Inference for a Single Proportion

Nate Wells

Math 141, 4/8/22

Outline

In this lecture, we will. . .

Outline

In this lecture, we will. . .

- Use theory to find the standard error for one sample proportions
- Calculate confidence intervals and perform hypothesis tests for proportions using the theory-based method
- Investigate the results of the lacroix taste-test

Section 1

Inference for a Single Proportion

The Sampling Distribution for Sample Proportion

- Consider a population variable that takes only two levels: A and B. Let p be the proportion of A's in the population.

The Sampling Distribution for Sample Proportion

- Consider a population variable that takes only two levels: A and B. Let p be the proportion of A's in the population.
- Suppose we randomly choose a single observation from a population, and define a random variable X to be 1 if the observation is an A, and 0 if it is a B.

The Sampling Distribution for Sample Proportion

- Consider a population variable that takes only two levels: A and B. Let p be the proportion of A's in the population.
- Suppose we randomly choose a single observation from a population, and define a random variable X to be 1 if the observation is an A, and 0 if it is a B.
  - The *mean* of X is p, and the *standard deviation* of X is $\sqrt{p(1-p)}$ (Why?)

The Sampling Distribution for Sample Proportion

- Consider a population variable that takes only two levels: A and B. Let p be the proportion of A's in the population.
- Suppose we randomly choose a single observation from a population, and define a random variable X to be 1 if the observation is an A, and 0 if it is a B.
    - The *mean* of X is p, and the *standard deviation* of X is $\sqrt{p(1-p)}$ (Why?)
- If we instead take an SRS of size n from the population, we can view the sample proportion $\hat{p}$ as a sample mean:

The Sampling Distribution for Sample Proportion

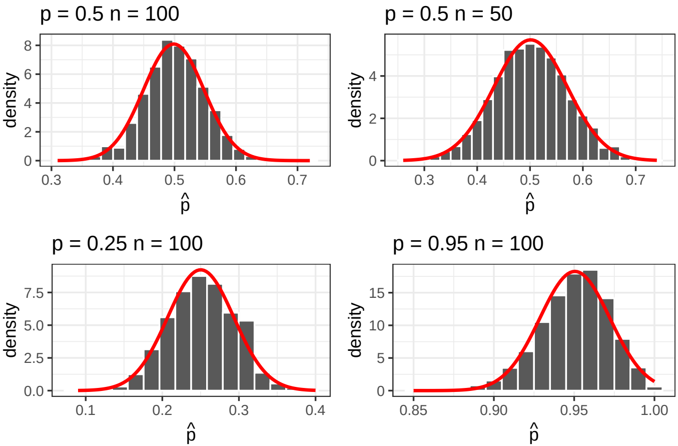
- Consider a population variable that takes only two levels: A and B. Let p be the proportion of A's in the population.
- Suppose we randomly choose a single observation from a population, and define a random variable X to be 1 if the observation is an A, and 0 if it is a B.
  - The *mean* of X is p, and the *standard deviation* of X is $\sqrt{p(1-p)}$ (Why?)
- If we instead take an SRS of size n from the population, we can view the sample proportion $\hat{p}$ as a sample mean:
  - Suppose each person in the sample has their own binary variable X_i. Then the sum $X_1 + \cdots + X_n$ is the number of A's in the sample, and the mean of the X_i is the proportion of A's.

The Sampling Distribution for Sample Proportion

- Consider a population variable that takes only two levels: A and B. Let p be the proportion of A's in the population.
- Suppose we randomly choose a single observation from a population, and define a random variable X to be 1 if the observation is an A, and 0 if it is a B.
  - The *mean* of X is p, and the *standard deviation* of X is $\sqrt{p(1-p)}$ (Why?)
- If we instead take an SRS of size n from the population, we can view the sample proportion $\hat{p}$ as a sample mean:
  - Suppose each person in the sample has their own binary variable X_i. Then the sum $X_1 + \cdots + X_n$ is the number of A's in the sample, and the mean of the X_i is the proportion of A's.
- By the Central Limit Theorem, if n is large, then $\hat{p}$ is approximately Normal, with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$

Examples

- Below are the sampling distributions for $\hat{p}$ for a variety of values of p and n, along with the approximating Normal curve:



Theory vs Simulation Methods

We now have two ways of making confidence intervals / performing hypothesis tests for p:

- Previously, we. . .

