Lecture 10: Hypothesis Test

Module 3: part 3

Spring 2025

Confidence Intervals: A Recap

Definition

Confidence Interval: A statistical procedure that produces an interval which, for a fixed proportion of the time (e.g., 95%), will contain the true parameter when applied to new data.

Confidence Intervals: A Recap

Definition

Confidence Interval: A statistical procedure that produces an interval which, for a fixed proportion of the time (e.g., 95%), will contain the true parameter when applied to new data.

- Represents a plausible range of values for the true parameter
- **Common Misconception**: "There's a 95% chance that the interval (1.5, 2.6) contains the true parameter"
- **Correct Interpretation**: Once computed, an interval either does or does not contain the true parameter
- Probability refers to the method's performance over hypothetical future samples

Confidence Interval Formula

For a new dataset, the $(1-\alpha)$ confidence interval for b_1 is:

$$\left(\hat{b}_1 - t_{\alpha/2,n-p-1}\sqrt{\widehat{\text{var}}(\hat{b}_1)}, \hat{b}_1 + t_{\alpha/2,n-p-1}\sqrt{\widehat{\text{var}}(\hat{b}_1)}\right)$$

.

Confidence Interval Formula

For a new dataset, the $(1-\alpha)$ confidence interval for b_1 is:

$$\left(\hat{b}_1 - t_{\alpha/2,n-p-1}\sqrt{\widehat{\text{var}}(\hat{b}_1)}, \hat{b}_1 + t_{\alpha/2,n-p-1}\sqrt{\widehat{\text{var}}(\hat{b}_1)}\right)$$

.

Simplified Notation

We can write:

$$\hat{b}_1 \pm t_{1-\alpha/2,n-p-1}\sqrt{\widehat{\text{var}}(\hat{b}_1)}$$

A Weather Forecasting Analogy

Scenario

Imagine you're a meteorologist predicting the average temperature for a city next summer. You estimate it to be 25°C (77°F) with a 95% confidence interval of 23°C to 27°C (73.4°F to 80.6°F).

A Weather Forecasting Analogy

Scenario

Imagine you're a meteorologist predicting the average temperature for a city next summer. You estimate it to be 25°C (77°F) with a 95% confidence interval of 23°C to 27°C (73.4°F to 80.6°F).

Correct Interpretation

- **Repeatability**: If you repeated this estimation many times, about 95% of the intervals would contain the true average temperature.
- **Fixed Outcome**: Once summer happens, the true average temperature is fixed; it either falls within your interval or it doesn't.

A Weather Forecasting Analogy

Scenario

Imagine you're a meteorologist predicting the average temperature for a city next summer. You estimate it to be 25°C (77°F) with a 95% confidence interval of 23°C to 27°C (73.4°F to 80.6°F).

Common Misinterpretations

- Incorrect: "There's a 95% chance the true average temperature is between 23°C and 27°C."
- Correct: The true average is fixed; it's either in the interval or not.
- Incorrect: "95% of daily temperatures will fall within this range."
- Correct: The interval is about the average temperature, not individual daily temperatures.
- Incorrect: "If we measure again next summer, there's a 95% chance it will fall in this range."
- Correct: This interval is about this specific estimation process, not future measurements.

A Weather Forecasting Analogy

Scenario

Imagine you're a meteorologist predicting the average temperature for a city next summer. You estimate it to be 25°C (77°F) with a 95% confidence interval of 23°C to 27°C (73.4°F to 80.6°F).

Conclusion

This analogy highlights that confidence intervals are about the reliability of the estimation method over many repetitions, not the specific estimate itself.

Questions

- All things equal, what will typically be wider? A 95% confidence interval or a 90% confidence interval?
- All things equal, what will typically be wider? A 95% confidence interval when $n = 100$ or when $n = 500$ where n is the number of observations?
- All things equal, what will typically be wider? A 95% confidence interval when $p = 5$ or when $p = 10$ where p is the number of predictors in the model?

Crash Course on Statistical Inference

Hypothesis Testing: A Basketball Analogy

The Claim

"I'm a really good basketball player and on average I score 20 points a game."

"In God we trust. All others must bring data."

W. Edwards Deming

The Challenge

How can we verify this claim using data?

