Math13 - Introduction to Statistics (code 40161), Fall 2015

Homework 2; September 21

NAME:

---

1) One survey produced the following summary of responses to the question "*What is your favorite color?*": blue, 42%; green, 14%; purple, 14%; red, 8%; black, 7%; orange, 5%; yellow, 3%; brown, 3%; gray, 2%; and white, 2%. Draw a bar chart of the percents and write a short summary of the major features of your graph.

2) The survey about color preferences reported the age distribution of the people who responded. Here are the age distribution of the results:

Age group (years)	1-18	19-24	25-35	36-50	51-69	70 and over
Count	10	97	70	36	14	5

a. Add the counts and compute the percents for each age group
b. Make a bar chart of the percents
c. Describe the distribution
d. Explain why your bar chart is not a histogram

3) The following table gives a breakdown of the materials that made up municipal solid waste (garbage):

Material	Weight (tons)	Percent
Food scraps	25.9	11.2
Glass	12.8	5.5
Metals	18.0	7.8
Paper	86.7	37.4
Plastics	24.7	10.7
Rubber & Textiles	15.8	6.8
Wood	12.7	5.5
Yard trimmings	27.7	11.9
Other	7.5	3.2
Total	*231.9*	*100.0*

a. Calculate the cumulative frequency of the percent values in decreasing order (i.e. ordering the values from tallest to shortest)

Material	Relative Frequency (percent)	Cumulative Frequency
Total	*100*	*100*

b. Make a bar chart with relative frequencies (from tallest to shortest) and add a pareto line of the cumulative percents.

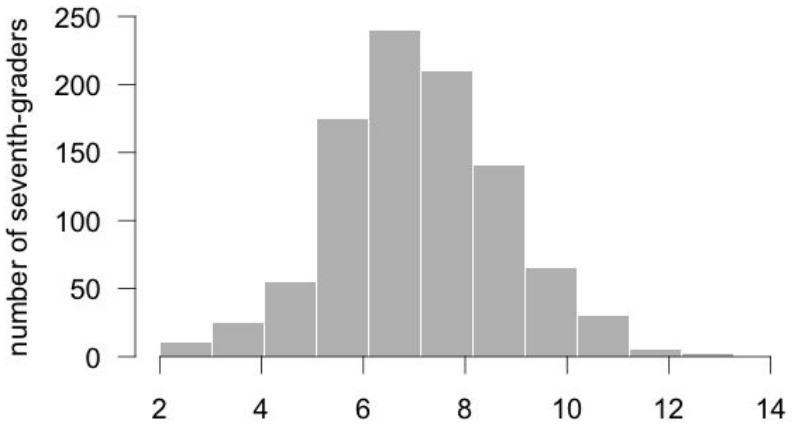
4) The following table gives the percentages of women among students seeking various graduate and professional degrees.

Degree	Percent
Master's in business administration	39.8
Master's in education	76.2
Other master of arts	59.6
Other master of science	53.0
Doctorate in education	70.8
Other PhD degree	54.2
Medicine (MD)	44.0
Law	50.2
Theology	20.2

a. Explain clearly why we cannot use a pie chart to display these data

b. What type of chart would you use to graph the data (explain why)

5) The next figure displays the scores of all 947 seventh-grade students, in the public schools of Springfield, on the vocabulary part of the Test of Basic Skills. Give a brief description of the overall pattern (shape, center, spread) of this distribution.



6) The states differ greatly in the kinds of severe weather that afflict them. The table below shows the average property damage caused by tornadoes per year over the period 1950 to 1999 in each of the fifty states.

State	Damage	State	Damage	State	Damage
Alabama	51.88	Louisiana	27.75	Ohio	44.36
Alaska	0.00	Maine	0.53	Oklahoma	81.94
Arizona	3.47	Maryland	2.33	Oregon	5.52
Arkansas	40.96	Massachusetts	4.42	Pennsylvania	17.11
California	3.68	Michigan	29.88	Rhode Island	0.09
Colorado	4.62	Minnesota	84.84	South Carolina	17.19
Connecticut	2.26	Mississippi	43.62	South Dakota	10.64
Delaware	0.27	Missouri	68.23	Tennessee	23.47
Florida	37.32	Montana	2.27	Texas	88.60
Georgia	51.68	Nebraska	30.26	Utah	3.57
Hawaii	0.34	Nevada	0.10	Vermont	0.24
Idaho	0.26	New Hampshire	0.66	Virginia	7.42
Illinois	62.94	New Jersey	2.94	Washington	2.37
Indiana	53.13	New Mexico	1.49	West Virginia	2.14
Iowa	49.51	New York	15.73	Wisconsin	31.33
Kansas	49.28	North Carolina	14.90	Wyoming	1.78
Kentucky	24.84	North Dakota	14.69		

a. What are the top five states for tornado damage?

b. What are the bottom five states for tornado damages?

c. Make a histogram of the data with classes "$0 \leq$ damage < 10", "$10 \leq$ damage < 20", "$20 \leq$ damage < 30", and so on. Describe the shape, center, and spread of the distribution. Which states may be outliers?

7) Students from a statistics class were asked to record their heights in inches. The heights (as recorded) were:

65	72	68	64	60	55	73	71	52	63	61	74
69	67	74	50	4	75	67	62	66	80	64	65

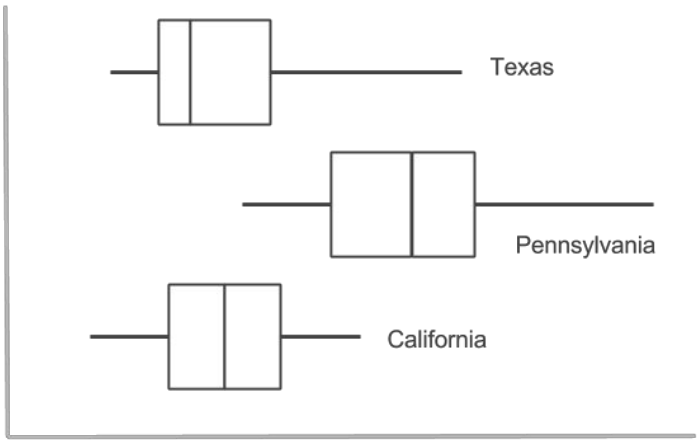
a. Make a box plot of the data

b. Find the first and third quartiles Q_1 and Q_3, and then obtain the value of the interquartile range (IQR)

c. Multiply the IQR by 1.5 and find the lower and upper limits

d. Are there any data values below the lower limit? above the upper limit? List any suspected outliers. What might be some explanations for the outliers?

8) *Consumer Reports* rated automobile insurance companies and listed annual premiums for top-rated companies in several states. The figure below shows box plots for annual premiums for urban customers in three states. The box plots in ent



a. Which state has the lowest premium?

b. Which state has the highest premium?

c. Which state has the highest median premium?

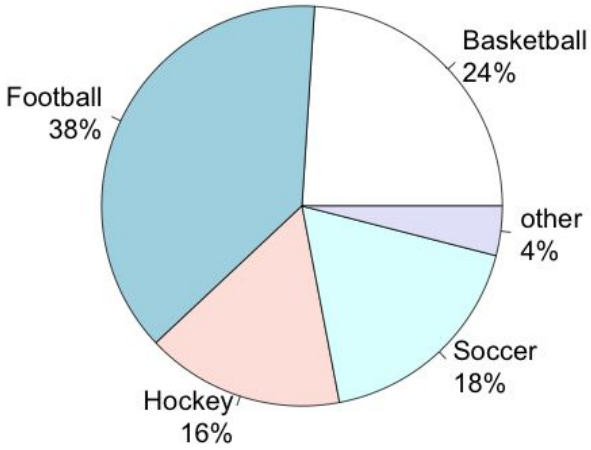
d. Which state has the smallest range of premiums?

e. Which states has the smallest interquartile range?

f. The table below shows the five-number summaries for the box plots. Match the summaries to the appropriate box plots.

A	B	C
n = 10 min = 2382 Q1 = 2758 Median = 2991 Q3 = 3652 max = 5715	n = 10 min = 3314 Q1 = 4326 Median = 5116.5 Q3 = 5801 max = 7527	n = 10 min = 2323 Q1 = 2801 Median = 3377.5 Q3 = 3966 max = 4482

9) A survey of 500 teenagers was taken to see which sport was their favorite to watch on television. The pie chart below displays the results. Choose the correct data (numbers of teenagers) from which the pie chart was constructed *(Explain your answer)*.



a. Basketball, 190; football, 120; hockey, 90; soccer, 80; other, 20
b. Basketball, 120; football, 190; hockey, 90; soccer, 80; other, 20
c. Basketball, 20; football, 90; hockey, 80; soccer, 190; other, 120
d. Basketball, 240; football, 380; hockey, 160; soccer, 180; other, 40
e. Basketball, 120; football, 190; hockey, 80; soccer, 90; other, 20

10) In Chemistry 400, weights are assigned to required activities as follows:

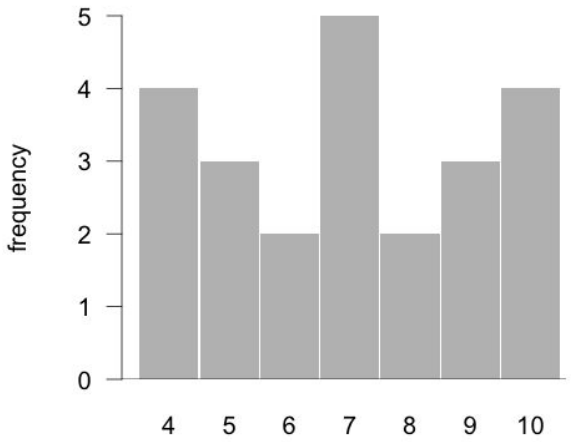
Class participation 15%; Exam 1, 20%; Exam 2, 20%; Exam 3, 20%; Laboratory 25%

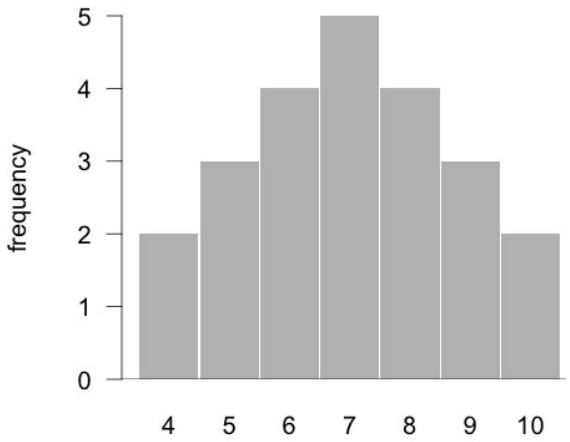
Each activity is graded on a 100-point scale. Mary earned 70 points on class participation, 80 points on exam 1, 64 points on exam 2, 77 points on exam 3, and 96 points on laboratory. Compute her overall weighted average in the Chemistry 400 class.

11) A data set has values ranging from a low 10 to a high 52. What's wrong with using the class limits: 10-19, 20-29, 30-39, 40-49 for a frequency table?

12) A data set with whole numbers (i.e. integers) has a minimum value of 20 and a maximum value 82. Find the class width and class limits for a frequency table with 7 classes.

13) Look at the two histograms below. Each involves the same number of data. The data are all whole (i.e. integer) numbers, so the height of each bar represents the number of values equal to the corresponding midpoint shown on the horizontal axis. Notice that both distributions are symmetric.





a. Estimate the mode, median, and the mean for each histogram

b. Which distribution has the larger standard deviation? Why?

14) The following table shows the calorie count for 22 vanilla-flavored ice cream bars.

342	377	319	353	295	234	294	286	377	182	310
439	111	201	182	197	209	147	190	151	131	151

a. Compute the five-number summary: minimum value, first quartile (Q_1), 2nd quartile or median, 3rd quartile (Q_3), and maximum value.
b. Make a box plot with the obtained summaries.

15) The following table displays the IQ scores of 60 fifth-grade students chosen at random from one school.

145	139	126	122	125	130	96	110	118	118
101	142	134	124	112	109	134	113	81	113
123	94	100	136	109	131	117	110	127	124
106	124	115	133	116	102	127	117	109	137
117	90	103	114	139	101	122	105	97	89
102	108	110	128	114	112	114	102	82	101

Make a frequency table and draw a histogram, by dividing the range of the data into the following classes:

- 75 to 84
- 85 to 94
- 95 to 104
- 105 to 114
- 115 to 124
- 125 to 134
- 135 to 144
- 145 to 154