

### **Problem 1**

The head circumference (in centimeters) of 15 college-age males was obtained, resulting in the following measurements:

55, 56, 56, 56.5, 57, 57, 57, 57.5, 58, 58, 58, 58.5, 59, 59, 63.

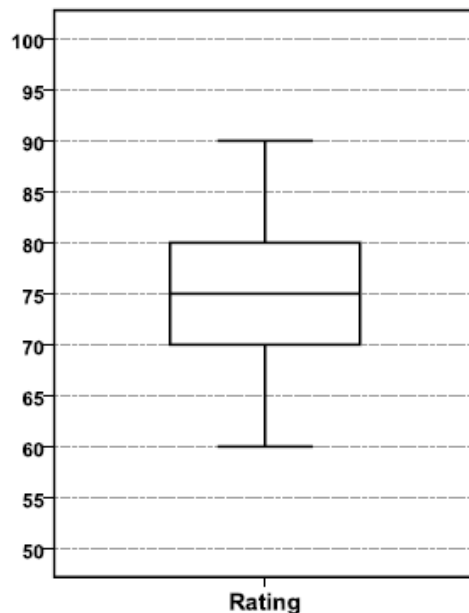
If the last measurement (63 cm's) were incorrectly recorded as 73, which one of the following statistics would change?

- a) Q1 (1st quartile)
- b) Standard deviation
- c) Median
- d) Q3 (3rd quartile)

### **Problem 2**

The following boxplot gives the distribution of the ratings of a new brand of peanut butter for 50 randomly selected consumers (100 points possible with higher points corresponding to a more favorable rating).

Identify the five-number summary measures and the description of each of them using the box plot.





### **Problem 3**

**This is a standard deviation contest. You must choose four numbers from the whole numbers 0 to 10, with repeats allowed.**

**3.1 Choose the data set that has the smallest standard deviation.**

- a) 7, 7, 7, 8
- b) 3, 5, 7, 9
- c) 1, 1, 1, 1
- d) 1, 2, 3, 4

**3.2 Choose the data set that has the largest standard deviation.**

- a) 0, 3, 6, 10
- b) 9, 9, 10, 10
- c) 1, 4, 7, 10
- d) 0, 0, 10, 10

### **Problem 4**

**Researchers are interested in how crime rates are different for southern states.**

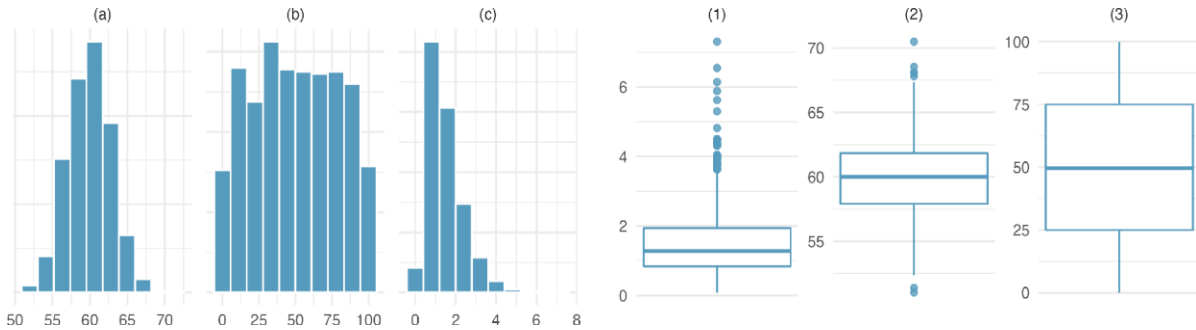
**The standard deviation of unemployment for 35 to 39-year-old for southern states is 8.01 and the standard deviation for non-southern states is 8.76. What do these standard deviations tell you about the southern and non-southern states?**

- a) The average unemployment rate for 35 to 39-year old in southern states is about the same as in non-southern states.
  - b) The average unemployment rate for 35 to 39-year old in southern states is slightly lower than in non-southern states.
  - c) The average spread from the mean rate for 35 to 39-year-old in southern states is lower than the non-southern states.
-



### Problem 5

Describe (in words) the distribution in the histograms below and match them to the box plots.

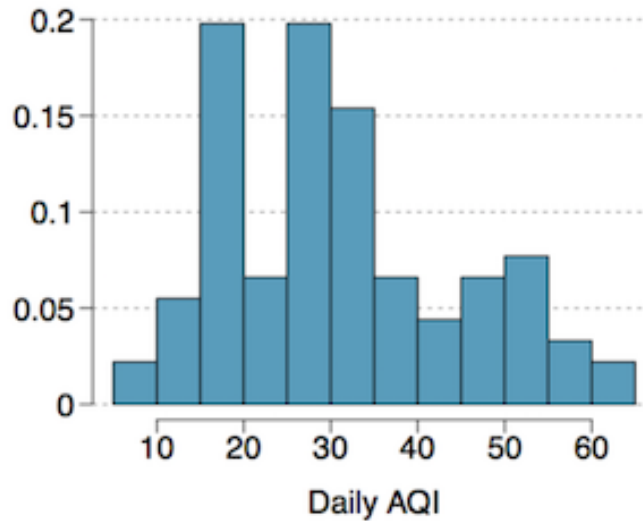


### Problem 6

College students in a statistics class were asked how many hours of television they watch per week, including online streaming services. This sample yielded an average of 8.28 hours, with a standard deviation of 7.18 hours. Is the distribution of the number of hours students watch television weekly symmetric? If not, what shape would you expect this distribution to have? Explain your reasoning.

