PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE19CS203 – STATISTICS FOR DATA SCIENCE

Unit-1 - Introduction to Data Science

QUESTION BANK

Data Visualization Techniques – Histogram

Exercises for Section 1.3

[Text Book Exercise – Section 1.3 – Q. No.[1 – 10] – Pg. No. [39 – 40]]

1) The weather in Los Angeles is dry most of the time, but it can be quite rainy in the winter. The rainiest month of the year is February. The following table presents the annual rainfall in Los Angeles, in inches, for each February from 1965 to 2006.

0.2	3.7	1.2	13.7	1.5	0.2	1.7
0.6	0.1	8.9	1.9	5.5	0.5	3.1
3.1	8.9	8.0	12.7	4.1	0.3	2.6
1.5	8.0	4.6	0.7	0.7	6.6	4.9
0.1	4.4	3.2	11.0	7.9	0.0	1.3
2.4	0.1	2.8	4.9	3.5	6.1	0.1

a. Construct a stem-and-leaf plot for these data. **(Exclude)**
b. Construct a histogram for these data.
c. Construct a dotplot for these data. **(Exclude)**
d. Construct a boxplot for these data. Does the boxplot show any outliers?

2) Forty-five specimens of a certain type of powder were analyzed for sulfur trioxide content. Following are the results, in percent. The list has been sorted into numerical order.

14.1	14.4	14.7	14.8	15.3	15.6	16.1	16.6	17.3
14.2	14.4	14.7	14.9	15.3	15.7	16.2	17.2	17.3
14.3	14.4	14.8	15.0	15.4	14.7	16.4	17.2	17.8
14.3	14.4	14.8	15.0	15.4	15.9	16.4	17.2	21.9
14.3	14.4	14.8	15.2	15.5	15.9	16.5	17.2	22.4

a. Construct a stem-and-leaf plot for these data. **(Exclude)**
b. Construct a histogram for these data.
c. Construct a dotplot for these data. **(Exclude)**
d. Construct a boxplot for these data. Does the boxplot show any outliers?

3) Following are measurements of soil concentrations (in mg/kg) of chromium (Cr) and nickel (Ni) at 20 sites in the area of Cleveland, Ohio. These data are taken from the article "Variation in North American Regulatory Guidance for Heavy Metal Surface Soil Contamination at Commercial and Industrial Sites" (A. Jennings and J. Ma, *J Environment Eng*, 2007:587–609).

Cr	34	1	511	2	574	496	322	424
	269	140	244	252	76	108	24	38
	18	34	30	191				

Ni	23	22	55	39	283	34	159	37
	61	34	163	140	32	23	54	837
	64	354	376	471				

a. Construct a histogram for each set of concentrations.
b. Construct comparative boxplots for the two sets of concentrations.
c. Using the boxplots, what differences can be seen between the two sets of concentrations?

4) A certain reaction was run several times using each of two catalysts, A and B. The catalysts were supposed to control the yield of an undesirable side product. Results, in units of percentage yield, for 24 runs of catalyst A and 20 runs of catalyst B are as follows:

Catalyst A			
4.4	3.4	2.6	3.8
4.9	4.6	5.2	4.7
4.1	2.6	6.7	4.1
3.6	2.9	2.6	4.0
4.3	3.9	4.8	4.5
4.4	3.1	5.7	4.5

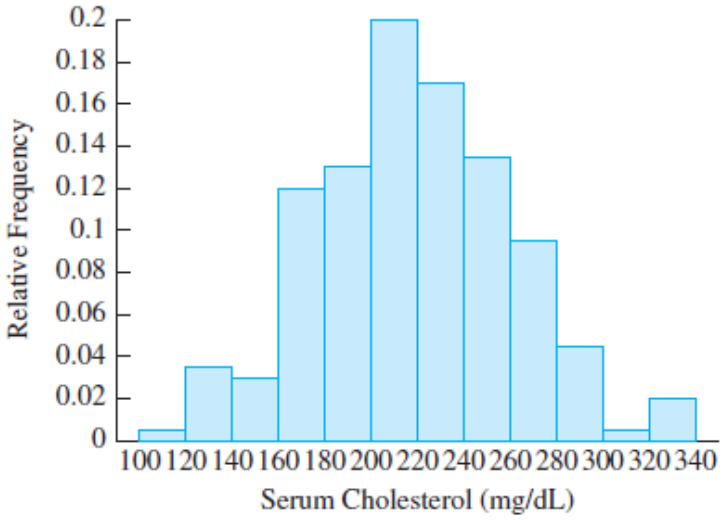
Catalyst B			
3.4	1.1	2.9	5.5
6.4	5.0	5.8	2.5
3.7	3.8	3.1	1.6
3.5	5.9	6.7	5.2
6.3	2.6	4.3	3.8

a. Construct a histogram for the yields of each catalyst.
b. Construct comparative boxplots for the yields of the two catalysts.
c. Using the boxplots, what differences can be seen between the results of the yields of the two catalysts?

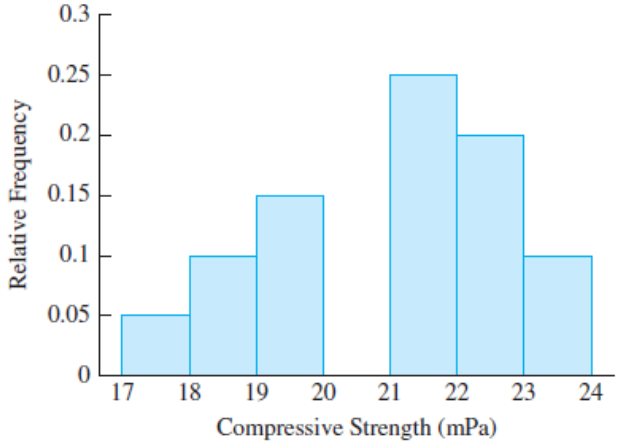
5) Sketch a histogram for which
a. The mean is greater than the median.

b. The mean is less than the median.
c. The mean is approximately equal to the median.

6) The figure below is a histogram showing the distribution of serum cholesterol level for a sample of men. Use the histogram to answer the following questions:
   a. Is the percentage of men with cholesterol levels above 240 mg/dL closest to 30%, 50%, or 70%?
   b. In which interval are there more men: 240–260 mg/dL or 280–340 mg/dL?



7) The histogram below presents the compressive strengths of a sample of concrete blocks hardened for 28 days. One rectangle from the histogram is missing. What is its height?



8) Refer to Table 1.4 (in Section 1.3).
   a. Using the class intervals in the table, construct a histogram in which the heights of the rectangles are equal to the frequencies.
   b. Using the class intervals in the table, construct a histogram in which the heights of the rectangles are equal to the densities.

c. Compare the histograms in parts (a) and (b) with the histogram in Figure 1.8, for which the heights are the relative frequencies. Are the shapes of the histograms the same?

9) Refer to Table 1.5 (in Section 1.3).
   a. Using the class intervals in the table, construct a histogram in which the heights of the rectangles are equal to the relative frequencies.
   b. Compare the histogram in part (a) with the histogram in Figure 1.9, for which the heights are the densities. Are the shapes of the histograms the same?
   c. Explain why the heights should not be set equal to the relative frequencies in this case.
   d. Which classes are visually exaggerated by making the heights equal to the relative frequencies?