Basic Statistics and Probability (2018-19)
Science & Engineering Program Boston University-Faculty of Science UAM

Chapter 2: Methods for Describing Sets of Data

The following problems are from McClave, J. and Sincich, T. (2017), *Statistics*, 13th. edition, Pearson. The number preceding the exercise is the corresponding one from this textbook. Some of the problems may also appear in previous editions of the book, possibly with different numbering.

2.9 Cheek teeth of extinct primates. The characteristics of cheek teeth (e.g., molars) can provide anthropologists with MoLARS information on the dietary habits of extinct mammals. The cheek teeth of an extinct primate species were the subject of research reported in the American Journal of Physical Anthropology (Vol. 142, 2010). A total of 18 cheek teeth extracted from skulls discovered in western Wyoming were analyzed. Each tooth was classified according to degree of wear (unworn, slight, light-moderate, moderate, moderate-heavy, or heavy). The 18 measurements are listed here.

Data on Degree of Wear	
Unknown	Slight
Unknown	Slight
Unknown	Heavy
Moderate	Unworn
Slight	Light-moderate
Unknown	Light-moderate
Moderate-heavy	Moderate
Moderate	Unworn
Slight	Unknown

a. Identify the variable measured in the study and its type (quantitative or qualitative).
b. Count the number of cheek teeth in each wear category.
c. Calculate the relative frequency for each wear category.
d. Construct a relative frequency bar graph for the data.
e. Construct a Pareto diagram for the data.
f. Identify the degree of wear category that occurred most often in the sample of 18 cheek teeth.

2.11 Estimating the rhino population. The International Rhino Foundation estimates that there are 28,933 rhinoceroses living in the wild in Africa and Asia. A breakdown of the number of rhinos of each species is reported in the accompanying table:

Rhino Species	Population Estimate
African Black	5,055
African White	20,405
(Asian) Sumatran	100
(Asian) Javan	40
(Asian) Greater One-Horned	3,333
Total	28,933

*Source: International Rhino Foundation, 2014.*

a. Construct a relative frequency table for the data.
b. Display the relative frequencies in a bar graph.
c. What proportion of the 28,933 rhinos are African rhinos? Asian?

2.31 Consider the stem-and-leaf display shown here:

Stem	Leaf
5	1
4	457
3	00036
2	1134599
1	2248
0	012

a. How many observations were in the original data set?
b. In the bottom row of the stem-and-leaf display, identify the stem, the leaves, and the numbers in the original data set represented by this stem and its leaves.
c. Re-create all the numbers in the data set, and construct a dot plot.

2.34 Irrelevant speech effects. In a psychological study of short-term memory, *irrelevant speech effects* refer to the degree to which the memorization process is impaired by irrelevant background speech (for example, trying to memorize a list of numbers while listening to a speech in an unfamiliar language). An analysis of irrelevant speech effects was carried out and published in *Acoustical Science & Technology* (Vol. 35, 2014). Subjects performed the memorization task under two conditions: (1) with irrelevant background speech and (2) in silence. The difference in the error rates for the two conditions–called the relative difference in error rate (RDER)–was computed for each subject. A MINITAB histogram summarizing the RDER values for 71 subjects is displayed here.
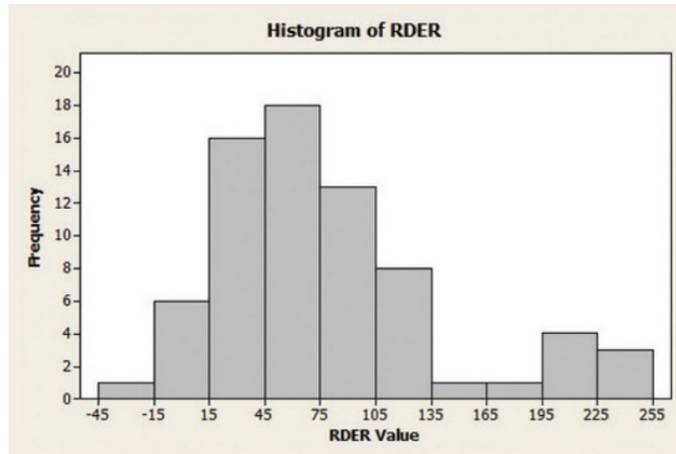
a. Convert the frequency histogram into a relative frequency histogram.
b. What proportion of the subjects had RDER values between 75 and 105?
c. What proportion of the subjects had RDER values below 15?

