

## Linear Regression Models: Active Learning Worksheet

*Follow-Along Questions for Video Learning*

Name: \_\_\_\_\_

Date: \_\_\_\_\_

### Learning Objectives

- Calculate a predicted response value using a linear regression model
- Understand the components of the linear regression equation  $\hat{y} = a + bx$
- Recognize the dangers of extrapolation in making predictions

## Part 1: Video 1 - Building Understanding

### Pre-Video Activation (0:00)

**Before watching:** What do you already know about linear equations from algebra? Write the slope-intercept form:

### Context Building (0:25-0:58)

1. **Summarize the real-world question:** What is Linda Saledo investigating about supermarkets in San Antonio?

2. **Identify the variables:**

- Explanatory variable (x): \_\_\_\_\_
- Response variable (y): \_\_\_\_\_

### Core Concepts (2:22-3:15)

3. **Compare and contrast:** Fill in the table below as you watch:

	Algebra	Statistics
Equation	$y = mx + b$	
Y-value represents	Exact value	
Slope symbol		
Y-intercept symbol		

4. **Critical thinking:** Why do we use  $\hat{y}$  (y-hat) instead of just  $y$  in statistics?

## Application (3:15-4:57)

5. **Record the model:** Write Linda's linear regression equation:
  6. **Practice prediction:** Show your work for predicting organic items when income = \$90,000:
  7. **Conceptual understanding:** The prediction gives 75.3 organic items. Explain why this decimal makes sense even though you can't have 0.3 of an item:

## Critical Analysis (4:57-5:36)

8. **Preview question:** The study predicts 100% of Americans will be overweight by 2048. What concerns you about this prediction? (We'll explore this more in Video 2)

## Part 2: Video 2 - Deepening Understanding

## Extrapolation Concept (0:18-2:29)

9. **Define in your own words:** What is extrapolation?
  10. **Explain the danger:** Why might the overweight prediction for 2048 be unreliable?
  11. **Apply to non-time data:** How does extrapolation apply to Linda's income/organic items study?

**AP Exam Practice - Swine Population Problem****Part A: Scatter Plot (3:17-5:20)**

12. **Annotation practice:** As you watch, identify:

- Motivation for the problem: \_\_\_\_\_
- Explanatory variable: \_\_\_\_\_
- Response variable: \_\_\_\_\_

13. **Scoring criteria (E-P-I):** List the three requirements for an "E" (Essentially correct) scatter plot:

- i. \_\_\_\_\_
- ii. \_\_\_\_\_
- iii. \_\_\_\_\_

**Part B: Correlation Interpretation (5:20-6:22)**

14. **Three components:** When interpreting  $r = 0.85$ , what three things must you include?

- i. \_\_\_\_\_
- ii. \_\_\_\_\_
- iii. \_\_\_\_\_

15. **Write your interpretation:** Using all three components for  $r = 0.85$ :

**Part C: Model Appropriateness (6:22-7:33)**

16. **Two-part justification:** Why is a linear model appropriate? Base your answer on:

- Scatter plot: \_\_\_\_\_
- Correlation coefficient: \_\_\_\_\_

**Part D: Prediction and Reliability (7:33-9:56)**

17. **Calculation check:** Why do we use 0.2 instead of 200 for the swine population?

18. **Prediction:**  $\hat{y} = \text{_____}$  parts per million

19. **Reliability assessment:**

- Is this prediction reliable? \_\_\_\_\_

- Why or why not? (Use the term "extrapolation" in your answer)

## Synthesis Questions

### Checking Understanding

Answer these questions after watching both videos:

1. **Compare differences:** What's the key difference between  $y$  in algebra and  $\hat{y}$  in statistics?
2. **Identify weaknesses:** Name two situations where a linear regression model might give unreliable predictions:
  - a. \_\_\_\_\_
  - b. \_\_\_\_\_
3. **Create your own example:** Write a real-world scenario where extrapolation would be dangerous:
4. **AP Exam strategy:** List three key strategies for free response questions:
  - i. \_\_\_\_\_
  - ii. \_\_\_\_\_
  - iii. \_\_\_\_\_

## Self-Assessment

### Reflect on Your Learning

Rate your understanding (1-5 scale, 5 being strongest):

- I can write and interpret the linear regression equation  $\hat{y} = a + bx$
- I can calculate predicted values using a regression model
- I understand why extrapolation is dangerous
- I can identify when a prediction is reliable vs. unreliable
- I know what to include for full credit on AP exam questions

**One question I still have:** \_\_\_\_\_

---

## Key Takeaways Box

### Remember the Statistician's Mantra:

"Be critical. Be cautious. Be compassionate. Avoid BS."

### Essential Formulas:

- Linear Regression Model:  $\hat{y} = a + bx$
- $a$  = y-intercept (predicted value when  $x = 0$ )
- $b$  = slope (change in  $\hat{y}$  per unit change in  $x$ )
- $\hat{y}$  = predicted response value (not exact!)