

Statistical Significance

How to make sense of results?

So far:

How to make sense of results?

So far:

- ▶ Focus on findings and computing differences.

How to make sense of results?

So far:

- ▶ Focus on findings and computing differences.
- ▶ but do these differences matter?

How to make sense of results?

So far:

- ▶ Focus on findings and computing differences.
- ▶ but do these differences matter?
 - ▶ How would we know?

How to make sense of results?

So far:

- ▶ Focus on findings and computing differences.
- ▶ but do these differences matter?
 - ▶ How would we know?

Today:

How to make sense of results?

So far:

- ▶ Focus on findings and computing differences.
- ▶ but do these differences matter?
 - ▶ How would we know?

Today:

- ▶ Systematic approach for understanding which differences are meaningful and which are not.

What is Statistical Significance?

Statistical significance is a measure that helps us determine whether the observed results in a study are likely due to chance or if they reflect a true effect.

What is Statistical Significance?

Statistical significance is a measure that helps us determine whether the observed results in a study are likely due to chance or if they reflect a true effect.

It helps researchers decide if their findings are reliable and can be generalized to a larger population.

Important Considerations

1. Statistical significance does not imply practical significance

Important Considerations

1. Statistical significance does not imply practical significance
2. Large differences are not always (or even often) significant

Important Considerations

1. Statistical significance does not imply practical significance
2. Large differences are not always (or even often) significant
3. Larger samples provide more reliable results

Important Considerations

1. Statistical significance does not imply practical significance
2. Large differences are not always (or even often) significant
3. Larger samples provide more reliable results
4. Significance does not imply causation

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Possible Hypotheses:

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Possible Hypotheses:

- ▶ “Artillery attacks have no effect on rebel activity”

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Possible Hypotheses:

- ▶ “Artillery attacks have no effect on rebel activity”
- ▶ “Artillery attacks decrease rebel activity by 5”

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Possible Hypotheses:

- ▶ “Artillery attacks have no effect on rebel activity”
- ▶ “Artillery attacks decrease rebel activity by 5”

But we need a formal framework:

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Possible Hypotheses:

- ▶ “Artillery attacks have no effect on rebel activity”
- ▶ “Artillery attacks decrease rebel activity by 5”

But we need a formal framework:

We call our starting hypothesis *the null hypothesis* (H_0)

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Possible Hypotheses:

- ▶ “Artillery attacks have no effect on rebel activity”
- ▶ “Artillery attacks decrease rebel activity by 5”

But we need a formal framework:

We call our starting hypothesis *the null hypothesis* (H_0)

- ▶ usually framed in the negative

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Possible Hypotheses:

- ▶ “Artillery attacks have no effect on rebel activity”
- ▶ “Artillery attacks decrease rebel activity by 5”

But we need a formal framework:

We call our starting hypothesis *the null hypothesis* (H_0)

- ▶ usually framed in the negative
- ▶ determine whether we can **reject it** based on our data.

Hypothesis Testing

Notation:

Hypothesis Testing

Notation:

H_0 (Null Hypothesis)

Hypothesis Testing

Notation:

H_0 (Null Hypothesis)

H_a (Alternative Hypothesis)

Hypothesis Testing

Exmaples:

Hypothesis Testing

Exmaples:

“Artillery attacks have no effect on rebel activity”

Hypothesis Testing

Exmaples:

“Artillery attacks have no effect on rebel activity”

“Artillery attacks changed rebel activity”

Hypothesis Testing

Exmaples:

“Artillery attacks have no effect on rebel activity”

“Artillery attacks changed rebel activity”

“Artillery attacks increase rebel activity”

Example: Generate a test statistic

- ▶ Calculate a **t-statistic**

Example: Generate a test statistic

- Calculate a **t-statistic**

$$t = \frac{\text{sample mean} - \text{population mean (given } H_0\text{)}}{\text{estimated standard error (of sample mean)}}$$

Example: Generate a test statistic

- Calculate a t-statistic

$$t = \frac{\text{sample mean} - \text{population mean (given } H_0\text{)}}{\text{estimated standard error (of sample mean)}}$$

$$t = \frac{\bar{x} - \mu_0}{s_x / \sqrt{n}} = \frac{0.19 - 0.10}{0.394 / \sqrt{100}} = \boxed{2.283}$$

What do we do with this?

Compare to critical value

What do we do with this?

Compare to critical value

- ▶ Given the two-sided null hypothesis we obtained $t = 2.283$

What do we do with this?

