Cumulative Review Exercises CR5.1 - CR5.19

CR5.1 The article "Rocker Shoe Put to the Test: Can it Really Walk the Walk as a Way to Get in Shape?" (*USA Today*, October 12, 2009) describes claims made by Skechers about Shape-Ups, a shoe line introduced in 2009. These curved-sole sneakers are supposed to help you "get into shape without going to the gym" according to a Skechers advertisement. Briefly describe how you might design a study to investigate this claim. Include how you would select subjects and what variables you would measure. Is the study you designed an observational study or an experiment?

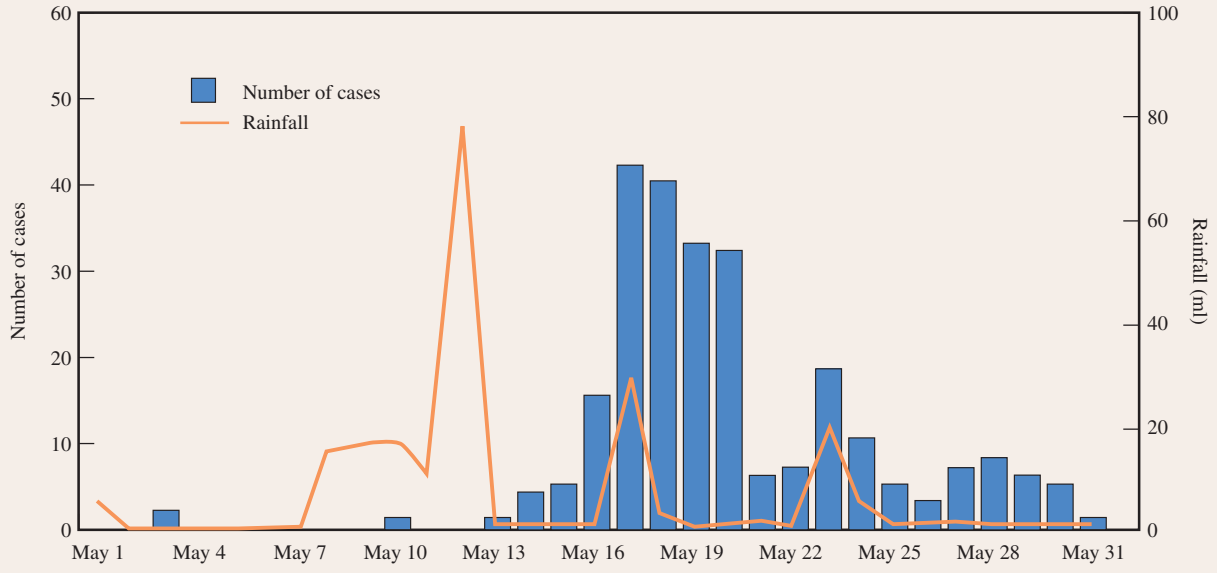
CR5.2 Data from a survey of 1046 adults age 50 and older were summarized in the *AARP Bulletin* (November 2009). The following table gives relative frequency distributions of the responses to the question, "How much do you plan to spend for holiday gifts this year?" for respondents age 50 to 64 and for respondents age 65 and older. Construct a histogram for each of the two age groups and comment on the differences between the two age groups. (Notice that the interval widths in the relative frequency distribution are not the same, so you shouldn't use relative frequency on the y-axis for your histograms.)

Amount Plan to Spend	Relative Frequency for Age Group 50 to 64	Relative Frequency for Age Group 65 and Older
less than $100	.20	.36
$100 to <$200	.13	.11
$200 to <$300	.16	.16
$300 to <$400	.12	.10
$400 to <$500	.11	.05
$500 to <$1000	.28	.22

CR5.3 The graph in Figure CR5.3 appeared in the report "Testing the Waters 2009" (Natural Resources Defense Council). Spend a few minutes looking at the graph and reading the caption that appears with the graph. Briefly explain how the graph supports the claim that discharges of polluted storm water may be responsible for increased illness levels.

