Follow-Along Worksheet: Topic 2.9

Analyzing Departures from Linearity

Name: ______________________________ Period: __________

Part 1: Influential Points (Video 1)

Opening (0:00–0:29)

1. What three main concepts will this video cover?

   a) ____________________________________
   b) ____________________________________
   c) ____________________________________

Context and Setup (0:30–1:41)

2. What relationship is being explored with the San Antonio grocery store data?

   x-variable: ______________________________
   y-variable: ______________________________

3. Record the regression equation parameters:

   - y-intercept $\approx$ _____
   - slope $\approx$ _____

4. The mean income is $__________ and the mean number of items is _____.

   Key Insight: This point $(\bar{x}, \bar{y})$ always lies ____________________.

Low vs. High Leverage Points (1:42–3:41)

5. **Low leverage points** are close to _______________ on the x-axis.

   When removed, what happens to the regression line? __________________________

6. **High leverage points** are far from _______________ on the x-axis.

7. When three high leverage points were removed:

   - The y-intercept changed from approximately _____ to _____
   - This is a shift of _____ units!

8. **Definition:** High leverage points have unusually __________ or __________ x-values.

Outliers in Regression (3:43–4:58)

9. Looking at the green point in the simple example:
   - Is it a high leverage point? ________
   - Why or why not? ____

10. An **outlier** in regression has an unusually high magnitude ________________________.

11. When the outlier was removed, what happened to:
    - Correlation (r): ________________________
    - Coefficient of determination (R^2): ________________________

Types of Influential Points (5:02–6:03)

12. Complete the table for the three types of influential points:

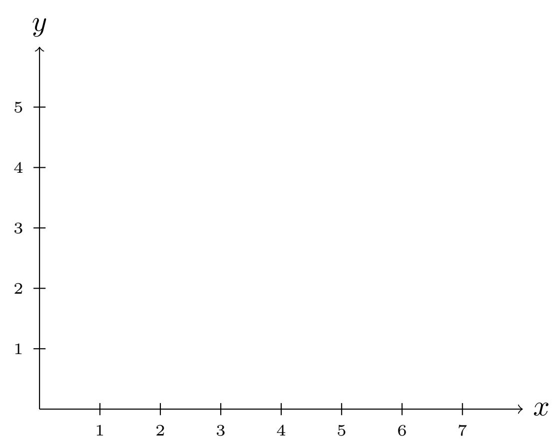
Type	**Primary Effect When Removed**
Outlier	Changes ____
High Leverage Point	Changes ____
Both Outlier AND High Leverage	Changes ____

13. **Synthesis:** An influential point is defined as any point that, if removed, substantially changes the ____________, ____________, and/or ____________.

Reflection Questions

14. Why might a point with an x-value close to $\bar{x}$ still be influential?

15. Sketch a scatter plot showing:
    - 5 points following a linear trend
    - 1 outlier that is NOT a high leverage point
    - 1 high leverage point that IS following the trend



Part 2: Transforming Data for Linearity (Video 2)

Opening and Context (0:00–1:14)

1. What will we learn about nonlinear relationships?
   a) What to do when data has a ____________________________
   b) Effects of _____________________ on datasets
   c) How to assess if transforming improved _____________________

2. The example explores the relationship between:
   - x: ____ (GDP per person)
   - y: ____

Initial Analysis (1:15–3:06)

3. Describe the association: ______________ (positive/negative/none)

4. Compare the United States and Japan:

Country	Income per Person	Life Expectancy
United States	$___________	_______ years
Japan	$___________	_______ years

   What does this comparison show? ____

5. For the original linear model:
   - $R^2 =$ _______%
   - The residual plot shows a _____________________ pattern (good/bad?)

6. What type of model does the speaker suggest might be more appropriate? _____________________

Data Transformation (3:07–4:11)

7. Why is income data typically right-skewed?

8. What does a log transformation do to high values?
   - Makes them _____________________
   - While preserving the _____________________ between values

9. The transformation applied was: $x_{new} =$ ______________

Comparing Models (4:12–5:14)

10. Complete the comparison table:

Measure	Untransformed	Log-Transformed
Scatter plot form	Curved	________________
Residual plot pattern	Shows pattern	________________
R^2 value	________%	________%

11. **Key Question:** How do we know the transformation improved the model? List two pieces of evidence:

    a) ____

    b) ____

Other Transformations (5:15–5:40)

12. Besides logarithm, what other transformations are mentioned?

    - ____________________
    - ____________________

13. In AP Statistics, will you typically transform data yourself? _______

    Instead, you'll assess model fit using ____________________ and _______ values.

Synthesis Questions

14. Why might a linear model be preferred even when data shows a curved pattern?

15. A student claims: "Since the R^2 went from 46.6% to 71.1%, the transformed model explains 71.1% of the variation in life expectancy."

    Is this interpretation correct? Explain carefully.

16. When would you choose to transform data? Circle all that apply:

    a) The residual plot shows a clear pattern

    b) The R^2 value is below 50%

    c) The original relationship appears nonlinear

    d) You want to use linear regression methods

    e) The data contains outliers

Practice Problem

17. A researcher studying the relationship between city population (x) and number of coffee shops (y) finds:

    - Original model: $\hat{y} = 12 + 0.003x$ with $R^2 = 0.42$
    - After log-transforming x: $\hat{y} = -85 + 42\log(x)$ with $R^2 = 0.78$

    a) Which model appears to be a better fit? Why?

    b) If a city has a population of 100,000, predict the number of coffee shops using the transformed model. (Use $\log(100000) \approx 5$)

Summary: Key Concepts Checklist

Check off each concept as you master it:

- ☐ I can identify high leverage points (unusual x-values)
- ☐ I can identify outliers in regression (large residuals)
- ☐ I can explain how influential points affect slope, y-intercept, and correlation
- ☐ I understand when to consider transforming data
- ☐ I can assess model fit using residual plots
- ☐ I can assess model fit using R^2 values
- ☐ I can interpret transformed regression models

Remember: Be critical, be cautious, be compassionate, and avoid bad statistics!