AP Statistics Unit 2.8: Least Squares Regression

Video Follow-Along Worksheet

Video 1: Determining the Least Squares Regression Line

Part A: Understanding Residuals (0:00 - 2:30)

1. **Context Connection:** At timestamp 0:33, the speaker discusses attendance and test scores. Before continuing:
   - What is the explanatory variable? ________________
   - What is the response variable? ________________
2. **Model Comparison (2:04):** When Models A and B are introduced, pause and predict:
   - Which model appears to fit better? Why?

     ____

     ____
3. **Critical Thinking (2:56):** Before the speaker reveals the problem with summing residuals:
   - What issue might arise if we simply add positive and negative residuals?

     ____

     ____

Part B: The Least Squares Method (3:00 - 5:00)

4. **Mathematical Understanding (3:57):** Why do we square the residuals instead of using absolute values?
   - List two mathematical advantages:
     1. ________________________________
     2. ________________________________
5. **Definition Check (4:27):** Complete the definition:

   The Least Squares Regression Line is the linear model that minimizes

   ____

Part C: Properties of the LSRL (5:08 - 7:05)

6. **Mean Point Property (5:08):** Calculate and verify:
   - Given: $\bar{x} = 86.4\%$ and $\bar{y} = 41.3$
   - The LSRL passes through the point: (__________, __________)
7. **Slope Formula (6:01):** Before calculating, write the formula:

$$b = \underline{\hspace{5cm}}$$

8. **Calculation Practice (6:22):** Using the given values:
   - $r = 0.95$, $s_y = 6.08$, $s_x = 10.2$
   - Show your work to find the slope:

$$\begin{aligned} b &= r \cdot \frac{s_y}{s_x} \\ &= \underline{\hspace{2.5cm}} \cdot \frac{\underline{\hspace{2.5cm}}}{\underline{\hspace{2.5cm}}} \\ &= \underline{\hspace{2.5cm}} \end{aligned}$$

Video 2: Interpreting Slope and Y-Intercept

Part A: Understanding the Equation (0:00 - 3:00)

1. **Equation Recognition (1:06):** Given $\hat{y} = -7.69 + 0.57x$
   - Identify the slope: __________
   - Identify the y-intercept: __________
2. **Notation Connection (3:06):** Match the statistical notation to algebraic notation:

Statistical	Algebraic
$\hat{y}$	__________
b (coefficient)	__________
a (constant)	__________

Part B: Interpretation in Context (3:55 - 5:35)

3. **Slope Interpretation (3:55):** Complete the interpretation:

> "For every 1 percentage point increase in ____________________,
> the model predicts an average _______________ of __________
> exam questions answered correctly."

4. **Y-Intercept Analysis (4:50):** Pause and think:
   - What would $x = 0$ mean in this context?

____

- Is this meaningful? Why or why not?

____

____

Video 3: Coefficient of Determination and Computer Output

Part A: Understanding R^2 (0:00 - 4:30)

1. **Model Comparison (1:20):** The "mean model" predicts $\hat{y} = 41.3$ for everyone.

   - Why is this called a model that "cannot use attendance"?

     ____

2. **Visual Analysis (3:07):** After seeing both models' squared residuals:

   - Which model has smaller squared residuals? ________________
   - What does this tell us? ________________________

3. R^2 **Interpretation (4:04):** Complete the interpretation:

   > "90.3% of the variation in ________________________
   > can be explained by the linear relationship with ____________________."

4. **Mathematical Connection (4:34):** If $R^2 = 0.903$, then:

   - $R = \pm\sqrt{\underline{\hspace{2cm}}} = \pm\underline{\hspace{2cm}}$
   - How do we determine the sign? ________________________

Part B: Reading Computer Output (5:19 - 7:26)

5. **Output Identification (5:48):** From the computer output table, locate:

   - The constant (y-intercept): __________
   - The attendance coefficient (slope): __________
   - The R^2 value: __________

6. **Equation Construction (6:24):** Write the complete regression equation from the output:

$$\hat{y} = \underline{\hspace{2cm}} + \underline{\hspace{2cm}} \cdot x$$

Synthesis Questions

I. **Properties Summary:** List three key properties of the LSRL:

   a. ____

   b. ____

   c. ____

II. **Strength Assessment:** If $R^2 = 0.25$, what does this tell us about the linear relationship?

____

____

III. **Real-World Application:** Describe a scenario where the y-intercept would be meaningful in context:

____

____

____

Key Takeaways Checklist

- ☐ I can explain why we minimize the sum of *squared* residuals
- ☐ I can calculate the slope using $b = r(s_y/s_x)$
- ☐ I can interpret slope and y-intercept in context
- ☐ I understand that the LSRL passes through $(\bar{x}, \bar{y})$
- ☐ I can interpret R^2 as proportion of variation explained
- ☐ I can extract regression components from computer output
- ☐ I can determine when a y-intercept is meaningful in context