CR5.4 The cost of Internet access was examined in the report "Home Broadband Adoption 2009" (pewinternet.org). In 2009, the mean and median amount paid monthly for service for broadband users was reported as $39.00 and $38.00, respectively. For

FIGURE CR5.3 Influence of Heavy Rainfall on Occurrence of *E. Coli* Infections



The graph shows the relationship between unusually heavy rainfall and the number of confirmed cases of *E. coli* infection that occurred during a massive disease outbreak in Ontario, Quebec, in May 2000. The incubation period for *E.Coli* is usually 3 to 4 days, which is consistent with the lag between extreme precipitation events and surges in the number of cases.

Bold exercises answered in back ● Data set available online ✦ Video Solution available

dial-up users, the mean and median amount paid monthly were $26.60 and $20.00, respectively. What do the values of the mean and median tell you about the shape of the distribution of monthly amount paid for broadband users? For dial-up users?

CR5.5 ● Foal weight at birth is an indicator of health, so it is of interest to breeders of thoroughbred horses. Is foal weight related to the weight of the mare (mother)? The accompanying data are from the paper "Suckling Behaviour Does Not Measure Milk Intake in Horses" (*Animal Behaviour* [1999]: 673–678):

Observation	Mare Weight (x, in kg)	Foal weight (y, in kg)
1	556	129
2	638	119
3	588	132
4	550	123.5
5	580	112
6	642	113.5
7	568	95
8	642	104
9	556	104
10	616	93.5
11	549	108.5
12	504	95
13	515	117.5
14	551	128
15	594	127.5

The correlation coefficient for these data is 0.001. Construct a scatterplot of these data and then write a few sentences describing the relationship between mare weight and foal weight that refer both to the value of the correlation coefficient and the scatterplot.

CR5.6 In August 2009, Harris Interactive released the results of the "Great Schools" survey. In this survey, 1086 parents of children attending a public or private school were asked approximately how much they had spent on school supplies over the last school year. For this sample, the mean amount spent was $235.20 and the median amount spent was $150.00. What does the large difference between the mean and median tell you about this data set?

CR5.7 ● Bidri is a popular and traditional art form in India. Bidri articles (bowls, vessels, and so on) are made by casting from an alloy containing primarily zinc along with some copper. Consider the following observations on copper content (%) for a sample of Bidri artifacts in London's Victoria and Albert Museum ("Enigmas of Bidri," *Surface Engineering* [2005]: 333–339), listed in increasing order:

2.0 2.4 2.5 2.6 2.6 2.7 2.7 2.8 3.0
3.1 3.2 3.3 3.3 3.4 3.4 3.6 3.6 3.6
3.6 3.7 4.4 4.6 4.7 4.8 5.3 10.1

a. Construct a dotplot for these data.
b. Calculate the mean and median copper content.
c. Will an 8% trimmed mean be larger or smaller than the mean for this data set? Explain your reasoning.

CR5.8 ● ✦ Medicare's new medical plans offer a wide range of variations and choices for seniors when picking a drug plan (*San Luis Obispo Tribune*, November 25, 2005). The monthly cost for a stand-alone drug plan can vary from a low of $1.87 in Montana, Wyoming, North Dakota, South Dakota, Nebraska, Minnesota, and Iowa to a high of $104.89. Here are the lowest and highest monthly premiums for stand-alone Medicare drug plans for each state:

State	$ Low	$ High
Alabama	14.08	69.98
Alaska	20.05	61.93
Arizona	6.14	64.86
Arkansas	10.31	67.98
California	5.41	66.08
Colorado	8.62	65.88
Connecticut	7.32	65.58
Delaware	6.44	68.91
District of Columbia	6.44	68.91
Florida	10.35	104.89
Georgia	17.91	73.17
Hawaii	17.18	64.43
Idaho	6.33	68.88
Illinois	13.32	65.04
Indiana	12.30	70.72
Iowa	1.87	99.90
Kansas	9.48	67.88
Kentucky	12.30	70.72
Louisiana	17.06	70.59
Maine	19.60	65.39
Maryland	6.44	68.91
Massachusetts	7.32	65.58
Michigan	13.75	65.69
Minnesota	1.87	99.90
Mississippi	11.60	70.59
Missouri	10.29	68.26
Montana	1.87	99.90
Nebraska	1.87	99.90
Nevada	6.42	64.63

(continued)

State	$ Low	$ High
New Hampshire	19.60	65.39
New Jersey	4.43	66.53
New Mexico	10.65	62.38
New York	4.10	85.02
North Carolina	13.27	65.03
North Dakota	1.87	99.90
Ohio	14.43	68.05
Oklahoma	10.07	70.79
Oregon	6.93	64.99
Pennsylvania	10.14	68.61
Rhode Island	7.32	65.58
South Carolina	16.57	69.72
South Dakota	1.87	99.90
Tennessee	14.08	69.98
Texas	10.31	68.41
Utah	6.33	68.88
Vermont	7.32	65.58
Virginia	8.81	68.61
Washington	6.93	64.99
West Virginia	10.14	68.61
Wisconsin	11.42	63.23
Wyoming	1.87	99.90

Which of the following can be determined from the data? If it can be determined, calculate the requested value. If it cannot be determined, explain why not.

a. the median premium cost in Colorado
b. the number of plan choices in Virginia
c. the state(s) with the largest difference in cost between plans
d. the state(s) with the choice with the highest premium cost
e. the state for which the minimum premium cost is greatest
f. the mean of the minimum cost of all states beginning with the letter "M"

CR5.9 *Note: This exercise requires the use of a computer.* Refer to the Medicare drug plan premium data of Exercise 5.8.

a. Construct a dotplot or a stem-and-leaf display of the lowest premium cost data.
b. Based on the display in Part (a), which of the following would you expect to be the case for the lowest cost premium data?
   i. the mean will be less than the median
   ii. the mean will be approximately equal to the median
   iii. the mean will be greater than the median
c. Compute the mean and median for the lowest cost premium data.
d. Construct an appropriate graphical display for the highest cost premium data.
e. Compute the mean and median for the highest cost premium data.

CR5.10 ● ✦ The paper "Total Diet Study Statistics on Element Results" (Food and Drug Administration, April 25, 2000) gave information on sodium content for various types of foods. Twenty-six tomato catsups were analyzed. Data consistent with summary quantities given in the paper were

Sodium content (mg/kg)

12,148 10,426 10,912 9116 13,226 11,663
11,781 10,680 8457 10,788 12,605 10,591
11,040 10,815 12,962 11,644 10,047 10,478
10,108 12,353 11,778 11,092 11,673 8758
11,145 11,495

Compute the values of the quartiles and the interquartile range.

CR5.11 ● The paper referenced in Exercise 5.10 also gave data on sodium content (in milligrams per kilogram) of 10 chocolate puddings made from instant mix:

3099 3112 2401 2824 2682 2510 2297
3959 3068 3700

a. Compute the mean, the standard deviation, and the interquartile range for sodium content of these chocolate puddings.
b. Based on the interquartile range, is there more or less variability in sodium content for the chocolate pudding data than for the tomato catsup data of Cumulative Exercise 5.10?

CR5.12 ● ✦ A report from Texas Transportation Institute (Texas A&M University System, 2005) on congestion reduction strategies looked into the extra travel time (due to traffic congestion) for commute travel per traveler per year in hours for different urban areas. Below are the data for urban areas that had a population of over 3 million for the year 2002.

