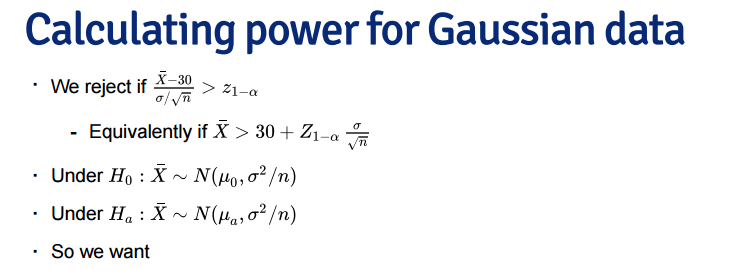
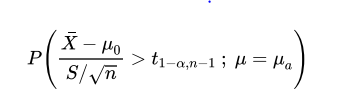
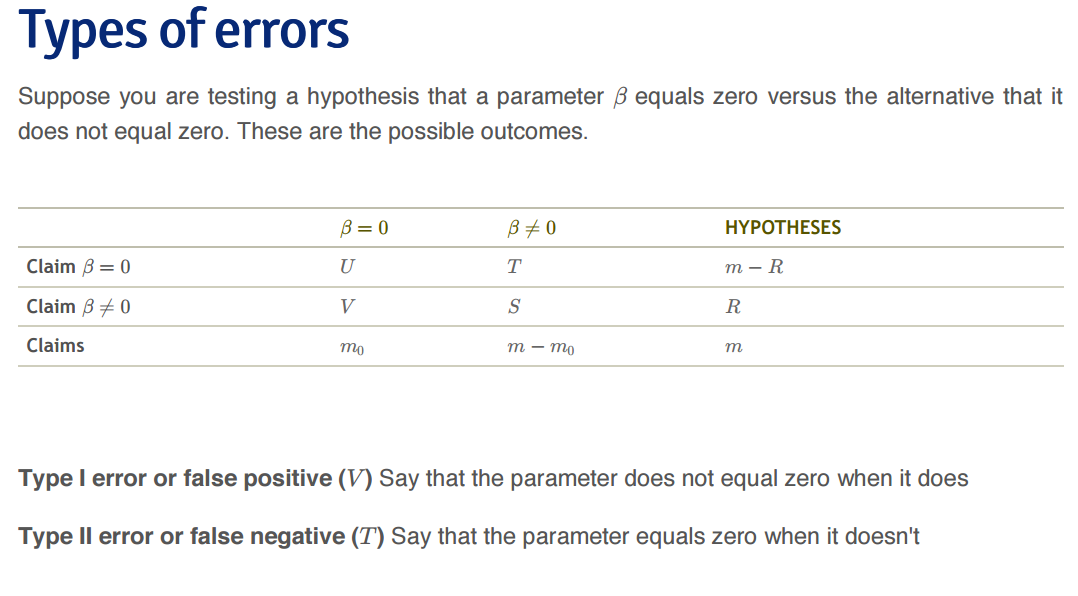
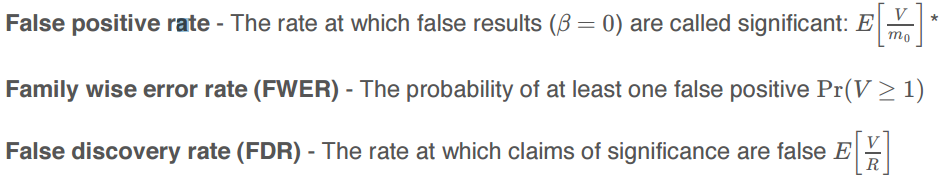
**Coursera Data Science Course 6 Week 4**

**Power**-power is the probability of rejecting the null hypothesis when it is false  
-ergo, power is a good thing; you want more power  
-a type II error (a bad thing, as its name would suggest) is failing to reject the null hypothesis when it is false; the probability of a type II error is called beta.  
-therefore, power = 1 – beta

-consider our previous example involving RDI  
-Ho: u = 30 vs Ha: u > 30  
-then power is   
-note that this is a function that depends on the specific value of ua!  
-notice as ua approaches 30 the power approaches alph 

Notes  
-the calculation for Ha: u < uo is similar  
-for Ha: u =/= uo calculate the one sided power using alpha/2 (this is only approximately right, it excludes the probability of getting a large TS in opposide direction of the truth)  
-power goes up as alpha gets larger  
-power of a one sided test is greater than the power of the associated two sided test  
-power goes up as u1 gets further away from uo  
-power goes up as n goes up  
-power doesn’t need ua, alpha or n  
-only need effect size, the difference in the means in standard deviation units  
-being unit free, it has some hope of interpretability across settings

