

Part A: Multiple Choice Select the best answer for each question.

_____ 1. If a distribution is skewed to the right, which of the following is true?

- (a) The mean must be less than the median.
- (b) The mean and median must be equal.
- (c) The mean must be greater than the median.
- (d) The mean is either equal to or less than the median.
- (e) It's impossible to tell which of the above statements is true without seeing the data.

_____ 2. Given the following stemplot of data.

```

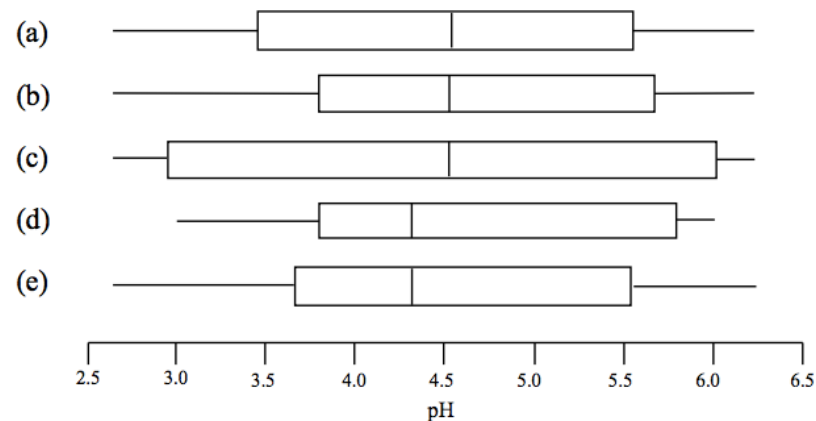
2 | 679
3 | 237789
4 | 1222446899
5 | 0556788
6 | 0233

```

Key: 3|7 = pH 3.7

The data ranged from pH 2.6 to pH 6.3.

Which of the following boxplots is a correct representation of the same distribution?



_____ 3. The area under the standard Normal curve corresponding to $-0.3 < Z < 1.6$ is

- (a) 0.3273
- (b) 0.4713
- (c) 0.5631
- (d) 0.9542
- (e) none of the above

_____ 4. Which of the following statements about standard deviation is true?

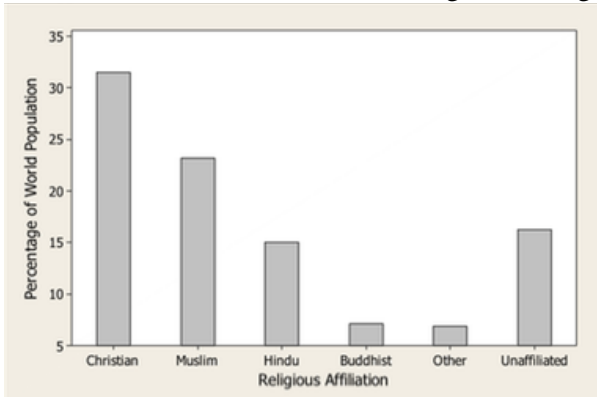
- (a) Standard deviation measures the typical distance data points in a distribution are from the median.
- (b) Regardless of what units the data are in, standard deviation has no units.
- (c) Standard deviation is always a non-negative number.
- (d) If the data is measured in hours, standard deviation would be measured in hours squared.
- (e) Standard deviation is resistant to outliers.

_____ 5. Which statement about influential points is true?

- I. Removal of an influential point changes the regression line.
- II. Data points that are outliers in the horizontal direction are more likely to be influential than points that are outliers in the vertical direction.
- III. Influential points have large residuals.

- (a) I only
- (b) I and II
- (c) I and III
- (d) II and III
- (e) I, II, and III

_____ 6. The bar graph below shows the distribution of religious affiliation around the world. Which of the following statement best describes what is wrong with this graph?



- (a) Since this is a bar graph, the bars should be touching each other.
- (b) The “other” category should be subdivided into all the religions included.
- (c) The bars should be in decreasing order of size.
- (d) Starting the vertical scale at 5 distorts the relative frequency of the religions.
- (e) The vertical scale should be in numbers of people, not percentages.

_____ 7. One hundred people work at Lagtime Internet Services. Suppose after a particularly successful year, the president of the company decided to double his own salary and not change anyone else’s. Assuming the president makes more than anyone else, which of the following statements about changes in measures of center and spread is true?