Urban Area	Extra Hours per Traveler per Year
Los Angeles	98
San Francisco	75
Washington DC	66
Atlanta	64

(continued)

Urban Area	Extra Hours per Traveler per Year
Houston	65
Dallas, Fort Worth	61
Chicago	55
Detroit	54
Miami	48
Boston	53
New York	50
Phoenix	49
Philadelphia	40

a. Compute the mean and median values for extra travel hours. Based on the values of the mean and median, is the distribution of extra travel hours likely to be approximately symmetric, positively skewed, or negatively skewed?
b. Construct a modified boxplot for these data and comment on any interesting features of the plot.

CR5.13 ● ✦ The paper "Relationship Between Blood Lead and Blood Pressure Among Whites and African Americans" (a technical report published by Tulane University School of Public Health and Tropical Medicine, 2000) gave summary quantities for blood lead level (in micrograms per deciliter) for a sample of whites and a sample of African Americans. Data consistent with the given summary quantities follow:

Whites	8.3	0.9	2.9	5.6	5.8	5.4	1.2
	1.0	1.4	2.1	1.3	5.3	8.8	6.6
	5.2	3.0	2.9	2.7	6.7	3.2	
African	4.8	1.4	0.9	10.8	2.4	0.4	5.0
Americans	5.4	6.1	2.9	5.0	2.1	7.5	3.4
	13.8	1.4	3.5	3.3	14.8	3.7	

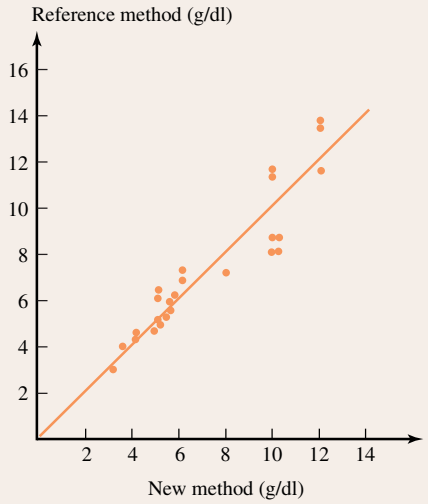
a. Compute the values of the mean and the median for blood lead level for the sample of African Americans. Which of the mean or the median is larger? What characteristic of the data set explains the relative values of the mean and the median?
b. Construct a comparative boxplot for blood lead level for the two samples. Write a few sentences comparing the blood lead level distributions for the two samples.

CR5.14 ● Cost-to-charge ratios (the percentage of the amount billed that represents the actual cost) for 11 Oregon hospitals of similar size were reported separately for inpatient and outpatient services. The data are shown in the table at the top of the next column.

	Cost-to-Charge Ratio	
Hospital	Inpatient	Outpatient
Blue Mountain	80	62
Curry General	76	66
Good Shepherd	75	63
Grande Ronde	62	51
Harney District	100	54
Lake District	100	75
Pioneer	88	65
St. Anthony	64	56
St. Elizabeth	50	45
Tillamook	54	48
Wallowa Memorial	83	71

a. Does there appear to be a strong linear relationship between the cost-to-charge ratio for inpatient and outpatient services? Justify your answer based on the value of the correlation coefficient and examination of a scatterplot of the data.
b. Are any unusual features of the data evident in the scatterplot?
c. Suppose that the observation for Harney District was removed from the data set. Would the correlation coefficient for the new data set be greater than or less than the one computed in Part (a)? Explain.

CR5.15 The accompanying scatterplot shows observations on hemoglobin level, determined both by the standard spectrophotometric method (y) and by a new, simpler method based on a color scale (x) ("A Simple and Reliable Method for Estimating Hemoglobin," *Bulletin of the World Health Organization* [1995]: 369–373):



a. Does it appear that x and y are highly correlated?
b. The paper reported that $r = .9366$. How would you describe the relationship between the two variables?
c. The line pictured in the scatterplot has a slope of 1 and passes through (0, 0). If x and y were always identical, all points would lie exactly on this line. The authors of the paper claimed that perfect correlation ($r = 1$) would result in this line. Do you agree? Explain your reasoning.