Theory vs Simulation Methods

We now have two ways of making confidence intervals / performing hypothesis tests for p:

- Previously, we...
  - constructed confidence intervals by approximating the sampling distribution through bootstrapping, computing quantiles in the bootstrap distribution to get confidence interval bounds.

Theory vs Simulation Methods

We now have two ways of making confidence intervals / performing hypothesis tests for p:

- Previously, we. . .
  - constructed confidence intervals by approximating the sampling distribution through bootstrapping, computing quantiles in the bootstrap distribution to get confidence interval bounds.
  - Performed hypothesis tests by approximating the null distribution through permutation/simulation, calculating p-values as proportion of simulated null statistics more extreme than the observed statistic.

Theory vs Simulation Methods

We now have two ways of making confidence intervals / performing hypothesis tests for p:

- Previously, we...
  - constructed confidence intervals by approximating the sampling distribution through bootstrapping, computing quantiles in the bootstrap distribution to get confidence interval bounds.
  - Performed hypothesis tests by approximating the null distribution through permutation/simulation, calculating p-values as proportion of simulated null statistics more extreme than the observed statistic.
- But now, using the Central Limit Theorem, we can...

Theory vs Simulation Methods

We now have two ways of making confidence intervals / performing hypothesis tests for p:

- Previously, we...
  - constructed confidence intervals by approximating the sampling distribution through bootstrapping, computing quantiles in the bootstrap distribution to get confidence interval bounds.
  - Performed hypothesis tests by approximating the null distribution through permutation/simulation, calculating p-values as proportion of simulated null statistics more extreme than the observed statistic.
- But now, using the Central Limit Theorem, we can...
  - Construct confidence intervals using the quantiles of the Normal distribution, which can be transformed into bounds for the confidence intervals.

Theory vs Simulation Methods

We now have two ways of making confidence intervals / performing hypothesis tests for p:

- Previously, we...
  - constructed confidence intervals by approximating the sampling distribution through bootstrapping, computing quantiles in the bootstrap distribution to get confidence interval bounds.
  - Performed hypothesis tests by approximating the null distribution through permutation/simulation, calculating p-values as proportion of simulated null statistics more extreme than the observed statistic.
- But now, using the Central Limit Theorem, we can...
  - Construct confidence intervals using the quantiles of the Normal distribution, which can be transformed into bounds for the confidence intervals.
  - Perform hypothesis tests by obtaining p-values from probabilities in the Normal distribution.

Theory vs Simulation Methods

We now have two ways of making confidence intervals / performing hypothesis tests for p:

- Previously, we...
  - constructed confidence intervals by approximating the sampling distribution through bootstrapping, computing quantiles in the bootstrap distribution to get confidence interval bounds.
  - Performed hypothesis tests by approximating the null distribution through permutation/simulation, calculating p-values as proportion of simulated null statistics more extreme than the observed statistic.
- But now, using the Central Limit Theorem, we can...
  - Construct confidence intervals using the quantiles of the Normal distribution, which can be transformed into bounds for the confidence intervals.
  - Perform hypothesis tests by obtaining p-values from probabilities in the Normal distribution.
- Why learn two methods?

