2.3: Organizing Quantitative Data

Recall: Types of Data

Qual

Nominal (categorical): Values are names or categories. Values do not measure the amount of any characteristic, and do not have a meaningful order.

could be either

Ordinal: Values have a meaningful order.

Quant

Interval: Values have a meaningful order, and also a meaningful difference (interval) between values. Values do not have a meaningful zero.

Ratio: Values have a meaningful order, a meaningful difference (interval) between values, and also have a meaningful zero. The zero value indicates the absence of the characteristic being measured.

Quantitative data can be classified as *discrete* or *continuous*.

Discrete: A *discrete* variable takes on either a finite number of values, or a countably infinite number of values. *Countable* means the values can be counted (1, 2, 3, …); countably infinite means that you can't ever finish counting the values. A discrete variable cannot take on every value in an interval.

Continuous: A *continuous* variable takes on every value in an interval. It does not make sense to count (assign the numbers 1, 2, 3, ….) to the values.

Examples of discrete variables:

# of cars in your family
# of people in a class
Integers

Examples of continuous variables:

height
weight
length of time a plant live
Real numbers (distance)

length of fingernails
density of hemoglobin

Histograms:

A histogram is a specific type of bar graph used to represent quantitative data. The values for the variable are on the horizontal axis, and the frequencies or relative frequencies are on the vertical axis. All values within the range of the variable are represented on the *x*-axis.

To create a histogram, the data (observations) are divided into groups. There are three common grouping methods:

(a) single-value grouping (when there are a small # of possible values)
(b) limit grouping ← usually used for discrete (countable) variables
(c) cutpoint grouping ← usually used for continuous variables

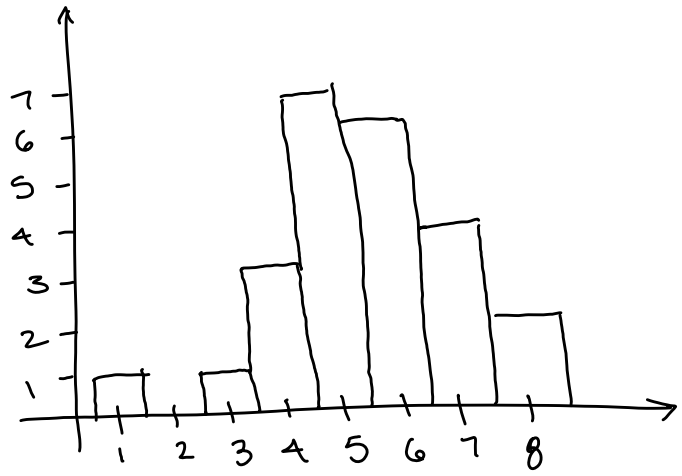
Example 1: Create a frequency distribution, relative frequency distribution, frequency histogram, and relative frequency histogram. Use single-value grouping.

2.82 Cottonmouth Litter Size. In the paper "The Eastern Cottonmouth (*Agkistrodon piscivorus*) at the Northern Edge of Its Range" (*Journal of Herpetology*, Vol. 29, No. 3, pp. 391–398), C. Blem and L. Blem examined the reproductive characteristics of the eastern cottonmouth, a once widely distributed snake whose numbers have decreased recently due to encroachment by humans. A simple random sample of 24 female cottonmouths in Florida yielded the following data on number of young per litter. Use single-value grouping.

8	6	7	7	4	3	1	7
5	6	6	5	6	8	5	5
7	4	6	6	5	5	5	4

# of young per litter	Frequency	Freq	Rel. Freq
1	1	1	1/24
2		0	0
3	1	1	1/24
4	III	3	3/24
5	𝍸 II	7	7/24
6	𝍸 I	6	6/24
7	IIII	4	4/24
8	II	2	2/24
		n = 24	1

Reduce or write as decimals



Example 2: Create a frequency distribution, relative frequency distribution, frequency histogram, and relative frequency histogram.

2.85 Early-Onset Dementia. Dementia is a person's loss of intellectual and social abilities that is severe enough to interfere with judgment, behavior, and daily functioning. Alzheimer's disease is the most common type of dementia. In the article "Living with Early Onset Dementia: Exploring the Experience and Developing Evidence-Based Guidelines for Practice" (*Alzheimer's Care Quarterly*, Vol. 5, Issue 2, pp. 111–122), P. Harris and J. Keady explored the experience and struggles of people diagnosed with dementia and their families. A simple random sample of 21 people with early-onset dementia gave the following data on age, in years, at diagnosis. Use limit grouping with a first class of 40–44 and a class width of 5.

60	58	52	58	59	58	51
61	54	59	55	53	44	46
47	42	56	57	49	41	43

Age	
40 - 44	
45 - 49	
50 - 54	
55 - 59	
60 - 64	
65 - 69	

Example 3: Create a frequency distribution, relative frequency distribution, frequency histogram, and relative frequency histogram. Use cutpoint grouping.

2.90 Fuel Tank Capacity. *Consumer Reports* provides information on new automobile models, including price, mileage ratings, engine size, body size, and indicators of features. A simple random sample of 35 new models yielded the following data on fuel tank capacity, in gallons. Use cutpoint grouping with 12 as the first cutpoint and classes of equal width 2.

17.2	23.1	17.5	15.7	19.8	16.9	15.3
18.5	18.5	25.5	18.0	17.5	14.5	20.0
17.0	20.0	24.0	26.0	18.1	21.0	19.3
20.0	20.0	12.5	13.2	15.9	14.5	22.2
21.1	14.4	25.0	26.4	16.9	16.4	23.0

12 – under 14

14 – under 16

16 – under 18

13.99 goes here

14.00 goes here

Stem-and-leaf plots:

A *stem-and-leaf plot* provides a similar summary display of relative frequencies as a bar chart or histogram, but preserves the individual data points. In a stem-and-leaf plot, each value is divided into a "stem" and a "leaf." For example, the number 29 could be represented as a stem of 20 and a leaf of 9. The number 15.7 could be represented as a leaf of 15 and a stem of 0.7.

Example 4:

2.100 San Francisco Giants. From the Baseball Almanac website, we found the heights, in inches, of the players on the 2012 World Series–winning San Francisco Giants baseball team.

76	75	77	75	72	74	73	71	75
71	76	74	72	72	76	75	76	72
70	76	76	74	74	73	71	74	73
77	72	74	73	73	76	76	71	70
71	71	72	73	74	74	71	76	73

a. Construct a stem-and-leaf diagram of these data with five lines per stem.

b. Why is it better to use five lines per stem here instead of one or two lines per stem?

Dot plots:

A dot plot is a simple graph in which each observation is represented by a dot. The dots are place in a column above the value they represent. A dot plot is used for discrete variables (not for continuous variables).

Example 1:

Time-series plots:

A *time-series plot* is used to analyze trends in data over time. The horizontal axis represents time; the vertical axis represents the value of the variable.

Example 2: