We have the following key definition:

A $(1-\alpha)\cdot 100\%$ confidence interval (CI) for p based on $\hat{p}$ is given by

$$\hat{p} - a\mathsf{SE}(\hat{p}) \leq p \leq \hat{p} + a\mathsf{SE}(\hat{p}),$$

where a is related to α by the standard normal distribution $1-\alpha = P(-a < Z < a)$ (or $P(Z > a) = \alpha/2$); and where $\mathsf{SE}(\hat{p})$, called the *standard error* of $\hat{p}$, is given by

$$\mathsf{SE}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

The quantity $a\mathsf{SE}(\hat{p})$ is known as the margin of error (MOE), and is commonly expressed in percentage points after multiplying by 100.

The value $(1-\alpha)\cdot 100\%$ is the confidence level.

The key assumption is that the binomial model applies to the random variable $X = \hat{p}n$. If the sampling is done at random with replacement from the population of interest this will be the case.

Excercise 1. A census bureau report from 2004 says that 13.2% of Americans live below the poverty line. As 2004 is not a census year, this is based on a sample. Suppose the sample was a random sample of 50,000 Americans. Compute a 99% confidence interval for p. Does your interval contain 13 percent?

Excercise 2. What percent of M&M's are brown? We consider a bag of M&M's to be a random sample from all M&M's produced. In a large bag of milk chocolate M&M's of 250 candies, 26% were brown. Find a 95% confidence interval for the population proportion. Does it include the stated proportion of 30%?

Excercise 3. Do Peanut M&Ms have a different proportion of brown candies? A a really large bag with 1,000 candies had 22% brown. Assume this sample of all candies can be considered a random sample. Find a 90% confidence interval for the population parameter. Does it include 30%? How about 20% (the stated proportion)?

Excercise 4. Baseball fans like to summarize a players performance by their batting average (which we call a sample proportion $\hat{p}$). Suppose David Wright bats .300 over his first 100 at bats. Find a 95% CI for his proportion of future successes assuming this performance is treated as a random sample.

How would your answer change, if this were an average over 400 atbats?

Excercise 5. We say in class that for a given margin of error (MOE) we could find a sample size that would produce a confidence interval with a margin or error that big or smaller. The formula for this came from solving

$$\text{MOE} = a\mathsf{SE}(\hat{p})$$

and was

$$n \leq \frac{1}{4}\left(\frac{a}{\text{MOE}}\right)^2.$$

How big a sample would be needed to produce CIs with margin of errors no bigger than 5 percentage points? 2 percentage points? 0.5 percentage points?

Excercise 6. An assumption is made that X, as defined above, is binomial and n is large enough so that X can be treated as normal. This isn't always the case as it requires a condition like $np(1-p) > 5$.

Suppose we are trying to figure out what percentage of the population will buy a size 9 running shoe. As both the size and style restrict the proportion we expect it to be quite small. Perhaps no more than 0.5 percentage points. How big must n be so that a 90% CI for p has a margin of error of no more than 1 percentage point, and the normal approximation is valid?

Excercise 7. Suppose that in order to measure the health of coral on a reef a grid is constructed that takes measurements every 5 feet. At each spot on the grid a person decides if the coral is healthy or not healthy. If the grid contains 1,000 sites, and 22 are found not healthy, find a 95% CI for the proportion of unhealthy areas of coral in total.

Excercise 8. Why would the coral researchers be worried about the grid sites being too close together? (Hint: it affects one of the assumptions made about $X = n\hat{p}$.)