- (a) The mean and standard deviation will increase, but the median and interquartile range will stay the same.
- (b) The mean and interquartile range will increase, but the median and standard deviation will stay the same.
- (c) The mean, standard deviation, and interquartile range will increase, but the median will stay the same.
- (d) The mean will increase, but the median, standard deviation, and interquartile range will stay the same.
- (e) The mean, median, standard deviation, and interquartile range will all increase.

_____ 8. The weights of adult male Labrador Retrievers are approximately Normally distributed with a mean of 87 pounds and a standard deviation of 8 pounds. Which of the following statements is true?

- (a) About 95% of adult male Labrador Retrievers weigh between 79 and 96 pounds.
- (b) The median weight of adult male Labrador Retrievers is more than 87 pounds.
- (c) The proportion of adult male Labrador Retrievers that weigh less than 80 pounds is approximately equal to the proportion that weigh more 94 pounds.
- (d) Almost all adult male Labrador retrievers weight more than 75 pounds.
- (e) The standardized weights of 80-pound and 94-pound adult male Labrador Retrievers are roughly equal.

_____ 9. The customer service department of an online store keeps track of the length of time customers wait on hold for a representative. The mean wait time is 5 minutes and the standard deviation is 2.3 minutes. Suppose the company wants to change the variable from minutes to seconds exceeding a target time of 120 seconds. That is, if X is wait time in minutes, the new variable is $Y = 60X - 120$. Which of the following are the mean and standard deviation of Y ?

- (a) $\mu_Y = 180$; $\sigma_Y = 2.3$
- (b) $\mu_Y = 180$; $\sigma_Y = 18$
- (c) $\mu_Y = 180$; $\sigma_Y = 138$
- (d) $\mu_Y = 300$; $\sigma_Y = 2.3$
- (e) $\mu_Y = 300$; $\sigma_Y = 18$

_____ 10. A forester studying oak trees finds that the correlation between $x =$ the ages (measured in years) and $y =$ height (in feet) of a sample of trees is 0.78. Which of the following statements must be true?

- (a) 78% of the variability in tree heights can be explained by variation on the trees' ages.
- (b) For every year a tree ages, it's height increases, on average, by 78%.
- (c) If we let $x =$ height of tree and $y =$ age of tree, then the correlation would be the reciprocal of 0.78.
- (d) If we measure the height in meters instead of feet, the correlation would still be 0.78.
- (e) The unit for correlation in this context is foot-years.

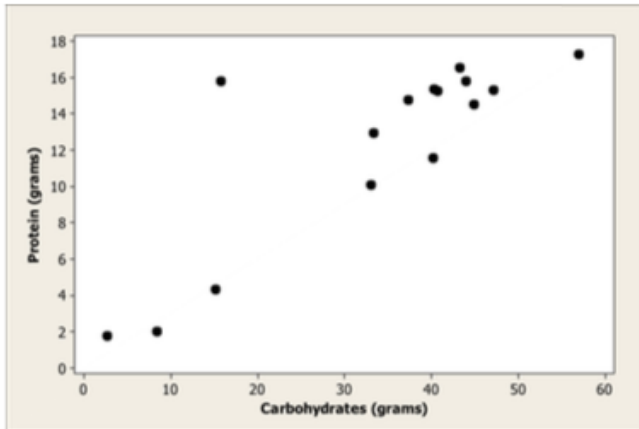
_____ 11. The protein content for the 15 bean varieties has a mean of 12.2 grams and a standard deviation of 5.3 grams. The mean carbohydrate content is 33.6 grams with a standard deviation of 15.7 grams. The correlation is 0.84. Which of the following expressions represents the slope of the least squares regression of $y =$ protein content on $x =$ carbohydrate content?

- (a) $\frac{(.84)(5.3)}{(15.7)}$
- (b) $\frac{(.84)(15.7)}{(5.3)}$
- (c) $\frac{(.84)(12.2)}{(33.6)}$
- (d) $\frac{(.84)(33.6)}{(12.2)}$
- (e) $\frac{(33.6)}{(.84)(12.2)}$

_____ 12. An airline has 10 daily flights from Philadelphia to Denver. To assess customer satisfaction, a random sample of 15 passengers from each flight on a single day are asked to fill out a survey about their experience on the flight. What type of sample is this?

- (a) Simple random sample
- (b) Stratified random sample
- (c) Multistage sample
- (d) Cluster sample
- (e) Convenience sample

_____ 13. The scatterplot below shows the relationship between carbohydrates and protein in one-cup servings of 15 different varieties of beans.



The unusual point in the upper left part of the plot is for navy beans, with 15.8 grams of protein and 15.8 grams of carbohydrates. Which of the following best describes how correlation would change if we removed navy beans from the data set?

