Math 150, Spring 2023
Jo Hardin
WU # 6
Thursday, 2/2/2023

Your Name: ________________________________

Names of people you worked with: ________________________________

Task: You remember from somewhere (maybe just this week) that we can approximate the distribution of the sample ln odds ratio using:

$$\ln(\widehat{OR}) \overset{\text{approx}}{\sim} N\left(\ln(OR), \sqrt{\frac{1}{n_1 \hat{p}_1(1-\hat{p}_1)} + \frac{1}{n_2 \hat{p}_2(1-\hat{p}_2)}} \right)$$

Which leads you to the following 95% CI for the true $\ln(OR)$:

$$\left(\ln(\widehat{OR}) - 1.96 \cdot \sqrt{\frac{1}{n_1 \hat{p}_1(1-\hat{p}_1)} + \frac{1}{n_2 \hat{p}_2(1-\hat{p}_2)}}, \ln(\widehat{OR}) + 1.96 \cdot \sqrt{\frac{1}{n_1 \hat{p}_1(1-\hat{p}_1)} + \frac{1}{n_2 \hat{p}_2(1-\hat{p}_2)}} \right)$$

Filling in the pieces, the above 95% CI for the true value of $\ln(OR)$ (the parameter) is (2.71, 4.02).

Argue that $(e^{2.71}, e^{4.02})$ is a 95% confidence interval for the true value of OR by making the following argument (show your work):

1. If you happen to have gotten one of the 95% of samples that lead to a CI which captures the true value, then the first CI capturing the true $\ln(OR)$ implies that the second interval **must** capture the true OR.

2. If you were unlucky and happened to have gotten one of the 5% of samples which do not lead to a capture of the true value, then the first CI will not capture $\ln(OR)$ which implies that the second interval will **not** capture the true OR.

3. Thus, if the first interval (the one for the $\ln(OR)$) has a capture rate of 95%, then the second interval (where the endpoints are exponentiated) **must** also have a capture rate for OR of 95%.

Solution:

1. If I happen to have gotten a sample with endpoints that capture the true $\ln(OR)$, then:

$$\begin{aligned} 2.71 \leq \ln(OR) &\leq 4.02 \\ e^{2.71} \leq e^{\ln(OR)} &\leq e^{4.02} \\ e^{2.71} \leq OR &\leq e^{4.02} \end{aligned}$$

   Therefore if $\ln(OR) \in (2.71, 4.02)$ then $OR \in (e^{2.71}, e^{4.02})$.

2. Without loss of generality, let's say the interval is completely below $\ln(OR)$ (same argument holds if the interval is completely above $\ln(OR)$).

$$\begin{aligned} 2.71 < 4.02 &\leq \ln(OR) \\ e^{2.71} < e^{4.02} &\leq e^{\ln(OR)} \\ e^{2.71} < e^{4.02} &\leq OR \end{aligned}$$

   Therefore if $\ln(OR) \notin (2.71, 4.02)$ then $OR \notin (e^{2.71}, e^{4.02})$.

3. A 95% confidence interval describes a procedure by which a range of values is calculated. If the process captures the true parameter in 95% of samples and not in 5% of samples, then it is a valid 95% confidence interval. As described above, the process by which we exponentiate the endpoints creates a valid 95% confidence interval.