### Inference in Simple Linear Regression

- Inference About the Slope
- Inference About the Intercept
- Mean Value of Response at Specific Value of Predictor
- Prediction of a New Response at Specific Value of Predictor

#### Four Types of Inference in Simple Linear Regression

1. **Inference about the slope**
   - Is there a _________________ between the predictor and response?

2. **Inference about the intercept**
   - What can we expect the response to be when the _________________?

3. **Mean value of $Y$ at a specified value of $X_0$**
   - What can we expect the _________________ to be at some value of the predictor?

4. **Prediction of a new value of $Y$ at $X_0$**
   - What can we expect an _________________ to be at some value of the predictor?

#### Example: Inferential Questions

- **Scenario:** Use ACT score of 29 college freshmen to describe their freshman year GPA.

- **Inferential Questions:**
  1. Is ACT score a significant predictor of GPA?
  2. What are plausible values for the slope of the regression line?
  3. Is .05 a plausible increase in GPA for an additional point in ACT score?
  4. Does the intercept of the regression line differ from zero?
  5. What are plausible values for the intercept of the regression line?
  6. What are plausible values for GPA for all students who scored 25 on the ACT?
  7. Is 3.30 a plausible GPA for all students who averaged 25 on the ACT?
  8. What are plausible values for GPA for a single student who scored 25 on the ACT?
Inference About the Slope

• Performing inference on the slope allows us to learn more about the relationship between \( X \) and \( Y \)

\[
Y = \beta_0 + E
\]

\( \beta_1 = 0 \)

\[
Y = \beta_0 + \beta_1 X + E
\]

\( \beta_1 \neq 0 \)

• Two types of tests:
  • **Test of Zero Slope**: Determines if \( X \) is a significant predictor of \( Y \)
    • Hypotheses: \( H_0: \beta_1 = 0 \) vs. \( H_A: \beta_1 \neq 0 \)
  • **Test of Nonzero Slope**: Determines if some value other than zero is a plausible value for the slope of the regression line
    • Hypotheses: \( H_0: \beta_1 = \beta_1^{(0)} \) vs. \( H_A: \beta_1 \neq \beta_1^{(0)} \)

Inference About the Slope

• To test if the regression line helps to describe the relationship between the predictor \( X \) and the response \( Y \):
  • **Test Statistic**: \( t = \frac{\hat{\beta}_1 - \beta_1^{(0)}}{\frac{S_{Y|X}}{s_X} \sqrt{n-1}} \) which has \( n-2 \) degrees of freedom
    • Note: In the test for zero slope, \( \beta_1^{(0)} = 0 \)
    • **Confidence Interval**: \( \hat{\beta}_1 \pm t_{n-2} \times \frac{S_{Y|X}}{s_X \sqrt{n-1}} \)

• If 0 is not a plausible value for the slope, then \( X \) and \( Y \) are said to have a **significant linear relationship**.

**Note**: It is most common to test if the slope of the regression line differs significantly from 0, but any hypothesized slope and any alternative (\(<\), \(>\), or \(\neq\)) can be tested.

Inferential Question #1: Test for Zero Slope

• **Scenario**: Use ACT score of 29 college freshmen (without outlier) to describe freshman year GPA.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total Count</th>
<th>Mean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT</td>
<td>29</td>
<td>24.172</td>
<td>4.277</td>
</tr>
<tr>
<td>GPA</td>
<td>29</td>
<td>2.974</td>
<td>0.625</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.987</td>
<td>0.570</td>
<td>1.73</td>
<td>0.095</td>
</tr>
<tr>
<td>ACT</td>
<td>0.0822</td>
<td>0.0232</td>
<td>3.53</td>
<td>0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( S )</th>
<th>R-sq</th>
<th>R-sq(adj)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5261</td>
<td>31.64%</td>
<td>29.11%</td>
</tr>
</tbody>
</table>

• **Task**: Test if the slope of the regression line differs from zero.

