Homework 4 Solutions - STAT 324

Question 7

Using the Cereal dataset, where Sugar is used to predict Calories without any transformation, examine the top 10 points ranked by Cook's distance (highest to lowest). You are shown only the 10 most influential points based on the model Calories ~ Sugar. Do you notice evidence of strong outliers, highly influential points, or high-leverage observations? Explain your reasoning based on the pattern you observe in the data, even though exact Cook's distance or leverage thresholds are not provided.

Solution (Instructor Notes):

- In the Cereal model (`Calories ~ Sugar`), the **top Cook's distance points** stand out but **are not extreme**. There are 3 points with Cook's distance above 0.111, these same points are considered to be outliers as well. Seven point have high leverage.
- Most points have moderate Cook's distances; **a few points show higher influence** but not enough to distort the overall model fit.
- **Reasoning:** The linear model is fairly stable; extreme outliers or highly influential leverage points are not dominant.

Rubric (Out of 1.75 Points)

Component	Points	Description
Recognizes that there are some influential points but no extreme distortion based on the the 0.111 thres hold	0.75	Correct observation about influence patterns.

Component	Points	Description
Notices 7 out of the points have high leverage	0.50	Student points out 0.05 is the threshold for high leverage.
Reasonable reasoning (model remains stable, residuals balanced, no major leverage issues)	0.50	Explains why influence is not a major concern.

Partial credit: If student identifies influence correctly but confuses Cook's distance interpretation a little.

Question 12

After transforming the wt variable to log10(wt), examine the top 10 points with the highest Cook's distance. You are shown only the 10 most influential points ranked by Cook's distance (highest to lowest). Did the transformation reduce the number of strong outliers, influential points, or high-leverage observations? Explain your reasoning based on patterns you observe, even though exact Cook's distance or leverage thresholds are not provided.

Solution (Instructor Notes):

- After transforming `wt` to `log10(wt)`, the frequency of **strong influential points** (those with high Cook's distance) slightly **decreased**.
- Cook's distances for the top 10 points are **lower** after the transformation (maximum dropped from ~0.532 to ~0.262). There are still points with high Cook's distance, but they are **less extreme**.
- There are **still points** with relatively high leverage (e.g., `.hat` values around 0.13 and 0.191), but they **are not extreme**.
- **Outliers in residuals** are slightly less extreme after transformation (e.g., std. residuals are smaller in magnitude).
- **Summary:** Transformation **reduced** the overall influence and outlier severity slightly, but did **not eliminate** all leverage or influential points completely.

Rubric (Out of 1.75 Points)

Component	Points	Description
Identifies that Cook's distance magnitudes decreased after transformation	0.50	Clearly states that top influence points have lower Cook's distances post-transformation.
Discusses leverage points appropriately (some leverage remains but less extreme)	0.50	Recognizes presence of moderate leverage values, but no major outliers.
Discusses outliers (residuals closer to zero, fewer large std.resid values)	0.50	Mentions reduction in extreme residuals (outliers smaller after transformation).
Clear and logical explanation overall	0.25	Organized, direct answer referencing frequency and change clearly.

Partial credit: If student mentions only influence or only leverage but not both.