Math 316

Section 10.7 Confidence Intervals for Difference between Probabilities

Recall for one population $SE_{\bar{Y}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{s^2}{n}}$ and for two populations $SE_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ and the confidence interval for the difference of the two population means is

$$\mu_1 - \mu_2 = \bar{y}_1 - \bar{y}_2 \pm t_{\alpha/2} \cdot SE_{\bar{Y}_1 - \bar{Y}_2} .$$

It's similar for finding the confidence interval for the difference of two population proportions. Recall that for one sample from one population, if we were finding a 95% confidence interval, then we would have

$$SE_{\tilde{P}} = \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} \quad \text{where } \tilde{p} = \frac{y+2}{n+4}$$

Similarly for two populations (still if wanting a 95% confidence interval)

$$SE_{\tilde{P}_1 - \tilde{P}_2} = \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+2}} \quad \text{where } \tilde{p}_i = \frac{y_i+1}{n_i+2}$$

and

$$p_1 - p_2 = \tilde{p}_1 - \tilde{p}_2 \pm 1.96 \cdot SE_{\tilde{P}_1 - \tilde{P}_2}$$

It's similar for other confidence levels.

See **Book Example 10.7.1** (from **Example 10.1.1**). Recall that the value 1.96 is used in finding the confidence interval is a z value, i.e. a t value with $df = \infty$, for 95% confidence. So based on the 95% confidence interval $0.042 < p_1 - p_2 < 0.464$, there is significant evidence that there is a difference between the real and the sham surgeries. We could also have found a one-side confidence interval. See **Class Example 1**.

In **Class Example 2** let's look at how things change if we double all values, so we have double the sample sizes, but with the same proportions in each outcome.

Section 10.8 Paired Data and 2×2 Tables

In Chapter 8 we learned about working with paired data with continuous (numerical) values. For example, recall **Book Example 8.1.1** and **Table 8.1.1**. There really are not two samples, each with 9 values. There is a single set of 9 values: the 9 differences in the final column.

In Section 10.8 we learn about working with paired data with categorical values. Consider **Book Example 10.8.1**. What are your first impressions about whether there is a difference in older vs. younger siblings getting HIV? In **Table 10.8.1** it would seem that there are 114 values in one sample and 114 values in the other. In reality, there are 114 total values for the 114 pairs of siblings. Somehow the way we list and analyze the data should reflect this. This is the point of **Table 10.8.2**, which is a better way to organize the data. What we want to investigate is: in pairs of siblings in which one sibling has HIV and the other does not, is it more likely that the younger one will have HIV when older one does not, or conversely? These are the values of 18 and 17 in Table 10.8.2. The other two values are when both are the same: 2 pairs of siblings in which both have HIV and 77 pairs in which neither has HIV. So we are really just interested in the categories of Yes/No and No/Yes. Let's look at **Class Example 3**. Compare our work to **the formulas given on page 420**. Note that $df = (2-1) \times (2-1) = 1$.

Regarding Table 10.8.1 vs. 10.8.2 and how best to organize/display the data, other possibilities for Table 10.8.2 that would have fit the values given in Table 10.8.1:

Table 10.8.2	$\chi_s^2 =$	P
$\begin{matrix} 0 & 19 \\ 20 & 75 \end{matrix}$	$\frac{1^2}{20+19} \approx .026$	$> .20$
$\begin{matrix} 1 & 18 \\ 19 & 76 \end{matrix}$	$\frac{1^2}{19+18} \approx .027$	$> .20$
$\begin{matrix} 2 & 17 \\ 18 & 77 \end{matrix}$	$\frac{1^2}{18+17} \approx .029$	$> .20$
$\vdots$	$\vdots$	$\vdots$
$\begin{matrix} 19 & 0 \\ 1 & 94 \end{matrix}$	$\frac{1^2}{1+0} \approx 1.000$	$> .20$

And suppose Table 10.8.1 itself were a little different, say $\begin{matrix} 19 & 30 \\ 95 & 84 \end{matrix}$. What is your intuition about Older vs. Younger in this case? Here are some possible values for Table 10.8.2 that would fit this new Table 10.8.1. (Remember that Table 10.8.1 is misleading. Table 10.8.2 more clearly communicates what is going on.)

Table 10.8.2	$\chi_s^2 =$	P
$\begin{matrix} 0 & 19 \\ 30 & 65 \end{matrix}$	$\frac{11^2}{30+19} \approx 2.469$	$.10 < P < .20$
$\begin{matrix} 1 & 18 \\ 29 & 66 \end{matrix}$	$\frac{11^2}{29+18} \approx 2.574$	$.10 < P < .20$
$\begin{matrix} 2 & 17 \\ 28 & 67 \end{matrix}$	$\frac{11^2}{28+17} \approx 2.689$	$.10 < P < .20$
$\vdots$	$\vdots$	$\vdots$
$\begin{matrix} 19 & 0 \\ 11 & 84 \end{matrix}$	$\frac{11^2}{11+0} = 11$	$.0001 < P < .001$