**T-test power**-consider calculating power for a Gossett’ss T test for our example  
-the power is  


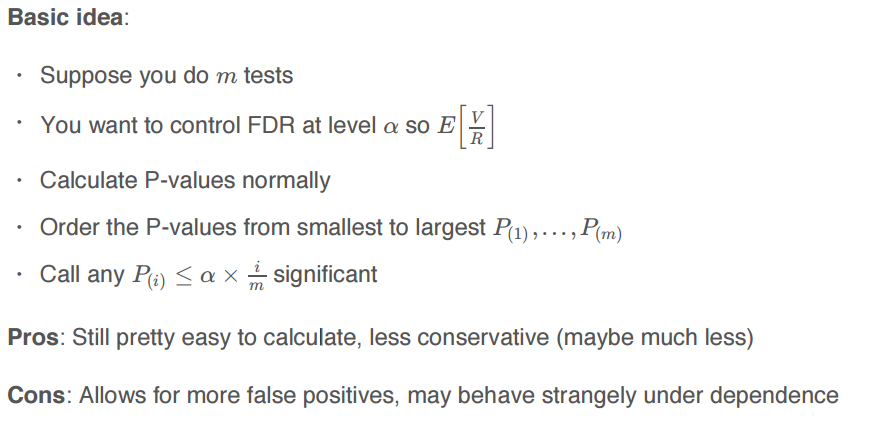
**Error Rates**

**Controlling….**

**1. False Positive Rate**-if P-values are correctly calculated calling all P < alpha significant will control the false positive rate at level alpha on average.  
P < 0.05 significant – therefore 10000 x 0.05 = 500 false positives.  
THAT’S A LOT

**2**. **Family-wise error rate (FWER)**

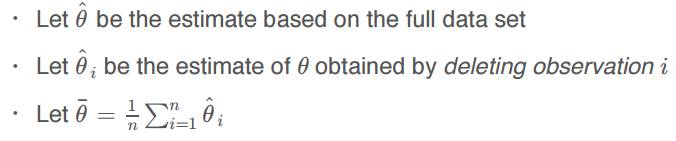
The Bonferroni correction is the oldest multiple testing correction.  
Suppose you do m tests  
-you want to control FWER at level alpha so Pr(V >= 1) < alpha  
-calculate P-values normally  
-set alpha(fwer) = alpha/m  
-call all P=values less than alpha(fwer) significant  
PRO: Easy to calculate, conservative but MIGHT BE TOO CONSERVATIVE

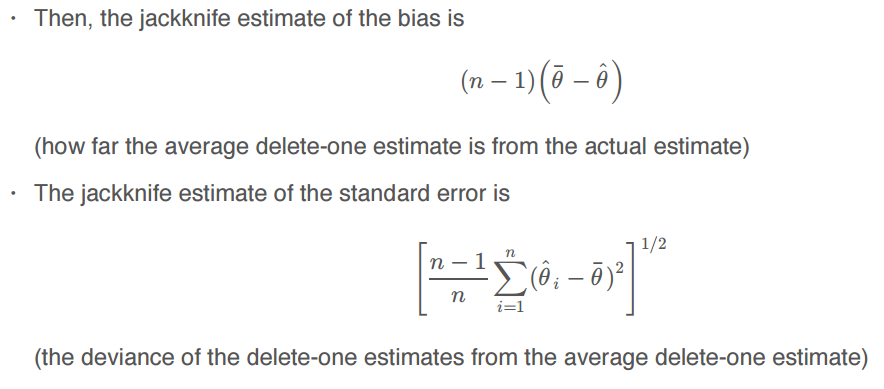
**3. False Discovery Rate (FDR)**

**Adjusted P Values**-one approach is to adjust to threshold alpha  
-a different approach is to calculate “adjusted p-values”  
-they are not p-values anymore  
-but they can be used directly without adjusting alpha

**The JACKKNIFE**-the jackknife is a tool for estimating standard errors and the bias of estimators  
-as its name suggests, the jackknife is a small, handy tool; in contrast to the bootstrap, which is then the moral equivalent of a giant workshop full of tools  
-both the jackknife and the bootstrap involve resampling data; that is, repeatedly creating new data sets from the original data

-The jackknife deletes each observation and calculates an estimate based on the remaining n-1 of them  
-it uses this collection of estimates to do things like estimate the biase and the standard deviation  
-note that estimating the bias and having a standard error are not needed for things like sample means, which we know are unbiased estimates of population means and what their standard errors are

-We’ll consider the jackknife of univariate data  
-Let X1 to Xn be a collection data used to estimate a parameter theta  




**The Bootstrap**tremendously useful tool for constructing confidence intervals and calculating standard errors for difficult statistics  
for example, how would one derive a confidence interval for the median?  
the bootstrap procedure follows from the so called bootstrap principle

**THE PRINCIPLE**  
-supposed I have a statistic that estimates some population parameter, but I don’t know its sampling distribution  
-the bootstrap principle suggests using the distribution defined by the data to approximate its sampling distribution

**The Bootstrap in Practice**-in practice, the bootstrap principle is always carried out using simulation  
-we will cover only a few aspects of bootstrap resampling  
-the general procedure follows by first simulating complete data sets from observed data with replacement  
 -this is approximately drawing from the sampling distribution of that statistic, at least as far as the data is able to approximate the true population distribution  
-calculate the statistic for each simulated data set  
-use the simulated statistics to either define a confidence interval or take a standard deviation to calculate a standard error

**Permutation tests**  
-consider the null hypothesis that the distribution of the observations from each group is the same  
-then the group labels are irrelevant  
-we then discard the group levels and permute the combined data  
-split the permuted data into two groups with nA and nB observations  
-evaluate the probability of getting a statistic as large or large than the one observed  
-an example statistic would be the difference in the averages between the two groups, one could also use a t-statistic