- (a) The correlation would be closer to 1, because the remaining data would have a stronger positive relationship.
- (b) The correlation would be closer to 1, because there would be fewer individuals in the data set.
- (c) The correlation would be closer to 0, because the data would more closely resemble a straight line.
- (d) The correlation would be closer to 0, because the standard deviation of the residuals would be smaller.
- (e) Correlation could no longer be calculated, because the remaining data would fall into two distinct groups.

_____ 14. A track coach wants to test the effectiveness of a new training program for distance runners. He selects his two fastest runners and, using a coin flip, randomly chooses one to participate in the new program. The other will continue with the standard training regimen. He then picks the next two fastest runners and randomly assigns one to each program. He repeats this process with all his runners, choosing the two fastest remaining each time. He then compares race times within each group of two runners. What is the name for this kind of experimental design?

- (a) Completely randomized design.
- (b) Randomized block design, but not matched pairs.
- (c) Stratified random design.
- (d) Non-random pairs design.
- (e) Matched pairs design.

_____ 15. If you flip three coins, the probability of getting three heads is 0.125. Which of the following statements follows from this?

- (a) If you flip three coins 1000 times, you will get three heads exactly 125 times.
- (b) If you get three heads in half of the first 100 sets of flips, the probability that the next set of flips will produce three heads is slightly lower.
- (c) If you flip three coins 20 times, then at least one set of flips will be three heads.
- (d) If you flip three coins 5000 times, the percentage of time you get three heads will be very close 12.5%.
- (e) If you get three heads four times in a row, the probability of getting three tails four times in a row increases.

Use the following for questions 16 – 18:

A group of 125 pick-up truck owners were asked what brand truck they owned and whether it had four-wheel drive. The results are given in the two-way table below.

		Four wheel drive?		Totals
		Yes	No	
Truck Brand	Ford	28	17	45
	Chevy	32	18	50
	Dodge	20	10	30
	Totals	80	45	125

_____ 16. You randomly select one truck owner. What is the probability that he owns a Dodge or has four wheel drive?

- (a) 20/80
- (b) 20/125
- (c) 80/125
- (d) 90/125
- (e) 110/125

_____ 17. You randomly select one truck owner. What is the probability that he owns a Chevy, given that he has four-wheel drive?

- (a) 32/50
- (b) 32/80
- (c) 32/125
- (d) 50/125
- (e) 80/125

_____ 18. You randomly select one truck owner. Which one of the following is true about the events “Owner has a Chevy” and “Owner’s truck has four-wheel drive”?

- (a) These two events are mutually exclusive and independent.
- (b) These two events are mutually exclusive, but not independent.
- (c) These two events are not mutually exclusive, but they are independent.
- (d) These two events are neither mutually exclusive nor independent.
- (e) These two events are mutually exclusive, but we do not have enough information to determine if they are independent.

_____ 19. An inspector examines a large shipment (several thousand) of laser pointers by randomly selecting a pointer, one at a time, until he finds one that is defective. Suppose that 1% of the laser points are defective. Which one of the following expressions describes the probability that the first defective pointer he finds is the sixth one that he selects?

- (a) $(0.01)^5 (0.99)$
- (b) $(0.01)^6 (0.99)$
- (c) $(0.01)^6$
- (d) $(0.99)^6 (0.01)$
- (e) $(0.99)^5 (0.01)$

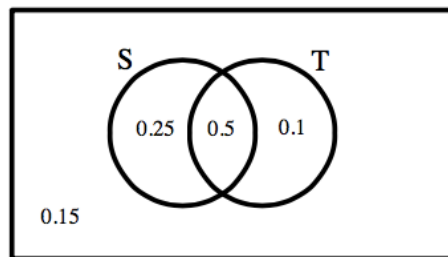
_____ 20. John maintains two separate lines of lobster traps in Penobscot Bay. The East Bay traps produce a mean of 12 pounds of lobsters per day with a standard deviation of 7 pounds. The West Bay traps produce a mean of 10 pounds of lobsters a day with a standard deviation of 4.5 pounds. Which of the following are the mean and approximate standard deviation of the total weight of lobsters John traps in a day?

- (a) Mean = 11; Standard deviation = 11.50
- (b) Mean = 11; Standard deviation = 8.32
- (c) Mean = 22; Standard deviation = 5.75
- (d) Mean = 22; Standard deviation = 8.32
- (e) Mean = 22; Standard deviation = 11.50

_____ 21. Capital Car Wash sells a membership card for \$25 per year that entitles the buyer to a \$2 discount on each car wash purchased during the year. The number of times Danica washes her car has a mean of 15 with standard deviation of 4. What are the mean and standard deviation of the net gain Danica receives if she buys a membership card?

