Lesson Overview
Statistics in Chemistry

Objective: The student will be able to analyze data using the statistical tools of standard deviation and correlation.

Statistics in Chemistry

Statistics is the study of the collection, analysis, interpretation, presentation, and organization of data.

There are a wide variety of techniques in statistics.

- **distributions**
- **averages**
- **standard deviation**
- **graphical methods**
- **matrices**

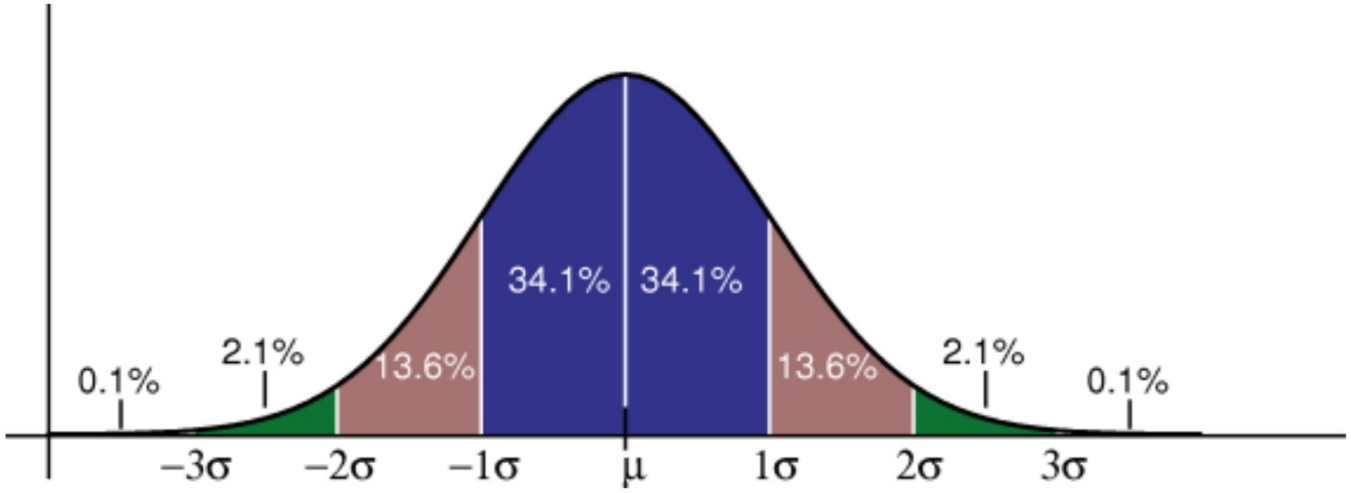
We will be using only a small sampling of these techniques!

Gaussian Distributions

You have by now seen or discussed the idea of a bell curve:

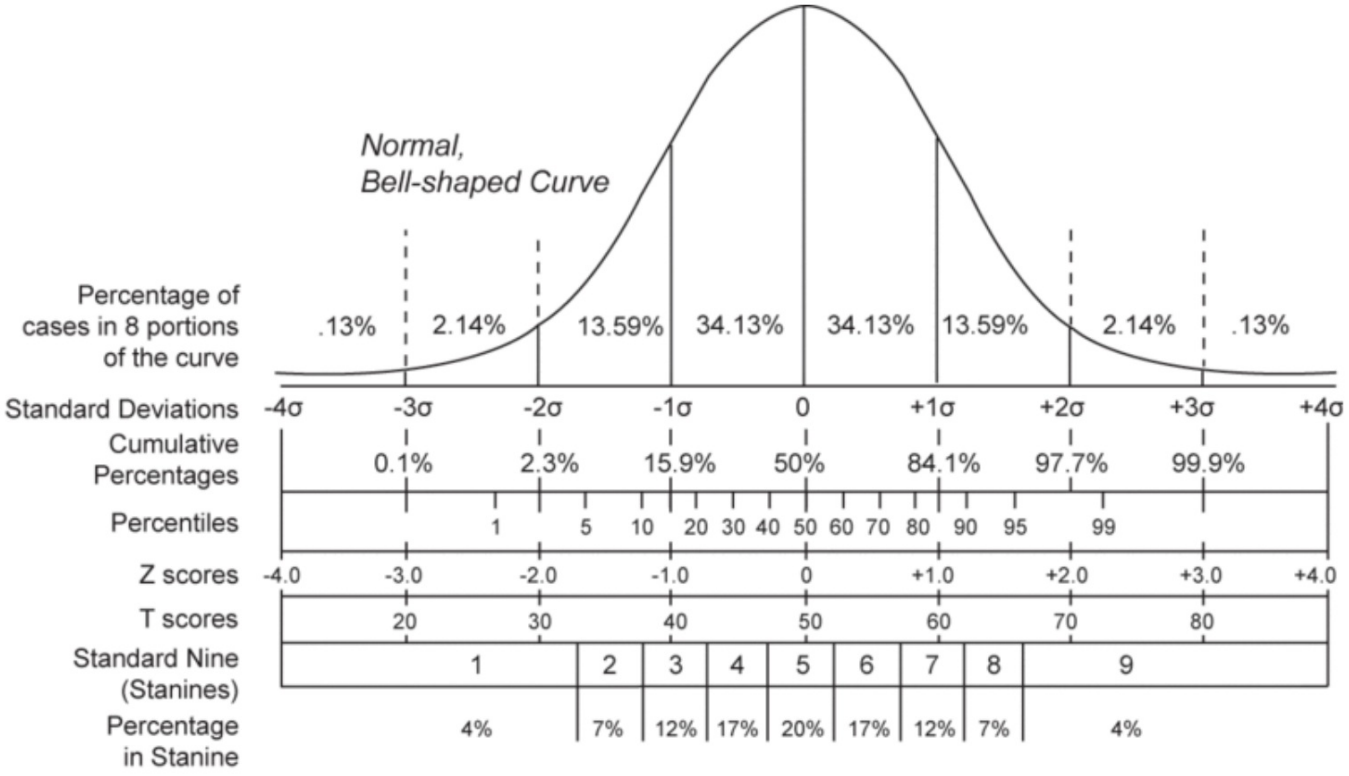
What do you know about this already?

Gaussian Distributions



Gaussian Distributions

Analysis of these distributions can get complex!



Standard Deviation (σ)

a quantity calculated to indicate the extent of deviation for a group as a whole

a measure that is used to quantify the amount of variation or dispersion of a set of data values.

Standard Deviation

In addition to expressing the variability of a population, the standard deviation is commonly used to measure confidence in statistical conclusions.

In science, researchers commonly report the standard deviation of experimental data, and only effects that fall much farther than two standard deviations away from what would have been expected are considered statistically significant

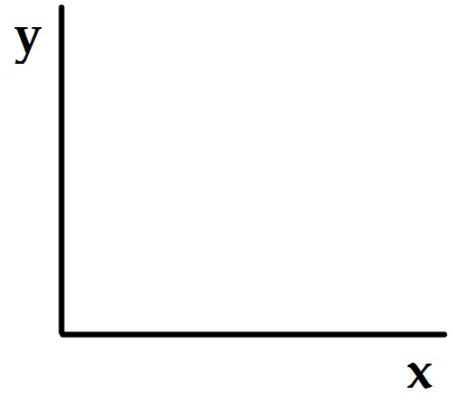
Standard Deviation

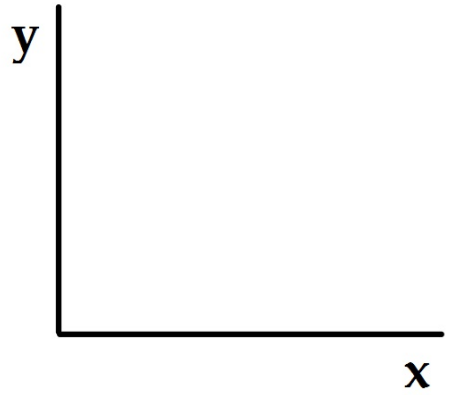
$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Sample Problem

Find the average and the standard deviation for 821, 783, 834, and 855.

Linear Regression
& Correlation





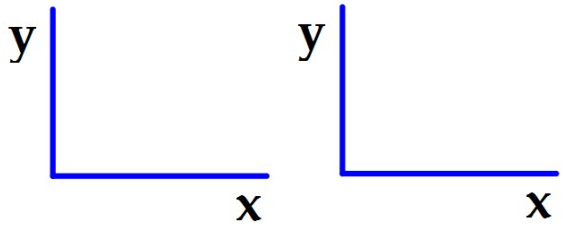
The strength of a linear coefficient is measured by the correlation coefficient "r".

discuss scatter plot, line of best fit, + and - correl.

