

## STATISTICS

### SECTION I

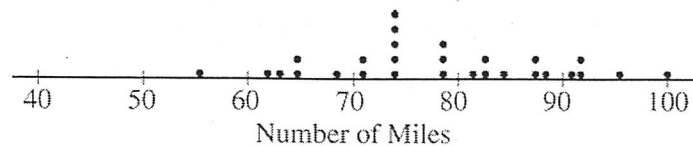
Time—1 hour and 30 minutes

Number of questions—40

Percent of total grade—50

**Directions:** Solve each of the following problems, using the available space for scratch work. Select the answer choice that is best in each case and place the letter of your choice in the corresponding box on the student answer sheet. Do not spend too much time on any one problem.

1. The dotplot below displays the total number of miles that the 28 residents of one street in a certain community traveled to work in one five-day workweek.

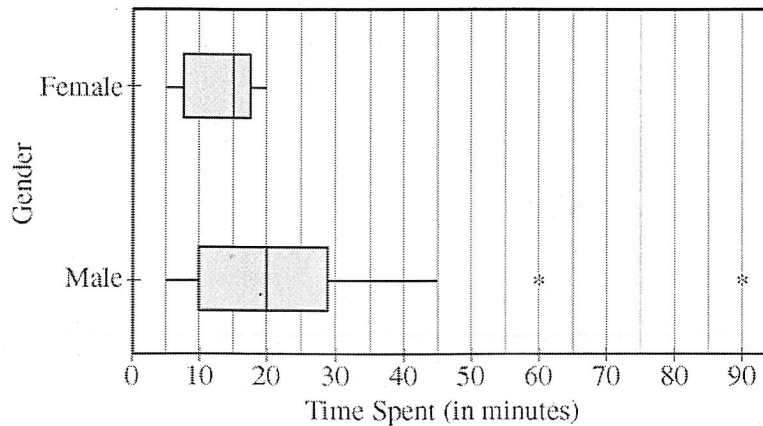


Which of the following is closest to the percentile rank of a resident from this street who traveled 85 miles to work that week?

- (A) 60
  - (B) 70
  - (C) 75
  - (D) 80
  - (E) 85
2. A large simple random sample of people aged nineteen to thirty living in the state of Colorado was surveyed to determine which of two MP3 players just developed by a new company was preferred. To which of the following populations can the results of this survey be safely generalized?
- (A) Only people aged nineteen to thirty living in the state of Colorado who were in this survey
  - (B) Only people aged nineteen to thirty living in the state of Colorado
  - (C) All people living in the state of Colorado
  - (D) Only people aged nineteen to thirty living in the United States
  - (E) All people living in the United States

GO ON TO THE NEXT PAGE.

3. Janelle collected data on the amount of time in minutes each person in a large sample of customers spent in a local store. The data also included recording the gender of each customer. These data were used to generate the boxplots shown below.



Which of the following statements is true?

- (A) The range in the amount of time in minutes males in the sample of customers spent in the store is approximately 40 minutes.
  - (B) The mean amount of time in minutes males in the sample of customers spent in the store is approximately 20 minutes.
  - (C) The third quartile of the amount of time in minutes males in the sample of customers spent in the store is approximately 45 minutes.
  - (D) The interquartile range of the amount of time in minutes females in the sample of customers spent in the store is 15 minutes.
  - (E) Approximately half of the males in the sample of customers spent at least as much time in the store as any female in the sample of customers.
4. The probability of obtaining a head when a certain coin is flipped is about 0.65. Which of the following is closest to the probability that heads would be obtained 15 or fewer times when this coin is flipped 25 times?
- (A) 0.14
  - (B) 0.37
  - (C) 0.39
  - (D) 0.60
  - (E) 0.65

GO ON TO THE NEXT PAGE.

