Chapter 2: Descriptive Statistics

ACSTA101 w/ Professor MG

Key Terms: Frequency Distributions

When analyzing or summarizing data, it can be difficult to see patterns. One way of making this easier is to create a frequency distribution

Frequency Distribution	**A table that shows intervals (called classes) of data entries with a count of the number of entries that are in each class**
Frequency (f)	**The number of entries in a given class**
Lower Class Limit	**The lowest number in a given class**
Upper Class Limit	**The highest number in a given class**
Class Width	**The difference between the upper (or lower) limit of consecutive classes.**
Range	**The difference between the highest and lowest values in a data set**

Constructing a Frequency Distribution

01

Determine # of Classes

Decide how many classes you want to include (typically between 5 and 20)

02

Find the Class Width

$$\text{width} = \frac{\text{range}}{\text{number of classes}}$$

You can round up to the nearest convenient number if needed.

03

Find the Class Limits

You can use the smallest entry for the first lower class limit. Add the class width to find the next lower class limit.

04

Find the Frequencies

Count the number of data entries in each class. It can be helpful to list the numbers in order.

Interactive Example:

Have everyone in class give you a number between 1-30. Use this data set to create a frequency distribution of the answers.

Example: The following data consists of how much money 30 adults spent on medicine in the past year. Construct a frequency distribution with 7 classes.

200	239	155	252	384	165	296	405	303	400
307	241	256	315	330	317	352	266	276	345
238	306	290	271	345	312	293	195	168	342

Class	Frequency, *f*
155–190	3
191–226	2
227–262	5
263–298	6
299–334	7
335–370	4
371–406	3

Note: A quick way of checking for counting errors is adding up the numbers in the *f* column. It should add up to *n* (the total amount of entries)

More Key Terms

Midpoint (aka Class Mark)	$$\text{midpoint} = \frac{(\text{lower limit}) + (\text{upper limit})}{2}$$
Relative Frequency	**The proportion or percentage describing how much of the data falls in a particular class.** $$\text{relative frequency} = \frac{\text{class frequency}}{\text{sample size}} = \frac{f}{n}$$
Cumulative Frequency	**The sum of a given class' frequency with all of the previous classes' frequencies.** **The last class should have cumulative frequency equal to *n***

Example: Using the previous data

Class	Frequency, *f*	Midpoint	Relative Frequency (approximately)	Cumulative Frequency
155–190	3	172.5	0.1	3
191–226	2	208.5	0.07	5
227–262	5	244.5	0.17	10
263–298	6	280.5	0.2	16
299–334	7	316.5	0.23	23
335–370	4	352.5	0.13	27
371–406	3	388.5	0.1	30

Note: The relative frequencies should approximately add up to 1

Frequency Histogram

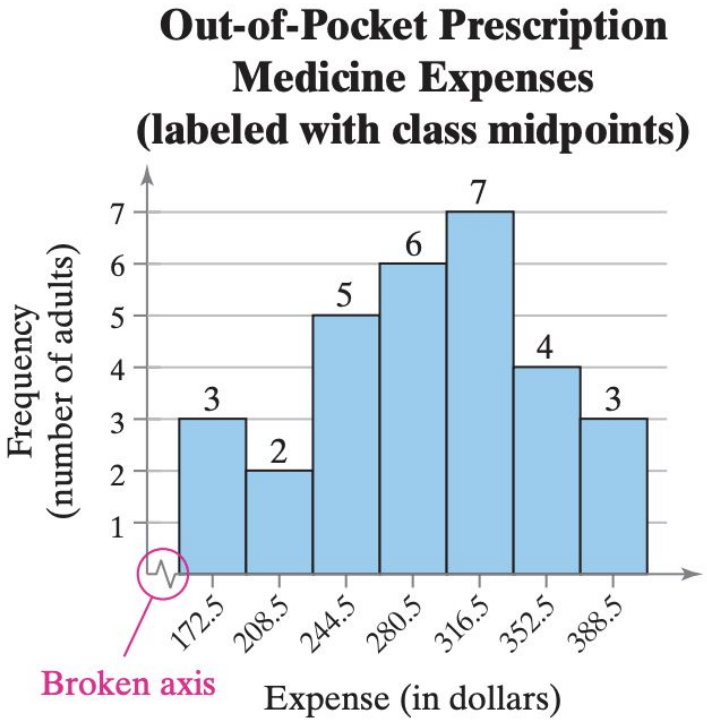
A frequency histogram is a bar graph that represents a frequency distribution. They have three properties:

