2: Frequency distributions 2/19/16

Chapter 2 Review Questions

1. How many *leaves* do you plot on a stemplot?
2. How many stem values should you use when drawing a stemplot?
3. Why would you use split stem-values on a stemplot?
4. What is the *depth* of an observation?
5. What is the purpose of the stem-multiplier on a stemplot?
6. Distributions with long tails toward the higher numbers are said to have a _____ skew.
7. The three elements of a distribution are shape, location, and _____.
8. What is the "middle point" of the ordered array called?
9. The median has a depth of _____.
10. This is a value that does not fit in with the general pattern of a distribution.
11. Define the following terms: frequency, relative frequency, cumulative relative frequency.
12. What is a distribution?
13. What is the full name for stemplot?
14. What do you look for on a stemplot?
15. Why must the stem and leaves of a stemplot align vertically and horizontally?
16. What do you do if a value in a data set has more significant digits that can fit on a Tukey stemplot?

The Exercises start on the next page

Exercises

2.1 *Irish healthcare websites*. The Irish Department of Health recommends a reading level at or below 12- to 14-years of age for health information aimed at the public. Reading levels for 46 Irish healthcare websites are shown below (O'Mahoney 1999).

Data. Reading levels of 46 Irish healthcare websites (O'Mahoney 1999).

08	10	11	11	12	13	13	13	13	14
14	15	15	15	15	15	15	15	16	16
16	16	17	17	17	17	17	17	17	17
17	17	17	17	17	17	17	17	17	17
17	17	17	17	17	17				

(A) Create a stemplot for these data. (Hints: Use a stem-multiplier of x1. The tenths place for all the data points is .0.)
(B) Narratively describe the distribution's shape, location, and spread. (Skews are described by the direction of their tails.)
(C) Create a frequency tale for the data. Include columns for frequency, relative frequency, and cumulative frequency, and cumulative relative frequency. (Hints: Include table cells with 0 counts. Frequency tables should always include a total row at the bottom. The total should sum to n. The relative frequency should sum to 100%. There is no total for relative frequencies and cumulative relative frequency.)
(D) What percentage of the website met the recommended level having a reading-level of 14 or below?

2.2 *Poverty in eastern states*. Poverty is associated with many health determinants. The table below shows the percentage of people living below the poverty line in the 26 states east of the Mississippi for the year 2000 (Delaker 2001). Create an informative stemplot of these data and discuss your findings.

Data for exercise 2.2. Percentage of people living below poverty level in 26 states east of the Mississippi, 2000.

Alabama	14.6	Indiana	08.2	Mississippi	15.5	Pennsylvania	09.9	West Virginia	15.8
Connecticut	07.6	Kentucky	12.5	New Hampshire	07.4	Rhode Is.	10.0	Wisconsin	08.8
Delaware	09.8	Maine	09.8	New Jersey	08.1	S. Carolina	11.9		
Florida	12.1	Maryland	07.3	New York	14.7	Tennessee	13.3		
Georgia	12.6	Massachusetts	10.2	North Carolina	13.2	Vermont	10.1		
Illinois	10.5	Michigan	10.2	Ohio	11.1	Virginia	08.1		

2.3 *Hospital duration data.* A study by <u>Townsend and co-worker (1979)</u> looked at duration of hospital stays (in days). A sample of 25 patients from this study showed the following hospital stay durations (days):

5	10	6	11	5	14	30	11	17	3
9	3	8	8	5	5	7	4	3	7
9	11	11	9	4					

(A) Create a stemplot for these data using a stem multiplier of ×10.
(B) Create a stemplot with split stem-values.
(C) Which of the above stemplots does a better job demonstrating the shape of the distribution?
(D) Describe the distribution's shape, location, and spread.
(E) Construct a frequency table of the data using 5-day class intervals (0 – 4 days, 5 – 9 days, etc.).
(F) What percentage of hospital stays were less than or equal to 5 days?
(G) What percentage were less than or equal to 15 days?
(H) What percentage were *at least* 15 days in length? (At least means "that much or more.")

2.4 *Outpatient wait time.* The length of time patients wait for attention in doctors' offices is an important consideration in health care. Waiting times (in minutes) for 25 patients at a public health clinic are:

35	22	63	6	49	19	16	31	24	29
23	32	72	13	51	45	77	16	33	55
10	42	28	72	13					

(A) Draw a stemplot of the data. Describe the distribution of values.
(B) Using 10-unit class intervals, create a frequency table showing frequency counts, relative frequencies, and cumulative relative frequencies.
(C) What percentage of wait times were less than 20 minutes?
(D) What percentage were *at least* 20 minutes?

2.5 *Percentage ideal body weight (n = 18).* This data set contains data made up of a variable that is equal to (actual body weight) ÷ (ideal body weight) x 100. A value of 100 represents 100% of ideal body weight, 120 represents 20% above ideal body weight, and so on. Data for *n* = 18 subjects are shown below. (Source: <u>Saudek 1989</u>; Pagano 1993, p. 208).

107	119	99	114	120	104	88	114	124	116
101	121	152	100	125	114	95	117		

(A) Construct and stem-and-leaf plot of these data. Describe the distribution.
(B) Construct a frequency table for these data using 20-unit class intervals. Report frequencies, relative frequencies, and cumulative relative frequencies.

2.6 *Seizures following bacterial meningitis* (Source unknown). The time between exposure to a causative agent and first symptoms is called the incubation or incubation period. A study examined induction periods in 13 seizure cases following bacterial meningitis. Data (months) are listed below. Using a stem-value multiplier of × 10, construct a stemplot for these data. Discuss your findings. [*Hint*: The value of 0.10 has a tens-place of 0 and ones-place of 0, so shows up as 0|0 on the plot with the ×10 stem-multiplier.]

```
0.10    0.25    0.50    4       12      12      24      24      31      36
42      55      96
```

2.7 *Children of physicians*. The number of offspring of physicians at a particular hospital are shown below.

