Biochemistry Lab
Quick Statistics

Statistics: Using statistics in biochemistry can be a very complicated and involved effort. Simply put, we use statistical analysis to interrogate data to, without bias, understand if one sample set or population of data is unique from another. How the data is going to be compared will define the type of test used. This document is a very simple introduction to statistical analysis and is not close to complete. This document is intended to help you understand the statistics you may encounter in papers, will help you when you start to use software to perform the appropriate test and you will be expected to use these terms and tools appropriately for your biochemistry research paper.

Keep or Throw Out? How do you know when to throw out a bad data point? This is a difficult question. More often than not, the experiment should be repeated. If the particular experiment allows, you might be able to repeat just the point in question. Sometimes the differences are obvious. If the data points should lie on a particular line or curve and one point is "way off" then that point should be repeated or discarded in drawing the line (although you still show the point on the graph). Experience and judgment is often the guide to these kinds of problems. Not all "bad" points are actually bad. Many important discoveries have been made when the results do not follow expected outcome.

When making several measurements of a single sample, there are several ways in which to know whether or not to reject the point. When a single measurement deviates substantially from a set of three or more replicate measurements, the researcher must decide whether the data point should be rejected. There are several methods to examine the data one of the simpler methods is the Q test, which will be discussed here. The Q-test is best applied when the number of observations is low. The difference between the questioned point and its nearest neighbor ($|X_? - X_{nearest}|$) is divided by the difference between the highest value and lowest value $|X_{max} - X_{min}|$, one of which is $X_?$.

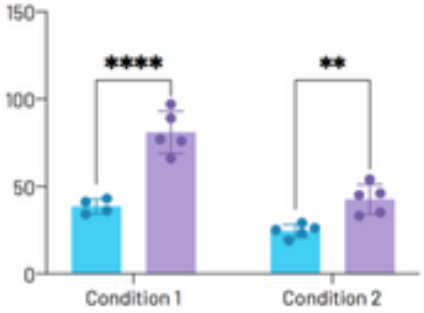
$$\frac{|X_? - X_{nearest}|}{|X_{max} - X_{min}|} = \text{Q test}$$

As an example consider the following data set of five replicate values: 5.0, 5.2, 5.7, 5.8, 7.5. Intuition suggests throwing out the 7.5 value, but an objective criteria is preferred, so we will use the Q-test. Q-test is compared to the standard Q critical values (90% confidence) in the table listed below.

Number of points	Q critical value
3	0.94
4	0.76
5	0.64
6	0.56
7	0.51
8	0.47
9	0.44
10	0.41

If the Q test > Q critical then you may reject the questionable point. In the example, the data yield Q test = 1.7 / 2.5 = 0.68 which is greater than the Q critical value for n=5 (0.64). The value in question, 7.5 can then be rejected.

Error Bars: Error bars represent the deviation or uncertainty of a measurement of some data set. Error bars often represent either the standard deviation or the standard error of the mean. You can see in the figure to the right, that both the individual data points and the error bars are included to display the variation in the data being analyzed. The stars indicate the level of statistical differences. Be warned, that excel gives an option button for "error bars" but it has NEVER worked that way. Work with your instructor on how to correctly add error bars to your graph using excel or other programs mentioned at the end of this handout.



Standard Deviation vs Error of the Mean: While in some ways both of these terms are similar, they are very different from each other.

Standard Deviation (SD or Std Dev) is a measurement of the variability or dispersion of data in relation to the mean. It is an estimate of the variability of the population from which the sample was taken. For normal data, ~95% of results will have values within two standard deviation of the mean. Or in other words, 95% of all observations will fall within this limit. Error bars of the standard deviation will basically give you a graphic understanding that if you take 100 repeated data points, with the experimental conditions used, 95 of the points will fall within that range. Standard deviation of an enter population (all of the possible data) is known as s (sigma) and is calculated using the square of the difference between each data point and the population mean, the sum of that data all divided by the N and then taking the square root of that variance.

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

- Standard deviation (SD) measures the dispersion of a dataset relative to its mean.
- Standard error of the mean (SEM) measured how much discrepancy there is likely to be in a sample's mean compared to the population mean.
- The SEM takes the SD and divides it by the square root of the sample size.

Standard error of the mean (SEM) measures the precision of the estimate of the sample from the sample mean. Or – how far your sample is likely to fall from the population mean (imagine if you could measure the same experiment an infinite amount of time). This is kind of a precision estimate. While many statistical experts will cringe if they read this, the SEM is often used for smaller data sets to give a better measure of the variance from a larger population sample. It is a way to provide a statistical inference based on sampling distribution. SEM is therefore the SD of the theoretical distribution of the sampling distribution (AKA how close is YOUR experiment to the TRUE mean of an infinite sized population).

In even simpler terms, the Standard Error of the Mean tells us how far away your mean is from the actual population mean. The smaller the SEM, the more your data represents the actual population. Thus SEM tells how far away your data scatter are from the actual population mean. And is often used for smaller data sets.

