4.4 Exercises

Basic Concepts

1. Describe the purpose of data subsetting.
2. Describe a data set where data subsetting should be implemented. What are the disadvantages of not subsetting the data?

Exercises

Data

stat.hawkeslearning.com
Discovering Business Statistics, Second Edition > Data Sets > Beers and Breweries

3. Suppose you are a craft beer lover taking a trip to Denver on business and you want to be sure to stop at one of the local breweries while you are there. Using the Beers and Breweries data set from the companion website, subset the data to only show beers brewed in Denver, Colorado, and answer the following questions.
   a. What level of measurement do each of the variables represent?
   b. What variables other than City could be used to subset the data?
   c. How many craft breweries are in Denver?
   d. Which craft beer has the highest Alcohol by Volume (ABV) of the beers brewed in Denver? Give the name of the beer and the brewery.
   e. For the Renegade Brewing Company, how many different IPA styles of beer do they make? What are they?
   f. What is the mean and standard deviation of the ABV values for the craft beers made by the Wynkoop Brewing Company?
   g. Calculate the coefficient of variation of the ABV values for both the Renegade and Wynkoop breweries. Which brewery has more consistent ABV values?

Data

stat.hawkeslearning.com
Discovering Business Statistics, Second Edition > Data Sets > Mount Pleasant Real Estate Data

4. Suppose you are looking for a house in Mount Pleasant, SC, which is near Charleston, and you have limited your search to three subdivisions: Park West, Dunes West, and Carolina Park. Using the Mount Pleasant Real Estate data set from the companion website, answer the following questions.
   a. What level of measurement do each of the variables represent?
   b. Which variables could be used to subset the data?
   c. How could you subset the data using quantitative variables such as List Price and Acreage?
   d. How many different house styles are represented in these three subdivisions? What are the styles?
   e. How many of the houses are newly built (2015–2017)? Which subdivision has the most new homes?
   f. What is the average price of new homes (2015–2017) in Carolina Park? Round your answer to the nearest whole dollar.
   g. For all new homes (2015–2017) in the three subdivisions, what is the minimum and maximum priced homes and in which subdivision are they?
   h. What is the price per square foot of the two homes in part **g.**?
   i. What variables do you think may contribute to the high price of the house with the maximum price?

5. You are asked to evaluate whether there are issues with how funds are distributed to individuals with developmental disabilities in California. There is a concern that expenditures may not be allocated equitably across various demographic groups. Using the California DDS Expenditures data set, answer the following questions:

Data

stat.hawkeslearning.com

Discovering Business Statistics, Second Edition > Data Sets > California DDS Expenditures

- **a.** What are the mean, mode, and median for the variable Expenditures of the entire data set?
- **b.** What are the variance, standard deviation, and range for the variable Expenditures of the entire data set?
- **c.** What are the 1st and 3rd quartile for the variable Expenditures of the entire data set?
- **d.** Which variables could be used to subset the data?
- **e.** Find the average Expenditure for each Age Group
- **f.** Which Age Group has the highest average Expenditure? Do you notice any trends by Age Group? What might account for differences that exist?
- **g.** Which age group has the highest level of dispersion in Expenditure as measured by the standard deviation and coefficient of variation? Do you notice any trends by Age Group? What might account for differences that exist?
- **h.** Find the average Expenditure for each Ethnicity.
- **i.** What proportion of Expenditures is allocated to each Ethnicity?
- **j.** Briefly discuss your findings based on the analysis in the previous sections.

6. You have been hired by a large company to investigate their employees' satisfaction level. There is some concern that there is high turnover with experienced employees. Using the Employee Satisfaction data set, answer the following questions:

Data

stat.hawkeslearning.com

Discovering Business Statistics, Second Edition > Data Sets > Employee Satisfaction

- **a.** What are the mean, mode, and median for the variable Satisfaction Level?
- **b.** What are the variance, standard deviation, and range for the variable Satisfaction Level?
- **c.** What are the 1st and 3rd quartile for the variable Satisfaction Level?
- **d.** Which variables could be used to subset the data?
- **e.** What is the correlation coefficient between employee satisfaction and the employee's last evaluation score? What does this correlation tell you about the relationship?
- **f.** Find the average Satisfaction Level for each Department?
- **g.** Which salary grouping (low, medium, high) has the highest level of dispersion for Satisfaction Level as measured by standard deviation and coefficient of variation?
- **h.** Find the average Satisfaction Level of each year of experience. Are there differences in Satisfaction Level based on years spent at the company?
- **i.** Briefly discuss your findings based on the analysis in the previous sections.

Data

stat.hawkeslearning.com
Discovering Business Statistics, Second Edition > Data Sets > San Francisco Salaries 2014

7. You have been applying for jobs in San Francisco. You want to research to understand what salary level you can expect to be offered. Using the San Francisco Salaries 2014 data set, answer the following questions:

a. What are the mean, mode, and median for the variable Total Pay and Benefits?

b. What are the variance, standard deviation, and range for the variable Total Pay and Benefits?

c. What are the 1st and 3rd quartile for the variable Total Pay and Benefits?

d. Which variables could be used to subset the data?

e. Construct a frequency distribution for Base Pay? Include the relative frequency of each class.

f. From part e. which pay group has the highest relative frequency? What trends do you notice? What might account for differences that exist?

g. Determine the percentage of jobs that have overtime.

h. Which group (overtime or no overtime) has the highest level of dispersion for total pay as measured by standard deviation and coefficient of variation?

i. Briefly discuss your findings based on the analysis in the previous sections.