Inference for Proportions

Prof. Wells

STA 209, 4/19/23

Outline

In this lecture, we will. . .

Outline

In this lecture, we will. . .

- Discuss the Central Limit Theorem and its role in statistics
- 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

Section 1

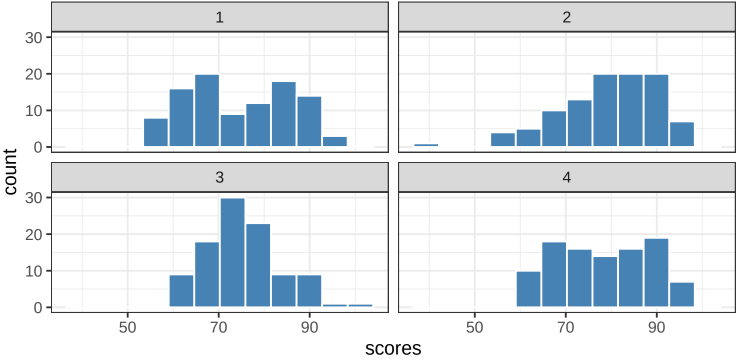
The Central Limit Theorem

Exam scores

Consider the following distributions for scores on a statistics exam for 4 classes of 100 students:

Exam scores

Consider the following distributions for scores on a statistics exam for 4 classes of 100 students:



Random Sample Means

Suppose we repeatedly take samples of 10 students from each class, and compute the average score $\bar{x}$ for each sample

Random Sample Means

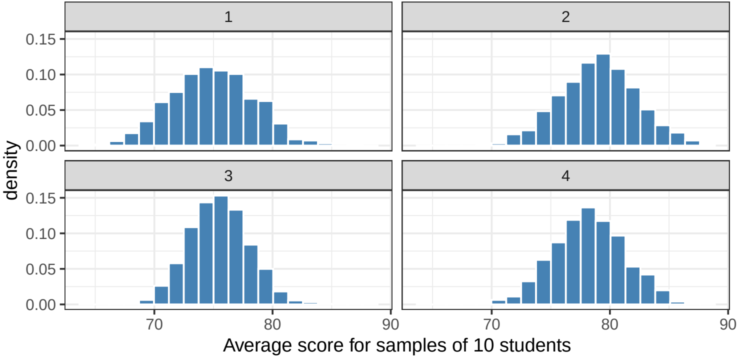
Suppose we repeatedly take samples of 10 students from each class, and compute the average score $\bar{x}$ for each sample

- What does the distribution of sample means $\bar{x}$ look like?

Random Sample Means

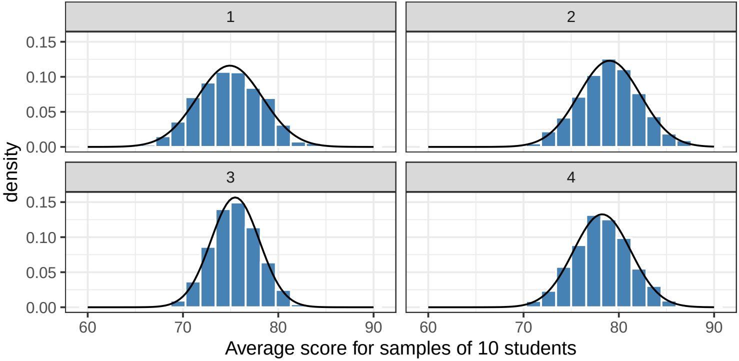
Suppose we repeatedly take samples of 10 students from each class, and compute the average score $\bar{x}$ for each sample

- What does the distribution of sample means $\bar{x}$ look like?



The Normal Distribution

- In the previous example, the sampling distribution for *each* class appeared approximately Normal, regardless of the shape of the population distribution.