Compare to critical value

- ▶ Given the two-sided null hypothesis we obtained $t = 2.283$
- ▶ Based on this t-statistic, can we reject the null hypothesis?

What do we do with this?

Compare to critical value

- ▶ Given the two-sided null hypothesis we obtained $t = 2.283$
- ▶ Based on this t-statistic, can we reject the null hypothesis?
- ▶ Another way of asking: what is the probability that we would observe a sample mean of 0.19 (given $N=100$) by *random chance*, if the population mean is 0.10?

What do we do with this?

Compare to critical value

- ▶ Given the two-sided null hypothesis we obtained $t = 2.283$
- ▶ Based on this t-statistic, can we reject the null hypothesis?
- ▶ Another way of asking: what is the probability that we would observe a sample mean of 0.19 (given $N=100$) by *random chance*, if the population mean is 0.10?
- ▶ We call this probability the **p-value**: the likelihood we would observe a value at least as extreme as our sample mean (if the null hypothesis were true), based on chance.

Interpreting p-values

- ▶ We can never be 100% certain that a hypothesis is true.

Interpreting p-values

- ▶ We can never be 100% certain that a hypothesis is true.
- ▶ **Null hypothesis:**

Interpreting p-values

- ▶ We can never be 100% certain that a hypothesis is true.
- ▶ **Null hypothesis:**
- ▶ “Shaking a candidates’ hand has no effect on voting behavior”

Interpreting p-values

- ▶ We can never be 100% certain that a hypothesis is true.
- ▶ **Null hypothesis:**
- ▶ “Shaking a candidates’ hand has no effect on voting behavior”
- ▶ Type 1 Error: Finding that shaking a hand has an effect on voting (rejecting the null hypothesis) when it in fact does not.

Interpreting p-values

- ▶ We can never be 100% certain that a hypothesis is true.
- ▶ **Null hypothesis:**
- ▶ “Shaking a candidates’ hand has no effect on voting behavior”
- ▶ Type 1 Error: Finding that shaking a hand has an effect on voting (rejecting the null hypothesis) when it in fact does not.
- ▶ Type 2 Error: Finding that shaking a hand has no effect on voting (retaining the null hypothesis), when in fact it does.

Interpreting p-values

- ▶ Generally, we try to minimize Type 1 Error (α). In the case of a 95% confidence interval, $(\alpha) = .05$, and our chance of a false positive is only 5%.

Interpreting p-values

- ▶ Generally, we try to minimize Type 1 Error (α). In the case of a 95% confidence interval, $(\alpha) = .05$, and our chance of a false positive is only 5%.
- ▶ If our p-value is $\leq \alpha$, we reject the null hypothesis

Interpreting p-values

- ▶ Generally, we try to minimize Type 1 Error (α). In the case of a 95% confidence interval, $(\alpha) = .05$, and our chance of a false positive is only 5%.
- ▶ If our p-value is $\leq \alpha$, we reject the null hypothesis
- ▶ So in our example, since $p = .0224$, we reject the hypothesis that the true population mean is 0.1

Interpreting p-values

- ▶ Generally, we try to minimize Type 1 Error (α). In the case of a 95% confidence interval, $(\alpha) = .05$, and our chance of a false positive is only 5%.
- ▶ If our p-value is $\leq \alpha$, we reject the null hypothesis
- ▶ So in our example, since $p = .0224$, we reject the hypothesis that the true population mean is 0.1

$$H_0 : \mu = \mu_0 (= 0.1)$$

Estimating the Error of a Sample Mean

- ▶ To assess the quality of a sample mean we will compute a measure of error for the estimate (standard error)

Estimating the Error of a Sample Mean

- ▶ To assess the quality of a sample mean we will compute a measure of error for the estimate (standard error)
- ▶ Intuitively, we expect the error to decrease as the sample size increases.

Estimating the Error of a Sample Mean

- ▶ To assess the quality of a sample mean we will compute a measure of error for the estimate (standard error)
- ▶ Intuitively, we expect the error to decrease as the sample size increases.
- ▶ We might also expect it to take longer for the sample mean to converge if there is a lot of variation in the underlying distribution.

