Friday Quiz 3 Solutions - STAT 324

Q3 - Point Analysis in Example 3

Question Text:
"Consider Example 3 (Delivery Times vs Customer Satisfaction). First, define what is meant by an outlier, a leverage point, and an influential point in the context of this customer service scenario.

Next, specifically evaluate the unusual data point mentioned in the text — the customer with a long delivery delay but high satisfaction. Discuss whether this point likely represents an outlier, has high leverage, and/or is influential on the relationship between delivery time and satisfaction.

Finally, briefly explain how the apparent number of data points (sample size) visible in the plot for Example 3 might affect your assessment of whether a single point like this one is considered influential."

Points Possible: 4

Solution Outline:

- **Definitions (contextualized):**
  - **Outlier:** A customer whose satisfaction score is unusual given their delivery time (i.e., large vertical distance from trend).
  - **Leverage Point:** A customer with an unusually long or short delivery time (i.e., extreme on x-axis).
  - **Influential Point:** A customer whose data changes the regression line substantially when included/excluded (often both an outlier and leverage point).
- **Evaluation of Specific Point:**

  - **Outlier:** Yes — it has a high satisfaction score despite a long delay, which deviates from the trend (negative residual).

  - **Leverage:** Maybe moderate — it's an extreme delivery time, but not extreme enough to dominate the x-axis range.

  - **Influence:** Likely influential — it's an outlier and possibly moderate leverage, so it may flatten the slope and affect regression interpretation.

- **Effect of Sample Size:**

  - Fewer points = greater influence per point. In a small dataset, one unusual point can shift the line more dramatically.

Rubric:

Component	Points
Clear definitions of outlier, leverage, and influential point	1
Correct outlier & leverage classification of the point	1
Insightful evaluation of influence	1
Discussion of sample size impact	1

Q7 - Normality Check of Residuals

Question Text:
"Referring to the histogram in Example 5 (Cholesterol Study), describe whether the residuals appear normally distributed or skewed, and justify your answer based on the shape of the histogram. Mention whether this histogram fits the normal assumption of the linear regression model."

Points Possible: 4

Solution Outline:

- **Distribution Shape:** The histogram is **right-skewed** (long tail to the right).
- **Justification:** Most residuals cluster around zero with a longer positive tail. The distribution is not symmetric or bell-shaped.
- **Normality Assumption:** Violated — residuals should be approximately normal for valid inference (confidence intervals, p-values, etc.).

- **Conclusion:** Since the skew is visible, model assumptions may not hold. This limits the reliability of regression outputs.

Rubric:

Component	Points
Identifies skewness (right/positive)	1
Provides shape-based justification	1
States this violates the normality assumption	1
Full answer is coherent and well-explained	1

Q10 - Log10 vs Square Root Transformation Interpretation

Question Text:
"Explain what happens to large and small values when a log10 or square root transformation is applied. In particular, discuss how transformations affect the interpretation of slopes and fitted values in models like those discussed in the examples."

Points Possible: 4

Solution Outline:

- **Compression of Values:**
  - Large values are compressed more than small ones. Log10 and $\sqrt{}$ reduce the range and skewness.
  - For log10, huge gaps become small (e.g., 10 to 1000 = just 1 to 3).
- **Slope Interpretation:**
  - **Transformed Y (log10(Y) ~ X):** A 1-unit X increase corresponds to a multiplicative change in Y.
  - **Transformed X (Y ~ log10(X)):** A 10× increase in X causes a unit change in Y (if log10).
  - **Both transformed:** Slope is an elasticity (percent change in Y per percent change in X).

- **Fitted Values:**
  - Model output is on the transformed scale.
  - To interpret in original units, apply inverse functions (e.g., $10^{\hat{y}}$ or $\hat{y}^2$).
  - Back-transformed predictions approximate the **median**, not the mean (due to Jensen's inequality).

Rubric:

Component	Points
Explains compression effect on values	1
Correct slope interpretation for at least one form	1
Discusses back-transformation of fitted values	1
Clarity and completeness of response	1