• **Hypotheses**: \( H_0: \) ________ vs. \( H_A: \) ________

• **Test Statistic**: ____________________________

• **Critical Values**: ________________; **P-Value**: __________

• **Conclusion**: __________ and conclude GPA and ACT score ________ ___________ (i.e. ________)

(i.e. ________)
Inferential Question #1: Test for Zero Slope

**Scenario:** Use ACT score of 29 college freshmen (without outlier) to describe freshman year GPA.

**Question:** What can we deduce about the relationship between $X$ and $Y$ based on the test for the slope?

**Answer:** One of the following is guaranteed to be true:

1. ACT score is a __________________ of GPA; that is, the regression line $\hat{Y} = .987 + .0822X$ is a better prediction of GPA than the __________.
2. There is evidence that a student's ACT is ________________ to their freshman year GPA, although this relationship may only be a __________ _________________.

Based on the scatterplot, it appears as if ____ is true.

**Question:** What happens in a situation where (2) is true?

Example: Curved Relationship

**Scenario:** Use income of husband (in thousands of dollars) to describe amount of time after marriage (in months) before birth of first child.

**Question:** What can we deduce about the relationship if testing $H_0: \beta_1 = 0$ vs. $H_A: \beta_1 \neq 0$?

**Answer:**

- Income is a __________________ of time to birth (__________________)
- The straight line model is __________________ due to the __________ relationship seen in the scatterplot

Inferential Question #2: Confidence Interval for the Slope

**Scenario:** Use ACT score of 29 college freshmen (without outlier) to describe freshman year GPA.

**Task:** Construct a 95% confidence interval for the slope.

**Critical Value:** __________________

**Confidence Interval:**

<table>
<thead>
<tr>
<th>Term</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>19.02</td>
<td>5.14</td>
<td>3.70</td>
<td>0.002</td>
</tr>
<tr>
<td>Income (Thousands)</td>
<td>0.1499</td>
<td>0.0688</td>
<td>2.18</td>
<td>0.043</td>
</tr>
</tbody>
</table>

**Question:** How should this interval be interpreted?

**Answer:** We are ____________ that for every additional point a student scores on the ACT, their ____________________________.
Inferential Question #2: Confidence Interval for the Slope

- **Scenario:** Use ACT score of 29 college freshmen (without outlier) to describe freshman year GPA.
- **Question:** What decision would be made if testing \( H_0: \beta_1 = 0 \) vs. \( H_A: \beta_1 \neq 0 \)?
- **Answer:** __________________
  - Confidence interval ____________________, implying that 0 is ___________ ________________ for the slope
- **Question:** What can we deduce about the relationship between \( X \) and \( Y \) based on the test/confidence interval for the slope?
- **Answer:** There is a ____________________________ between ACT score and freshman year GPA.

Inferential Question #3: Test for Nonzero Slope

- **Scenario:** Use ACT score of 29 college freshmen (without outlier) to describe freshman year GPA.
- **Task:** Test if the slope of the regression line is greater than .05.

<table>
<thead>
<tr>
<th>Term</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.987</td>
<td>0.570</td>
<td>1.73</td>
<td>0.095</td>
</tr>
<tr>
<td>ACT</td>
<td>0.0822</td>
<td>0.0232</td>
<td>3.53</td>
<td>0.001</td>
</tr>
</tbody>
</table>

- **Hypotheses:** \( H_0: \) __________ vs. \( H_A: \) __________
- **Test Statistic:** ____________________________
- **Critical Values:** ____________ ; **P-Value:** ____________
- **Conclusion:** ____________ and conclude GPA ____________ ____________ for every additional ACT point.

Inference About the Intercept

- **Hypothesized Intercept:** \( \beta_0^{(0)} \)
- To test where the regression line passes through the Y-axis:
  - **Hypotheses:** \( H_0: \beta_0 = \beta_0^{(0)} \) vs. \( H_A: \beta_0 \neq \beta_0^{(0)} \)
  - **Test Statistic:** \( t = \frac{\hat{\beta}_0 - \beta_0^{(0)}}{SE(\beta_0)} \) which has \( n - 2 \) degrees of freedom
  - **Confidence Interval:** \( \hat{\beta}_0 \pm t_{n-2,1-\alpha/2} \times SE(\beta_0) \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{(n-1)s_X^2}} \)

**Note:** It is most common to test if the intercept of the regression line differs significantly from 0, but any hypothesized intercept and any alternative (<, >, or ≠) can be tested.
Inferential Question #4: Test for Zero Intercept

**Scenario:** Use ACT score of 29 college freshmen (without outlier) to describe freshman year GPA.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Count</th>
<th>Mean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT</td>
<td>29</td>
<td>24.172</td>
<td>4.277</td>
</tr>
<tr>
<td>GPA</td>
<td>29</td>
<td>2.974</td>
<td>0.625</td>
</tr>
</tbody>
</table>

**Task:** Test if the intercept of the regression line differs from zero.

**Hypotheses:**
- $H_0$: ______ vs. $H_A$: ______

**Test Statistic:** __________________________

**Critical Values:** ___________; P-Value: ______

**Conclusion:** __________ and conclude _____________

Inferential Question #5: Confidence Interval for Intercept

**Scenario:** Use ACT score of 29 college freshmen (without outlier) to describe freshman year GPA.

**Task:** Construct a 95% confidence interval for the intercept.

**Critical Value:** __________

**Confidence Interval:** ______________________________________________________________________________

**Question:** Is this result consistent with the hypothesis test?

**Answer:** ______
- **Hypothesis Test:** ____________ → ____________
- **Confidence Interval:** ________________ → ________________

Inference for the Regression Line

**Regression line provides the best estimate of the response $Y$ at a specific value of $X_0$ by calculating $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_0$**

**Due to natural variability in the population, not every response will be equal to $\hat{Y}$**

**Often desirable to do inference to answer two different questions:**

1. **What can we expect the average response to be for all observations at a specific value of the predictor $X_0$?**
   - Find a **confidence interval** for $\mu_{Y|X_0}$

2. **What can we expect an individual response to be at a specific value of the predictor $X_0$?**
   - Find a **prediction interval** for $\mu_{Y|X_0}$
Mean Value of $Y$ at a Specified Value of $X$

- To estimate the average response $\bar{Y}$ for all observations at a specific value $X_0$:
  - **Confidence Interval**: $\hat{Y} \pm t_{n-2} \times S_{Y|X} \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{(n-1)S_X^2}}$

- To test if $\mu_{Y|X}$ is a plausible value for the average response $\bar{Y}$ at $X_0$:
  - **Hypotheses**: $H_0: \mu_{Y|X} = \mu_{Y|X}^{(0)}$ vs. $H_A: \mu_{Y|X} \neq \mu_{Y|X}^{(0)}$
  - **Test Statistic**: $t = \frac{\bar{Y}_{X_0} - \mu_{Y|X}^{(0)}}{S_{Y|X} \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{(n-1)S_X^2}}}$ which has $n - 2$ degrees of freedom

Inferential Question #6: CI for Mean Value of $Y$ at $X_0$

- **Scenario**: Use ACT score of 29 college freshmen (without outlier) to describe freshman year GPA.
  - **Term** | **Coef** | **SE Coef**
  - Constant | 0.987 | 0.570
  - ACT | 0.0822 | 0.0232

- **Question**: What is a 95% confidence interval for the average GPA for all students who scored 25 on the ACT?
- **Predicted Value**: _____________________________
- **Confidence Interval**: _____________________________

Inferential Question #7: Test for Mean Value of $Y$ at $X_0$

- **Scenario**: Use ACT score of 29 college freshmen (without outlier) to describe freshman year GPA.
  - **Term** | **Coef** | **SE Coef**
  - Constant | 0.987 | 0.570
  - ACT | 0.0822 | 0.0232