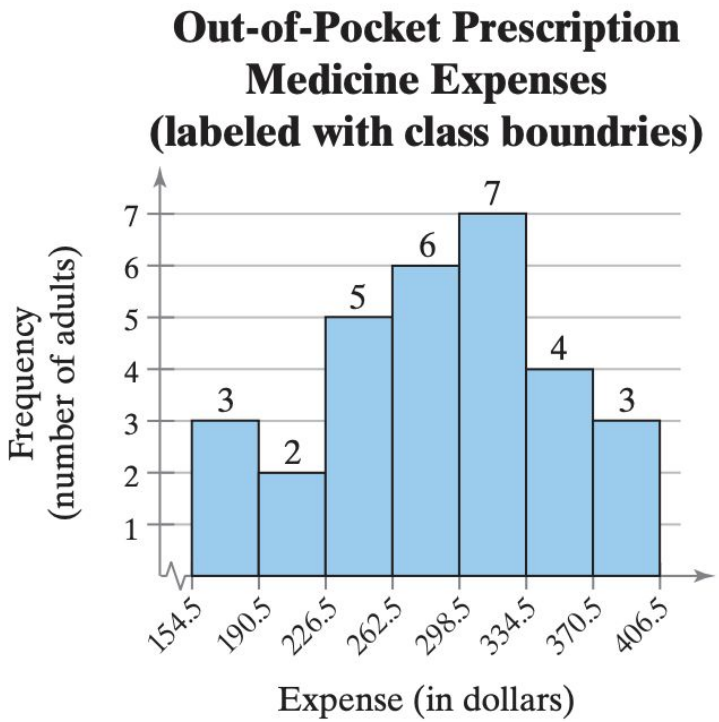
- **The horizontal axis (x-axis) represents the quantitative data values**
- **The vertical axis (y-axis) represents the frequency**
- **Consecutive bars must touch**

The x-axis can be labeled with class midpoints or class boundaries.

Class boundaries are found by subtracting 0.5 from the lower class limit and adding 0.5 to the upper class limit