6. The buyer for an electronics store wants to estimate the proportion of defective wireless game controllers in a shipment of 5,000 controllers from the store's primary supplier. The shipment consists of 200 boxes each containing 25 controllers. The buyer numbers the boxes from 1 to 200 and randomly selects six numbers in that range. She then opens the six boxes with the corresponding numbers, examines all 25 controllers in each of these boxes, and determines the proportion of the 150 controllers that are defective. What type of sample is this?

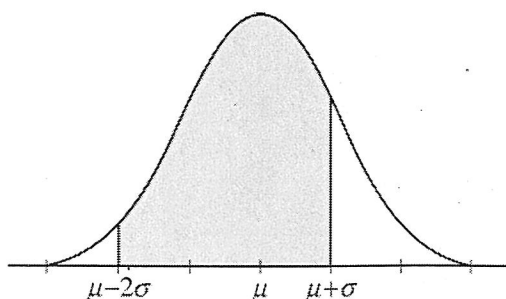
- (A) Biased random sample
- (B) Nonrandom sample
- (C) Simple random sample
- (D) Stratified random sample
- (E) Cluster random sample

---

7. A distribution of test scores is not symmetric. Which of the following is the best estimate of the  $z$ -score of the third quartile?

- (A) 0.67
- (B) 0.75
- (C) 1.00
- (D) 1.41
- (E) This  $z$ -score cannot be estimated from the information given.

9. A certain type of remote-control car has a fully charged battery at the time of purchase. The distribution of running times of cars of this type, before they require recharging of the battery for the first time after its period of initial use, is approximately normal with a mean of 80 minutes and a standard deviation of 2.5 minutes. The shaded area in the figure below represents which of the following probabilities?



- (A) The probability that the running time of a randomly selected car of this type, before it requires recharging of the battery for the first time after its period of initial use, is between 75 minutes and 82.5 minutes.
- (B) The probability that the running time of a randomly selected car of this type, before it requires recharging of the battery for the first time after its period of initial use, is between 75 minutes and 85 minutes.
- (C) The probability that the running time of a randomly selected car of this type, before it requires recharging of the battery for the first time after its period of initial use, is between 77.5 minutes and 82.5 minutes.
- (D) The probability that the running time of a randomly selected car of this type, before it requires recharging of the battery for the first time after its period of initial use, is between 77.5 minutes and 85 minutes.
- (E) The probability that the running time of a randomly selected car of this type, before it requires recharging of the battery for the first time after its period of initial use, is between 77.5 minutes and 87.5 minutes.
11. Publishers of a magazine wish to determine what proportion of the magazine's 50,000 subscribers are pleased with their subscription. The publishers intend to mail a survey to 1,000 subscribers randomly selected from those who have received the magazine for 5 years or more. This introduces selection bias, since long-subscribing customers are more likely to be pleased with their subscription. Which of the following would best eliminate selection bias?
- (A) Mail surveys to 2,000 subscribers randomly selected from those who have received the magazine for 5 years or more.
- (B) Mail surveys to 1,000 subscribers randomly selected from those who have received the magazine for 1 year or less.
- (C) Mail surveys to 1,000 subscribers randomly selected from all subscribers.
- (D) Mail surveys to 1,000 subscribers randomly selected from those who have received the magazine for 5 years or more and 1,000 subscribers randomly selected from those who have received the magazine for 1 year or less.
- (E) Mail surveys to 500 subscribers randomly selected from a group who have received a free six-month subscription within the past year.

GO ON TO THE NEXT PAGE.

12. An environmental scientist wants to test the null hypothesis that an antipollution device for cars is not effective. Under which of the following conditions would a Type I error be committed?
- (A) The scientist concludes that the antipollution device is effective when it actually is not.
  - (B) The scientist concludes that the antipollution device is not effective when it actually is.
  - (C) The scientist concludes that the antipollution device is effective when it actually is.
  - (D) The scientist concludes that the antipollution device is not effective when it actually is not.
  - (E) A Type I error cannot be committed in this situation.