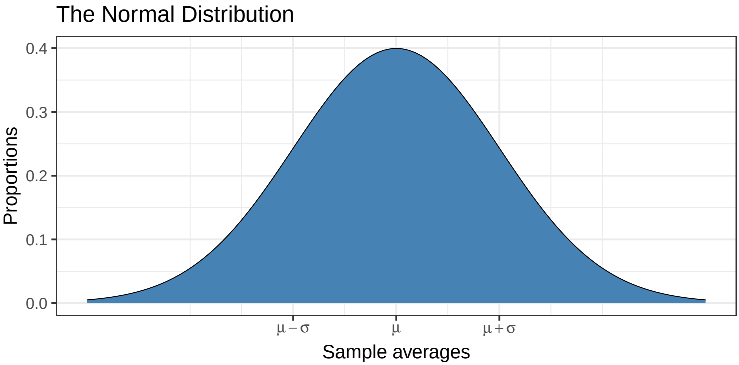


The Normal Distribution

- In the previous example, the sampling distribution for *each* class appeared approximately Normal, regardless of the shape of the population distribution.

The Normal Distribution

- In the previous example, the sampling distribution for *each* class appeared approximately Normal, regardless of the shape of the population distribution.

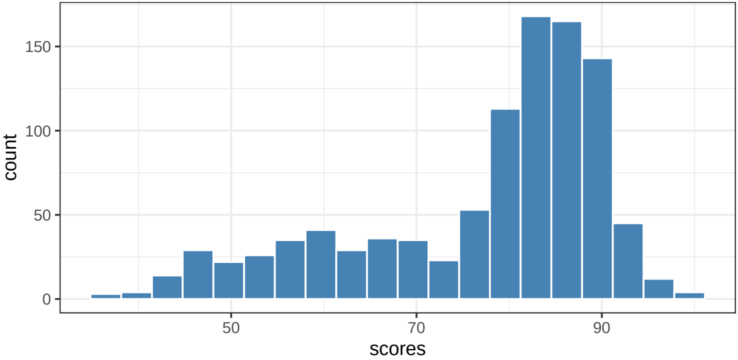


Effect of Sample Size

Suppose we have a class of 1000 students with the following score distribution

Effect of Sample Size

Suppose we have a class of 1000 students with the following score distribution

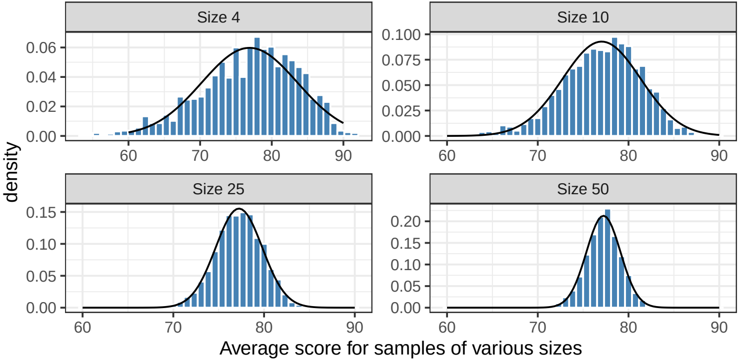


Effect of Sample Size II

What happens to the distribution of sample means as we increase the size of each sample (keeping the number of samples drawn constant)?

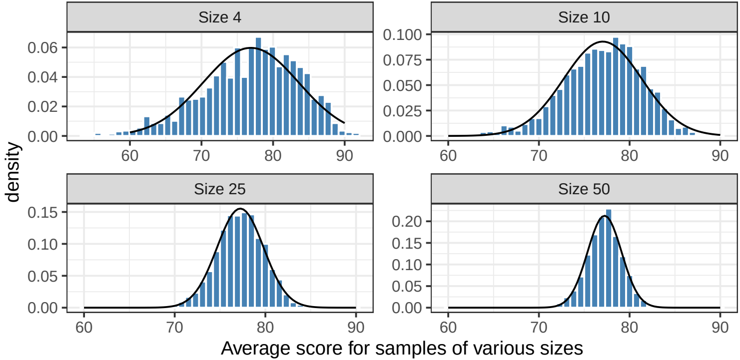
Effect of Sample Size II

What happens to the distribution of sample means as we increase the size of each sample (keeping the number of samples drawn constant)?