Theory vs Simulation Methods

We now have two ways of making confidence intervals / performing hypothesis tests for p:

- Previously, we...
  - constructed confidence intervals by approximating the sampling distribution through bootstrapping, computing quantiles in the bootstrap distribution to get confidence interval bounds.
  - Performed hypothesis tests by approximating the null distribution through permutation/simulation, calculating p-values as proportion of simulated null statistics more extreme than the observed statistic.
- But now, using the Central Limit Theorem, we can...
  - Construct confidence intervals using the quantiles of the Normal distribution, which can be transformed into bounds for the confidence intervals.
  - Perform hypothesis tests by obtaining p-values from probabilities in the Normal distribution.
- Why learn two methods?
  - The Theory-based method works best when modeling assumptions are true
  - Simulation-based methods can perform well in a variety of circumstances, but sometimes lack precision

Section 2

Hypothesis Testing Procedures

z-Scores

- The **z-score** for a test statistic x with standard error SE and mean μ under the Null hypothesis is

$$z = \frac{x - \mu}{SE}$$

z-Scores

- The **z-score** for a test statistic x with standard error SE and mean μ under the Null hypothesis is

$$z = \frac{x - \mu}{SE}$$

- Suppose X is approximately Normal with mean μ and standard deviation σ. Then

$$Z = \frac{X - \mu}{\sigma}$$

is approximately standard Normal (mean of 0, st. dev. of 1).

z-Scores

- The **z-score** for a test statistic x with standard error SE and mean μ under the Null hypothesis is

$$z = \frac{x - \mu}{SE}$$

- Suppose X is approximately Normal with mean μ and standard deviation σ. Then

$$Z = \frac{X - \mu}{\sigma}$$

  is approximately standard Normal (mean of 0, st. dev. of 1).

- By location-scale invariance,

$$P(X > x) = P\left(Z > \frac{x - \mu}{\sigma}\right)$$

z-Scores

- The **z-score** for a test statistic x with standard error SE and mean μ under the Null hypothesis is

$$z = \frac{x - \mu}{SE}$$

- Suppose X is approximately Normal with mean μ and standard deviation σ. Then

$$Z = \frac{X - \mu}{\sigma}$$

  is approximately standard Normal (mean of 0, st. dev. of 1).

- By location-scale invariance,

$$P(X > x) = P\left(Z > \frac{x - \mu}{\sigma}\right)$$

- If we want to compute a P-Value for test statistic x, we can instead compute a P-value for its z-score z:

$$\begin{array}{lcll} \text{P-value} & = & P(Z > z) & \text{if } H_a \text{ is one-sided right} \\ \text{P-value} & = & P(Z < z) & \text{if } H_a \text{ is one-sided left} \\ \text{P-value} & = & 2 \cdot P(Z > |z|) & \text{if } H_a \text{ is two-sided} \end{array}$$

Hypothesis Tests

By the central limit theorem, if $H_0 : p = p_0$ is true, then for large n, $\hat{p}$ is approximately Normal, with the standard error

$$SE(\hat{p}) = \sqrt{\frac{p_0(1-p_0)}{n}}$$

Hypothesis Tests

By the central limit theorem, if $H_0 : p = p_0$ is true, then for large n, $\hat{p}$ is approximately Normal, with the standard error

$$SE(\hat{p}) = \sqrt{\frac{p_0(1-p_0)}{n}}$$

Theorem

To test $H_0 : p = p_0$ against $H_a : p \neq p_0$ (or the one-sided alternative) we use the standardized test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

If n is large enough so that both $n\hat{p}$ and $n(1-\hat{p})$ are at least 10, then the p-value for the test is computed using the standard Normal distribution.

Taste Test

- On Wednesday, Math 141 students participated in an experiment to determine whether the typical Reed student can distinguish between two different flavors of carbonated water.

Taste Test

- On Wednesday, Math 141 students participated in an experiment to determine whether the typical Reed student can distinguish between two different flavors of carbonated water.
    - Each student was provided 3 cups; 2 of the cups had the same flavor, and the other cup had a different flavor. Students were asked to identify the cup that was different.

Taste Test

- On Wednesday, Math 141 students participated in an experiment to determine whether the typical Reed student can distinguish between two different flavors of carbonated water.
    - Each student was provided 3 cups; 2 of the cups had the same flavor, and the other cup had a different flavor. Students were asked to identify the cup that was different.
- *Null Hypothesis:* flavors cannot be distinguished.
- *Alternative Hypothesis:* flavors can be distinguished.