13. A data set of test scores is being transformed by applying the following rule to each of the raw scores.

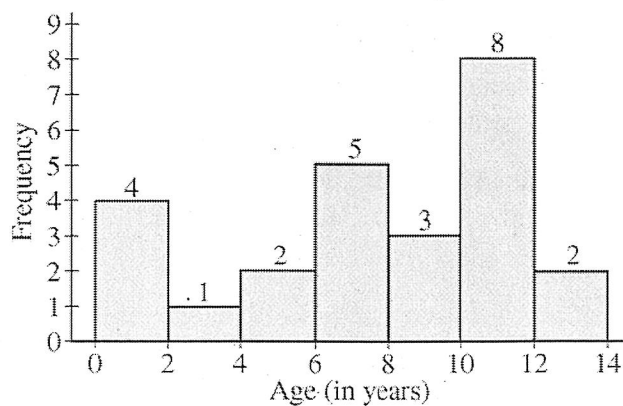
$$\text{Transformed score} = 3.5(\text{raw score}) + 6.2$$

Which of the following is NOT true?

- (A) The mean transformed score equals  $3.5(\text{the mean raw score}) + 6.2$ .
- (B) The median transformed score equals  $3.5(\text{the median raw score}) + 6.2$ .
- (C) The range of the transformed scores equals  $3.5(\text{the range of the raw scores}) + 6.2$ .
- (D) The standard deviation of the transformed scores equals  $3.5(\text{the standard deviation of the raw scores})$ .
- (E) The IQR of the transformed scores equals  $3.5(\text{the IQR of the raw scores})$ .

**GO ON TO THE NEXT PAGE.**

14. A random sample of 25 households from the Mountainview School District was surveyed. In this survey, data were collected on the age of the youngest child living in each household. The histogram below displays the data collected in the survey.



In which of the following intervals is the median of these data located?

- (A) 0 years old to less than 2 years old
- (B) 4 years old to less than 6 years old
- (C) 6 years old to less than 8 years old
- (D) 8 years old to less than 10 years old
- (E) 10 years old to less than 12 years old

GO ON TO THE NEXT PAGE.

16. Ten percent of all Dynamite Mints candies are orange and 45 percent of all Holiday Mints candies are orange. Two independent random samples, each of size 25, are selected—one from Dynamite Mints candies and the other from Holiday Mints candies. The total number of orange candies in the two samples is observed. What are the expected total number of orange candies and the standard deviation for the total number of orange candies, respectively, in the two samples?

- (A) 7 and 2.905
- (B) 7 and 3.987
- (C) 13.75 and 2.233
- (D) 13.75 and 2.905
- (E) 13.75 and 3.987

---

17. In a random sample of 60 shoppers chosen from the shoppers at a large suburban mall, 36 indicated that they had been to a movie in the past month. In an independent random sample of 50 shoppers chosen from the shoppers in a large downtown shopping area, 31 indicated that they had been to a movie in the past month. What significance test should be used to determine whether these data provide sufficient evidence to reject the hypothesis that the proportion of shoppers at the suburban mall who had been to a movie in the past month is the same as the proportion of shoppers in the large downtown shopping area who had been to a movie in the past month?

- (A) One proportion  $z$ -test
- (B) One sample  $z$ -test for a mean
- (C) One sample  $t$ -test for a mean
- (D) Two proportion  $z$ -test
- (E) Two sample  $t$ -test for means

**GO ON TO THE NEXT PAGE.**

Questions 18-19 refer to the following scenario and set of data.

In the 1830s, land surveyors began to survey the land acquired in the Louisiana Purchase. Part of their task was to note the sizes of trees they encountered in their surveying. The table of data below is for bur oak trees measured during the survey.