Evaluating the Claim: Scenarios

Claim: I score 20 points a game on average

- **Scenario 1:** You observe 2 games, average score = 18 points

Evaluating the Claim: Scenarios

Claim: I score 20 points a game on average

- **Scenario 1:** You observe 2 games, average score = 18 points
- **Scenario 2:** You observe 2 games, average score = 10 points

Evaluating the Claim: Scenarios

Claim: I score 20 points a game on average

- **Scenario 1:** You observe 2 games, average score = 18 points
- **Scenario 2:** You observe 2 games, average score = 10 points
- **Scenario 3:** You observe 25 games, average score = 10 points

Evaluating the Claim: Scenarios

Claim: I score 20 points a game on average

- **Scenario 1:** You observe 2 games, average score = 18 points
- **Scenario 2:** You observe 2 games, average score = 10 points
- **Scenario 3:** You observe 25 games, average score = 10 points

Evaluating the Claim: Scenarios

Claim: I score 20 points a game on average

- **Scenario 1:** You observe 2 games, average score = 18 points
- **Scenario 2:** You observe 2 games, average score = 10 points
- **Scenario 3:** You observe 25 games, average score = 10 points

Key Questions

- How do you decide whether to reject the claim?
- How much evidence is enough?
- What constitutes a "significant" difference from the claim?

The Logic of Hypothesis Testing

Claim: I score 20 points a game on average

Considerations:

- Initial trust in the claim
- "Normal range" of outcomes for a 20-point average player
- Observed data vs. expected outcomes

Decision Process:

- If observed outcome is **plausible**: Cannot rule out the claim
- If observed outcome is **very unlikely**: May reject the claim

Key Concept

Hypothesis testing involves comparing observed data against expected outcomes to make informed decisions about claims.

Hypothesis Testing Framework

1. **Define Hypotheses**
   - Null hypothesis (H_0): "status quo"
   - Alternative hypothesis (H_1): typically what the researcher aims to prove
2. **Select Test Statistic**
   - Distribution known under H_0
   - Distribution under H_1 not necessary
3. **Collect Data & Calculate Statistic**
4. **Calculate p-value**
   - Compare observed statistic to H_0 distribution
   - P-value: probability of observing a statistic as or more extreme than observed, assuming H_0 is true
5. **Draw Conclusions**
   - Reject H_0 if p-value $< \alpha$
   - Fail to reject H_0 if p-value $\geq \alpha$

Key Concept

Hypothesis testing is a systematic approach to making statistical decisions based on sample data and probability theory.

Hypothesis Testing Errors

Decision Outcomes

- We either reject or fail to reject the null hypothesis
- We don't confirm or accept the null hypothesis

	Null is rejected	Null is not rejected
Null is true	Type I error	✓
Null is false	✓	Type II error

Hypothesis Testing Errors

Decision Outcomes

- We either reject or fail to reject the null hypothesis
- We don't confirm or accept the null hypothesis

	Null is rejected	Null is not rejected
Null is true	Type I error	✓
Null is false	✓	Type II error

Key Points

- Type I errors are typically considered more "costly" than Type II errors
- We limit the probability of Type I error to a certain **significance level** (α)
- Conventionally, $\alpha = 0.05$, but this choice is somewhat arbitrary

Remember

The significance level (α) is the probability of rejecting the null hypothesis when it is actually true.

Type I and Type II Errors: Examples

Example: Rock Climbing Equipment

- Type I Error (False Positive):
  - Frank thinks his equipment is unsafe when it's actually safe
- Type II Error (False Negative):
  - Frank thinks his equipment is safe when it's actually unsafe

Example: Roulette Wheel

- Type I Error:
  - Concluding the wheel is biased when it's actually fair
- Type II Error:
  - Failing to detect that the wheel is biased when it actually is

Question for Reflection

In a medical diagnosis scenario:

- What would be a Type I error?
- What would be a Type II error?

Hypothesis Testing in Linear Models

Common Tests in Linear Regression

We often test hypotheses about:

1. **Individual Coefficients**
   - H_0: $\beta_i = 0$ (no effect)
   - H_1: $\beta_i \neq 0$ (significant effect)
2. **Categorical Variables**
   - H_0: All category coefficients $= 0$
   - H_1: At least one category coefficient $\neq 0$
3. **Model Comparison**
   - H_0: Additional variables don't improve the model
   - H_1: Additional variables significantly improve the model

Key Concept

These tests help us determine which variables are statistically significant predictors in our model.

Testing a Single Coefficient in Linear Regression

Model Assumption

Assume the following linear model:

$$Y_i = b_0 + b_1 X_1 + b_2 X_2 \dots b_p X_p + \varepsilon_i$$

where $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$

Research Question

Does X_1 have a significant association with Y when controlling for $X_2, X_3, \dots, X_p$?

Hypotheses:

$$H_0 : b_1 = 0$$
$$H_1 : b_1 \neq 0$$

Decision Rule:

- Reject H_0 if $|\hat{b_1}|$ is "large enough"
- But how large is "large enough"?

Key Consideration

We need to determine a threshold for $|\hat{b_1}|$ that balances Type I and Type II errors.

