Math 102
Statistics Handout 3

Measures of Dispersion:
Last time, we discussed the different ways we can go about deciding where the "middle" of a given data set is. That is, what the "average" a the data set is. Now, we will consider a different aspect of the "shape" of a data set. Namely, how "wide" a data set is. There are several ways to measure the "width" of a data set.

1. One way of representing the "width" of a numerical data set is to compute the **range** of the data. To do this, we subtract the largest and smallest values in the data set. One disadvantage of using the range is that, like the mean, it is sensitive to "outliers". That is, one very large or very small data value changes the range quite a bit.

   Example: Given the data set $\{4, 8, 12, 15, 11\}$, the range of this data set is $15 - 4 = 11$

2. Another way of computing the "width" of a numerical data set is to compute the **standard deviation** of the data. Computing the standard deviation requires several steps:

   (a) **Step 1:** We start by computing the **deviation from the mean** for *each* element in the data set. That is, given a data value x, we measure how far it is from the mean $\overline{x}$ by computing $x - \overline{x}$

   Example: Again using the data set $\{4, 8, 12, 15, 11\}$, we first compute the mean $\overline{x}$ of the data:
   $\overline{x} = \dfrac{4+8+12+15+11}{5} = \dfrac{50}{5} = 10$

   Next, we compute the deviation from the mean for each data value in the following table:

x	$x - \overline{x}$
4	
8	
11	
12	
15	

   Notice that this computation only tells us how far each individual data value is from the mean. What we want is to figure out how far, on average, all data values in this data set are from the mean. However, if we merely average the deviations of the individual data points, they sum to zero (this is not a surprise since we expect the mean to be "halfway up" the data set).

   (b) **Step 2:** Since just adding up the deviations does not give us useful information, what we do instead is to add up the *square* of all the deviations. This allows us to combine all the deviations in such a way that they don't all cancel each other out. If we take the sum of the *squared deviations* and compute an "average" by dividing by **one less than** the total number of data points, we get a statistic called the **variance** of the data set.

   A formula representing this computation is: Variance $= \dfrac{\sum(x-\overline{x})^2}{n-1}$, where n is the total number of data values.

   Note: You probably think it is a bit strange that when we compute the "average" squared deviation, we divide by $n - 1$ rather than n. There are technical reasons behind this the would be discussed in a more advanced statistics class.

   Example: Again using the data set $\{4, 8, 12, 15, 11\}$, we continue our previous example. To compute the variance, we first fill in the following table:

x	$x - \overline{x}$	$(x - \overline{x})^2$
4	-6	
8	-2	
11	1	
12	2	
15	5	

   Then the variance is $= \dfrac{\sum(x-\overline{x})^2}{n-1} = \dfrac{\quad}{} =$

(c) **Step 3:** The final step in computing the **standard deviation** of a data set is to take the square root of the Variance. This makes sense, since the Variance is the average of the *squared* deviations, and we want a statistic that gives the *average deviation* for the data.

The Standard Deviation $S = \sqrt{\text{Variance}} = \sqrt{\frac{\sum(x-\overline{x})^2}{n-1}} = \sqrt{\qquad} \approx$

Example 2: We will now do an example of computing the standard deviation of a data set that is given to us in a frequency table.

x	frequency	$x \cdot f$	$x - \overline{x}$	$(x-\overline{x})^2$	$(x-\overline{x})^2 \cdot f$
4	2				
7	3				
9	3				
12	2				
Total:					

To find the mean, we fill in the column $x \cdot f$ and then compute $\overline{x} = \frac{}{}$

The Variance is: $\frac{}{} \approx$

The Standard Deviation is: $\sqrt{\qquad} \approx$

3. A third way of computing the "width" of a numerical data set is to compute the **coefficient of variance** of the data. The coefficient of variance combines the mean and standard deviation as follows:

$$CV = \frac{S}{\overline{x}} \cdot 100\%$$

Examples: Consider two data sets: one with mean 19 and standard deviation 5, the other with mean 58 and standard deviation 5. Find the coefficient of variance for each of there data sets:

Notice that the standard deviation of both data sets is the same, but since the means are quite different, a deviation of 5 is larger percentage of the mean data value in the first data set than it is in the second data set.

In general, we think of the coefficient of variance as measuring the **volatility** of the data set.