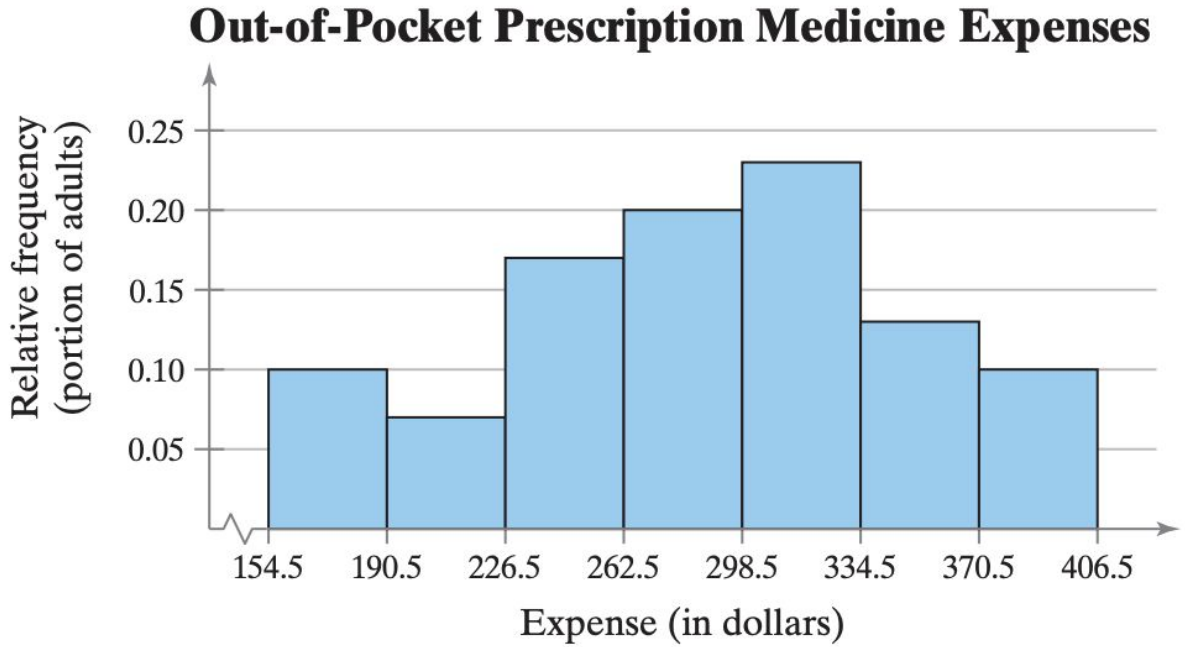
Frequency Histogram w/ Example





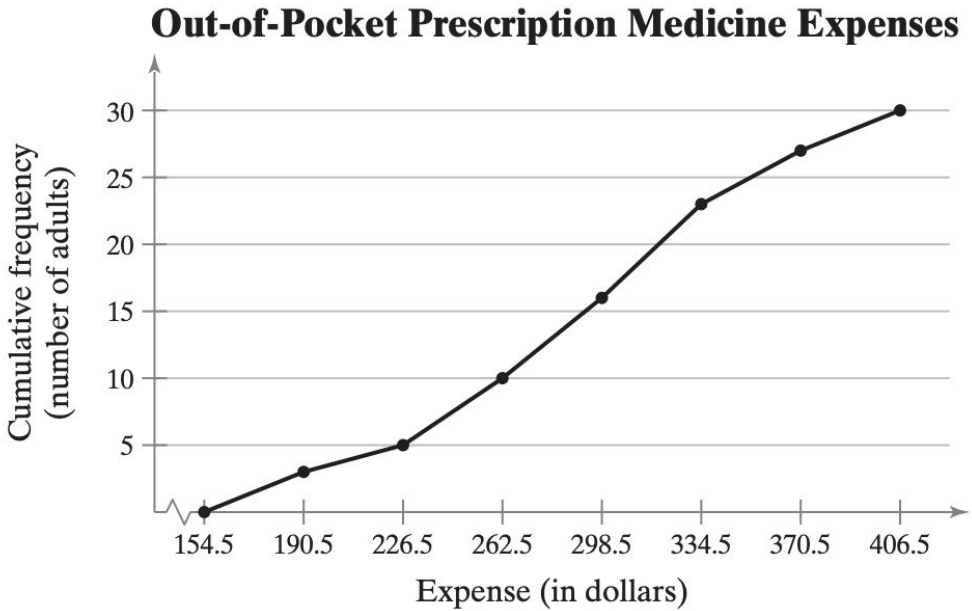
Relative Frequency Histogram

This is the same as the frequency histogram, except the y-axis indicates relative frequencies instead of frequencies. Both frequency and relative frequency histograms have the same shape.



Cumulative Frequency Graph (or Ogive)

A cumulative frequency graph is a line graph that displays the cumulative frequency of each class. The class upper boundaries are marked on the x-axis with the cumulative frequencies marked on the y-axis.



2.2

More Graphs and Displays

Stem-and-Leaf Plots

Stem

The stem of a number is the leftmost digit(s)

Leaf

The leaf of a number is the rightmost digit

How many digits included in the stem and leaf of a data set depends on how large the data entries are

Example: The data on the left indicates the number of text messages sent in one day by 50 cell phone users.

Number of text messages sent				
76	49	102	58	88
122	76	89	67	80
66	80	78	69	56
76	115	99	72	19
41	86	48	52	28
26	29	33	26	20
33	24	43	16	39
29	32	29	29	40
23	33	30	41	33
38	24	53	30	149

