Chapter 1: Exploring Data

Key Vocabulary:

* individual
* variable
* frequency table
* relative frequency table
* distribution
* pie chart
* bar graph
* two-way table
* marginal distributions
* conditional distributions
* side-by-side bar graph
* association
* dotplot
* stemplot
* histogram
* SOCS
* outlier
* symmetric
* 
* 
* spread
* variability
* median
* quartiles
* Q1, Q3
* IQR
* five-number summary
* minimum
* maximum
* boxplot
* resistant
* standard deviation
* variance

 

Data Analysis: Making Sense of Data (pp.2-6)

1. *Individuals* are…
2. A *variable* is …
3. When you first meet a new data set, ask yourself:
* Who…
* What…
* Why, When, Where and How…
1. Explain the difference between a *categorical* variable and a *quantitative* variable. Give an example of each.
2. Give an example of a categorical variable that has number values.
3. Define *distribution*:
4. What are the four steps to *exploring data*?
* Begin by….
* Study relationships…
* Start with a …
* Then add…
1. Answer the two questions for the *Check Your Understanding* on page 5*:*
2. Define *inference.*
	1. Analyzing Categorical Data (pp.8-22)
3. A *frequency* table displays…
4. A *relative frequency table* displays…
5. What type of data are *pie charts* and *bar graphs* used for?
6. *Categories* in a bar graph are represented by \_\_\_\_\_\_\_\_\_\_\_ and the *bar heights* give the category \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.
7. What is a *two-way table*?
8. Define *marginal distribution*.
9. What are the two steps in examining a marginal distribution?
10. Answer the two questions for the *Check Your Understanding* on page 14.
11. What is a *conditional distribution*? Give an example demonstrating how to calculate one set of conditional distributions in a two-way table.
12. What is the purpose of using a *segmented bar graph*?
13. Answer question one for the *Check Your Understanding* on page 17.
14. Describe the four steps to organizing a statistical problem:
* State…
* Plan…
* Do…
* Conclude…
1. Explain what it meant by an *association* between two variables.
	1. Analyzing Categorical Data (pp.27-42)
2. What is a *dotplot*? Draw an example.
3. When examining a distribution, you can describe the overall pattern by its

**S\_\_\_\_\_ O\_\_\_\_\_ C\_\_\_\_\_ S\_\_\_\_\_**

1. If a distribution is *symmetric*, what does it look like?
2. If a distribution is *skewed to the right*, what does it look like?
3. If a distribution is *skewed to the left*, what does it look like?
4. Describe and illustrate the following distributions:
5. Unimodal
6. Bimodal
7. Multimodal
8. Answer questions 1-4 for the *Check Your Understanding* on page 31.
9. How are a *stemplot* and a *histogram* similar?
10. When is it beneficial to *split the stems* on a stemplot?
11. When is it best to use a *back-to-back stemplot*?
12. List the three steps involved in making a histogram.
13. Why is it advantageous to use a relative frequency histogram instead of a frequency histogram?
14. Answer questions 2-4 for the *Check Your Understanding* on page 35.
	1. Analyzing Categorical Data (pp.50-67)
15. What is the most common *measure of center*?
16. Explain how to calculate the *mean,* .
17. What is the meaning of ∑?
18. Explain the difference between  and μ.
19. Define *resistant measure*.
20. Explain why the mean is not a resistant measure of center.
21. What is the *median* of a distribution? Explain how to find it.
22. Explain why the median is a resistant measure of center?
23. How does the shape of the distribution affect the mean and median?
24. What is the *range*?
25. Is the range a resistant measure of spread? Explain.
26. How do you find *first quartile* Q1 and *third quartile* Q3?
27. What is the *Interquartile Range* (IQR)?
28. Is the IQR and the quartiles a resistant measure of spread? Explain.
29. How is the IQR used to identify *outliers*?
30. What is the *five-number summary* of a distribution?
31. Explain how to use the five-number summary to make a *boxplot*.
32. What does the *standard deviation* measure? How do we calculate it?
33. What is the relationship between *variance* and *standard deviation*?
34. What are the *properties* of the standard deviation as explained on page 64?
35. How should one go about choosing measures of center and spread?