Taste Test

- On Wednesday, Math 141 students participated in an experiment to determine whether the typical Reed student can distinguish between two different flavors of carbonated water.
  - Each student was provided 3 cups; 2 of the cups had the same flavor, and the other cup had a different flavor. Students were asked to identify the cup that was different.
- *Null Hypothesis:* flavors cannot be distinguished.
- *Alternative Hypothesis:* flavors can be distinguished.
- Let p denote the true proportion of the population who can correctly identify the cup that is different.

Taste Test

- On Wednesday, Math 141 students participated in an experiment to determine whether the typical Reed student can distinguish between two different flavors of carbonated water.
  - Each student was provided 3 cups; 2 of the cups had the same flavor, and the other cup had a different flavor. Students were asked to identify the cup that was different.
- *Null Hypothesis:* flavors cannot be distinguished.
- *Alternative Hypothesis:* flavors can be distinguished.
- Let p denote the true proportion of the population who can correctly identify the cup that is different.
  - If H_0 is true, what is the corresponding value of p?

Taste Test

- On Wednesday, Math 141 students participated in an experiment to determine whether the typical Reed student can distinguish between two different flavors of carbonated water.
  - Each student was provided 3 cups; 2 of the cups had the same flavor, and the other cup had a different flavor. Students were asked to identify the cup that was different.
- *Null Hypothesis:* flavors cannot be distinguished.
- *Alternative Hypothesis:* flavors can be distinguished.
- Let p denote the true proportion of the population who can correctly identify the cup that is different.
  - If H_0 is true, what is the corresponding value of p?
  - If H_a is true, how does the true value of p compare to the null value?

Taste Test

- On Wednesday, Math 141 students participated in an experiment to determine whether the typical Reed student can distinguish between two different flavors of carbonated water.
  - Each student was provided 3 cups; 2 of the cups had the same flavor, and the other cup had a different flavor. Students were asked to identify the cup that was different.
- *Null Hypothesis:* flavors cannot be distinguished.
- *Alternative Hypothesis:* flavors can be distinguished.
- Let p denote the true proportion of the population who can correctly identify the cup that is different.
  - If H_0 is true, what is the corresponding value of p?
  - If H_a is true, how does the true value of p compare to the null value?

$$H_0 : p = \frac{1}{3} \qquad H_a : p > \frac{1}{3}$$

Taste Test Results

- Of 59 students who performed experiment, 29 students correctly identified the different cup (blue cup).
  - Our sample statistic is $\hat{p} = \frac{29}{59} = 0.49$

Taste Test Results

- Of 59 students who performed experiment, 29 students correctly identified the different cup (blue cup).
    - Our sample statistic is $\hat{p} = \frac{29}{59} = 0.49$
- If H_0 is true, the standard error for $\hat{p}$ is

$$SE(\hat{p}) = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.33(1-0.33)}{59}} = 0.061$$

Taste Test Results

- Of 59 students who performed experiment, 29 students correctly identified the different cup (blue cup).
    - Our sample statistic is $\hat{p} = \frac{29}{59} = 0.49$
- If H_0 is true, the standard error for $\hat{p}$ is

$$SE(\hat{p}) = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.33(1-0.33)}{59}} = 0.061$$

- The z-score for $\hat{p}$ is therefore

$$z = \frac{\hat{p} - p_0}{SE} = \frac{0.49 - 0.33}{0.061} = 2.578$$

Taste Test Results

- Of 59 students who performed experiment, 29 students correctly identified the different cup (blue cup).
  - Our sample statistic is $\hat{p} = \frac{29}{59} = 0.49$
- If H_0 is true, the standard error for $\hat{p}$ is

$$SE(\hat{p}) = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.33(1-0.33)}{59}} = 0.061$$

- The z-score for $\hat{p}$ is therefore

$$z = \frac{\hat{p} - p_0}{SE} = \frac{0.49 - 0.33}{0.061} = 2.578$$

  - That is, the observed $\hat{p}$ was 2.5 standard errors above the mean.
  - This seems unlikely to occur, if the null hypothesis were true (remember, 95% of all observations are within 2 standard errors of mean)