Effect of Sample Size II

What happens to the distribution of sample means as we increase the size of each sample (keeping the number of samples drawn constant)?



- As sample size increases, sampling distribution becomes **more** Normal, with **decreasing** variance

The Central Limit Theorem

Theorem

Suppose an SRS of size n is drawn from a population with mean μ and standard deviation σ. When n is large, the sample mean $\bar{x}$ is approximately Normally distributed, with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

The Central Limit Theorem

Theorem

Suppose an SRS of size n is drawn from a population with mean μ and standard deviation σ. When n is large, the sample mean $\bar{x}$ is approximately Normally distributed, with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

A proof of the CLT requires more advanced techniques in probability (STA 335/336). But intuitively. . .

The Central Limit Theorem

Theorem

Suppose an SRS of size n is drawn from a population with mean μ and standard deviation σ. When n is large, the sample mean $\bar{x}$ is approximately Normally distributed, with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

A proof of the CLT requires more advanced techniques in probability (STA 335/336). But intuitively. . .

- A sample mean is obtained by adding together INDEPENDENT values from the population.
- In order to get a very large or very small value, nearly ALL of the independent values need to be extreme.
- To get a moderate value, many can be extreme in the opposite direction, or many can be moderate (or several variations in between).
- There are more ways to obtain moderate values in an average than to obtain extreme values

Implications for Statistics

- **Regardless** of the underlying population distribution, when sample size is large, the distribution of sample means is predictable, and variance in means decreases as sample size increases

Implications for Statistics

- **Regardless** of the underlying population distribution, when sample size is large, the distribution of sample means is predictable, and variance in means decreases as sample size increases
- We can use properties of the Normal distribution to determine probabilities of obtaining extreme sample statistics

Implications for Statistics

- **Regardless** of the underlying population distribution, when sample size is large, the distribution of sample means is predictable, and variance in means decreases as sample size increases
- We can use properties of the Normal distribution to determine probabilities of obtaining extreme sample statistics
- Statistical inference can be performed using theoretical density functions, in addition to using simulation and bootstrapping

Theory vs Simulation Methods

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

- Previously, we. . .

Theory vs Simulation Methods

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

- Previously, we. . .
  - Constructed confidence intervals by approximating the sampling distribution through bootstrapping, used percentile method to get interval bounds.

Theory vs Simulation Methods

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

- Previously, we. . .
  - Constructed confidence intervals by approximating the sampling distribution through bootstrapping, used percentile method to get interval bounds.
  - Performed hypothesis tests by creating the null distribution through randomization, calculating p-values as proportions of extreme null statistics.

Theory vs Simulation Methods

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

- Previously, we...
  - Constructed confidence intervals by approximating the sampling distribution through bootstrapping, used percentile method to get interval bounds.
  - Performed hypothesis tests by creating the null distribution through randomization, calculating p-values as proportions of extreme null statistics.
- But now, using the Central Limit Theorem, we can...

Theory vs Simulation Methods

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

- Previously, we...
  - Constructed confidence intervals by approximating the sampling distribution through bootstrapping, used percentile method to get interval bounds.
  - Performed hypothesis tests by creating the null distribution through randomization, calculating p-values as proportions of extreme null statistics.
- 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 have two ways of making confidence intervals / performing hypothesis tests:

- Previously, we. . .
  - Constructed confidence intervals by approximating the sampling distribution through bootstrapping, used percentile method to get interval bounds.
  - Performed hypothesis tests by creating the null distribution through randomization, calculating p-values as proportions of extreme null statistics.
- 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 have two ways of making confidence intervals / performing hypothesis tests:

