

Inference for the Regression Slope



Motivation

- Goal: Estimate a Linear Model for the Relationship Y versus X

$$Y = \beta_0 + \beta_1 X + \epsilon$$

- Strategy: Get Sample of 100 Observations and Fit Line

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X = 3 - 5X$$

- Question: *What would happen if we got a sample of a different 100 observations, and refit the line?*

Uncertainty in Estimated Slope

- Standard Error of the Slope

$$SE_{\hat{\beta}_1} = \sqrt{\frac{1}{n-2} * \frac{\sum(y_i - \hat{y})^2}{\sum(x_i - \bar{x})^2}} = \frac{\hat{\sigma}_\epsilon}{\sqrt{\sum(x_i - \bar{x})^2}}$$

- Measures How Much We Expect the Estimated Slope to Vary from One Sample to the Next
- Measures Our Uncertainty About $\hat{\beta}_1$ as an Estimate of β_1

Confidence Interval for Slope

- Goal: Give Range of Possible Values for β_1

- Formula for 95% Confidence Interval:

$$\hat{\beta}_1 \pm t_{0.025, n-2} * SE_{\hat{\beta}_1}$$

- Critical Value $t_{0.025, n-2}$ is the 97.5 percentile on the t-Distribution with $n - 2$ Degrees of Freedom
- This Positive and Negative Versions of the Critical Value Divide the **Appropriate** t-Distribution into Middle 95% and Outside 5%

t-Test for Slope

- Step 1: Choose a Significance Level α
 - Assume 0.05 if Not Specified
- Step 2: State Hypotheses (Null and Alternative)
 - $H_0: \beta_1 = 0$
 - $H_a: \beta_1 \neq 0$
- Step 3: Acquire Data and Perform Simple Linear Regression
 - $\hat{\beta}_1 = \text{Estimated Slope}$
 - $\hat{\sigma}_\epsilon = \text{Standard Error of Regression}$

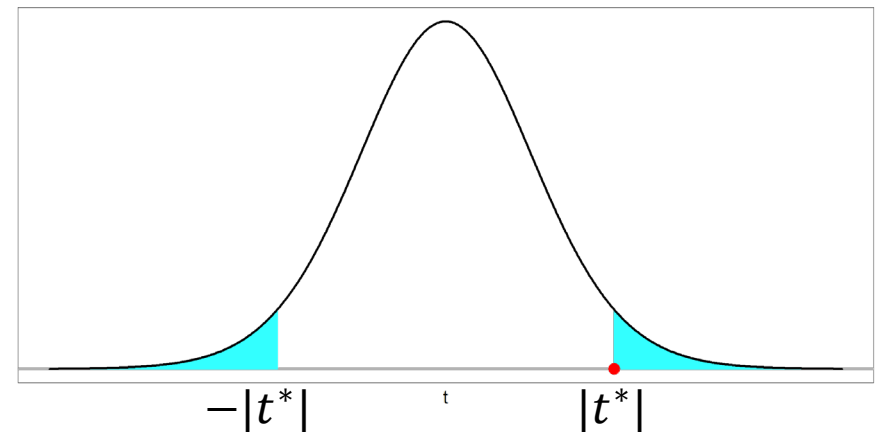
t-Test for Slope

- Step 4: Test Statistic

- $t^* = \widehat{\beta}_1 / SE_{\widehat{\beta}_1}$

- Step 5: P-value

- Use t-Distribution with $n - 2$ Degrees of Freedom
- Two-Sided or Non-Directional
- P-value = Shaded Area



t-Test for Slope

- Step 6: Decision
 - $P\text{-value} < 0.05$, then Reject Null and Accept Alternative
 - $P\text{-value} > 0.05$, then Fail to Reject the Null
- Step 7: Interpret Results to Audience
 - Interpret for People with Background in Basic Math
 - Should Use Words that are Based off the Context of the Data
 - What Does “Statistically Significant” Mean to the Audience

t-Test for Slope

- Step 6: Decision
 - $P\text{-value} < 0.05$, then Reject Null and Accept Alternative
 - $P\text{-value} > 0.05$, then Fail to Reject the Null
- Step 7: Interpret Results to Audience
 - Interpret for People with Background in Basic Math
 - Should Use Words that are Based off the Context of the Data
 - What Does “Statistically Significant” Mean to the Audience

Conclusion

- Learning What Is Classically Meant When a Person States a Predictor Variable to be “Statistically Significant”
- We Can Use the Confidence Interval to Conduct a Hypothesis Test
- A Confidence Interval is Superior to a Hypothesis Test
- **Always** Supplement Your Conclusion from a Hypothesis Test with a Confidence Interval

Make Reasonable Decisions

