An Introduction to Hypothesis Testing

I. Hypothesis Testing as an Inference

One of the roles of inferential statistics is to make inferences about the relationships between variables (usually individual/social characteristics).

We do this by following 3 basic steps:

1) stating hypotheses about these relationships
2) quantifying these relationship with statistics like X^2, r, and t
   - The type of statistic we use depends on the type of variables that we have.
   - For a single nominal-discrete variable or two nominal-discrete variables we use X^2 goodness of fit test, or Pearson's X^2 test for independence.
   - For Two continuous (interval or ratio) variables we use *Pearson's r* (the most common of the correlation coefficients).
   - For nominal-discrete independent variable and a continuous dependent variable we use either t (the t test) or F (The One Way Anova), depending on whether you have two group or two or more groups, respectively.
   - There are many other statistics that can be used, depending on the type of data and the number of independent and dependent variables you want to consider at any given time.
3) making inferences about whether the values we get for each of these statistics is likely to have occurred by chance alone, which is expressed using p values and is conceptualized as α (alpha).
   - Sometimes alpha is referred to as Type I Error. Where type one error is the likely hood that you will say a relationship exists between two variables, when in fact that relationship does not exist.

II. Stating Hypotheses about relationships.

- When we state hypotheses about relationships we do it in two ways.
  1) The Null Hypothesis = Ho
  2) The Alternative Hypothesis (Research Hypothesis) = H_A

-The Null Hypothesis states that there is no relationship between the two variables.

-The Alternative Hypothesis states that there is a relationship.

- Our job as scientists is to find evidence that will allow us to reject one or the other hypotheses.
  - We are never allowed to accept a hypothesis, we can only fail to reject it.
  - We never say that a hypothesis is true or that it can be proven to be true.
- This is because our decisions to reject or fail to reject are based on probabilities, there is always a chance that our decision not to reject a hypothesis could be wrong.
- This helps keep us honest.

III. The Null and Alternative hypotheses and the statistics we select

- The form that the Null hypothesis and Alternative Hypothesis takes depends on which statistics we are talking about, which in turn depends on the type of data we have, and ultimately depends on the type of question we are asking.
- As different statistics are presented in this class and other classes you will see the null hypothesis and the alternative hypothesis presented in their various forms.
- For example: The first statistic we will cover is the Goodness of Fit X^2 (Chi Square) test.
  - This test compares the frequency of occurrence of different nominal groups (e.g. male vs. female) in an obtained sample, with the frequency of occurrence expected by chance (or what is found in the population). If the sample does not differ from what is expected to be found in the population then the Observed frequencies would be equal to the Expected Frequencies.

This is expressed in the Null Hypothesis below. Conversely, if the sample and population do differ, then the Observed and Expected frequencies would differ (the Alternative Hypothesis)

Statistical Hypotheses: $Ho : O_f = E_f$
$H_A : O_f$ =\= E_f

IV. Inferences about the likelihood of occurrence.

- When we evaluate the value of a statistic that tests a hypothesis we are ultimately making an inference about the population from which the sample was drawn. However, when we draw samples from populations there will always be a certain amount of error (sampling error). Just through random chance the sample statistics are going to be different from the true population parameter. Our evaluation of the hypothesis test must account for this error.
- We account for sampling error by using statistical tables to convert our hypothesis testing statistic into a probability value (*p*) which reflects the likelihood that value of our statistic is the result of sampling error, rather than reflecting a true relationship between our variables of interest.
- This probability is often referred to as α (alpha) level.
- We also sometimes call this Type I error rate, which is the likelihood that you will say there is a significant relationship between two variables, when in reality (in the population) the variable are unrelated.

- In the Social Sciences we must be at least 95% sure that our statistic reflects a true (non-random) relationship before we can say it is significant. Inversely, we must have Type I Error Rate (α) of 5% or less, to infer a significant relationship between the variables of interest.
    - With respect to *p*, we must have $p \leq .05$ in order to infer a significant relationship.
    - p (100) = % .05(100) = 5%.

- If $p \leq .05$, then Reject Ho and Fail to Reject H_A
- If $p > .05$, then Reject H_A and Fail to Reject Ho

V. Type I vs. Type II Errors

- Again, Type I Errors occurs when we reject the Null Hypothesis (Fail to reject the Alternative Hypothesis) when in fact the Null Hypothesis is true. That is, we say the variables ***are*** significantly related, when really they ***are not*** related.

- It would seem reasonable to think that it would be best to never commit a Type I Error.
- However, complete protection against type I error would come at the cost of protection against Type II Error

- Type II Error occurs when we Fail to reject the Null Hypothesis (Reject the Alternative Hypothesis) when in fact the Null Hypothesis is false. Specifically, we say the variables ***are not*** significantly related, when in reality they ***are*** related.

Gynocentric Example = Pregnancy Tests

a) 100 % Type I error & 0% Type II - everyone who takes it comes up pregnant (according to test), regardless of whether they are pregnant or not.

b) 0% Type I error & 100% Type II - everyone who takes it come up not-pregnant regardless of whether they are or not.

- What type of error is preferred? It depends...
    - on whether you want to spare non-pregnant people anxiety associated with a false alarm (Test b is best)
    - on whether you want to protect the fetus from accidental exposure to teratogens when a mother is pregnant and thinks she is not (Test a is best).