Calculate P-Value

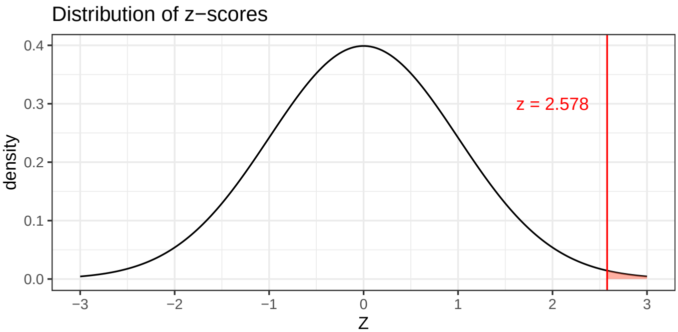
- If H_0 is true, the z-score should be Normally distributed, with mean 0 and st. dev.

Calculate P-Value

- If H_0 is true, the z-score should be Normally distributed, with mean 0 and st. dev.
  - The p-value is the probability that a standard Normal variable is larger than $z = 2.578$

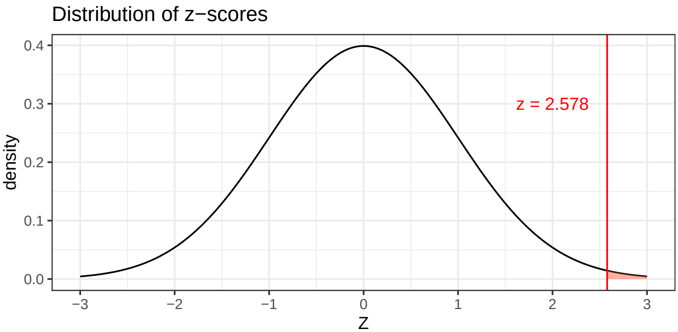
Calculate P-Value

- If H_0 is true, the z-score should be Normally distributed, with mean 0 and st. dev.
  - The p-value is the probability that a standard Normal variable is larger than $z = 2.578$



Calculate P-Value

- If H_0 is true, the z-score should be Normally distributed, with mean 0 and st. dev.
  - The p-value is the probability that a standard Normal variable is larger than $z = 2.578$



- The exact p-value is

```
1-pnorm(q=2.578, mean = 0, sd = 1)
```

```
## [1] 0.0049687
```

Conclusions

- If the two types of carbonated water were indistinguishable, we would expect that approximately 33% of students would identify the correct cup due by random guessing.

Conclusions

- If the two types of carbonated water were indistinguishable, we would expect that approximately 33% of students would identify the correct cup due by random guessing.
  - Moreover, we would observe a sample proportion greater than or equal to 49% only 0.5% of the time ($\text{p-value} = 0.0049687$)

Conclusions

- If the two types of carbonated water were indistinguishable, we would expect that approximately 33% of students would identify the correct cup due by random guessing.
  - Moreover, we would observe a sample proportion greater than or equal to 49% only 0.5% of the time ($\text{p-value} = 0.0049687$)
- At a liberal significance level of $\alpha = 0.1$, since $\text{p-value} < \alpha$, we reject the null hypothesis in favor of the alternative.

Conclusions

- If the two types of carbonated water were indistinguishable, we would expect that approximately 33% of students would identify the correct cup due by random guessing.
  - Moreover, we would observe a sample proportion greater than or equal to 49% only 0.5% of the time ($\text{p-value} = 0.0049687$)
- At a liberal significance level of $\alpha = 0.1$, since $\text{p-value} < \alpha$, we reject the null hypothesis in favor of the alternative.
  - This experiment provides evidence that the two flavors are indeed distinguishable

Conclusions

- If the two types of carbonated water were indistinguishable, we would expect that approximately 33% of students would identify the correct cup due by random guessing.
  - Moreover, we would observe a sample proportion greater than or equal to 49% only 0.5% of the time ($\text{p-value} = 0.0049687$)
- At a liberal significance level of $\alpha = 0.1$, since $\text{p-value} < \alpha$, we reject the null hypothesis in favor of the alternative.
  - This experiment provides evidence that the two flavors are indeed distinguishable
- How does this compare to the simulation results?

