Review of Statistics – Hypothesis Testing

1. Why is it important to understand statistics?

   Often used to explain biological systems
   Used to estimate the probability of a response occurring
   Provides a measure of the variation in response
   Increases the likelihood that a recommendation is dependable

2. Scientists have an idea about the response to treatments in biological systems. This can generally be defined as a hypothesis. A foundation of statistical analyses is stating a hypothesis and then developing a viable experiment to test this hypothesis. Including appropriate controls is important in the experimental procedure.

3. Example of hypothesis testing: The researcher needs to determine if there is a difference in yield when comparing varieties. The researcher has a hunch that yield of the varieties will differ, and in order to make a recommendation about the varieties to practitioners (farmers, extension agents, consultants,) an experiment is established comparing the varieties. Using appropriate data collected from the experiment, statistics are used to determine whether or not practitioners can expect differences in yield of the varieties (in situations outside of the experiment.)

4. The null hypothesis in this case is: *Yield does not differ when comparing these varieties*

   The alternative hypothesis (rejection of the null hypothesis) in this case is: *Yield of at least one of the varieties is different from yield of at least one of the other varieties*

5. Statistics are used to help the researcher make a solid recommendation about yield of the varieties in question. When extrapolating to other situations (the farmer’s field,) the researcher wants to reduce the likelihood of making an error in the recommendation. There are two types of error the researcher can make.

Type I. Reject null hypothesis when the null hypothesis is in fact true.

The researcher tells practitioners that yield of varieties differs when in fact yield of varieties is the same

Type III. Accept the null hypothesis when in fact the null hypothesis is false.

The researcher tells practitioners that there is no difference in yield among varieties when there is a difference in yield among varieties

6. The researcher, given the nature of the comparison, must make a decision about which of these errors is the most critical (i.e. which error is most costly and should be guarded against or avoided.)

   Generally, the researcher would protect against the Type III error. Theoretically, it is more costly to recommend that there is no difference among yield of varieties when in fact there is a difference in yield among varieties.

   A test statistic, for example the F statistic along with a probability value (p-value,) is used to establish a level of probability of the Type III error being made (saying all varieties yield the same when in fact at least two varieties yield differently.) The F statistic is associated with partitioning of variance associated with the treatments that are being compared in contrast with the variation associated with the experimental procedure (statistics courses go into great depth on this concept.) There is a need to “compartmentalize” or “know” or “address” experimental error (associated with variation not resulting from treatment effects) in order to make conclusions from an experiment. For example, if the p value is 0.05 (based on the analysis of variance,) this means that there is a 5% probability that the F value (a higher F value suggests a difference among treatments) will be greater (hence 5 times out of 100.) The researcher is comfortable in that each time the treatments (varieties) are compared, there is a 5% chance (5 times out of 100) that a type III error will be made (saying that there is no difference in yield among varieties when in fact there is a difference in yield between at least two varieties.) This is different from stating that if you conduct the experiment many many times that 95% of the time you will get the same result (that you recommended.) This implies that 5% of the time you will be incorrect. You may never be incorrect, but each time the treatments (varieties) are compared there is a 5% chance that you will make a wrong conclusion about the null hypothesis. Or, if you decided to plant one of several varieties previously tested in a field in a given growing season, there is a 5%probabilty that even though you thought yields would be the same (regardless of which variety you selected) at least one of the varieties will yield differently from one of the other varieties. The test statistic focuses on and applies to “the next time” the treatments are compared irrespective of how the treatments compared the last 5 or 50 or 100 times they were in question. Recently, researchers in the agricultural and life science disciplines have reported differences among treatments at the 10% and 20% levels (protecting from Type I or III errors.) This is an example of how history influences decisions. Prior to computing capabilities we now enjoy and take for granted, probability tables were developed “by hand.” The 5% level was assumed to be reasonable and for that reason 5% (0.05) became the norm because the tables were prepared. With modern computation using other probability levels that are reasonable and less restrictive in protection against error have become more popular.

7. Using hypothesis testing and statistics, the researcher gains an appreciation of the variance of the biological system when comparing treatments. By gaining this appreciation, the researcher can make an informed recommendation which carries the probability of a similar response occurring under conditions outside of the experiment. Bottom line: The practitioner wants to implement a strategy that has a predictable outcome – this predictability greatly increases the practitioner's ability to manage the biological system and consequently manage soil and crop resources.

   When considering the land grant system, the role of statistics in interpreting data and extending a recommendation cannot be over-emphasized. While statistics are correctly considered "a tool" and "not-an-end" in the scientific process, they are none-the-less extremely valuable in predicting response in a biological system (and other systems) in order to facilitate an appropriate recommendation.