- (a) Mean = 5; Standard deviation = 4
- (b) Mean = 5; Standard deviation = 8
- (c) Mean = 30; Standard deviation = 4
- (d) Mean = 30; Standard deviation = 8
- (e) There is not enough information to determine both values.

_____ 22. Select a random student at Lakeville High School. Let Event S = The student owns a smartphone, and Event T = The student sends an average of more than 50 texts per day. The Venn diagram below summarizes the probabilities associated with these two events.

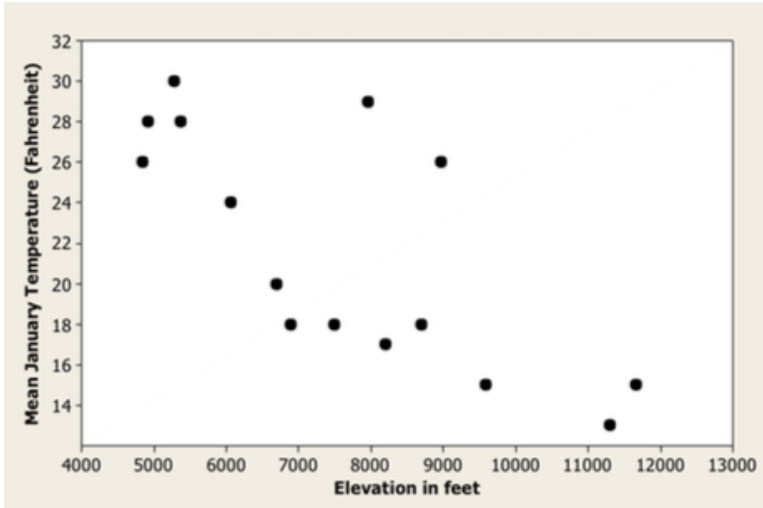


Which of the following is the probability that a randomly selected student does not send an average of more than 50 texts per day, given that they own a smartphone?

- (a) 0.4
- (b) 0.333
- (c) 0.294
- (d) 0.2
- (e) 0.133

Part B: Please answer all parts and show all work for full credit.

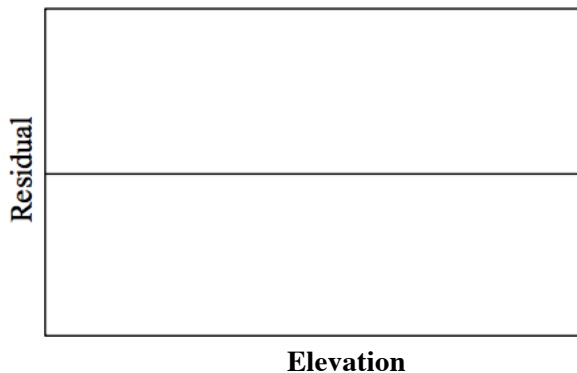
23. The scatterplot below gives the mean January temperature (in degrees Fahrenheit) and elevation for 15 towns along Interstate 70 in Minnesota.



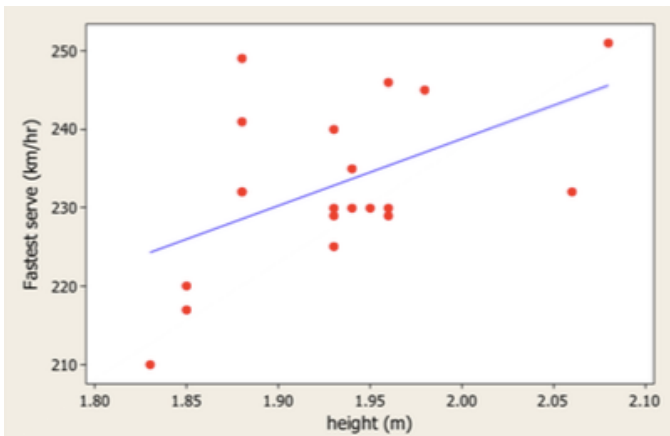
(a) Based on this scatterplot, describe the relationship between elevation and mean January temperature in this region.

(b) One of the points in the plot represents the town of St. Cloud, which has an elevation of 6069 ft. and a mean January temperature of 24 F°. Find the percentile for St. Cloud's elevation among the towns in this data set.