Conclusions

- If the two types of carbonated water were indistinguishable, we would expect that approximately 33% of students would identify the correct cup due by random guessing.
  - Moreover, we would observe a sample proportion greater than or equal to 49% only 0.5% of the time ($\text{p-value} = 0.0049687$)
- At a liberal significance level of $\alpha = 0.1$, since $\text{p-value} < \alpha$, we reject the null hypothesis in favor of the alternative.
  - This experiment provides evidence that the two flavors are indeed distinguishable
- How does this compare to the simulation results?

```
set.seed(48)
lacroix %>% specify(response = correct, success = "yes") %>%
  hypothesize(null = "point", p = 1/3) %>%
  generate(reps = 5000, type = "simulate") %>%
  calculate(stat = "prop") %>%
  get_p_value(obs_stat = .5, direction = "right")
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1  0.0038
```

Section 3

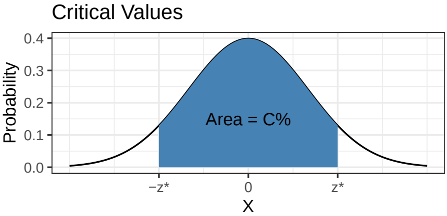
Confidence Intervals

Critical Values

- The **critical value** z^* for a $C\%$ confidence interval is the value so that $C\%$ of area is between $-z^*$ and z^* in the standard Normal distribution

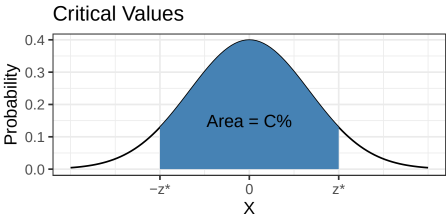
Critical Values

- The **critical value** z^* for a $C\%$ confidence interval is the value so that $C\%$ of area is between $-z^*$ and z^* in the standard Normal distribution



Critical Values

- The **critical value** z^* for a $C\%$ confidence interval is the value so that $C\%$ of area is between $-z^*$ and z^* in the standard Normal distribution



- For Normal distributions, approximately 95% of observations are within 2 standard deviations of the mean.
  - So the critical value for 95% confidence is approximately

$$z^* = 2 \qquad \text{(exact value is } z^* = 1.96\text{)}$$

Confidence Intervals

When a sample statistic is approximately Normally distribution, the $C\%$ confidence interval is

$$\text{statistic} \pm z^* \cdot SE$$

where z^* is the critical value for $C\%$ confidence and SE is the standard error for the statistic.

Confidence Intervals

When a sample statistic is approximately Normally distribution, the $C\%$ confidence interval is

$$\text{statistic} \pm z^* \cdot SE$$

where z^* is the critical value for $C\%$ confidence and SE is the standard error for the statistic.

- The standard error for a sample proportion $\hat{p}$ is $SE = \sqrt{\frac{p(1-p)}{n}}$. Since we don't know p, we estimate it in the SE formula with $\hat{p}$.

Confidence Intervals

When a sample statistic is approximately Normally distribution, the $C\%$ confidence interval is

$$\text{statistic} \pm z^* \cdot SE$$

where z^* is the critical value for $C\%$ confidence and SE is the standard error for the statistic.

- The standard error for a sample proportion $\hat{p}$ is $SE = \sqrt{\frac{p(1-p)}{n}}$. Since we don't know p, we estimate it in the SE formula with $\hat{p}$.

Theorem

Suppose an SRS of size n is collected from a population with parameter p. If n is large enough so that both $n\hat{p}$ and $n(1-\hat{p})$ are at least 10, then the confidence interval for p is

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Taste Test Continued

- Suppose we are interested in estimating the value of p, the proportion of the population who will correctly identify the different cup.

Taste Test Continued

- Suppose we are interested in estimating the value of p, the proportion of the population who will correctly identify the different cup.
  - Create a 90% confidence interval for this parameter.