Number of Text Messages Sent

Key: 10|2 = 102

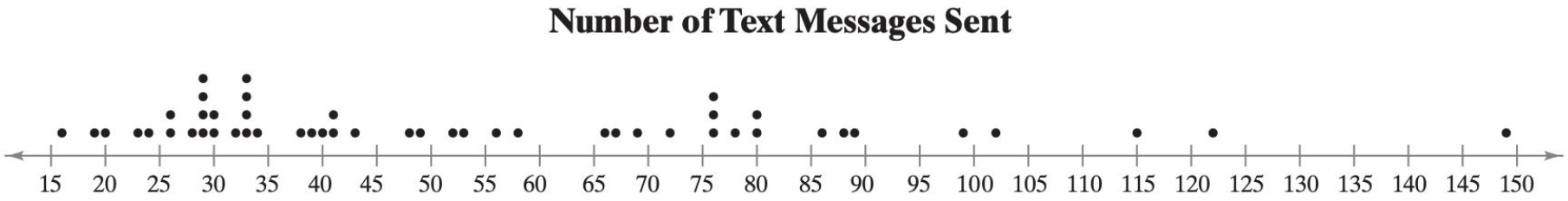
1	6 9
2	0 3 4 6 6 8 9 9 9 9
3	0 0 2 3 3 3 3 4 8 9
4	0 1 1 3 8 9
5	2 3 6 8
6	6 7 9
7	2 6 6 6 8
8	0 0 6 8 9
9	9
10	2
11	5
12	2
13	
14	9

Dot Plots

In a dot plot, each entry is plotted, using a point (or dot), above a horizontal axis.

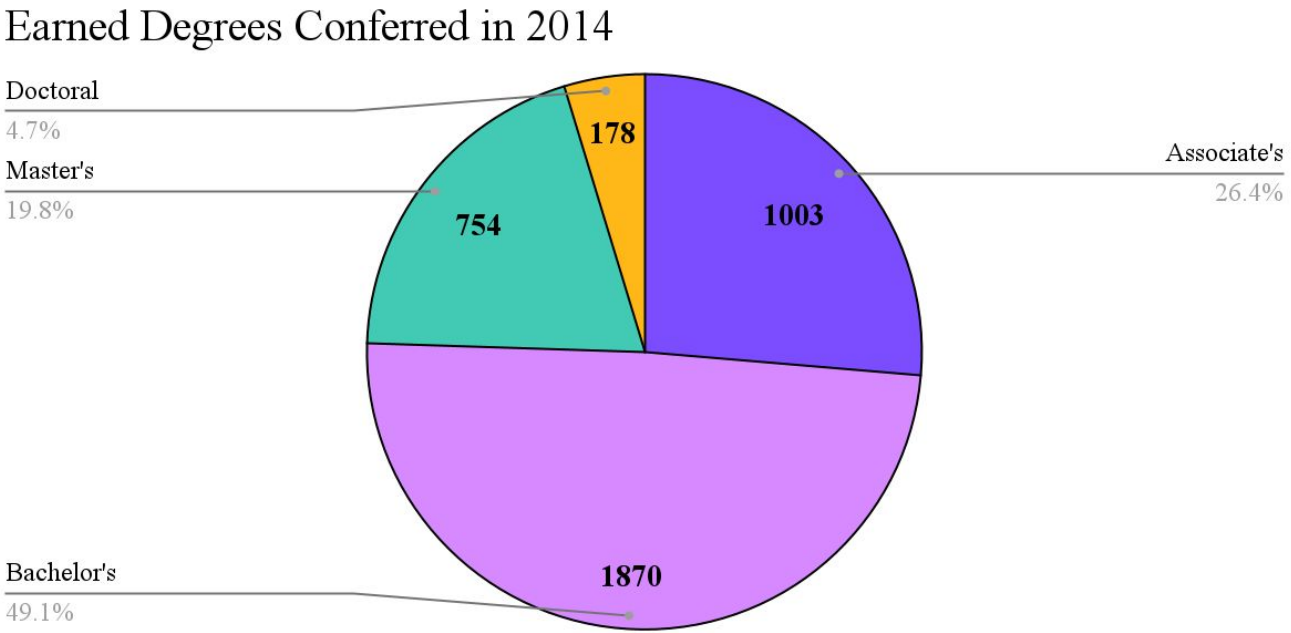
Like a stem-and-leaf plot, dot plots allow for the user to see how the data is distributed while retaining the numerical values.

Example Dot Plot (text message data)



Pie Charts

A pie chart is a circle that is divided into sectors that represent categories. The area of each sector is proportional to the frequency of that category.



Pareto Charts

A pareto chart is like a frequency or relative frequency histogram except the vertical bars are positioned from tallest to shortest.

These types of charts are used frequently in business.

Example: In 2014, these were the leading causes of death in the United States

- **Accidents: 136,053**
- **Cancer: 591,699**
- **Chronic lower respiratory disease: 147,101**
- **Heart disease: 614,348**
- **Stroke (cerebrovascular diseases): 133,103**