Tree Diameter (in inches)	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
4	2	0.005	2	0.005
5	5	0.012	7	0.016
6	19	0.044	26	0.060
7	5	0.012	31	0.072
8	41	0.095	72	0.167
9	12	0.028	84	0.195
10	53	0.123	137	0.318
11	3	0.007	140	0.325
12	64	0.148	204	0.473
13	4	0.009	208	0.483
14	46	0.107	254	0.589
15	11	0.026	265	0.615
16	48	0.111	313	0.726
17	2	0.005	315	0.731
18	47	0.109	362	0.840
19				
20	30	0.070	392	0.910
21				
22	4	0.009	396	0.919
23				
24	18	0.042	414	0.961
25				
26	1	0.002	415	0.963
27				
28	2	0.005	417	0.968
29				
30	8	0.019	425	0.986
31				
32				
33				
34	1	0.002	426	0.988
35	1	0.002	427	0.991
36	3	0.007	430	0.998
37				
38	1	0.002	431	1.000

GO ON TO THE NEXT PAGE.



18. Which of the following differences in cumulative relative frequencies gives the proportion of trees that are 12 inches to 16 inches, inclusive, in diameter?

(A)  $0.615 - 0.325$

(B)  $0.615 - 0.473$

(C)  $0.726 - 0.325$

(D)  $0.726 - 0.473$

(E)  $0.731 - 0.325$

---

19. An outlier may be defined as a data point that is more than 1.5 times the interquartile range below the lower quartile or is more than 1.5 times the interquartile range above the upper quartile. According to this definition, what is the diameter, in inches, of the smallest tree that is an outlier?

(A) 4

(B) 28

(C) 30

(D) 34

(E) 36

GO ON TO THE NEXT PAGE.

23. A marketing company wants to estimate the proportion of consumers in a certain region of the country who would react favorably to a new marketing campaign. Further, the company wants the estimate to have a margin of error of no more than 5 percent with 90 percent confidence. Of the following, which is closest to the minimum number of consumers needed to obtain the estimate with the desired precision?

- (A) 136
- (B) 271
- (C) 385
- (D) 542
- (E) 769

---

24. Traffic data revealed that 35 percent of automobiles traveling along a portion of an interstate highway were exceeding the legal speed limit. Using highway cameras and license plate registrations, it was also determined that 52 percent of sports cars were also speeding along the same portion of the highway. What is the probability that a randomly selected car along the same portion of the highway was a speeding sports car?

- (A) 0.870
- (B) 0.673
- (C) 0.182
- (D) 0.170
- (E) It cannot be determined from the information given.

**GO ON TO THE NEXT PAGE.**

25. In a large school district, 16 of 85 randomly selected high school seniors play a varsity sport. In the same district, 19 of 67 randomly selected high school juniors play a varsity sport. A 95 percent confidence interval for the difference between the proportion of high school seniors who play a varsity sport in the school district and high school juniors who play a varsity sport in the school district is to be calculated. What is the standard error of the difference?

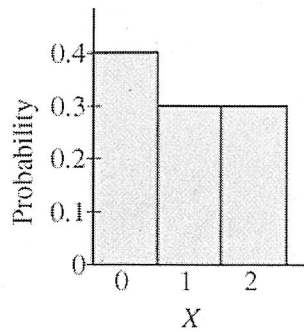
- (A) 0.0347
- (B) 0.0695
- (C) 0.1362
- (D) 0.9800
- (E) 1.6900

21. Ninety percent of the people who have a particular disease will have a positive result on a given diagnostic test. Ninety percent of the people who do not have the disease will have a negative result on this test. If 5 percent of a certain population has the disease, what percent of that population would test positive for the disease?

- (A) 4.5%
- (B) 5%
- (C) 10%
- (D) 14%
- (E) 90%

GO ON TO THE NEXT PAGE.

26. The number of points,  $X$ , scored in a game has the probability distribution below.



The number of points obtained in one game is independent of the number of points obtained in a second game. When the game is played twice, the sum of the number of points for both times could be 0, 1, 2, 3, or 4. If  $Y$  represents the sampling distribution of the sum of the scores when the game is played twice, for which value of  $Y$  will the probability be greatest?

