**Coursera Data Science Course 6 Week 3**

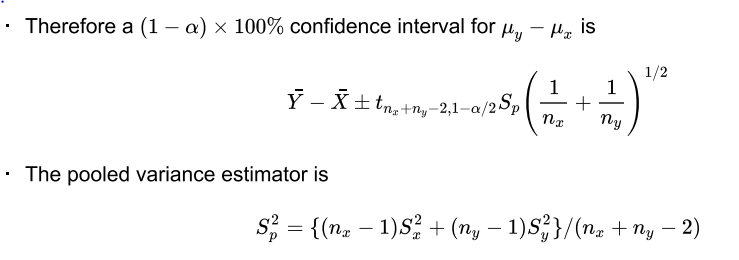
**T Confidence Intervals**

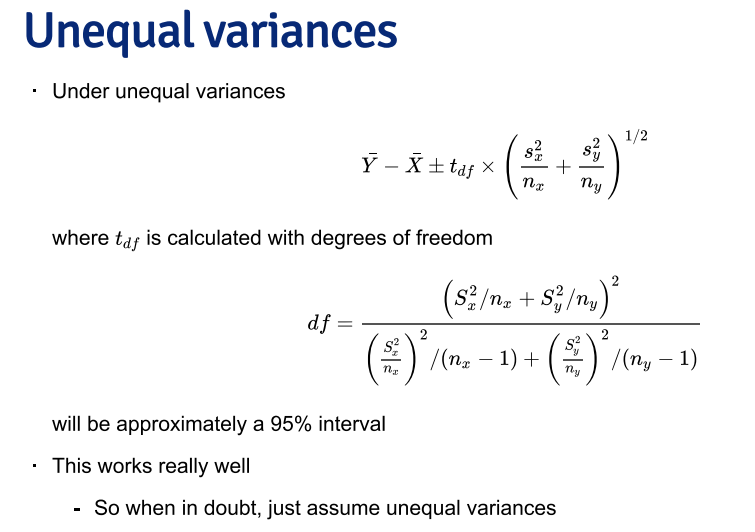
-Some methods for small samples.  
-They are of the form Est +/- TQ x SEEst-these are some of the handiest of intervals  
-if you want a rule between whether to use a t interval or normal interval, just always use the t interval  
-one vs two group versions

**Gosset’s t distribution**  
-thicker tails than the normal  
-indexed by a degrees of freedom, gets more like a standard normal as df gets larger  
-it assumes that the underlying data are iid Gaussian with the result that (X – u) /(S/rootn)  
follows Gosset’s t distribution with n – 1 degrees of freedom  
-(if we replaced s by sigma the statistic would be exactly standard normal)  
-interval is X +/- tn-1S/ (rootn) where tn-1 is the relevant quantile

**…. Some notes**  
- the t interval technically assumes that the data are iid normal, though it is robust to this assumption  
-it works well whenever the distribution of the data is roughly symmetric and mound shaped  
-paired observations are often analyzed using the t interval by taking differences  
-for large degrees of freedom, t quantiles become the same as standard normal quantiles; therefore this interval converges to the same interval as the CLT yielded  
-for skewed distributions, the spirit of the t interval assumptions are violated  
-but WHY would you even center the interval at the mean for skewed distributions!?!?!?  
-consider taking logs or using a different summary like the median instead  
-for highly discrete data, like binary, other intervals should be used

**Independent group t confidence intervals**-supposed that we want to compare the mean blood pressure between two groups in a randomized trial; those who received the treatment vs a placebo  
-we cannot use the paired t test because the groups are independent and may have different sample sizes  
-what do we do then!? -> methods for comparing independent groups!!!



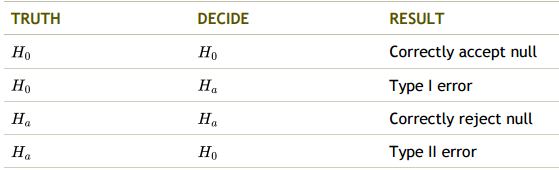
**Comparing other kinds of data**

-for binomial data, there’s lots of way to compare two groups

- relative risk, risk difference, odds ratio  
 - chi-square tests, normal approximations, exact tests  
-for count data, there’s also Chi-squared tests and exact tests  
-we’ll leave the discussions for comparing groups of data for binary and count data until covering glms in the regression class