Testing a Single Coefficient: The t-Statistic

Challenge

We can't directly compute $\frac{\hat{b}_1}{\sqrt{\text{var}(\hat{b}_1)}}$ as σ_ε^2 is unknown.

Solution: t-Statistic

Using $\hat{\sigma}_\varepsilon^2$ as an estimate, we define: $t = \frac{\hat{b}_1}{\sqrt{\widehat{\text{var}}(\hat{b}_1)}}$

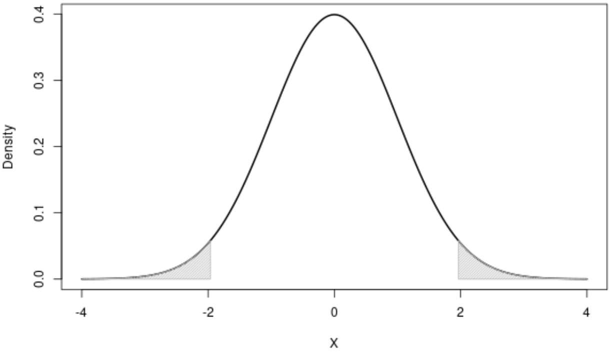
- t follows a T distribution with $n - p - 1$ degrees of freedom (T_{n-p-1})
- We define a **rejection region** for the test statistic
- The cut-off should be "extreme" enough to control Type I error
- We choose a cut-off such that t exceeds it only a predetermined proportion of the time (e.g., 5% for $\alpha = 0.05$)

Key Concept

The t-statistic allows us to standardize our estimate and compare it to a known distribution, enabling hypothesis testing.

Testing a single coefficient

Find the appropriate cut-off for a T_{n-p-1} so that each shaded region has area $\alpha/2$



Understanding P-values

Definition

The **p-value** is the probability of observing a test statistic as extreme as, or more extreme than, the one calculated from the sample data, assuming the null hypothesis is true.

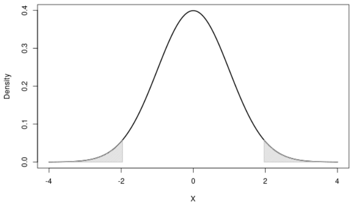
Understanding P-values

Definition

The **p-value** is the probability of observing a test statistic as extreme as, or more extreme than, the one calculated from the sample data, assuming the null hypothesis is true.

Practical Procedure:

1. Calculate the test statistic from sample data
2. Compute the p-value
3. Compare p-value to predetermined significance level (α)
4. Reject H_0 if p-value $< \alpha$



Understanding P-values

Definition

The **p-value** is the probability of observing a test statistic as extreme as, or more extreme than, the one calculated from the sample data, assuming the null hypothesis is true.

Common Misconception

Incorrect: The p-value is the probability that the null hypothesis is true.

Key Point

The p-value is about the data under H_0, not about the probability of H_0 itself.

Hypothesis test for single coefficient

Testing the hypothesis:

$$H_0 \ : \ b_1 = \beta \text{ (other coefficients are arbitrary)}$$
$$H_A \ : \ b_1 \neq \beta \text{ (other coefficients are arbitrary)}$$

1. Set a level for the hypothesis test (typically .05)
2. Fit regression and estimate $\hat{b}_1$, $\widehat{\text{var}}(b_1)$
3. Calculate
$$t = \frac{\hat{b}_1 - \beta}{\sqrt{\widehat{\text{var}}(b_1)}}$$
4. Calculate the p-value: the probability that an observation as or more extreme than t would occur from a T_{n-p-1}
5. Reject the null hypothesis if the p-value is less than the predetermined level (i.e., the observed value is very unlikely to happen under the null hypothesis)

Why P-value $\neq$ Probability of Null Hypothesis

Common Misconception

"The p-value is the probability that the null hypothesis is true."

Example: Coin Toss Experiment

Suppose we're testing if a coin is fair ($H_0 : p = 0.5$) or biased ($H_1 : p \neq 0.5$).

Scenario:

- We toss the coin 100 times
- We observe 60 heads
- This yields a p-value of 0.057

Interpretation:

- Incorrect: There's a 5.7% chance the coin is fair
- Correct: If the coin were fair, there's a 5.7% chance of observing 60 or more heads in 100 tosses

Why It's False

- The p-value assumes H_0 is true in its calculation
- It doesn't consider prior probabilities or alternative hypotheses
- The true state of H_0 is fixed (either true or false), not probabilistic

Wrapup

- Statistical Hypothesis testing starts with a null hypothesis
- P-value: what is the probability of occurrence for the data we actually observed
- Typically interested in testing if $b_k = 0$; i.e., is covariate X_k associated with Y conditional on all other covariates