Lecture 9: Hypothesis Testing 2

Jacob M. Montgomery

Quantitative Political Methodology

Lecture 9: Hypothesis Testing 2

Roadmap

Last class:

- What is a hypothesis test?
- The five steps of hypothesis testing.

Roadmap

Last class:

- What is a hypothesis test?
- The five steps of hypothesis testing.

This class:

- Hypothesis tests with small samples
- Types of errors
- Discussion of one-sided/two-sided tests
- Relationship between CI and NHPT

Small sample significance testing for quantitative variables

Step 1: Assumptions

- Random sampling
- Quantitative data

Small sample significance testing for quantitative variables

Step 1: Assumptions

- Random sampling
- Quantitative data
- **Population is distributed normally**

Step 2: State hypotheses

- $H_0 : \mu = \mu_0$ (e.g., $\mu = 12$)
- $H_a : \mu \neq \mu_0$
- This is a “two-sided test,” but it may be a “one-sided.”

Step 3: Calculate a test statistic

- $t^* = \dfrac{\bar{Y} - \mu_0}{\sigma_{\bar{Y}}}, df = (n - 1)$
- Just as before, this comes from the sampling distribution of $\bar{Y}$

Step 3: Calculate a test statistic

- $t^* = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{Y}}}, df = (n-1)$
- Just as before, this comes from the sampling distribution of $\bar{Y}$

Step 4: P-Value

- $p = Pr\Big(|t| \geq |\frac{\bar{Y}-\mu_0}{\sigma_{\bar{Y}}}|\Big) = 2 \times Pr(t \geq |\frac{\bar{Y}-\mu_0}{\sigma_{\bar{Y}}}|)$

Step 3: Calculate a test statistic

- $t^* = \dfrac{\bar{Y} - \mu_0}{\sigma_{\bar{Y}}}, df = (n-1)$
- Just as before, this comes from the sampling distribution of $\bar{Y}$

Step 4: P-Value

- $p = Pr\Big(|t| \geq |\frac{\bar{Y}-\mu_0}{\sigma_{\bar{Y}}}|\Big) = 2 \times Pr(t \geq |\frac{\bar{Y}-\mu_0}{\sigma_{\bar{Y}}}|)$
- Make sure you are using the right degrees of freedom.

Step 3: Calculate a test statistic

- $t^* = \dfrac{\bar{Y} - \mu_0}{\sigma_{\bar{Y}}}, df = (n-1)$
- Just as before, this comes from the sampling distribution of $\bar{Y}$

Step 4: P-Value

- $p = Pr\Big(|t| \geq |\frac{\bar{Y}-\mu_0}{\sigma_{\bar{Y}}}|\Big) = 2 \times Pr(t \geq |\frac{\bar{Y}-\mu_0}{\sigma_{\bar{Y}}}|)$
- Make sure you are using the right degrees of freedom.
- We use both tails, because we want to find the probability of error in both directions.
- `2*pt(abs(`t^*`), df = n-1, lower.tail=F)`

Step 5: Draw a conclusion

- If $p \leq \alpha$ we conclude that the evidence supports H_a
- But always report the p-value

Example: State spending on education

Assume that the theory is that states are spending less than 5% of their income on education. The data indicate that:

- $\bar{Y} = 4.7$, $S = 0.0922$
- $t^* = \frac{4.7-5}{0.09/\sqrt{50}} = -2.279$, $df = 49$
- P-value=`2*pt(2.279, df=49, lower.tail = F)` $= 0.027$

Type 1 and Type II Error

		Jury decision	
		Guilty	Innocent
Truth	Guilty	Correct	Type II
	Innocent	Type 1	Correct

Imagine we put the null hypothesis on trial:

		Researcher Conclusion	
		Reject Null	Don't reject
Truth	Null is False	Correct	Type II
	Null is True	Type 1	Correct

Imagine we put the null hypothesis on trial:

		Researcher Conclusion	
		Reject Null	Don't reject
Truth	Null is False	Correct	Type II
	Null is True	Type 1	Correct

- Type I error is when we reject a null hypothesis when the null it is actually true.

Imagine we put the null hypothesis on trial:

		Researcher Conclusion	
		Reject Null	Don't reject
Truth	Null is False	Correct	Type II
	Null is True	Type 1	Correct

- Type I error is when we reject a null hypothesis when the null it is actually true.
- Type II error is when we fail to reject a null hypothesis when the null is actually false.

Imagine we put the null hypothesis on trial:

		Researcher Conclusion	
		Reject Null	Don't reject
Truth	Null is False	Correct	Type II
	Null is True	Type 1	Correct

- Type I error is when we reject a null hypothesis when the null it is actually true.
- Type II error is when we fail to reject a null hypothesis when the null is actually false.
- We tend to prioritize reducing Type I error, although there are tradeoffs.

Notes on Type I and Type II error

How to calculate each?

- Type I:
  - $Pr(Reject\ H_0 | H_0\, True)$

Notes on Type I and Type II error

How to calculate each?

- Type I:
  - $Pr(Reject\ H_0 | H_0\, True) = \alpha$
  - This is our primary focus

Notes on Type I and Type II error

How to calculate each?

- Type I:
  - $Pr(Reject\ H_0 | H_0 True) = \alpha$
  - This is our primary focus

- Type II (hard):
  - $Pr(Fail\ to\ reject | H_0\ False)$
  - H_0 can be false for false for many values

Notes on Type I and Type II error

How to calculate each?

- Type I:
  - $Pr(Reject\ H_0|H_0 True) = \alpha$
  - This is our primary focus

- Type II (hard):
  - $Pr(Fail\ to\ reject|H_0\ False)$
  - H_0 can be false for false for many values
  - "Power" of a test is $1 - Pr(Type\ II\ error)$

Notes on Type I and Type II error

How to calculate each?

- Type I:
  - $Pr(Reject\ H_0|H_0\,True) = \alpha$
  - This is our primary focus
- Type II (hard):
  - $Pr(Fail\ to\ reject|H_0\ False)$
  - H_0 can be false for false for many values
  - "Power" of a test is $1 - Pr(Type\ II\ error)$
  - Leave this for a more advance classes
  - There is a trade-off between Type I and Type II error