

## Video 1: Calculating and Interpreting Residuals

### Part A: Building Understanding (0:00 - 1:37)

- Context Connection:** The study examines supermarkets in San Antonio, Texas. Before watching, predict: Would you expect stores in wealthier neighborhoods to have more or fewer organic food options? Why?
- Data Interpretation (1:13-1:37):** As you view the scatter plot:
  - The regression equation is  $\hat{y} = 0.0007x + 2.76$ . What does the slope of 0.0007 tell us in context?
  - For every \$10,000 increase in average household income, how many additional organic items would we expect?

### Part B: Understanding Residuals (1:37 - 3:35)

- Definition Focus (1:37-1:57):** Complete the formula:

$$\text{Residual} = \underline{\hspace{2cm}} - \underline{\hspace{2cm}} = y - \hat{y}$$

- Calculation Practice (2:28-3:23):** Work along with the example:

- Given: Income = \$66,703, Actual organic items = 84
- Show the prediction calculation:

$$\hat{y} = 0.0007(\underline{\hspace{2cm}}) + 2.76 = \underline{\hspace{2cm}}$$

- Calculate the residual:

$$\text{Residual} = \underline{\hspace{1cm}} - \underline{\hspace{1cm}} = \underline{\hspace{1cm}}$$

- Interpretation (3:23-3:53):** Complete this interpretation table:

Residual Sign	Meaning	Actual vs. Predicted
Positive (+)	Model _____	Actual ____ Predicted
Negative (-)	Model _____	Actual ____ Predicted

- Critical Thinking:** A store has a residual of -15.2 organic items. Write a complete sentence interpreting this in context.

**Part C: Constructing Residual Plots (3:53 - 5:22)**

7. **Axes Understanding (4:03-4:42):** In a residual plot:

- (a) What variable is on the x-axis? \_\_\_\_\_
- (b) What is plotted on the y-axis? \_\_\_\_\_
- (c) Where is the reference line drawn? At  $y =$  \_\_\_\_\_

8. **Plot Analysis:** Sketch what you expect to see in the residual plot space below. Mark the horizontal reference line clearly.



9. **Prediction (5:16-5:22):** Before watching Video 2, what might it mean if all the residuals were above the zero line? What about if they showed a curved pattern?

## Video 2: Assessing Model Fit with Residual Plots

### Part D: Recognizing Good Fit (1:26 - 2:23)

10. **Pattern Recognition:** When a linear model is a good fit, the residual plot shows:
- Apparent \_\_\_\_\_ in the residual plot
  - Scatter centered at \_\_\_\_\_
  - No clear \_\_\_\_\_
11. **Conceptual Understanding:** Why does “random noise” in a residual plot indicate a good model? (Hint: What has the model captured vs. what remains?)

### Part E: Recognizing Poor Fit (2:23 - 3:26)

12. **Pattern Detection (2:39-3:05):** Describe the pattern seen in the “bad fit” example:
13. **Diagnostic Skills:** Match each residual plot description with what it suggests:

Residual Plot Shows	Suggests
Random scatter around zero	A. Curvature in the relationship
U-shaped or inverted U pattern	B. Linear model is appropriate
Fan shape (spreading out)	C. Changing variability

### Part F: Application and Analysis (3:26 - 4:31)

14. **Real Data Analysis (3:26-3:57):** Looking at the grocery store residual plot:
- (a) What evidence suggests the model might be adequate?
- (b) What subtle patterns raise concerns about the fit?
15. **Decision Making:** You’ve created a linear model and the residual plot shows a clear curved pattern. What are your options? List at least two approaches:

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

**Part G: Synthesis and Reflection**

16. **Essential Knowledge Check:** Write the complete process for assessing a linear model:

Step 1: Fit the linear model and find  $\hat{y} =$  \_\_\_\_\_

Step 2: Calculate residuals using: \_\_\_\_\_

Step 3: Create a residual plot with \_\_\_\_\_ on x-axis and \_\_\_\_\_ on y-axis

Step 4: Look for \_\_\_\_\_ to confirm good fit, or \_\_\_\_\_ to identify poor fit

17. **Real-World Connection:** Why might it matter ethically if grocery stores in lower-income areas have fewer healthy food options? How could this analysis help address the issue?

18. **Extension Thinking:** If you were advising a grocery chain, what other variables besides income might predict organic food availability? List two and explain why:

**Quick Check (Complete after both videos)**

I. A residual of zero means:

- A. The model is perfect
- B. The predicted value equals the actual value for that point
- C. The linear model is inappropriate
- D. The point is an outlier

II. Random scatter in a residual plot is:

- A. Evidence the model is wrong
- B. Evidence of measurement error
- C. Evidence the linear model is appropriate
- D. Evidence we need more data

III. If residuals show a pattern, we should:

- A. Ignore it and use the model anyway

- B. Consider a different type of model
- C. Collect more data
- D. Remove the outliers

**Key Formulas to Remember:**

- Residual = Actual - Predicted =  $y - \hat{y}$
- Positive residual  $\rightarrow$  Model underestimated
- Negative residual  $\rightarrow$  Model overestimated
- Random residuals  $\rightarrow$  Good linear fit
- Patterned residuals  $\rightarrow$  Consider non-linear model