Descriptive Statistics
COR1-GB.1305 – Statistics and Data Analysis

Types of Data

1. The class survey asked each respondent to report the following information: gender; birth date; GMAT score; undergraduate major; time spent studying per week; interest level in the course; industry; job type; expected starting salary; number of dinners out per month; number of pairs of shoes; cups of coffee consumed per week; number of websites visited per day; political party; and presidential vote.
   (a) Which of the variables measured by the survey are categorical/qualitative?

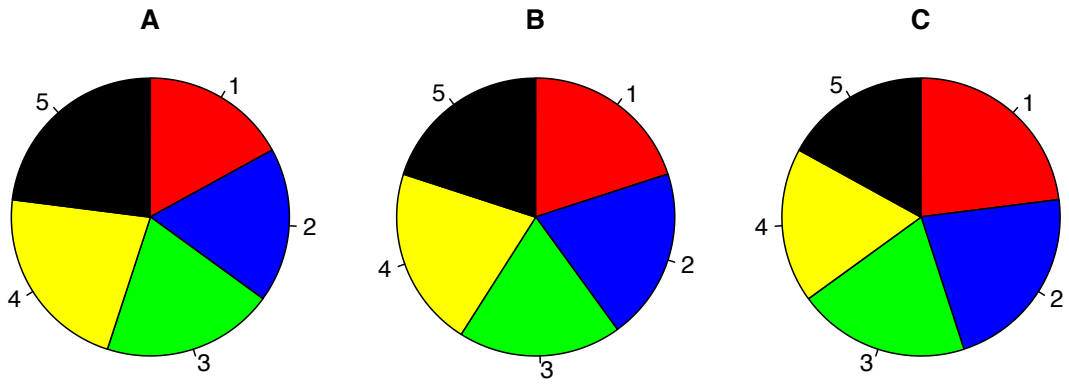
   (b) Which of the variables measured by the survey are numerical/quantitative?

2. What type of variable is the answer to the phone prompt "Enter '1' for English, '2' for Spanish."? Why?

3. Each Yelp restaurant includes a star rating (1–5). What type of variable is the star rating?

Describing Categorical (Qualitative) Data

4. Use the following pie charts to rank the categories (1–5) by size.



5. List two methods to describe the reported undergraduate majors of the class survey respondents.

6. Draw what you think the bar chart for the birth months of the survey respondents will look like.

Describing Numerical (Quantitative) Data

7. Draw what you think the histogram for "Websites Visited per Day" will look like.

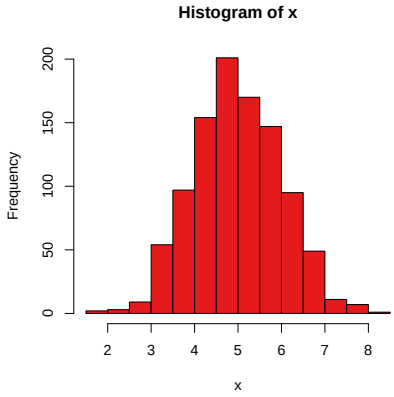
8. Draw what you think the histogram for "Dinners per Month" will look like.

9. Draw what you think the histogram for "Interest in this Class" will look like.

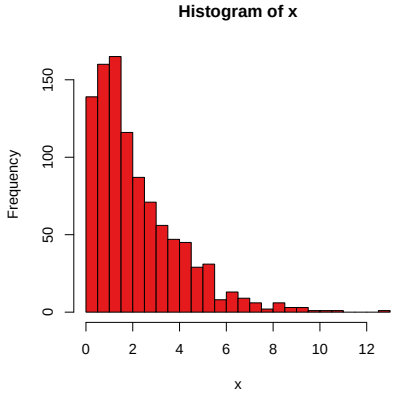
Measures of Central Tendency

10. Here are some histograms. Estimate the mean and median of the data.

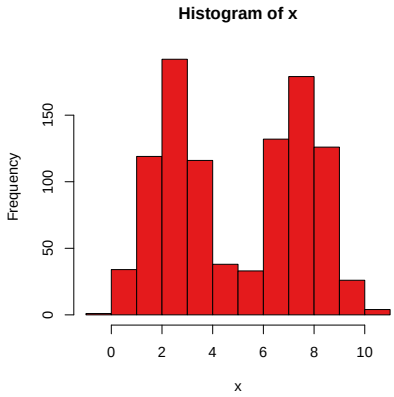
    (a) Symmetric and mound-shaped data.



    (b) Skewed data.



    (c) Bimodal data.



11. For the examples (a)–(c) of the previous problem, which is appropriate, the mean or the median?

Standard Deviation and The Empirical Rule

12. Thirty-three respondents to the class survey reported their GMAT scores. The mean score was 720, and the standard deviation was 30. What can you say about the range of scores reported? Assume that the distribution of reported scores is symmetric and mound-shaped.

13. The mean reported expected starting salary was $\$125K$ and the standard deviation was $\$50K$.

    (a) Complete the following statement with appropriate values for X and Y: "Approximately 95% of the survey respondents have expected starting salaries between X and Y."

    (b) What assumptions do you need to make for the statement in (a) to be correct? Do you think these assumptions are plausible? How could you check this?

    (c) What can we do if the assumptions needed in part (b) are not satisfied?

z-scores

14. Your company has an annual profit of \$60MM with a standard deviation of \$5MM. Assume that the distribution of your annual profits is symmetric and mound-shaped.
    (a) Would it be unusual for your company to have an annual profit of \$52MM?

    (b) Would it be unusual for your company to have an annual profit of \$83MM?

15. Thirty-five respondents from the class survey reported their expected starting salaries. The histogram of these responses was approximately bell-shaped. The mean and standard deviation (in \$1K/year) was was $\bar{x} = 125$ and $s = 50$. How many standard deviations above or below the mean are the following values?
    (a) A starting salary of \$300K.

    (b) A starting salary of \$100K.

    (c) A starting salary of \$200K.

16. In the previous problem, which of the values are unusual?