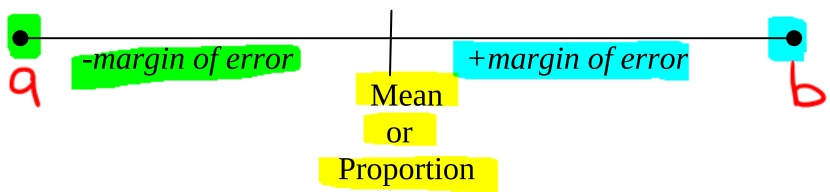
Section 7.2
Confidence Interval for a Proportion

Before any inferences can be made about a proportion, certain conditions must be satisfied:

1. The sample must be an SRS from the population of interest.
2. The population must be at least 10 times the size of the sample.
3. The number of successes must be $n\hat{p} \geq 10$, and the number of failures must be $n(1-\hat{p}) \geq 10$.

The sample statistic for a population proportion is $\hat{p}$, so based on the formula for a CI, we have $\hat{p} \pm \text{margin of error}$.

---------------Width of the Confidence Interval-------------------



$\Rightarrow [a, b]$

So if you have a confidence interval with width 0.22, then what is the margin of error?

$\frac{.22}{2} = \boxed{.11 = M.E.}$

How do we find the margin of error if it is not given to us? z^*

The margin of error is computed using the **critical value** (a number based on our level of confidence) and the **standard deviation** (**standard error**) of the statistic.

Critical Value: When the distribution is assumed to be normal, our critical value is found using qnorm in R. If it is not normal, we will use the t distribution (discussed later).

The formula is: $z^* = \text{qnorm}\left(\frac{1+\text{confidence level}}{2}\right)$

Standard Deviation/Error: When working with proportions, the standard deviation of the statistic $\hat{p}$ is $\sqrt{\frac{p(1-p)}{n}}$. Since p is unknown, we will use the standard error. To calculate the standard error of $\hat{p}$, use the formula $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

So, $\hat{p} \pm \text{margin of error} = \hat{p} \pm z^* \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Facts about Confidence Intervals

- For smaller n, the confidence interval becomes wider.
- For larger n, the confidence interval becomes narrower.
- Increasing the variance or increasing the confidence level will increase the width of the confidence interval and vice versa.
- Decreasing the variance or decreasing the confidence level will decrease the confidence interval and vice versa.

Example 1: In the first eight games of this year's basketball season, Lenny made 25 free throws in 40 attempts.
a. What is $\hat{p}$, Lenny's sample proportion of successes?

$$\hat{p} = \frac{25}{40} = 0.625$$

Let's quickly check the conditions:
1. We have an SRS. Even if it's not stated in the problem, we can assume it's an SRS.
2. The population is at least 10 times the sample (assume he won't get hurt and make many more free throws).
3. $n\hat{p} = 40 \cdot 25/40 = 25 \geq 10$ and $n(1-\hat{p}) = 40(1 - 25/40) = 15 \geq 10$

b. Find and interpret the 90% confidence interval for Lenny's proportion of free-throw success.

critical value

$z^* = \text{qnorm}\left(\frac{1 + \text{confidence level}}{2}\right)$ $= \text{qnorm}\left(\frac{1+.9}{2}\right) = 1.645$

Then $\hat{p} \pm z^* \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $= 0.625 \pm 1.645\sqrt{\frac{0.625 * (1-0.625)}{40}}$

Confidence Interval: [0.4991, 0.7509]

Interpretation:
90% confident that the true proportion of made free throws is between 49.91% and 75.09%

Example 2: Mars Inc. claims that they produce M&Ms with the following distributions:

Claim

Brown	30%	Red	20%	Yellow	20%
Orange	10%	Green	10%	Blue	10%

A bag of M&Ms was randomly selected from the grocery store shelf, and the color counts were:

One sample bag

Brown	22	Red	22	Yellow	22
Orange	12	Green	15	Blue	15

total = 108

Find the 95% confidence interval for the proportion of yellow M&Ms in that bag.

a. What is the proportion of yellow M&Ms in this bag?

$$\hat{p} = \frac{22}{108} = \frac{11}{54}$$

b. Find and interpret the 95% confidence interval for the proportion of yellow M&Ms in this bag.

$$z^* = \text{qnorm}\left(\frac{1+\text{confidence level}}{2}\right) = \text{qnorm}\left(\frac{1+.95}{2}\right) = 1.96$$

Then $\hat{p} \pm z^* \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{11}{54} \pm 1.96\sqrt{\frac{11/54 * (1 - 11/54)}{108}}$

95% Confidence Interval: [0.1277, 0.2797]

95% confident that true proportion of yellow M&M lies between 12.77% and 27.97%

n

Sometimes we are asked to find the minimum sample size needed to produce a particular margin of error given a certain confidence level. When working with a one-sample proportion, we can use the formula: Maximum ME $\geq z^* \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ In these problems, we'll be looking for n so we could simply plug into: $n \geq \frac{\hat{p}(1-\hat{p})}{\left(ME/z^*\right)^2}$

If $\hat{p}$ is unknown, use an estimate of p. If p is unknown, just assume 50, 50, so use $p = 0.5$.

Example 3: It is believed that 35% of all voters favor a particular candidate. How large of a simple random sample is required so that the margin of error of the estimate of the percentage of all voters in favor is no more than 3% at the 95% confidence level? n - ?

ME. C.L.

$p = .35$

Max ME = 0.03

$$z^* = \text{qnorm}\left(\frac{1+\text{confidence level}}{2}\right) = \text{qnorm}\left(\frac{1+.95}{2}\right) = 1.959964 \approx 1.96$$

$$n \geq \frac{\hat{p}(1-\hat{p})}{\left(ME/z^*\right)^2} \qquad n \geq \frac{0.35(1-.35)}{(0.03/1.96)^2}$$

$n \geq 971.0711$

Round up!

$\boxed{n = 972}$

Example 4: An oil company is interested in estimating the true proportion of female truck drivers based in five southern states. A statistician hired by the oil company must determine the sample size needed in order to make the estimate accurate to within 1% of the true proportion with 89% confidence. What is the minimum number of truck drivers that the statistician should sample in these southern states in order to achieve the desired accuracy?

$p =$ 0.5

ME = 0.01

$$z^* = \text{qnorm}\left(\frac{1+\text{confidence level}}{2}\right) = \text{qnorm}\left(\frac{1+.89}{2}\right) = 1.598$$

$$n \geq \frac{\hat{p}(1-\hat{p})}{(ME/z^*)^2} \qquad n \geq \frac{0.5(1-0.5)}{(0.01/1.598)^2}$$

$$n \geq 6384.01$$

$$n = 6385$$