- Previously, we...
  - Constructed confidence intervals by approximating the sampling distribution through bootstrapping, used percentile method to get interval bounds.
  - Performed hypothesis tests by creating the null distribution through randomization, calculating p-values as proportions of extreme null statistics.
- 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 have two ways of making confidence intervals / performing hypothesis tests:

- Previously, we. . .
  - Constructed confidence intervals by approximating the sampling distribution through bootstrapping, used percentile method to get interval bounds.
  - Performed hypothesis tests by creating the null distribution through randomization, calculating p-values as proportions of extreme null statistics.
- 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; they also require access to computing technology

Section 2

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 $\sqrt{p(1-p)}$

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 $\sqrt{p(1-p)}$
- 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

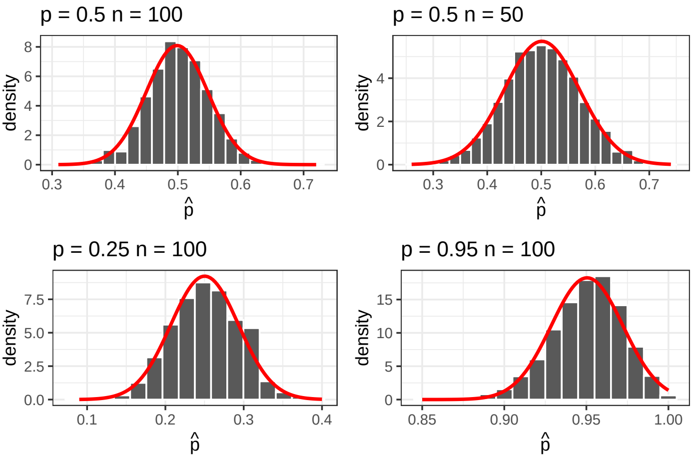
- Consider a population variable that takes only two levels: A and B. Let p be the proportion of A's in the population.
- Suppose we randomly choose a single observation from a population, and define a random variable X to be 1 if the observation is an A, and 0 if it is a B.
  - The *mean* of X is p,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



Section 3

Hypothesis Tests

z-Scores

- The **z-score** for a statistic x with standard error SE and mean μ is

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

z-Scores

- The **z-score** for a statistic x with standard error SE and mean μ is

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

- A z-score tells us how extreme an observed statistic is (i.e. how far it is from its mean), relative to its standard deviation.

z-Scores

- The **z-score** for a statistic x with standard error SE and mean μ is

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

- A z-score tells us how extreme an observed statistic is (i.e. how far it is from its mean), relative to its standard deviation.
  - Positive (negative) z-scores indicate statistics larger (smaller) than the mean

z-Scores

- The **z-score** for a statistic x with standard error SE and mean μ is

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

- A z-score tells us how extreme an observed statistic is (i.e. how far it is from its mean), relative to its standard deviation.
  - Positive (negative) z-scores indicate statistics larger (smaller) than the mean
- For approximately Normally distributed data, 68-95-99.7% of observations are within 1-2-3 standard deviations of their mean.

z-Scores

- The **z-score** for a statistic x with standard error SE and mean μ is

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

- A z-score tells us how extreme an observed statistic is (i.e. how far it is from its mean), relative to its standard deviation.
  - Positive (negative) z-scores indicate statistics larger (smaller) than the mean
- For approximately Normally distributed data, 68-95-99.7% of observations are within 1-2-3 standard deviations of their mean.
  - Thus, 68% of observed samples have z-scores between -1 and 1, 95% of samples have z-scores between -2 and 2, and 99.7% of samples have z-scores between -3 and 3.

z-Scores

- The **z-score** for a statistic x with standard error SE and mean μ is

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

- A z-score tells us how extreme an observed statistic is (i.e. how far it is from its mean), relative to its standard deviation.
  - Positive (negative) z-scores indicate statistics larger (smaller) than the mean
- For approximately Normally distributed data, 68-95-99.7% of observations are within 1-2-3 standard deviations of their mean.
  - Thus, 68% of observed samples have z-scores between -1 and 1, 95% of samples have z-scores between -2 and 2, and 99.7% of samples have z-scores between -3 and 3.
- Suppose a statistic X is approximately Normal with mean μ and standard deviation SE. Then

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

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

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.

