This handout is designed to help you with the exercises due Wednesday, July 7. See the syllabus at www.stt.msu.edu/~lepage for that assignment.

From Key 8 you are asked to solve the first bulleted exercise after dropping the first element of data. So you must determine the mean, median and mode for each of {8, 33, 2, 20} and {8, 58, 2, 20}. For the former we have mean (8+33+2+20)/4 = 63/4 = 15.75; median = *middle of ordered list {2, 8, 20, 33} = (8+20)/2 = 14*; modes are all of the four values since each occurs equally often (some might say there is no mode at all).

For the second bulleted exercise confirm that adding or multiplying the same constant to each value will do the same to the mean, MEDIAN AND MODE.

For the third bulleted exercise note that the variance is the mean of values $(x - \text{Mean}[x])^2$ none of which is negative. Since the mean of non-negative numbers can only be zero if each of those non-negative numbers is zero it follows that every value x must equal the mean "Mean[x]" of the list.

For the last bulleted exercise you are asked to first determine the mean μ and variance σ^2 for each of the lists {9, 11, 22} and {7, 15} (having dropped the first entry from each of the lists given in the original exercise). For the former (totals shown at the bottom of the table)

x	(x-Mean[x])^2	x^2
9	25	81
11	9	121
22	64	484
_	_	_
42	98	686

So we read off n = 3, $\mu = \frac{42}{3}$ and $\sigma^2 = \frac{98}{3}$. The standard deviation is the square root of the variance or $\sqrt{\frac{98}{3}} \sim$ 5.71548. **Check that you can obtain the population standard deviation directly from your calculator's built-in routine for σ.**

Confirm ALSO that (population) variance may be calculated in the following way:

$$\sigma^2 = \text{mean of squares} - \text{square of mean} = \frac{686}{3} - \left(\frac{42}{3}\right)^2.$$

Caution: Although mathematically exactly the same as population variance, the second method of calculating can be sensitive to rounding errors. For example, the list

0.999999
1.000001

has mean 1 and variance 10^{-12} as shown in the last row of the first and second columns below.

$$\begin{pmatrix} x & (x\text{-Mean}[x])^2 & x^2 \\ 0.999999 & 1.\times 10^{-12} & 0.999998 \\ 1. & 1.\times 10^{-12} & 1. \\ \_ & \_ & \_ \\ 2. & 2.\times 10^{-12} & 2. \end{pmatrix}$$

However the table (much like your calculator screen) is set up to display some fixed number of significant digits. My computer knows the value x = 1.000001 but shows it as 1. If I were to trust the display I would calculate the mean of x as 1.999999/2 instead of its correct value 2/2 = 1. The problem is worse for variance because of squares. Trusting the display we would have mean of squares = 2/2 = 1 and mean of x = 2/2 = 1. Plugging those into the second method for calculating variance would give variance 0. That is 100% inaccuracy in calculating the variance using the second method from values reported in the table (or your calculator screen) for the mean and mean of squares. Squares of x-Mean[x] can be more within the accuracy limitations of computing devices than squares of x.

Bonus. Cook up a list of two numbers for which σ, as calculated from the built-in routine *of your calculator*, differs from the result obtained using the alternative method $\sigma = \sqrt{\text{mean of squares} - \text{square of mean}}$ (using your calculator to obtain the root of the difference between mean of squares and square of mean).

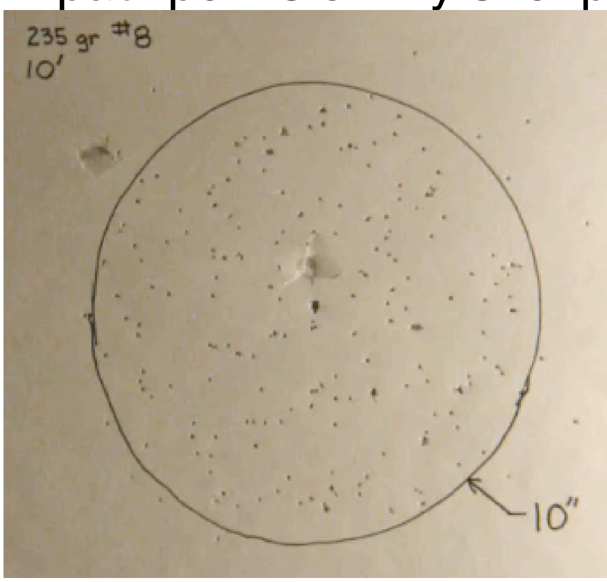
For the rest of the last bulleted exercise of Key 8 you are asked to look at the list of ALL DIFFERENCES x-y between elements of the x and y lists. In our case this is the list of all differences of {9, 11, 22}, {7, 15}:

{9-7, 11-7, 22-7, 9-15, 11-15, 22-15}.

Calculate the population variance σ^2_{x-y} of this list of all differences and verify that it is indeed exactly the same as the SUM $\sigma^2_x + \sigma^2_y$.

Remarkable consequences. This remarkable fact just described was used to experimentally estimate Avagadro's number by experimentally measuring the variance of distances traveled by tiny particles in consequence of countless molecule-induced bumps, following up on an early paper by A. Einstein in which was established *the mathematical relationship between* variance and Avagodro's number. It was one of several methods whose mutual agreement first established Avagadro's number.

In another example of variance at work here is a picture of impact points of tiny shot pellets.



In spite of being randomly scattered, their density (which declines with distance from the center) seems to define a disk. What disk? Its location and size are related in a precise way to mean and the standard deviations of the coordinates of the impact points. Water droplets from a machined nozzle in still air behave much the same.

A related effect may be seen in the shape of a smoke plume. Tiny particles leaving a smoke stack in a gentle horizontal breeze with no stray currents are seen to define a parabolic plume. In principle the plume might look, *from the side view*, like a wedge but it does not. It is instead a parabola whose shape is precisely (in the ideal) related to the standard deviations of up and down motions. If you would take a section of the plume it would appear as a disk.

Variance and standard deviation exert a controlling hand on random phenomena and serve to explain well defined patterns seen in data. We will harness that understanding for statistical work (you will see).

The fact that the variance of x-y is equal to the SUM of the respective variances of x and y (when all pairs of x, y are considered) is deep and relevant for statistics (you will see).