- **Task**: Test if the average GPA for all students who scored 25 on the ACT is greater than 3.30.
- **Hypotheses**: $H_0$: _________________ vs. $H_A$: _________________
- **Test Statistic**: _____________________________
Inferential Question #7: Test for Mean Value of Y at \( X_0 \)

- **Scenario:** Use ACT score of 29 college freshmen (without outlier) to describe freshman year GPA.
- **Task:** Test if the average GPA for all students who scored 25 on the act is greater than 3.30.
- **Critical Value:** ______________________; **P-Value:** __________
- **Conclusion:** __________ and conclude that the ____________ of students ________________ is ________________.
- **Question:** Is this result consistent with the confidence interval?
- **Answer:** ______
  - **Confidence Interval:** ______________________ → ______________________
  - **Hypothesis Test:** Concluded that ______________________

Prediction of a New Value of Y at \( X_0 \)

- To estimate the average response \( Y \) for all observations at a specific value \( X_0 \):
  - **Prediction Interval:** \( \hat{Y} \pm t_{n-2,1-\alpha/2} \times S_{Y|X} \sqrt{1 + \frac{1}{n} + \frac{(X_0-\bar{X})^2}{(n-1)S_X^2}} \)

**Notes:**
- Because predicting an individual response has more variability than the mean of all observations, the extra ‘1’ gets added under the radical.
- Because we are estimating an individual response \( Y \) (rather than a parameter \( \mu_{Y|X} \)), the interval must be called a prediction interval.

Inferential Question #8: P.I. for New Value of Y at \( X_0 \)

- **Scenario:** Use ACT score of 29 college freshmen (without outlier) to describe freshman year GPA.

<table>
<thead>
<tr>
<th>Term</th>
<th>Coef</th>
<th>SE Coef</th>
<th>S</th>
<th>R-sq</th>
<th>R-sq(adj)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.987</td>
<td>0.570</td>
<td>0.5261</td>
<td>31.64%</td>
<td>29.11%</td>
</tr>
<tr>
<td>ACT</td>
<td>0.0822</td>
<td>0.0232</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total</th>
<th>Count</th>
<th>Mean</th>
<th>StdDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT</td>
<td>29</td>
<td>24.172</td>
<td>4.277</td>
<td></td>
</tr>
<tr>
<td>GPA</td>
<td>29</td>
<td>2.974</td>
<td>0.625</td>
<td></td>
</tr>
</tbody>
</table>

- **Question:** What is a 95% prediction interval for the GPA of a single student a single student who scored 25 on the ACT?
- **Predicted Value:** __________________________
- **Prediction Interval:** __________________________
Example: Interpretation of Intervals

- **Scenario:** Use ACT score of 29 college freshmen (without outlier) to describe freshman year GPA.
- **Question:** What does the confidence interval mean in context?
- **Answer:** We are 95% confident that the ____________________________
  ________________________________________________________________.
- **Question:** What does the prediction interval mean in context?
- **Answer:** We are 95% confident that the ____________________________
  ________________________________________________________________.

Confidence Bands and Prediction Bands

- **Confidence Band:** a curved line drawn on a scatterplot representing the set of all confidence intervals for a regression
- **Prediction Band:** a curved line drawn on a scatterplot representing the set of all prediction intervals for a regression

Widths of Confidence Bands and Prediction Bands

1. Both confidence intervals and prediction intervals get _____ as \( X_0 \) gets _______ from \( \bar{X} \)

   \[
   \text{CI: } \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{(n-1)s_X^2}} \quad \text{PI: } \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{(n-1)s_X^2}}
   \]

2. Prediction intervals are always _____ than confidence intervals at the _____________ of \( X_0 \)

   \[
   \text{CI: } \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{(n-1)s_X^2}} \quad \text{PI: } \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{(n-1)s_X^2}}
   \]

Extra ‘1’ makes margin of error ________