P-Values

- Recall that the *p*-value of a sample is the probability of obtaining a sample more extreme than the observed sample, if the null hypothesis were true.

P-Values

- Recall that the *p*-value of a sample is the probability of obtaining a sample more extreme than the observed sample, if the null hypothesis were true.
- To obtain the *p*-value of an observed sample $\hat{p}$, we first calculate the corresponding z-score:

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

P-Values

- Recall that the *p*-value of a sample is the probability of obtaining a sample more extreme than the observed sample, if the null hypothesis were true.
- To obtain the *p*-value of an observed sample $\hat{p}$, we first calculate the corresponding *z*-score:

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

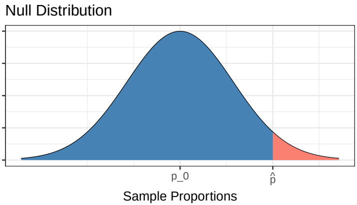
- The *p*-value of the sample is the probability of obtaining a sample with *z*-score more extreme than the observed *z*-score.

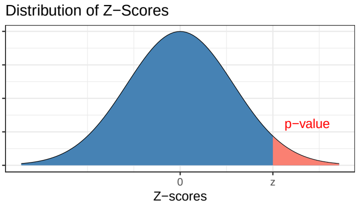
P-Values

- Recall that the *p*-value of a sample is the probability of obtaining a sample more extreme than the observed sample, if the null hypothesis were true.
- To obtain the *p*-value of an observed sample $\hat{p}$, we first calculate the corresponding *z*-score:

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

- The *p*-value of the sample is the probability of obtaining a sample with *z*-score more extreme than the observed *z*-score.
- By the Central Limit Theorem, these *z*-scores are approximately *standard* Normal. We can compute desired probabilities using the `pnorm()` function in R.





Taste Test

- Are these the same?

Taste Test

- In a previous year, intro stat students participated in an experiment to determine whether they can distinguish between two different flavors of carbonated water.

Taste Test

- In a previous year, intro stat students participated in an experiment to determine whether they 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

- In a previous year, intro stat students participated in an experiment to determine whether they 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 to some extent.

Taste Test

- In a previous year, intro stat students participated in an experiment to determine whether they 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 to some extent.
- Let p denote the true proportion of the population who can correctly identify the cup that is different.

Taste Test

- In a previous year, intro stat students participated in an experiment to determine whether they 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 to some extent.
- 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

- In a previous year, intro stat students participated in an experiment to determine whether they 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 to some extent.
- 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

- In a previous year, intro stat students participated in an experiment to determine whether they 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 to some extent.
- 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
    - 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
  - 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
  - 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
  - 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

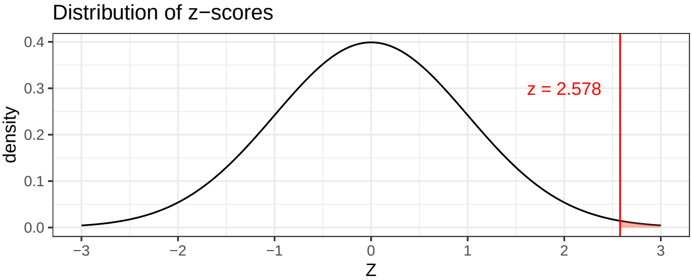
- If H_0 is true, the z-score should be Normally distributed, with mean 0 and st. dev.

