Types of Outliers

Mini-Assignment - MTH 361 A/B - Spring 2023

Instructions:

- Please provide complete solutions for each problem. If it involves mathematical computations, explanations, or analysis, please provide your reasoning or detailed solutions.
- Note that some problems have multiple solutions or ways to solve it. Make sure that your solutions are clear enough to showcase your work and understanding of the material.
- Creativity and collaborations are encouraged. Use all of the resources you have and what you need to complete the mini-assignment. Each student must take personal responsibility and submit their work individually. Please abide by the University of Portland Academic Honor Principle.
- **Please save your work as one pdf file, don't put your name in any part of the document, and submit it to the Teams Assignments for this course. Your document upload will correspond to your name automatically in Teams.**
- If you have questions or concerns, please feel free to ask the instructor.

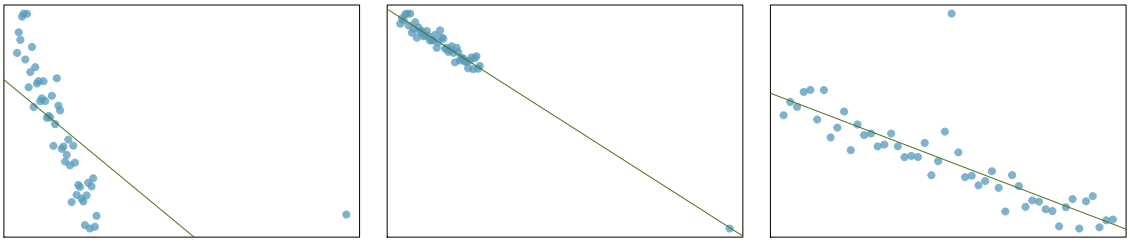
I. Outliers in Linear Regression

Materials

The exercises below are derived from the textbook OpenIntro Statistics (4th edition) by David Diez, Mine Cetinkaya-Rundel, and Christopher Barr.

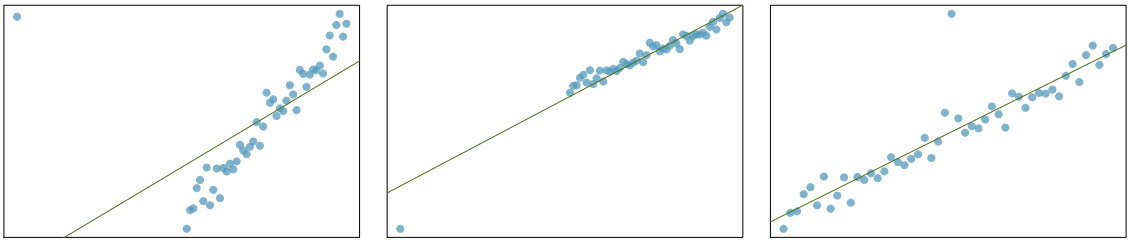
Exercises

1. **Outliers, Part I.** Identify the outliers in the scatterplots shown below, and determine what type of outliers they are. Explain your reasoning.



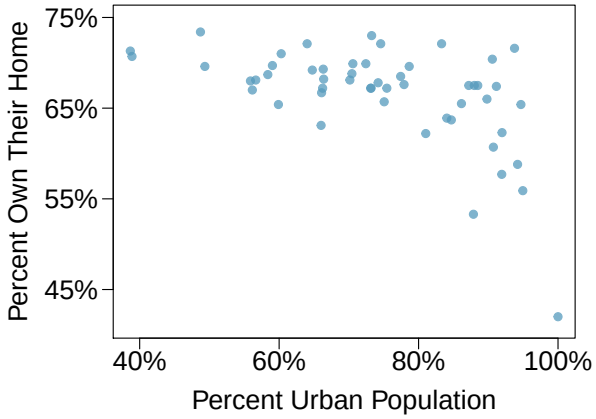
(a) (b) (c)

2. **Outliers, Part II.** Identify the outliers in the scatterplots shown below and determine what type of outliers they are. Explain your reasoning.



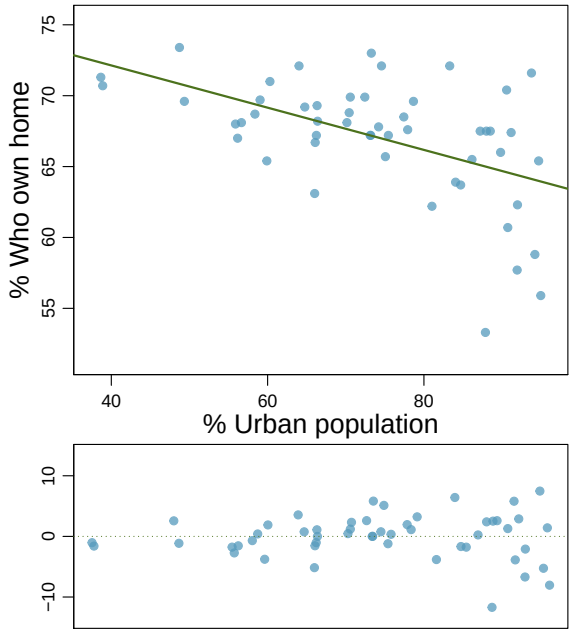
(a) (b) (c)

3. (Outstanding Question) **Urban homeowners** The scatterplot below shows the percent of families who own their home vs. the percent of the population living in urban areas. ("2010 Census Urban and Rural Classification and Urban Area Criteria," n.d.) There are 52 observations, each corresponding to a state in the US. Puerto Rico and District of Columbia are also included.



   a. Describe the relationship between the percent of families who own their home and the percent of the population living in urban areas.
   b. The outlier at the bottom right corner is District of Columbia, where 100% of the population is considered urban. What type of an outlier is this observation?

Below is a similar scatterplot, excluding District of Columbia, as well as the residuals plot. There were 51 cases.



   c. For these data, $R^2 = 0.28$. What is the correlation? How can you tell if it is positive or negative?
   d. Examine the residual plot. What do you observe? Is a simple least squares fit appropriate for these data?

References

2010 census urban and rural classification and urban area criteria. (n.d.). In *United States Census Bureau, Housing Characteristics: 2010*.