We consider binary responses, that is, responses that can be yes (1) or no (0).

If $Y$ is a binary response variable, let $\pi$ denote $E(Y)$, the probability of a yes. Then $\text{Var}(Y) = \pi (1 - \pi)$.

Let’s assume we take a SRS of size $n$ and estimate $\pi$ using $\hat{\pi}$ = the proportion of 1s.

Then simple calculations give $E(\hat{\pi}) = \pi$ and $\text{Var}(\hat{\pi}) = \pi (1 - \pi)/n$.

Furthermore, $\hat{\pi}$ has approximately a normal distribution for large enough $n$.

We can use these facts to draw inferences about $\pi$ (confidence intervals and hypothesis tests).
Difference between two proportions

We may be interested in comparing proportions $\pi_1$ and $\pi_2$ from two populations. One way to compare is by estimating the difference $\pi_1 - \pi_2$ based on independent samples of size $n_1$ and $n_2$ that give $\hat{\pi}_1$ and $\hat{\pi}_2$.

Because variances of independent random variables like $\hat{\pi}_2$ and $-\hat{\pi}_1$ add, the variance of $\hat{\pi}_2 - \hat{\pi}_1$ is

$$\frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2}$$

We obtain $\text{SE}(\hat{\pi}_2 - \hat{\pi}_1)$ by replacing the $\pi$’s by $\hat{\pi}$’s and taking the square root:

$$\text{SE}(\hat{\pi}_2 - \hat{\pi}_1) = \sqrt{\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}}$$

If we’re testing a null hypothesis, usually it’s that $\pi_1 = \pi_2$, in which case we replace all the $\hat{\pi}_1$ and $\hat{\pi}_2$ with $\hat{\pi}_c$, where $c$ stands for combined, to get

$$\text{SE}_0(\hat{\pi}_2 - \hat{\pi}_1) = \sqrt{\frac{\hat{\pi}_c(1 - \hat{\pi}_c)}{n_1} + \frac{\hat{\pi}_c(1 - \hat{\pi}_c)}{n_2}}$$

In any case, we use the approximate normality of $\hat{\pi}_2 - \hat{\pi}_1$ for large enough $n_1$ and $n_2$ to produce confidence intervals and tests.
From proportions to odds

**Definition:** If a population proportion is \( \pi \), then the odds are \( \omega = \pi / (1 - \pi) \).
Similarly, if the sample proportion is \( \hat{\pi} \), then the sample odds are \( \hat{\omega} = \hat{\pi} / (1 - \hat{\pi}) \).

Note that odds are not defined if \( \pi = 0 \) or \( \pi = 1 \). Also, odds can be any value greater than zero (unlike a proportion, which must be \( \leq 1 \)).

To convert odds to a proportion, use \( \pi = \omega / (1 + \omega) \).

Consider now the possibility that we have two populations of interest and we wish to compare their proportions. We’ve discussed doing this by considering \( \pi_2 - \pi_1 \), but does this really make sense? (Not always!)

Instead, let’s consider the ratio of odds \( \omega_2 / \omega_1 \). This **odds ratio** should be 1 if \( \pi_2 = \pi_1 \), greater than 1 if \( \pi_2 > \pi_1 \), and less than 1 if \( \pi_2 < \pi_1 \).

To perform inference about the odds ratio \( \omega_2 / \omega_1 \), we can use the following facts:

1. \( \log(\hat{\omega}_2 / \hat{\omega}_1) \) has approximate mean \( \log(\omega_2 / \omega_1) \).
2. \( \log(\hat{\omega}_2 / \hat{\omega}_1) \) has approximate variance \( 1 / [n_1 \pi_1 (1 - \pi_1)] + 1 / [n_2 \pi_2 (1 - \pi_2)] \).
3. \( \log(\hat{\omega}_2 / \hat{\omega}_1) \) is approximately normal for large \( n_1, n_2 \).