Help Us Translate

0:04 In this video, we'll be talking about estimating a difference in two population proportions with confidence. To start off with, we're looking at a poll that was conducted by C. S. Mott Children's Hospital. They conduct a lot of national polls on issues in children's health. In this particular poll, they were asking about water safety and about swimming lessons. The research question that we'll be looking at is, what is the difference of population proportions of parents reporting that their children aged 6-18 have had some swimming lessons between white children and black children? So, our populations of interest, we have two of them. We'll be looking at all parents of white children age 6-18 and all parents of black children age 6-18. Now, in this case, our parameter, since we're looking at two different population proportions will be the difference in our two population proportions. In notation, this will be p_1 minus p_2. Now, because we have two proportions in here, we need to specify which group is what. We use the subscripts, the group one and group two labels to identify that we have two different population proportions here. In this case, you have to specify which is group one and which is group two. In this video, we'll be using group one as the parents of white children and group two as the parents of black children. Now, our goal for this video is constructing a 95 percent confidence interval for the difference in population proportions of parents reporting that their child has had swimming lessons between the parents of white children and the parents of black children. All right. So, let's go ahead and look at the survey results. We took a sample of 247 parents of black children, and we found that 91 said that their child has had some swimming lessons. We also took a sample of 988 parents of white children, and we found that 543 said that their child has had some swimming lessons. Now, it's hard to look at just those numbers and come up with a really clear answer to our question of what is the difference in the population proportions of children who have taken swimming lessons. So, what we'll want to do is look at confidence interval basics. So, if you remember, we have our best estimate plus or minus the margin of error. Now, we do know how to calculate confidence intervals for each of these population proportion separately. But what would we do once we've calculated those two confidence intervals to create a joint confidence interval for the difference in those two population proportions? We'll stick to this formula and modify it for our use with the difference in population proportions. To start off with, let's think about what our best estimate might be. Our perimeter remember is p_1 minus p_2. So, we could estimate it with the sample proportions instead of the population proportions and get an estimate for our perimeter that way. We might look at p hat one minus p hat two or the difference in the two sample proportions. Here, we still have that plus or minus the margin of error. If you remember in the past, we used our margin of error being a few standard errors. So, we'll substitute n for the margin of error, a few standard errors. Now, since we're looking at a 95 percent confidence interval, we'll use 1.96 as "a few", and our standard error formula will be given to you here, but within that standard error formula, it looks pretty similar to the standard error formula that we had for a single portion. So here, you notice that we have p hat one times one minus p hat one over n_1, and we add that to the same thing for group two. But that looks really similar to the standard error for a single population proportion found in that formula. Here, we'll square root again. So, first let's think about what is the best estimate of our parameter. We'll use p hat one, we can calculate that sample proportion pretty easily. It's 543 divided by 988, and so we find that in our sample, 55 percent of the parents reported that their child has had some swimming lessons of those parents of white children. Remembering that group two is the parents of black children, we can calculate the same sample proportion, and so we find that it's 91 over 247. We see that 37 percent of the parents of black children report that their child has had some swimming lessons. So, we can go ahead and calculate the simple difference between these two and we find that the parents of white children report that their child has had swimming lessons 18 percent more than the parents of black children. So, we have our simple estimate 0.18, but we also want to estimate with some sort of confidence. So, we can't forget about that margin of error. Looking back at that formula for the confidence interval for the difference in two proportions, we can go ahead and substitute in our p hat numbers and our sample size numbers. So, we know what p hat one and p hat two are, we know n_1 and n_2 are. We can go ahead and substitute those numbers in, go ahead and plug and chug, and we'll find that our confidence interval goes from about 11 percent to about 25 percent. So, let's think about what that confidence interval means. If you remember a confidence interval is a range of reasonable values for our parameter. Our parameter is the difference in the two population proportions, and so our interval will look something like this. With 95 percent confidence, the population proportion of parents with white children who have taken swimming lessons is roughly 11 percent to roughly 25 percent higher than the population proportion of parents with black children who have taken swimming lessons. So, we've covered the actual calculation of the confidence interval as well as a little bit of the interpretation of the confidence interval. In the next video, we'll talk more about how you interpret confidence intervals as well as some of the underlying assumptions that you should check before you actually calculate that confidence interval.