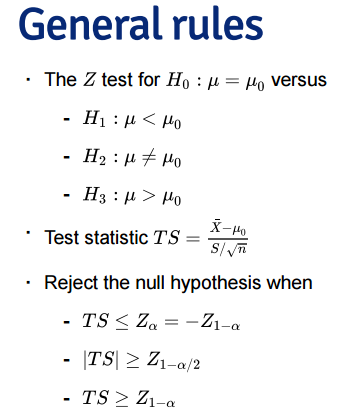
**Hypothesis testing**

-hypothesis testing is concerned with making decisions using data  
-a null hypothesis is specified that represents the status quo, usually labeled Ho  
-the null hypothesis is assumed true and statistical evidence is required to reject it in favor of a research or alternative hypothesis

-can be <, > or not equal  
****

**Some Pointers… ; how**-consider a court of law, the null hypothesis is that the defendant is innocent  
-we require a standard on the available evidence to reject the nulypothesis  
-if we set a low standard, then we would increase the percentage of innocent people convicted (type I errors); however we would also increase the percentage of guilty people convicted (correctly rejecting the null)  
-if we set a high standard, then we increase the percentage of innocent people let free (correctly accepting the null) while we would also increase the percentage of guilty people let free (type II errors)

-In general we don’t convert C back to the original scale  
-we would just reject because the Z-score; which is how many standard errors the sample mean is above the hypothesize mean



**NOTES**-we have fixed alpha to be low, so if we reject Ho (either our model is wrong), or there is a low probability that we have made an error  
-we have not fixed the probability of a type II error, beta; therefore we tend to say “Fail to reject Ho” rather than accepting Ho  
-statistical significance is not the same as scientific significance  
-the region of TS values for which you reject Ho is called the rejection region

MORE NOTES  
-the Z test requires the assumptions of the CLT and for n to be large enough for it to apply  
-if n is small, then a Gossett’s T test is performed exactly in the same way, with the normal T-the probability of rejecting the null hypothesis when it is false is called power  
-power is used to calculate sample sizes for experiments

Two sided test  
-suppose that we would reject the null hypothesis if in fact the mean was too large or too small  
-that is, we want to test the alternative Ha: u =/= 30  
-we will reject if the test static, 0.8, is either too large or too small  
-then we want the probability of rejecting under the null to be 5%, split equally as 2.5% in the upper tail and 2.5% in the lower tail.  
-thus we reject if our test statistic is larger than qt or smaller than qt  
-so we fail to reject the two sided test as well

**Connections with confidence intervals**  
-consider testing Ho: u = uo vs Ha: u =/= uo  
-take the set of all possible values for which you fail to reject Ho, this set is a (1-alpha)100% confidence interval for u

-For two group intervals, rejection rules are the same  
-test Ho: u1 = u2

Notes:  
-its impossible to get an exact 5% level test for this case due to the discreteness of the binomial.  
 -the closest is the rejection region  
 -any alpha level lower than 0.0039 is not attainable  
-for larger sample sizes, we could do a normal approximation, but you already knew this  
-two sided isn’t obvious  
 -given a way to do two sided tests, we could take the set of values of po for which we fail to reject to get an exact binomial confidence interval (called the Clopper/Pearson interval)

**P-values**-most common measure of “statistical significance”

Suppose nothing is going on- how unusual is it to see the estimate we got

The “hows”  
1. Define the hypothetical distribution of a data summary (statistic) when “nothing is going on” (null hypothesis)  
2. Calculate the summary/statistic with the data we have (test statistic)  
3. Compare what we calculated to our hypothetical distribution and see if the value is “extreme” (p-value)

-the P-value is the probability under the null hypothesis of obtaining evidence as extreme or more extreme than would be observed by chance alone  
-if the P-value is small, then either Ho is true and we have observed a rare event or Ho is false   
-in our example the T statistic was 0.8  
-what’s the probability of getting a T statistic as larger as 0.8?

-therefore, the probability of seeing evidence as extreme or more extreme than that actually obtained under Ho is 0.2181

**The attained significance level**  
-our test statistic was 2 for Ho: uo = 30 vs Ha: u > 30  
-notice that we rejected the one sided test when alpha = 0.05, would we reject if alpha = 0.01, how about 0.001?  
-the smallest value for alpha that you still reject the null hypothesis is called the attained significance level  
-this is equivalent, but philosophically a little different from, the P-value

Notes  
-by reporting a P-value the reader can perform the hypothesis test wat whatever alpha level he or she chooses  
-if the P-value is less than alpha you reject the null hypothesis  
-for two sided hypothesis test, double the smaller of the two one sided hypothesis test Pvalues