Math 373 Hw 2 Comments and worked examples.

Hw. 53: 2.2ab. 61: 2.14abcd. 68: 2.18abc. 80: 2.29, 2.30ab. Rec. 53: 2.3abc. 61: 2.13abcd. 68: 2.17abc. 80: 2.29, 2.31.

Calculator

Once again, for this course you will need a calculator which can do "two variable statistics". You will need it for the very first exam. You can buy such calculators for a little around $20 at the bookstore. But be sure that it lists "two variable statistics" in its specifications. Calculators which only do one variable statistics won't do.

DEFINITION. For a data set $x_1, x_2, ..., x_n$ of n numbers:

$\bar{x} = \mu = \Sigma x_i / n$ is the *mean* or *average*. Use $\bar{x}$ for samples, μ for populations. "Mean" is also called "arithmetic mean".

After listing the numbers x_i in ascending order, the *median* is the middle number or the average of the two middle numbers.

The *mode* (if any) is the most frequently occurring measurement.

The *range* is the largest number minus the smallest.

$\sigma^2 = \Sigma(x_i - \bar{x})^2/n$ is the *variance* of a population;

$s^2 = \Sigma(x_i - \bar{x})^2/(n-1)$ is the *variance* of a sample.

The *standard deviation* of the sample or population is the (positive) square root of the variance: $s = \sqrt{s^2}$, $\sigma = \sqrt{\sigma^2}$.

- For the 9-element set of data: {1,2,3,6,6,6,9,10,11}:
  The arithmetic mean = (1+2+3+4+6+6+6+9+10+11)/9 = 6
  .5(n+1) = .5(9+1) = .5×10 = 5 = the *median position*
  Median = 6 = the element at the median position.
  Sample standard deviation = 3.54, variance = 12.5 .
  Population standard deviation = 3.33, variance = 11.11 .
  Mode = 6. Range = 11−1 = 10.

Learn to do these with your calculator. Instead of using the formula for the mean, find how to calculate the mean directly. Enter the data once. Then there should be a button for getting the mead, variance and std. dev. without reentering the data.

- In a group of 21 people, 11 are unemployed and receive $0 / month and 10 are employed and receive $2100 / month. Does the mean or the median best describe the "average" wage for this group?

Answer: mean =(11×0+10×2100)/21 = $100.
The middle position = .5(21+1) = .5×22 = 11. The median = the salary of the guy in position 11 = $0.

Here the median is rather unstable. Moving 1 person from unemployed to employed changes the median from 0 to 2100. The mean changes less and thus is better in this case.

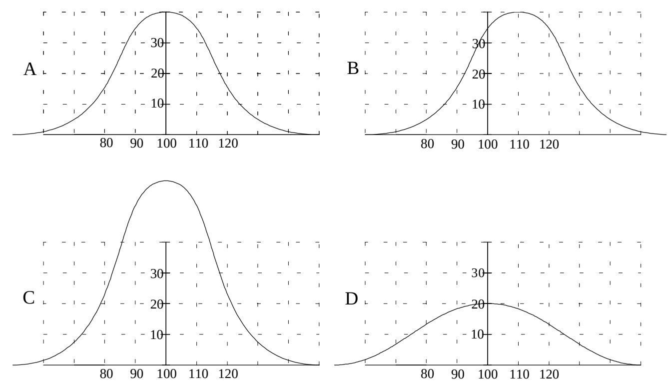
- In a group of 11 people, 10 receive $0 / month, 1 receives $1100 /month. Does the mean or the median best describe the wage earned by the "typical" member of this group?

Answer: Median = 0. Mean = (10×0+1100)/11 = 100.

Median is often the better measure of "typical". Clearly the 1100/mo. guy is an outlier and is not typical. If the outlier's salary rises or falls, so does the mean salary. But changes in the outlier's salary do not affect either the median salary or the salary of the typical guy. Hence the median salary is a better measure of the typical salary in this case.

- The following graphs are smoothed out line graphs of the frequency distributions of IQ scores of four different groups of people: groups A, B, C, and D.

Which group has the most people?	C
Which group has the highest average IQ?	B
Which group has the largest standard deviation?	D



EMPIRICAL BELL RULE. In <u>normal</u> distributions pictured above (bell or mound-shaped curves), about 68% of the data are within 1 std. dev. of the mean. About 95% are within 2 std. dev. About 99% are within 3 std. dev.

The *z-score* of x measures how many std. devs. x is from the mean $\bar{x}$. Thus $x = \bar{x} + zs$. Solving for z gives, the z-score of x: $z = (x - \bar{x})/s$.

- The mean is 50, the std. dev. is 10. Find the z-scores:

z-score of 30 = -2. z-score of 55 = .5. z-score of 50 = 0.

If there are n data items,
the median is the value (interpolated)at the position .5(n +1).
Q_1 is the value (interpolated) at the position .25(n +1).
Q_3 is the value (interpolated) at the position .75(n +1).
IQR, the *interquartile range*, = $Q_3 - Q_1$.

- Six data items: 0, 4, 5, 8, 9, 9. Find the median, Q_1, Q_2.

position	*1*	*2*	*3*	*4*	*5*	*6*
data	0	4	5	8	9	9

n = 6. middle position = .5(n+1) = .5(7) = 3.5.
median = value at position 3.5 = (*5*+*8*)/2 = *6.5*

Think of it this way, we want the value half way between *5* and *8*. The distance between *5* and *8* is (*8*−*5*) = *3*. To get the midpoint, start at *5* and go half way to *8*.
∴ median = *5* + .5(*8*−*5*) = *5*+.5×*3* = *5*+*1.5* = *6.5*

Now find Q_1. By definition, it is the value at position .25(n+1) = .25(7) = 1.75.

The value at position 1 is *0*. The value at position 2 is *4*. To get the value at 1.75, start at *0* and go .75 of the way to *4*. The distance between *0* and *4* = *4*−*0*.

Q_1 = *0*+(.75)(*4*−*0*) = *3*.
Q_3 position = .75(n+1) = .75(6+1) = 5.25,
Value at position 5 = 9, value at position 6 = 9.
Q_3 = 9+(.25)(9−9) = 9. IQR= $Q_3 - Q_1$ = 9−3 = 6.