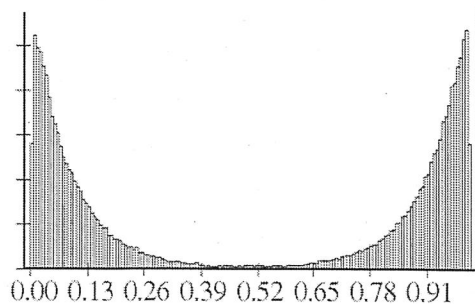
- (A) 0  
(B) 1  
(C) 2  
(D) 3  
(E) 4
29. A 90 percent confidence interval is to be created to estimate the proportion of television viewers in a certain area who favor moving the broadcast of the late weeknight news to an hour earlier than it is currently. Initially, the confidence interval will be created using a simple random sample of 9,000 viewers in the area. Assuming that the sample proportion does not change, what would be the relationship between the width of the original confidence interval and the width of a second 90 percent confidence interval that is created based on a sample of only 1,000 viewers in the area?
- (A) The second confidence interval would be 9 times as wide as the original confidence interval.  
(B) The second confidence interval would be 3 times as wide as the original confidence interval.  
(C) The width of the second confidence interval would be equal to the width of the original confidence interval.  
(D) The second confidence interval would be  $\frac{1}{3}$  as wide as the original confidence interval.  
(E) The second confidence interval would be  $\frac{1}{9}$  as wide as the original confidence interval.

GO ON TO THE NEXT PAGE.

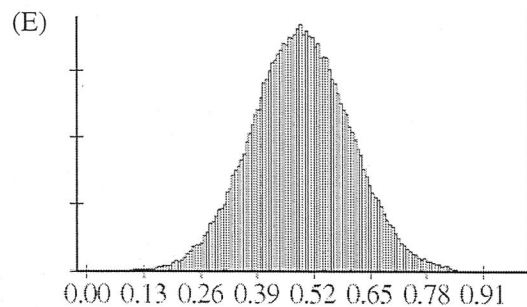
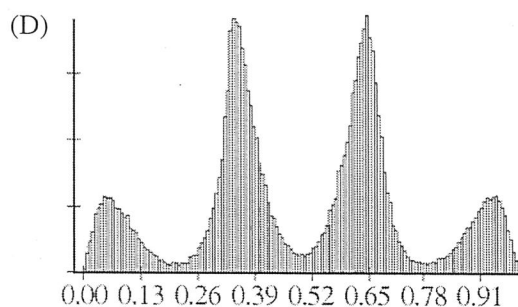
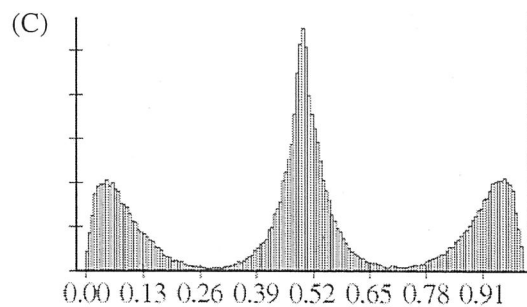
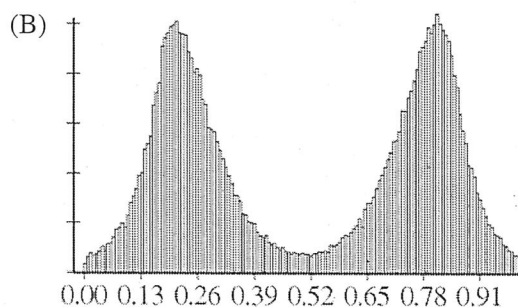
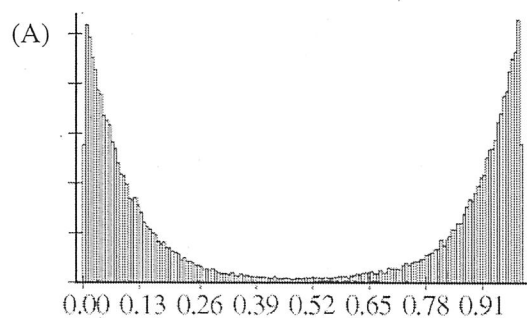
30. A new restaurant is interested in determining the best time-temperature combination for roasting a five-pound cut of lamb. The times to be tested are 45 minutes, 60 minutes, and 90 minutes at temperatures of 350 degrees Fahrenheit and 425 degrees Fahrenheit for each time, with the exception of the 90 minute–425 degree combination. That combination is being eliminated because it will overcook the lamb, which leaves five combinations remaining. From 10 identical cuts of lamb, 2 are randomly selected to roast using each of the time-temperature combinations in the same oven. The quality of the finished product is evaluated for each roast. Which of the following is true?
- (A) The explanatory variable is the quality of the finished lamb.
  - (B) The response variable is the roasting temperature for the lamb.
  - (C) If the experiment is repeated, identical results will be expected.
  - (D) There should be a control group (i.e., a group in which no treatment is given).
  - (E) The two cuts that are being roasted for each time-temperature combination are an example of replication.
31. A local real estate magazine used the median instead of the mean when it reported the SAT score of the average student who attends Groveland High School. A graphical display of SAT scores of students who attend Groveland High School indicated that the data were strongly skewed to the right. Which of the following explains why, in this situation, the median is a more accurate indicator of the SAT score of the average student than the mean is?
- (A) The mean is affected by the skewness, whereas the median is not.
  - (B) The median is always the preferred statistic.
  - (C) The mean will be less than the median when the data are strongly skewed to the right.
  - (D) The mean should be used only when data are strongly skewed to the left.
  - (E) The median is equal to one-half the sum of the maximum and minimum SAT scores at Groveland High.

