INS 099 Data Analysis Summer 2007

Day 3: More Data Plotting and Standard Deviations **Due Thurs. July 19, 2007**

Part 1

Minitab

We will be using the program Minitab to plot the data temperature versus optical density (O.D.) found in the P:/ drive labeled as "INS_099_2_Rob"

1. **Define the dependent and independent variables.**

2. **What do the three columns all labeled as "O.D." represent?**

3. **How would you plot this data? How many different ways can you think of to plot this data? Print out all the different plots you make. Do at least three different plots that represent all the data.**

4. **Define the "range" and "optima" for this data.**

Some definitions you got on Mon.

Standard Deviation and Variance

The variance and the closely-related standard deviation are measures of how spread out a distribution is. In other words, they are measures of variability.

The variance is computed as the average squared deviation of each number from its mean. For example, for the numbers 1, 2, and 3, the mean is 2 and the variance is:

$$\sigma^2 = \frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3} = 0.667$$

The formula (in summation notation) for the variance in a population is

$$\sigma^2 = \frac{\sum (X-\mu)^2}{N}$$

where μ is the mean and N is the number of scores.
When the variance is computed in a sample, the statistic

$$S^2 = \frac{\sum (X-M)^2}{N}$$

(where M is the mean of the sample) can be used. S^2 is a biased estimate of σ^2, however. By far the most common formula for computing variance in a sample is:

$$s^2 = \frac{\sum (X-M)^2}{N-1}$$

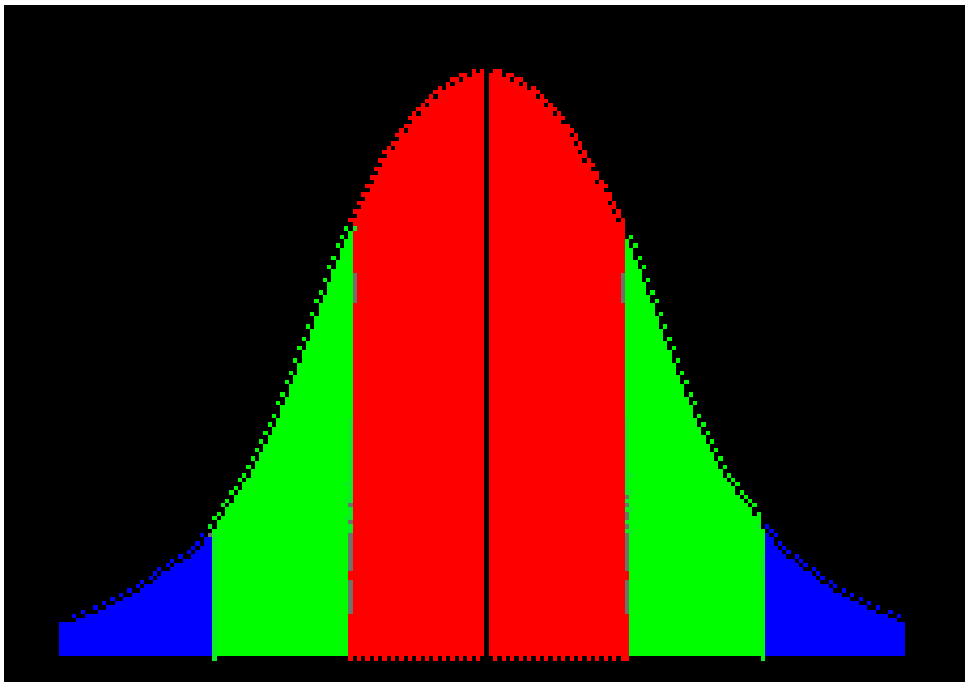
which gives an unbiased estimate of σ^2. Since samples are usually used to estimate parameters, s^2 is the most commonly used measure of variance. Calculating the variance is an important part of many statistical applications and analyses. It is the first step in calculating the standard deviation.

Standard deviation

Standard deviation is simply the square root of the variance, or σ. It is the most commonly used measure of spread.

An important attribute of the standard deviation as a measure of spread is that if the mean and standard deviation of a **normal distribution** are known, it is possible to compute the percentile rank associated with any given score. **How can you determine if you have a normal distribution?** In a normal distribution, about 68% of the scores are within one standard deviation of the mean and about 95% of the scores are within two standard deviations of the mean. These numbers are often referred to as **confidence intervals**.

The standard deviation has proven to be an extremely useful measure of spread in part because it is mathematically tractable. Many formulas in inferential statistics use the standard deviation.



Part 2

For some further practice, you will also be plotting some data that you get involving the effects of temperature and pH on microbial growth.

Effect of temperature on microbial growth:

http://www.brooklyn.cuny.edu/bc/ahp/CellBio/Growth/MGTemp.html

Effect of pH on microbial growth:

http://www.brooklyn.cuny.edu/bc/ahp/CellBio/Growth/MGpH.html

In this simulation, you will classify 8 microorganisms based on their pH and temperature ranges and optima. You will turn in the plots of your data that support the conclusions that you have drawn. Make sure your axes are labeled appropriately to define what your data are.

1. **Which microorganisms are thermophiles? psychrophiles? mesophiles? Define the temperature range and optimum for each microorganism. Make a table.**

2. **Which microorganisms are alkaliphiles? acidophiles? neutrophiles? Define the pH range and optimum for each microorganism. Make a table.**

3. **Which microorganism has the broadest pH range? pH optima?**

4. **Which microorganism has the narrowest pH range? pH optima?**

5. **Which microorganism has the broadest temperature range? temperature optima?**

6. **Which microorganism has the narrowest temperature range? temperature optima?**

7. Where might you expect to find each microorganism? Give specific examples of habitats.

Based upon material found at: http://davidmlane.com/hyperstat/A16252.html, http://www.robertniles.com/stats/stdev.shtml