LLO 8200: Beyond Summary

Youtube Generated subtitles

now that we've reviewed some of the basic summary statistics let's move on to the two major ways in we which we will be addressing statistical problems hypothesis testing and parameter estimation so hypothesis testing is simply is there an effect or not simple yes or no when a paper or an article says that a study found a statistically significant difference this is what we're talking about we're talking about a hypothesis test this just means that a test statistic the result of some test was above a particular threshold usually resulting in a probability of no effect being less than .05 parameter estimation is instead going to look at how large a relationship is okay it's going to quantify the size of the effect the difference the association without saying whether or not it is likely statistically okay you can use you can use hypothesis testing by itself you can use parameter estimation by itself but obviously it's much more powerful when you use both you say an effect is present or not and how large that effect is if it is in fact there so four hypothesis testing again it is testing whether there is an effect or not because of the way that we structure our statistical tests we are actually using a falsification criteria so instead of testing whether or not an effect is present we test whether or not there is no difference or whether a lack of an effect is likely or unlikely okay if we reject that hypothesis then we conclude that there is a difference that there is an effect okay the null hypothesis is the name that we give to this claim that there is no effect if our test statistic is too small and it produces a measured probability that is greater than our Alpha then we will fail to reject the null hypothesis if this test statistic is great enough it will produce a p-value that is less than our Alpha with so we will reject the null hypothesis okay just a couple quick things our Alpha in this case is going to be our false positive rate this is how many false positives out of 100 we're willing to tolerate well expressed as a proportion so typically it's set to .05 but there are other values that it can be set to as well okay um now one thing to note here is that there is that inverse relationship between the test statistic and the p-value for the most part the larger the test statistic the smaller the p-value so one thing that I do want to emphasize here is note the way that I said it okay I said that we failed to reject the null hypothesis not that we accept it we never accept that there is no difference we just fail to reject their um we just failed to reject there being no difference that is to say that a lack of a difference is unlikely okay

most hypothesis testing will produce a statistically significant difference with enough data for example sex is actually related to latitude so tropical climates will yield sex ratios so that 51.1 percent of boys are born as opposed to sub-arctic latitudes where the percentage of boys that are born is actually 51.3 so where you are on the planet in terms of latitude actually has an effect on the likelihood that you are born male or born female okay the effect size though the Practical significance of this difference is fairly trivial so it's not something that we would necessarily make policy over but it is something that we would find a statistically significant difference if we gathered a sufficiently large sample okay this is why we can't rely solely on hypothesis testing and we do need to fall back on parameter estimation as well to get some kind of an indication about that practical significance

so hypothesis testing for us is most likely going to come into the form of a ratio of some kind between what we observe and what we expect okay take for example the Z and the znt tests okay what we're doing is we are comparing uh meaning dividing the distance between our calculated mean and our hypothetical mean and we are comparing that to the what we expect to see from the sampling distribution the sampling error the standard error okay and that's in the case of both the Z test and the t-test okay when this ratio exceeds a critical value we say a lack of an effect is unlikely we we found something we reject the null hypothesis of there being no effect

strictly speaking the for the T statistic the critical value is going to change as a function of the degrees of freedom or the sample size but when that sample size is greater than about 60 the critical value is 2 and then above 60 is less than 2 approaching 1.96 so as a sort of shorthand as a heuristic we will often just use plus or minus 2 as a cut off when we're sort of making decisions looking at a statistical analysis rather than going and having to read all of the very specific p-values especially when they're expressed sometimes in scientific notation if we look Instead at the F ratio okay we compare the variance of the variance that the model can account for and we divide that by the by the residual by the variance of the residual or the noise um it's going to be the same in the case of the classic Anova and the linear regression the only difference is going to be how we calculate those degrees of freedom okay in a classic Anova we have K groups and then we use a one degree of freedom to calculate the grand mean or the mean of the means for the um for for the data set that we have so we have the degree of freedom for the model or between is going to be K minus 1. and the residual therefore is going to be n minus K or the degree of Freedom within will be n minus k and then in the case of uh in the case of regression our degree of freedom for the model or the degree of Freedom that is being explained is going to be k for the number of predictors in the model and the residual what's left over is going to be n minus K minus 1. we use up one additional degree of freedom for estimating the um estimating the intercept so a critical part here is that we are looking at variances right these mean squared are variances right sums of squares divided by something that's n flavored right the degrees of freedom are very n reminiscent very end like so so the mean squares are essentially variances so this analysis of variance is in fact a ratio and F ratio is an analysis of the variance accounted for by the model divided by the variance that is unaccounted for by the model or the residual that's left over okay and I used model and residual for both the classic Anova and the regression case you learned for Anova you learned that this was between and within but conceptually it's the same thing okay