32. The director of a fitness center wants to examine the effects of two exercise classes (spinning and aerobics) on body fat percentage. A six-week spinning class and a six-week aerobics class are offered at the same time and on the same days, so that a person can enroll in only one of them. A new class of each is about to begin, and each class has 25 people in it. Ten people are randomly selected from each class. Each person's body fat percentage is measured at the beginning and again at the end of the six-week class. Using the change in body fat percentage as the response variable and conducting a test at the  $\alpha = 0.01$  level, the director determines that there is a significant difference between the treatment means. Which of the following is a confounding variable in the study?
- (A) The director's choice of spinning and aerobics classes as the types for use in the study
  - (B) The random sample of 10 people from each class
  - (C) The participants' choice of which class to take
  - (D) The use of body fat percentage as the measure of effectiveness of the treatment
  - (E) The fact that both classes were conducted three times each week at the same time of day and for the same amount of time each day
38. A dog food company wishes to test a new high-protein formula for puppy food to determine whether it promotes faster weight gain than the existing formula for that puppy food. Puppies participating in an experiment will be weighed at weaning (when they begin to eat puppy food) and will be weighed at one-month intervals for one year. In designing this experiment, the investigators wish to reduce the variability due to natural differences in puppy growth rates. Which of the following strategies is most appropriate for accomplishing this?
- (A) Block on dog breed and randomly assign puppies to existing and new formula groups within each breed.
  - (B) Block on geographic location and randomly assign puppies to existing and new formula groups within each geographic area.
  - (C) Stratify on dog breed and randomly sample puppies within each breed. Then assign puppies by breed to either the existing or the new formula.
  - (D) Stratify on geographic location of the puppies and randomly sample puppies within each geographic area. Then assign puppies by geographic area to either the existing or the new formula.
  - (E) Stratify on gender and randomly sample puppies within gender groups. Then assign puppies by gender to either the existing or the new formula.

40. The histogram below represents data obtained after the census of an entire population was conducted.



The sampling distribution of the sample mean based on samples of size 2 for the population was simulated, and a histogram of the results was produced. Which of the following histograms is most likely the histogram of that sampling distribution?



GO ON TO THE NEXT PAGE.