Taste Test Continued

- Suppose we are interested in estimating the value of p, the proportion of the population who will correctly identify the different cup.
  - Create a 90% confidence interval for this parameter.
- As before, our sample statistic is $\hat{p} = \frac{29}{59}$.

Taste Test Continued

- Suppose we are interested in estimating the value of p, the proportion of the population who will correctly identify the different cup.
  - Create a 90% confidence interval for this parameter.
- As before, our sample statistic is $\hat{p} = \frac{29}{59}$.
- The critical value for a 90% confidence interval is the number z^* so that 90% area is between $-z^*$ and $z*$. It is the 0.95 **quantile**

Taste Test Continued

- Suppose we are interested in estimating the value of p, the proportion of the population who will correctly identify the different cup.
    - Create a 90% confidence interval for this parameter.
- As before, our sample statistic is $\hat{p} = \frac{29}{59}$.
- The critical value for a 90% confidence interval is the number z^* so that 90% area is between $-z^*$ and $z*$. It is the 0.95 **quantile**

```
qnorm(p = .95, mean = 0, sd = 1)
```

```
## [1] 1.644854
```

Taste Test Continued

- Suppose we are interested in estimating the value of p, the proportion of the population who will correctly identify the different cup.
  - Create a 90% confidence interval for this parameter.
- As before, our sample statistic is $\hat{p} = \frac{29}{59}$.
- The critical value for a 90% confidence interval is the number z^* so that 90% area is between $-z^*$ and $z*$. It is the 0.95 **quantile**

```
qnorm(p = .95, mean = 0, sd = 1)
```

```
## [1] 1.644854
```

- The standard error for $\hat{p}$ is

$$SE(\hat{p}) \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.49(1-0.49)}{59}} = 0.065$$

An Example

- The theory-based confidence interval takes the form

$$\hat{p} \pm z^* \cdot SE$$

An Example

- The theory-based confidence interval takes the form

$$\hat{p} \pm z^* \cdot SE$$

- In this case,

$$0.49 \pm 1.64 \cdot 0.065 \qquad \text{or} \qquad 0.49 \pm 0.1066$$

An Example

- The theory-based confidence interval takes the form
$$\hat{p} \pm z^* \cdot SE$$
- In this case,
$$0.49 \pm 1.64 \cdot 0.065 \qquad \text{or} \qquad 0.49 \pm 0.1066$$
- That is, a plausible range of values for p is 0.38 to 0.60, with confidence 90%.

An Example

- The theory-based confidence interval takes the form

$$\hat{p} \pm z^* \cdot SE$$

- In this case,

$$0.49 \pm 1.64 \cdot 0.065 \qquad \text{or} \qquad 0.49 \pm 0.1066$$

- That is, a plausible range of values for p is 0.38 to 0.60, with confidence 90%.
- How does this compare to the bootstrap method?

An Example

- The theory-based confidence interval takes the form
$$\hat{p} \pm z^* \cdot SE$$
- In this case,
$$0.49 \pm 1.64 \cdot 0.065 \qquad \text{or} \qquad 0.49 \pm 0.1066$$
- That is, a plausible range of values for p is 0.38 to 0.60, with confidence 90%.
- How does this compare to the bootstrap method?

```
set.seed(84)
lacroix %>% specify(response = correct, success = "yes") %>%
  generate(reps=5000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = .9, type = "percentile")
```

```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1    0.390    0.593
```

Reflections on Experiment Design

- *Reflection on Experiment Design:*

Reflections on Experiment Design

- *Reflection on Experiment Design:*
  - Why did we ask students to identify which of 3 cups was different, rather than giving 2 unmarked cups (1 lemon, 1 lime) and asking students to identify which is lime?

Reflections on Experiment Design

- *Reflection on Experiment Design:*
  - Why did we ask students to identify which of 3 cups was different, rather than giving 2 unmarked cups (1 lemon, 1 lime) and asking students to identify which is lime?
  - Why would observing $\hat{p} < 0.33$ be unlikely under both the null and alternative hypotheses?