### Problem 7

Daily air quality is measured by the air quality index (AQI) reported by the Environmental Protection Agency. This index reports the pollution level and what associated health effects might be a concern. The index is calculated for five major air pollutants regulated by the Clean Air Act and takes values from 0 to 300, where a higher value indicates lower air quality. AQI was reported for a sample of 91 days in 2011 in Durham, NC. The histogram below shows the distribution of the AQI values on these days.



7.1 Estimate the median AQI value of this sample.

7.2 Would you expect the mean AQI value of this sample to be higher or lower than the median? Explain your reasoning.

7.3 Estimate Q1, Q3, and IQR for the distribution.

7.4 Would any of the days in this sample be considered to have an unusually low or high AQI? Explain your reasoning.

### Problem 8

In a class of 25 students, 24 of them took an exam in class and 1 student took a make-up exam the following day. The professor graded the first batch of 24 exams and found an average score of 74 points with a standard deviation of 8.9 points. The student who took the make-up the following day scored 64 points on the exam.

8.1 Does the new student's score increase or decrease the average score?

8.2 What is the new average?

8.3 Does the new student's score increase or decrease the standard deviation of the scores?



---

### **Problem 9**

**In each of the following situations, which is the explanatory variable and which is the response variable? Are they categorical or quantitative (quantitative means "numerical")?**

**9.1 The typical number of calories a person consumes per day and that person's percent of body fat.**

- a) Number of calories consumed per day: response, quantitative. Percent of body fat: explanatory, quantitative.
- b) Number of calories consumed per day: explanatory, quantitative. Percent of body fat: response, quantitative.
- c) Number of calories consumed per day: response, quantitative. Percent of body fat: explanatory, categorical.
- d) Number of calories consumed per day: explanatory, categorical. Percent of body fat: response, categorical.

**9.2 Water temperature is controlled at different levels and growth (measured by weight) of corals in aquariums.**

- a. Water temperature: response, quantitative. Growth: explanatory, categorical.
- b. Water temperature: explanatory, categorical. Growth: response, categorical.
- c. Water temperature: response, categorical. Growth: explanatory, quantitative.
- d. Water temperature: explanatory, quantitative. Growth: response, quantitative

---

### **Problem 10**

**Coffee is a leading export from several developing countries. When coffee prices are high, farmers often clear forests to plant more coffee trees. Here are data on prices paid to coffee growers in Indonesia and the rate of deforestation in a national park that lies in a coffee-producing region, for five years:**



---

Price (cents per pound)	Deforestation (percent)
29	0.49
40	1.59
54	1.69
55	1.82
72	3.10

---

**Coffee is currently priced in dollars. If it were priced in euros, and the dollar prices in the above table were translated into the equivalent prices in euros, would the correlation between coffee price and percent deforestation change?**

- (a) The correlation would remain zero, because the two variables are independent
  - (b) Yes, units affect correlation
  - (c) No, units do not affect correlation
  - (d) It is impossible to calculate the correlation, because coffee price is categorical.
- 

### **Problem 11**

**A study shows that there is a positive correlation between the size of a hospital (measured by its number of beds (x)) and the median number of days (y) that patients remain in the hospital.**

**What lurking variable could be present in this study?**

- a) cost: it's more expensive to run larger hospitals.
  - b) severity of disease: since large hospitals have better facilities and more doctors to cope with severe illness.
  - c) number of visitors: since larger hospitals receive more visitors.
  - d) facilities: since larger hospitals have better facilities, patients choose to stay longer
-

**Problem 12**

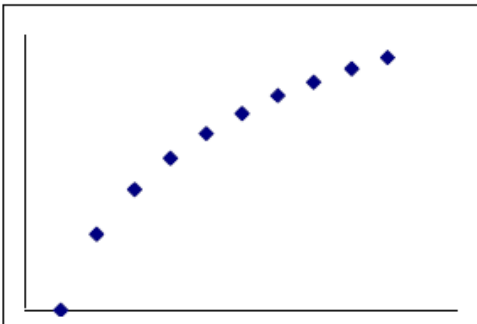
Milk use is positively correlated to cancer rates. While this is not a popular finding within the milk industry, there is a moderately positive correlation with drinking milk and getting cancer (Paulos, 1990). Milk consumption is greater in wealthier countries. In wealthier countries people live longer. Greater longevity means people live long enough to eventually get some type of cancer.

- 10.1 Which is the response and explanatory variable?
- 10.2 Identify the presence of a lurking variable, if any?
- 10.3 Will you conclude that drinking more milk increases the chance of getting cancer? Explain your reasoning.

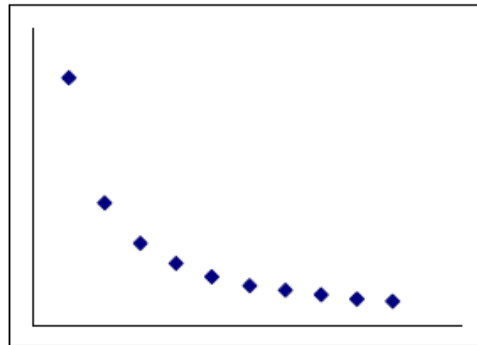
**Problem 13**

Which of the following plots will have a correlation coefficient of 0.85?

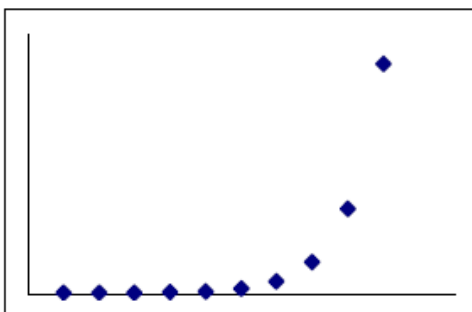
A.



B.



C.



D.