Estimating the Error of a Sample Mean

- ▶ To assess the quality of a sample mean we will compute a measure of error for the estimate (standard error)
- ▶ Intuitively, we expect the error to decrease as the sample size increases.
- ▶ We might also expect it to take longer for the sample mean to converge if there is a lot of variation in the underlying distribution.
- ▶ We can estimate the **standard error** of a sample mean using the following equations:

Estimating the Error of a Sample Mean

- ▶ To assess the quality of a sample mean we will compute a measure of error for the estimate (standard error)
- ▶ Intuitively, we expect the error to decrease as the sample size increases.
- ▶ We might also expect it to take longer for the sample mean to converge if there is a lot of variation in the underlying distribution.
- ▶ We can estimate the **standard error** of a sample mean using the following equations:

$$s.e. \text{ of sample mean} = \frac{\text{standard deviation of sample}}{\sqrt{\text{sample size } (n)}}$$

Estimating the Error of a Sample Mean

- ▶ To assess the quality of a sample mean we will compute a measure of error for the estimate (standard error)
- ▶ Intuitively, we expect the error to decrease as the sample size increases.
- ▶ We might also expect it to take longer for the sample mean to converge if there is a lot of variation in the underlying distribution.
- ▶ We can estimate the **standard error** of a sample mean using the following equations:

$$s.e. \text{ of sample mean} = \frac{\text{standard deviation of sample}}{\sqrt{\text{sample size } (n)}}$$

$$s.e. \text{ of sample proportion} = \sqrt{\frac{p(1-p)}{n}}$$

Standard Error

- ▶ The 'standard error' approximates the standard deviation of the sampling distribution.

Standard Error

- ▶ The 'standard error' approximates the standard deviation of the sampling distribution.
- ▶ Recall that the standard deviation is a measure of the dispersion of **observations** around the **sample mean**.

Standard Error

- ▶ The 'standard error' approximates the standard deviation of the sampling distribution.
- ▶ Recall that the standard deviation is a measure of the dispersion of **observations** around the **sample mean**.
- ▶ The standard error of the mean is a measure of the dispersion of **sample means** around the **population mean**.

Standard Error

- ▶ The 'standard error' approximates the standard deviation of the sampling distribution.
- ▶ Recall that the standard deviation is a measure of the dispersion of **observations** around the **sample mean**.
- ▶ The standard error of the mean is a measure of the dispersion of **sample means** around the **population mean**.
- ▶ Importantly, we can calculate it for a hypothetical sampling distribution, even if we are unable to repeatedly sample from a population.

Standard Error

- ▶ The 'standard error' approximates the standard deviation of the sampling distribution.
- ▶ Recall that the standard deviation is a measure of the dispersion of **observations** around the **sample mean**.
- ▶ The standard error of the mean is a measure of the dispersion of **sample means** around the **population mean**.
- ▶ Importantly, we can calculate it for a hypothetical sampling distribution, even if we are unable to repeatedly sample from a population.
- ▶ This is useful, because in many cases we cannot conduct repeated samples, and we still want to know how "likely" our estimate reflects the true population mean.

Generalizing Standard Error

For every estimate we have calculated in this course we can compute a standard error.

Generalizing Standard Error

For every estimate we have calculated in this course we can compute a standard error.

- ▶ So, for every mean and regression coefficient we have computed.

Generalizing Standard Error

For every estimate we have calculated in this course we can compute a standard error.

- ▶ So, for every mean and regression coefficient we have computed.

But what happens if we are not drawing from one underlying distribution (centered around a population mean), but two?

Generalizing Standard Error

For every estimate we have calculated in this course we can compute a standard error.

- ▶ So, for every mean and regression coefficient we have computed.

But what happens if we are not drawing from one underlying distribution (centered around a population mean), but two?

- ▶ For instance: for an average treatment effect, we estimate:

Generalizing Standard Error

For every estimate we have calculated in this course we can compute a standard error.

- ▶ So, for every mean and regression coefficient we have computed.

But what happens if we are not drawing from one underlying distribution (centered around a population mean), but two?

- ▶ For instance: for an average treatment effect, we estimate:
 - ▶ $\text{Mean}(\text{Treatment}) - \text{Mean}(\text{Control})$

Generalizing Standard Error

For every estimate we have calculated in this course we can compute a standard error.

- ▶ So, for every mean and regression coefficient we have computed.

But what happens if we are not drawing from one underlying distribution (centered around a population mean), but two?

- ▶ For instance: for an average treatment effect, we estimate:
 - ▶ $\text{Mean}(\text{Treatment}) - \text{Mean}(\text{Control})$
- ▶ Even if we run an experiment only once, we have two (hypothetical) sampling distributions, one for the treatment and one for the control.