2.57 Calculate the mean for samples for which

a. $n = 10, \sum x = 85$
b. $n = 16, \sum x = 400$
c. $n = 45, \sum x = 35$
d. $n = 18, \sum x = 242$

2.58 Construct one data set consisting of five measurements, and another consisting of six measurements, for which the medians are equal.

2.68 Mongolian desert ants. The *Journal of Biogeography* (Dec. 2003) published an article on the first comprehensive study of ants in Mongolia (Central Asia). Botanists placed seed baits at 11 study sites and observed the ant species attracted to each site. Some of the data recorded at each study site are provided in the table below and in file `GOBIANT.txt`.

Site	Region	Annual Rainfall (mm)	Max. Daily Temp. (C)	Total Plant Cover (%)	Number of Ant Species	Species Diversity Index
1	Dry Steppe	196	5.7	40	3	.89
2	Dry Steppe	196	5.7	52	3	.83
3	Dry Steppe	179	7	40	52	1.31
4	Dry Steppe	197	8	43	7	1.48
5	Dry Steppe	149	8.5	27	5	.97
6	Gobi Desert	112	10.7	30	49	.46
7	Gobi Desert	125	11.4	16	5	1.23
8	Gobi Desert	99	10.9	30	4	
9	Gobi Desert	125	11.4	56	4	.76
10	Gobi Desert	84	11.4	22	5	1.26
11	Gobi Desert	115	11.4	14	4	.69

a. Find the mean, median, and mode for the number of ant species discovered at the 11 sites. Interpret each of these values.
b. Which measure of central tendency would you recommend to describe the center of the number-of-ant- species distribution? Explain.
c. Find the mean, median, and mode for the percentage of total plant cover at the five Dry Steppe sites only.
d. Find the mean, median, and mode for the percentage of total plant cover at the six Gobi Desert sites only.
e. On the basis of the results of parts **c** and **d**, does the cen- ter of the distribution for total plant cover percentage appear to be different at the two regions?

2.80 Calculate the range, variance, and standard deviation for the following samples:

a. 4,2,1,0,1
b. 1,6,2,2,3,0,3
c. 8, -2,1,3,5,4,4,1,3
d. 0,2,0,0, -1,1, -2,1,0, -1,1, -1,0, -3, -2, -1,0,1

2.98 The following is a sample of 25 measurements.

7	6	6	11	8	9	11	9	10	8	7	7	5
9	10	7	7	7	7	9	12	10	10	8	6	

a. Compute $\bar{x}$, s^2, and s for this sample.
b. Count the number of measurements in the intervals $\bar{x} \mp s$, $\bar{x} \mp 2s$, and $\bar{x} \mp 3s$. Express each count as a percentage of the total number of measurements.
c. Compare the percentages found in part **b** with the percentages given by the empirical rule and Chebyshevs rule.
d. Calculate the range and use it to obtain a rough approximation for s. Does the result compare favorably with the actual value for s found in part **a**?
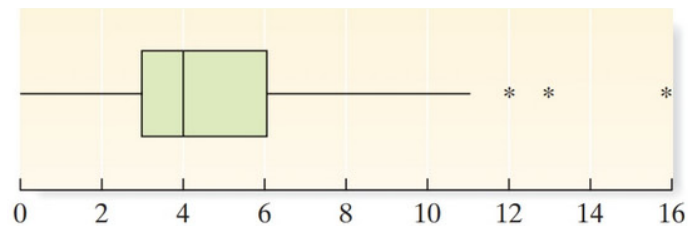
2.115 Give the percentage of measurements in a data set that are above and below each of the following percentiles:

a. 75th percentile
b. 50th percentile
c. 20th percentile
d. 84th percentile

2.117 Compute the z-score corresponding to each of the following values of x:

a. $x = 40$, $s = 5$, $\bar{x} = 30$
b. $x = 90$, $\bar{x} = 89$, $s = 2$
c. $\bar{x} = 50$, $s = 5$, $x = 50$
d. $s = 4$, $x = 20$, $\bar{x} = 30$

2.137 Consider the following horizontal box plot:

a. What is the median of the data set (approximately)?
b. What are the upper and lower quartiles of the data set (approximately)?
c. What is the interquartile range of the data set (approximately)?
d. Is the data set skewed to the left, skewed to the right, or symmetric?
e. What percentage of the measurements in the data set lie to the right of the median? To the left of the upper quartile?
f. Identify any outliers in the data.

2.139 Consider the following sample data set. Construct a box plot for the data and use it to identify any outliers.

121	171	158
173	184	163
157	85	145
165	172	196
170	159	172
161	187	100
142	166	171