(c) Suppose a least-squares linear regression line were calculated for these data. In the box below, make rough sketch of the residual plot associated with this regression line.



24. Can we predict how fast a tennis player can hit a serve from the player's height? The following computer output and scatterplot describe the regression of fastest serve speed (in kilometers per hour) on the height (in meters) of 21 current male professional tennis players.



Predictor	Coef	SE Coef	T	P
Constant	68.81	60.77	1.13	0.272
height (m)	84.98	31.51	2.70	0.014

S = 8.93380 R-Sq = 27.7% R-Sq(adj) = 23.9%

(a) Write the equation of the least-squares regression line.

(b) Interpret the value “R-sq = 27.7%.”

(c) Predict the fastest serve of a professional tennis player whose is 1.7 meters tall. Comment on the reliability of this prediction.

25. David Palmer, a psychology professor at the University of Virginia, studies the effect of “personal narratives” on people’s behavior. For example, if the “story” you have about your academic ability is positive, you do better in school than if your “story” is negative. Something as simple as hearing older students describe overcoming challenges similar to yours can help you change your story and improve your performance.

(a) Suppose you have 40 college freshman who have volunteered to be part of a study. Design a completely randomized experiment that tests the hypothesis that hearing older students talk about overcoming challenges improves academic performance. Be sure your design addresses how randomization will be incorporated.

(b) Suppose you have reason to believe that students with a history of strong academic performance respond differently than those with a history of modest performance. Describe how you would incorporate blocking to address this difference in response.

(c) Can this experimental design be carried out in a double blind manner? Explain.

26. On the Centerville High School debate team, 20% of the members are Sophomores, 35% are Juniors and 45% are Seniors. A team member is selected randomly to give the closing argument for the team. If a sophomore gives the closing argument, the team has a probability of 0.25 of winning the debate. If a junior gives the closing argument, the probability is 0.6. If a senior closes, the probability rises to 0.85.

(a) Find the probability that the teams wins the debate.

(b) Given that the team loses the debate, what is the probability that it was a junior who gave the closing argument?

27. Andrea and Ashley are middle-distance runners for their school's track team. Andrea's times in the 400-meter race are approximately Normally distributed with a mean of 62 seconds and a standard deviation of 0.8 seconds. Ashley's times are approximately Normally distributed with a mean of 62.8 seconds and a standard deviation of 1 second.

(a) The school record for the 400-meter is 60.0 seconds. What is the probability that Andrea breaks the record in her next race?

(b) Assuming that Andrea and Ashley's times in any race are independent, what is the probability that Ashley beats Andrea in their next race?

28. A company that sells bicycles online maintains a telephone help line to assist customers who are assembling bicycles after they have been delivered. To determine how many “helpers” are needed, they keep detailed records of the percentage of bicycle purchasers who call in for help. They have determined that 18% of all buyers call the help line.

(a) Suppose we select a random sample of 25 buyers. What are the mean and standard deviation of the count of buyers among the 25 who call in for help?

(b) What is the probability that exactly 8 of these 25 buyers call in for help?

(c) Suppose we select a random sample of 200 buyers and calculate the proportion of buyers in the sample who call in for help. Describe the sampling distribution for this sample proportion.

(d) What is the probability that more than 20% of the buyers in this sample of 200 call in for help?

29. The student council at Greenwood High school (a large urban high school) wants to know what proportion of the student body favors banning plastic water bottles from the school buildings and grounds. A simple random sample of 152 students finds that 85 support banning plastic bottles.

(a) Construct and interpret a 95% confidence interval for the proportion of students at Greenwood High School who support banning plastic water bottles.

(b) The student council will ban the bottles if they are convinced that the majority of students favor it. Does this confidence interval provide evidence for a ban? Explain.

*Extra credit question: (c) It turns out that the council's simple random sample was originally 165 students, but 13 individuals in the sample didn't respond because they were on an AP Biology field trip. Could this change your answer to part (b)? Explain your reasoning.

30. A 2011 Pew Research poll found that 18% of cell phone owners use their phone for conducting transactions with their bank. Let's assume that 0.18 is the true proportion (in 2011) of cell phone owners who use their cell phone for banking.

(a) If you selected a random sample of 150 people in 2011, what is the probability that more than 20% of them use their cell phone for banking?

(b) You take a random sample of 150 current cell phone owners and find that 38 of them use their cell phone for banking. Is this evidence that since 2011 there has been an increase from 0.18 in the proportion of cell phone owners who use their cell phone for banking? Justify your conclusion with a complete significance test.