Standard Error for a Difference in Means

We need to take variation (error) in the sampling distribution of both groups into account:

Standard Error for a Difference in Means

We need to take variation (error) in the sampling distribution of both groups into account:

$$s.e. \text{ of difference} = \sqrt{\text{Variance}(\textit{Treatment})/n + \text{Variance}(\textit{Control})/n}$$

Standard Error for a Difference in Means

We need to take variation (error) in the sampling distribution of both groups into account:

$$s.e. \text{ of difference} = \sqrt{\text{Variance}(\text{Treatment})/n + \text{Variance}(\text{Control})/n}$$

or

$$s.e. \text{ of difference} = \sqrt{\frac{sd(y_{\text{treatment}})^2}{n_{\text{treatment}}} + \frac{sd(y_{\text{control}})^2}{n_{\text{control}}}}$$

Standard Error for a Difference in Means

We need to take variation (error) in the sampling distribution of both groups into account:

$$s.e. \text{ of difference} = \sqrt{\text{Variance}(\text{Treatment})/n + \text{Variance}(\text{Control})/n}$$

or

$$s.e. \text{ of difference} = \sqrt{\frac{sd(y_{\text{treatment}})^2}{n_{\text{treatment}}} + \frac{sd(y_{\text{control}})^2}{n_{\text{control}}}}$$

- But do not despair—there is an easy way to do this with the help of R.

Confidence Interval

- ▶ So if the standard error tells us the uncertainty inherent in our sampling distribution/estimate, how do we interpret this?

Confidence Interval

- ▶ So if the standard error tells us the uncertainty inherent in our sampling distribution/estimate, how do we interpret this?
- ▶ In practice, we usually translate standard errors into **confidence intervals** using the normal distribution

Confidence Interval

- ▶ So if the standard error tells us the uncertainty inherent in our sampling distribution/estimate, how do we interpret this?
- ▶ In practice, we usually translate standard errors into **confidence intervals** using the normal distribution
- ▶ Confidence intervals give a range of values that are likely to include the true value of the parameter.

Confidence Interval

We choose defined points on the normal distribution, called **critical values**, as an upper or lower bound.

Confidence Interval

We choose defined points on the normal distribution, called **critical values**, as an upper or lower bound.

For instance, a critical value of 1.96 allows us to obtain a range of values that covers 95% of the normal distribution:

Confidence Interval

We choose defined points on the normal distribution, called **critical values**, as an upper or lower bound.

For instance, a critical value of 1.96 allows us to obtain a range of values that covers 95% of the normal distribution:

► Lower bound = $\text{mean}(\text{sample}) - 1.96 * \text{se}$

Confidence Interval

We choose defined points on the normal distribution, called **critical values**, as an upper or lower bound.

For instance, a critical value of 1.96 allows us to obtain a range of values that covers 95% of the normal distribution:

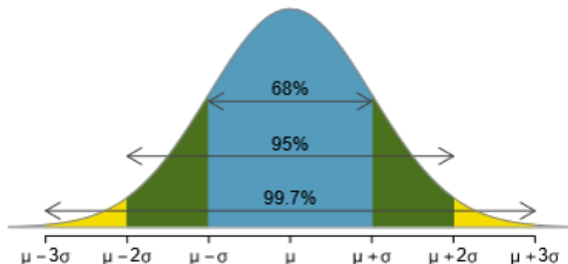
- ▶ Lower bound = $\text{mean}(\text{sample}) - 1.96 * \text{se}$
- ▶ Upper bound = $\text{mean}(\text{sample}) + 1.96 * \text{se}$

Confidence Interval

We choose defined points on the normal distribution, called **critical values**, as an upper or lower bound.

For instance, a critical value of 1.96 allows us to obtain a range of values that covers 95% of the normal distribution:

- ▶ Lower bound = $\text{mean}(\text{sample}) - 1.96 * \text{se}$
- ▶ Upper bound = $\text{mean}(\text{sample}) + 1.96 * \text{se}$



Example 1

Scenario:

Testing a new drug.

Hypotheses:

H_0 : The new drug is no more effective than the current drug.

H_a : The new drug is more effective than the current drug.

Results:

P-value: 0.03. Significance Level (α): 0.05.

Conclusion:

Since $0.03 < 0.05$, we reject H_0 and conclude that the new drug is more effective.

Example 2

Scenario:

A political scientist investigates whether exposure to a particular news channel affects support for a new policy.

Hypotheses:

H_0 : Exposure to the news channel does not affect support for the policy.

H_a : Exposure to the news channel increases support for the policy.

Results:

Control Group Support: 45%. Treatment Group Support: 48%. P-Value: 0.12.

Significance Level (α): 0.05.

Conclusion:

Since the p-value (0.12) is greater than the significance level (0.05), we do not reject the null hypothesis. The result is not statistically significant, indicating that exposure to the news channel does not significantly increase support for the policy.