3	2	0	1	4	7	3	2	4	1
0	2	5	6	2	1	2	1	0	0
3	6	2	1						

(A) Plot these data as a stemplot and narratively describe the distribution.
(B) Construct a frequency table for these data with columns for frequency, relative frequency, and cumulative relative frequency. What percentage of physicians at this hospital have 3 *or fewer* children? What percentage have *at least* 2 children?

2.8 *Surgical times for artificial hearts*. Durations of surgeries (hours) for 15 patients receiving artificial hearts are shown below (Kitchens, 1998, p. 139). Create a stem-and-leaf plot of these data. Are there any outliers?

```
7.0    6.5    3.5    3.8    3.1    2.8    2.5    2.6    2.4    2.1
1.8    2.3    3.1    3.0    2.5
```

2.9 *Grad student ages*. Ages of 36 graduate students are listed below and are stored. Explore these data with a stemplot and describe your results

```
29    25    29    27    27    29    32    30    28    30
28    26    25    25    27    29    25    24    28    28
24    28    25    28    29    26    27    30    28    16
29    30    33    29    32    31
```

2.10 *Health insurance coverage*. Data for the percentage of people without health insurance by state for the period 2002 through 2004 are shown below. Create an informative stemplot and then describe the distribution. [Note: Plots with too many or too few stem values are not acceptable, as they will fail to show the shape of the distribution.]

STATE	% w/out	STATE	% w/out	STATE	% w/out	STATE	% w/out	STATE	% w/out
Alabama	13.5	Hawaii	09.9	Michigan	11.4	North Carolina	16.6	Utah	13.4
Alaska	18.2	Idaho	17.3	Minnesota	08.5	North Dakota	11.0	Vermont	10.5
Arizona	17.0	Illinois	14.2	Mississippi	17.2	Ohio	11.8	Virginia	13.6
Arkansas	16.7	Indiana	13.7	Missouri	11.7	Oklahoma	19.2	Washington	14.2
California	18.4	Iowa	10.1	Montana	17.9	Oregon	16.1	West Virginia	15.9
Colorado	16.8	Kansas	10.8	Nebraska	11.0	Pennsylvania	11.5	Wisconsin	10.4
Connecticut	10.9	Kentucky	13.9	Nevada	19.1	Rhode Island	10.5	Wyoming	15.9
Delaware	11.8	Louisiana	18.8	New Hamp	10.6	South Carolina	13.8		
DC	13.5	Maine	10.6	New Jersey	14.4	South Dakota	11.9		
Florida	18.5	Maryland	14.0	New Mexico	21.4	Tennessee	12.7		
Georgia	16.6	Massach.	10.8	New York	15.0	Texas	25.1		

2.11 *UNICEF low birth weight data*. A weight at birth of less than 2,500 grams (about 5.5 pounds) qualifies as "low birth weight" according to standard conventions. Low birth-weight rates (per 100 births) for the year 1991 for 109 countries are stored in unicef.sav. There are 129 records in the data file, but 20 values are missing. Therefore, $n = 109$ for this analysis.

(A) Download unicef.sav and create a stemplot of the low birth weight rates.
(B) Determine the low birth weight rate for the US. (Sort the data by country and find the value for the USA. The menu commands are Data > Sort cases > Sort by.)
(C) Create a frequency table for low birth weights. Where does the USA stand in this table?

2.12 *Growth of U.S. Hispanic population*. The percent of residents in the United States who self-identified as Spanish, Hispanic, or Latino in the 2000 census is shown below. [Data also available in `per-hisp.sav`.] Create the best possible stemplot for these data and narrative describe the distribution.

STATE	PERCENT	STATE	PERCENT	STATE	PERCENT
Alabama	1.5	Louisiana	2.4	Ohio	1.9
Alaska	4.1	Maine	.7	Oklahoma	5.2
Arizona	25.3	Maryland	4.3	Oregon	8.0
Arkansas	2.8	Massachusetts	6.8	Pennsylvania	3.2
California	32.4	Michigan	3.3	Rhode Island	8.7
Colorado	17.1	Minnesota	2.9	South Carolina	2.4
Connecticut	9.4	Mississippi	1.3	South Dakota	1.4
Delaware	4.8	Missouri	2.1	Tennessee	2.0
Florida	16.8	Montana	2.0	Texas	32.0
Georgia	5.3	Nebraska	5.5	Utah	9.0
Hawaii	7.2	Nevada	19.7	Vermont	0.9
Idaho	7.9	New Hampshire	1.7	Virginia	4.7
Illinois	10.7	New Jersey	13.3	Washington	7.2
Indiana	3.5	New Mexico	42.1	West Virginia	0.7
Iowa	2.8	New York	15.1	Wisconsin	3.6
Kansas	7.0	North Carolina	4.7	Wyoming	6.4
Kentucky	1.5	North Dakota	1.2		

2.13 *East Boston Respiratory Disease Survey* (FEV). Download `fev.sav` and open the file in *SPSS*. Create a stemplot of the AGE variable. Notice that each leaf in the plot represents 2 cases. Also notice frequencies are reported to the left of the plot. Describe the shape, location, and spread of the age distribution you see.

2.14 *Student weights*. The data set body-weight.sav contains weights of 53 students (pounds). Download the dataset and plot the results as a stemplot. Describe the distribution. (Data from presentation slides prepared by J. Mays for Moore 2004, Chapter 2).