



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 – Box Plot

Exercises for Section 1.3

[Text Book Exercise – Section 1.3 – Q. No.[1,2,11 – 15,17] – Pg. No. [39 – 42]]

- 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?

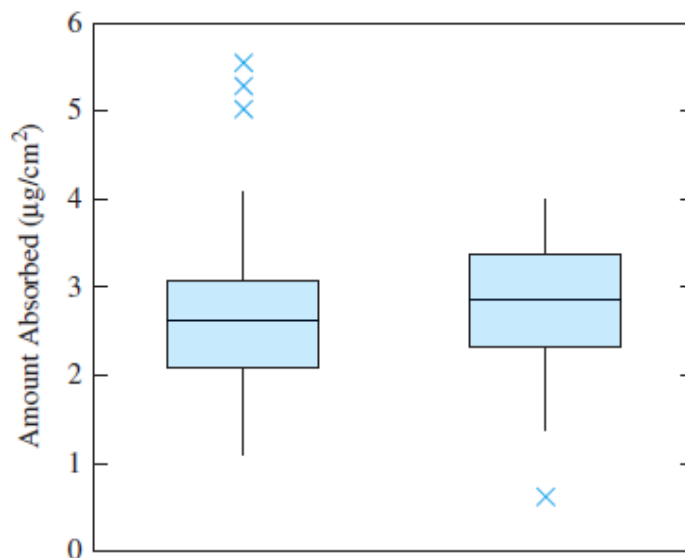
- 3) The following table presents the number of students absent in a middle school in northwestern Montana for each school day in January 2008.

Date	Number Absent	Date	Number Absent	Date	Number Absent
Jan 2	65	Jan 14	59	Jan 23	42
Jan 3	67	Jan 15	49	Jan 24	45
Jan 4	71	Jan 16	42	Jan 25	46
Jan 7	57	Jan 17	56	Jan 28	100
Jan 8	51	Jan 18	45	Jan 29	59
Jan 9	49	Jan 21	77	Jan 30	53
Jan 10	44	Jan 22	44	Jan 31	51
Jan 11	41				

- Construct a boxplot.
 - There was a snowstorm on January 27. Was the number of absences the next day an outlier?
- 4) Which of the following statistics cannot be determined from a boxplot?
- The median
 - The mean
 - The first quartile
 - The third quartile
 - The interquartile range
- 5) A sample of 100 resistors has an average resistance of $50\ \Omega$ and a standard deviation of 5Ω . A second sample of 100 resistors has an average resistance of 100Ω and a standard deviation of 5Ω . If the two samples are combined, the standard deviation of all 200 resistances will be .
- less than 5Ω
 - greater than 5Ω
 - equal to 5Ω
 - can't tell from the information given

(Hint: Don't do any calculations. Just try to sketch, very roughly, histograms for each sample separately, then for the combined sample.)

- 6) Following are boxplots comparing the amount of econazole nitrate (in $\mu\text{g}/\text{cm}^2$) absorbed into skin for a brand name and a generic antifungal ointment (from the article "Improved Bioequivalence Assessment of Topical Dermatological Drug Products Using Dermatopharmacokinetics." B. N'Dri-Stempfer, W.Navidi, R. Guy, and A. Bunge, Pharmaceutical Research, 2009:316–328).



True or false:

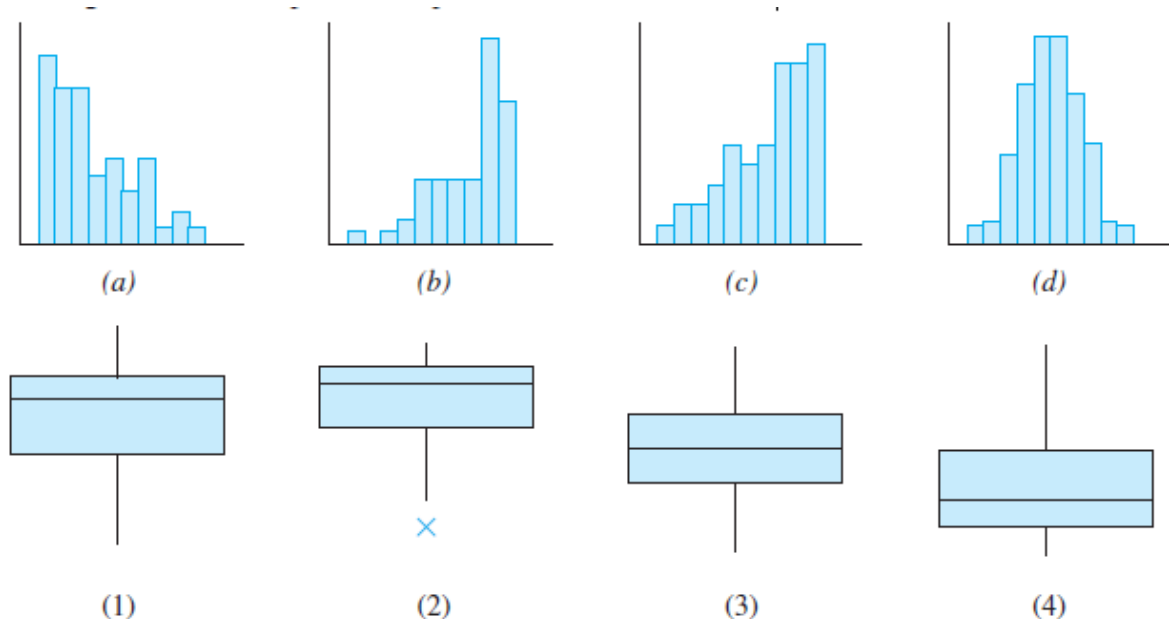
- The median amount absorbed for the brand name drug is greater than the 25th percentile of the amount absorbed for the generic drug.
- The median amount absorbed for the brand name drug is greater than the median amount absorbed for the generic drug.
- About half the sample values for the brand name drug are between 2 and 3.
- There is a greater proportion of values outside the box for the brand name drug than for the generic drug.
- Both samples are skewed to the right.
- Both samples contain outliers.

7) Following are summary statistics for two data sets, A and B.

	A	B
Minimum	0.066	-2.235
1 st Quartile	1.42	5.27
Median	2.60	8.03
3 rd Quartile	6.02	9.13
Maximum	10.08	10.51

- Compute the interquartile ranges for both A and B.
- Do the summary statistics for A provide enough information to construct a boxplot? If so, construct the boxplot. If not, explain why.
- Do the summary statistics for B provide enough information to construct a boxplot? If so, construct the boxplot. If not, explain why.

8) Match each histogram to the boxplot that represents the same data set.



17. Refer to the asphalt data in Example 1.12 (page 18).

- Construct a boxplot for the asphalt data.
- Which values, if any, are outliers?
- Construct a dotplot for the asphalt data.
- For purposes of constructing boxplots, an outlier is defined to be a point whose distance from the nearest quartile is more than 1.5 IQR. A more general, and less precise, definition is that an outlier is any point that is detached from the bulk of the data. Are there any points in the asphalt data set that are outliers under this more general definition, but not under the boxplot definition? If so, which are they?