CR5.16 In the article "Reproductive Biology of the Aquatic Salamander *Amphiuma tridactylum* in Louisiana" (*Journal of Herpetology* [1999]: 100–105), 14 female salamanders were studied. Using regression, the researchers predicted y = clutch size (number of salamander eggs) from x = snout-vent length (in centimeters) as follows:

$$\hat{y} = -147 + 6.175x$$

For the salamanders in the study, the range of snout-vent lengths was approximately 30 to 70 cm.

a. What is the value of the y intercept of the least-squares line? What is the value of the slope of the least-squares line? Interpret the slope in the context of this problem.
b. Would you be reluctant to predict the clutch size when snout-vent length is 22 cm? Explain.

CR5.17 Exercise CR5.16 gave the least-squares regression line for predicting y = clutch size from x = snout-vent length ("Reproductive Biology of the Aquatic Salamander *Amphiuma tridactylum* in Louisiana," *Journal of Herpetology* [1999]: 100–105). The paper also reported $r^2 = .7664$ and SSTo = 43,951.

a. Interpret the value of r^2.
b. Find and interpret the value of s_e (the sample size was $n = 14$).

CR5.18 ● A study, described in the paper "Prediction of Defibrillation Success from a Single Defibrillation Threshold Measurement" (*Circulation* [1988]: 1144–1149) investigated the relationship between defibrillation success and the energy of the defibrillation shock (expressed as a multiple of the defibrillation threshold) and presented the following data:

Energy of Shock	Success (%)
0.5	33.3
1.0	58.3
1.5	81.8
2.0	96.7
2.5	100.0

a. Construct a scatterplot of y = success percentage and x = energy of shock. Does the relationship appear to be linear or nonlinear?
b. Fit a least-squares line to the given data, and construct a residual plot. Does the residual plot support your conclusion in Part (a)? Explain.
c. Consider transforming the data by leaving y unchanged and using either $x' = \sqrt{x}$ or $x'' = \log(x)$. Which of these transformations would you recommend? Justify your choice by appealing to appropriate graphical displays.
d. Using the transformation you recommended in Part (c), find the equation of the least-squares line that describes the relationship between y and the transformed x.
e. What would you predict success percentage to be when the energy of shock is 1.75 times the threshold level? When it is 0.8 times the threshold level?

CR5.19 ● The paper "Population Pressure and Agricultural Intensity" (*Annals of the Association of American Geographers* [1977]: 384–396) reported a positive association between population density and agricultural intensity. The following data consist of measures of population density (x) and agricultural intensity (y) for 18 different subtropical locations:

x	1.0	26.0	1.1	101.0	14.9	134.7
y	9	7	6	50	5	100
x	3.0	5.7	7.6	25.0	143.0	27.5
y	7	14	14	10	50	14
x	103.0	180.0	49.6	140.6	140.0	233.0
y	50	150	10	67	100	100

a. Construct a scatterplot of y versus x. Is the scatterplot compatible with the statement of positive association made in the paper?

b. The scatterplot in Part (a) is curved upward like segment 2 in Figure 5.38, suggesting a transformation that is up the ladder for x or down the ladder for y. Try a scatterplot that uses y and x^2. Does this transformation straighten the plot?

c. Draw a scatterplot that uses $\log(y)$ and x. The $\log(y)$ values, given in order corresponding to the y values, are 0.95, 0.85, 0.78, 1.70, 0.70, 2.00, 0.85, 1.15, 1.15, 1.00, 1.70, 1.15, 1.70, 2.18, 1.00, 1.83, 2.00, and 2.00. How does this scatterplot compare with that of Part (b)?

d. Now consider a scatterplot that uses transformations on both x and y: $\log(y)$ and x^2. Is this effective in straightening the plot? Explain.