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 $z = 2.578$

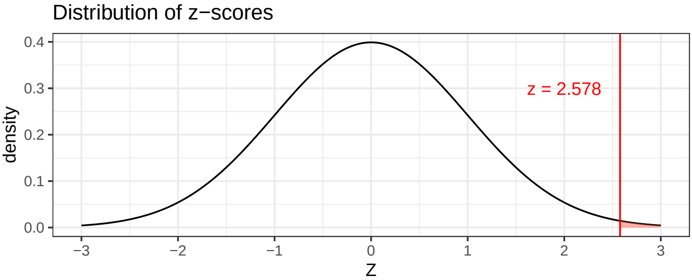
Calculate P-Value

- 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 4

Confidence Intervals

Critical Values

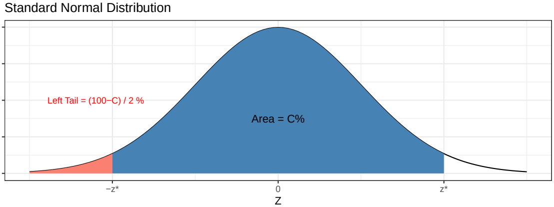
- The **critical value** z^* for a $C\%$ confidence interval is the value so that $C\%$ of area is between $-z^*$ and z^* in the standard Normal distribution

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
  - That is, the critical value of $C\%$ confidence is the $C + \frac{1-C}{2}$ percentile of the standard Normal distribution

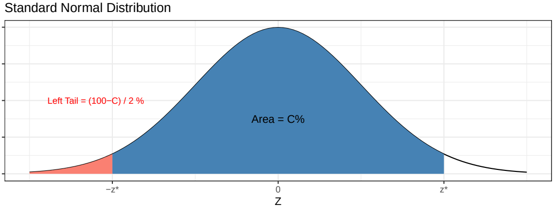
Critical Values

- 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
  - That is, the critical value of $C\%$ confidence is the $C + \frac{1-C}{2}$ percentile of the standard Normal distribution



Critical Values

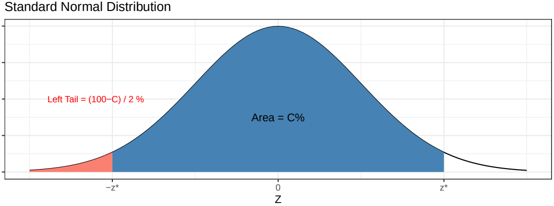
- The **critical value** z^* for a $C\%$ confidence interval is the value so that $C\%$ of area is between $-z^*$ and z^* in the standard Normal distribution
  - That is, the critical value of $C\%$ confidence is the $C + \frac{1-C}{2}$ percentile of the standard Normal distribution



- The critical value for 95% confidence is the $95 + \frac{100-95}{2} = 97.5$ percentile

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
  - That is, the critical value of $C\%$ confidence is the $C + \frac{1-C}{2}$ percentile of the standard Normal distribution



- The critical value for 95% confidence is the $95 + \frac{100-95}{2} = 97.5$ percentile

```
qnorm(.975, mean = 0, sd = 1) # The 97.5 percentile is the .975 quantile
```

```
## [1] 1.959964
```

Confidence Intervals

If the sample statistic is approximately Normal, the $C\%$ confidence interval is

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

where z^* is the critical value confidence and SE is the standard error of the statistic

Confidence Intervals

If the sample statistic is approximately Normal, the $C\%$ confidence interval is

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

where z^* is the critical value confidence and SE is the standard error of the statistic

- The standard error for a sample proportion $\hat{p}$ is $SE = \sqrt{\frac{p(1-p)}{n}}$.

Confidence Intervals

If the sample statistic is approximately Normal, the $C\%$ confidence interval is

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

where z^* is the critical value confidence and SE is the standard error of the statistic

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

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

Confidence Intervals

If the sample statistic is approximately Normal, the $C\%$ confidence interval is

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

where z^* is the critical value confidence and SE is the standard error of the statistic

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

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

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

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

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
```