now parameter estimation we said that hypothesis testing is going to take care of identifying whether or not an effect is present so now parameter estimation is saying okay regardless of whether or not there is an effect how big is it how big is the difference how big is the association how big is the relationship okay independent of its presence or absence there are several different methods there are several different methods of parameter estimation but they are going to tend to fall into one of a few different types the big types that we will be seeing are going to be distance Association and variance explained okay you already know examples of each of these Cohen's D is a measure of a distance right it measures the distance between two means in the number of standard deviation units right it says if I have one mean a mean for a control group and a mean for a treatment group and I divide that by the standard deviation the pooled standard deviation that will tell me how many standard deviations away the means are from each other okay so this tells us a distance you also learned about several types of correlation well I mean it's really all the same correlation but with varied inputs right we have the Pearson's R which appears we have the Pierce we have the traditional Pearson score correlation in which the the inputs the variables are both on an interval scale and it will allow us to sort of determine whether or not there is linearity between two variables we have spearman's rank order Row in which the two inputs are ordinal and this allows us to describe or test or check for monotonicity that is to say whether or not the data points move together positively move positively together or negatively together regardless of the linearity aspect they Point by serial correlation in which one of the variables is interval and one of them is dichotomous that is to say nominal with two cases zero or one and this will allow us to determine sort of discriminability right how much um how much do the two values of this dichotomous variable sort of separate into clear one State versus the other and then we have the Phi coefficient which is um two dichotomous variables both nominal variables and it will just test sort of a base association between the two variables

but outside of correlation you can even think about regression okay regression coefficients when you calculated a regression you got a regression equation okay that that equation contained an intercept it contained a slope for each of your predictors and then we said our outcome or a predicted outcome was going to be expressed by this sort of Y hat okay so these B weights these regression coefficients I mean they're basically covariances they're like unstandardized correlations they're like unstandardized partial correlations but they allow you to sort of predict a level of the outcome variable based on the the measured levels of each of the individual predictors uh you can also see it with the the standardized version right the beta weights we calculated the beta weights right we took all of we took all of the all of the inputs all of the independent variables and the dependent variable we standardize them and we ran the regression and we said oh look this is the standardized score it allowed us to compare the relative magnitude of these partial regression coefficients in a way that allowed us to literally compare apples and oranges and then the last one is going to be the variance explained this is going to go back to that F ratio specifically this relationship between the sums of squares total the sums of squares model and the sums of square residual right when we take the sums of squares as predicted by the model and we divide it by the total that's going to tell us the proportion of the variance it's accounted for by that model right this is the sums of squares of my model it is whatever my predicted value was minus the mean and we're dividing that by the observation minus the mean right the The observed The observed total sums of squares that's going to be true in both the regression case and the Anova case the a Anova case we can do some additional juggling because special cases with categorical variables but it's the same general idea we're taking this model sums of squares dividing it by the total and that is going to tell us what proportion of the variance is accounted for by this model in this case it's what's accounted for by the predictors in the regression model and in this case it's going to be what is predicted by the this set of group means by group membership of this categorical variable okay that'll apply for both R and 8 r squared and ETA squared now this one's also really nice because r squared is going to actually allow us to not just look at sort of this overall variance accounted for but by deriving this r squared and this ETA squared we can actually look at sort of the net Association strength of a set of variables through this multiple r that you were able to find using using Excel previously okay there are other statistics that we use to measure effect size but they're mostly just going to be more nuanced or sophisticated versions of the ones that we're showing here this is a vastly oversimplified review of hypothesis testing and parameter estimation but these are the things that I really want you to be to be cognizant of this idea that what we are doing here is we are make we're just making comparisons okay sometimes those comparisons are distances sometimes those comparisons are are variances what we need to be thinking about as we are going along and as we are um as we are reading studies and as we are talking about these other issues like as we talk about error as we talk about variance components um we're going to need to stop for a second and we're going to need to think about these individual parts and what it means for us to say that something has an r squared of uh of 0.5 or what does it mean that a correlation was a 0.3 or a 0.6 what does it mean that we had statistical significance with a p-value of 0.05 but not a 0.01