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$$

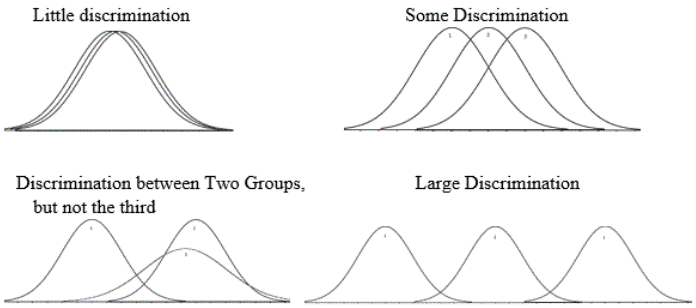
The calculation for the SEM is pretty simple. The mean is calculated from the sample and the sum of the differences of each data point subtracted by the sample mean squared and divided by the sample size (minus one). There are other types of standard errors but we will leave things at this.

How does one test to see if two populations of data different? Student T-test vs Analysis of Variance (ANOVA): When it is time to test if one averaged result is different from another. Say Km of wild-type MDH vs a mutant MDH. You would have measured the Km of both enzymes using the same condition three or more times. This means a full Km curve measured three times. What you are really asking is if the null hypothesis is valid or not. The null hypothesis for our MDH experiment would be that there is no

difference in Km between wild-type MDH and some mutation MDH. And that if the null hypothesis is not valid, then there is a low percent that the Kms are the same. It is kind of an odd way to think of finding the likely difference between Km.

Student t-Test: The simplest way to measure if a mutation gives a statistically significant (different) Km than a mutant is to perform a student t-test. The student t-test is used to compare the means of two data sets. Once not multiple times. For statistical reasons beyond this handout, you can only use a data set once with a student t-test. In other words, if you have two mutants and a wild-type MDH you can only compare wild-type to one of the mutants. BUT you can NOT use the wild-type data to compare to the other mutant. Don't ask. Statisticians just know… It is called a Type I error. This error is about repeating error from a single sample mean. If you are only asking if the Km is larger, a one-tailed t-test is used. If you are interested in knowing if the Km is higher or lower, then a two tailed t-test should be used.

ANOVA: When asking if several sets of mutants are different from each other or the wild-type MDH an analysis of variance (ANOVA) is used. There are lots of different ways of analyzing your data using and ANOVA test. Basically if the curve of a population overlaps with another experimental design, there won't be a significant difference between two populations. Imagine you measured Km values for the wild-type MDH and two mutants 4 or more times (more data – more better) and if you graphed each Km vs the number of times that value occurred it should look like a gaussian curve. If the curves overlap – no differences. If they minimally overlap – ta da, the mean populations are statistically significant from the other. Your mutation did something you can trust. See the figure from Psychstat – Missouri state to the right.



There are different types of ANOVAs. A "*one way ANOVA*" will test if any of the means are significantly different from each other but not which one. This is kind of a an extension of the student t-test and used when the data is expected to be evenly distributed. The one way ANOVA is a way to test the null-hypothesis. An "F value" tells if one of the groups are different from any of the others. The one way ANOVA will inform you if tell which group is different from another, a test of your ANOVA called a post hoc test must be run. One of the most common post hoc test is the Tukey or Dunnett test allowing for comparisons of all data groups. There are others, the Tukey is most commonly used in biochemistry related work.

A "*two way ANOVA*" is often used when you are looking to compare two independent variables of several experimental groups. This type of test is used to know how two independent variables in combination, affect the dependent variable. You might be measuring wild-type MDH vs three mutations each assayed in two different possible inhibitors. Then you could tell if the mutation and inhibitor have an effect on average Km values. To test and tell which variables are unique from each other, a post hoc analysis such as the Tukey or Bonferroni test will need to be run.

P-value: The probability or p-values often seen after a post hoc ANOVA and even a student t-test indicates the statistical difference between pairs of means. If a p-value is less than 0.05, the null hypothesis is rejected meaning there is a good likelihood that the samples are different. This is most correctly described as "significantly different". This is the kind of careful language scientists use vs "proved different". The smaller the p-value the greater the statistical difference between groups, meaning a measured sample is incompatible with the null hypothesis. Using our example a larger p-value for our MDH Km experiment would mean that there is no difference between the Km of a mutation and another mutation or the wild-type protein. This is usually p-value at or greater than 0.05.

A p-value is often demonstrated in graphs with "*" and associated with a numerical value.

- A p-value ≤ 0.05 is statistically significant, provides evidence against the null hypothesis as there is less than a 5% chance of random positive results. This is graphically shown using a single asterisks "*".
- P-values ≤ 0.01 will be given a ** symbol.
- While three asterisks, ***, indicates a P ≤ 0.001.

How do you calculate ANOVAs, post hoc tests and determine F and p-values? Take a statistics course! Alternatively, find a friend who can use the program R or use software program PRISM Graph Pad. There is a limited life-time free student trial available online.