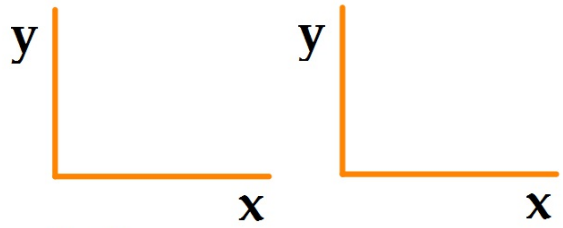
Types of Correlation

Correlation values (r) range from -1 to +1!

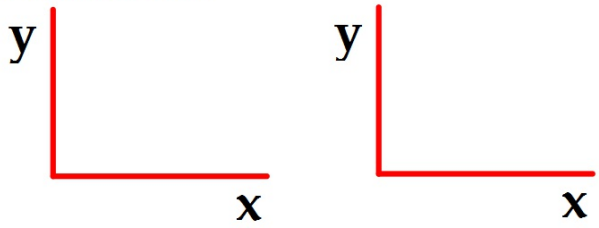
- perfect positive correlation



- no linear correlation

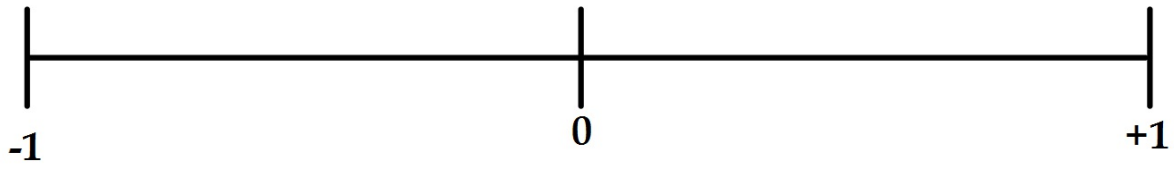


- perfect negative correlation

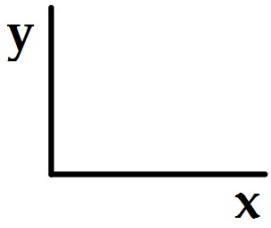


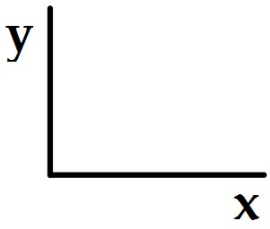
Degrees of Correlation

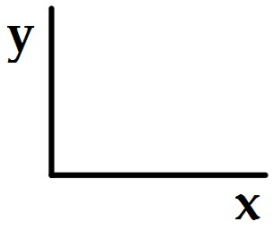
How do we interpret the range of correlation values?



Does each scatter plot show a positive or negative correlation? Is it strong, moderate, or weak?







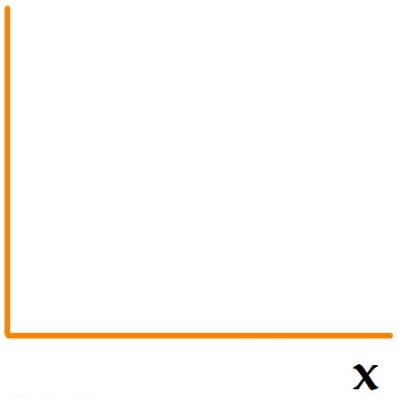
Interpretations of Correlation

Model: If x is shoe size and y is height, interpret each r value:

r value	Interpretation	Analysis (Knowing shoe size…)
-0.2		
0.8		
1		
0		

Coefficient of Determination (r^2)

r^2 is the percent of variation in "y" that is explained by its relationship with "x".



Example:

If r = 0.866, then $r^2 = (0.866)^2 = 0.750$

This means 75% of the variation in height is explained by the variation in shoe size.

What r-squared is not:

This does not mean that 75% of the time height can be predicted from shoe size.

Equation for Correlation Coefficient

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

Sample Problem

Determine the r and r^2 for the following data set and draw conclusions from the statistical treatments:

x	70	115	105	82	93	125	88
y	3	45	21	7	16	62	12

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$