3 - Q_1$ = 11

70 72 72 | 73 75 76 | 82 83 83 | 84 84 85 97 98

The standard deviation is a bit more complicated....

STANDARD DEVIATION = sq. root of (average square distance from the mean) = $\sqrt{\frac{\Sigma(x-\mu)^2}{n}}$ ← μ *or* $\bar{x}$, ← $n-1$ for a sample

22. Compute the s.d. (s or σ) for the data: 1 2 3 4 5

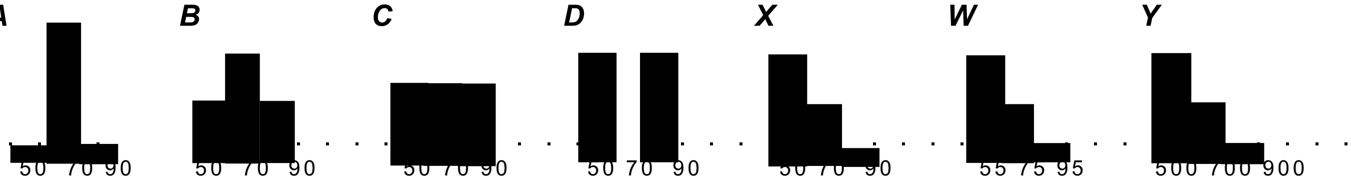
mean is 3.

x	x – mean distance	(distance)^2
1		
2		
3		
4		
5		

25. Consider our simple grouped data examples— compute standard deviation:

a	f	b	f	c	f	d	f	x	f	w	f	y	f
50	1	50	3	50	4	50	6	50	6	55	6	500	6
70	8	70	6	70	4	70	0	70	3	75	3	700	3
90	1	90	3	90	4	90	6	90	1	95	1	900	1

A *B* *C* *D* *X* *W* *Y*



mean:	$\bar{x}$ = 70	$\bar{x}$ = 70	$\bar{x}$ = 70	$\bar{x}$ = 70	$\bar{x}$ = 6	$\bar{w}$ = 16	$\bar{y}$ = 60
	sd ≈ 9.4	sd ≈ 14.8	sd ≈ 17.1	sd ≈ 20	sd ≈ 1.41	sd ≈ 1.41	sd ≈ 14.1

26. Would you say the first four distributions are "clustered about the mean" to the same degree?
How is this reflected by the standard deviation?

27. What kind of data would have a standard deviation of 0? ...a negative standard deviation?

28. The values in the *W* distribution are each 5 more than those in the *X* distribution, as noted earlier.
How does the standard deviation for W compare with the std. deviation for X? Does this make sense?

29. The values in the Y distribution are 10 times those in the *X* distribution, as noted earlier.
How does the standard deviation for Y compare with the std. deviation for X? Does this make sense?

If every value in a sample or population is MULTIPLIED by a factor "r", then

the mean is

the standard deviation is

If an amount "a" is ADDED to every value in a sample or population, then

the mean is

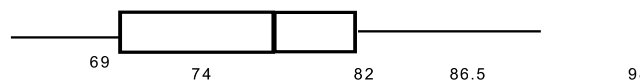
the standard deviation is

Range = highest value – the lowest value = the "width" of the data
Interquartile range = third quartile – first quartile = the "width of the middle 50%" of the data
Standard deviation = square root of average square distance from the mean (almost)

69	70	72	72	73	73	74	74	75	76	77	79	80	82
82	83	83	84	84	84	85	88	88	96	97	97	97	98

30. Find the median and the quartile marks of the combined data *of #3*.
(The **QUARTILE MARKS** consist of the first quartile, Q_1, and the third quartile, Q_3, along with the second quartile mark, which is also the median. These three values divide the data into quarters, thus the name.)

This is a box plot for data with median 82, Q_1= 74, Q_3= 86.5, lowest value 69, and highest value 98 (Of course this box plot needs a title.)



31. What are the IQRs– the **INTERQUARTILE RANGES** (Q_3 - Q_1 of the Classes in #4? Draw **BOX PLOTs** for both.
Step 1: Draw a box across the middle 50%– thus from Q_1 to Q_3, divided at the median.
Step 2: Determine any **OUTLIERS**– data more than 1½ IQRs outside of interval from Q_1 to Q_3.
Step 3: Draw “whiskers” from the box outward to the highest and lowest data that are not outliers.
Step 4: Add asterisks to the line plot for any outliers. (Label, with values at all important points.)