1. In the scatterplot of *y* versus *x* shown above, the least squares regression line is superimposed on the plot. Which of the following points has the largest residual?

    (A) *A*
    (B) *B*
    (C) *C*
    (D) *D*
    (E) *E*

2. Under which of the following conditions is it preferable to use stratified random sampling rather than simple random sampling?

    (A) The population can be divided into a large number of strata so that each stratum contains only a few individuals.
    (B) The population can be divided into a small number of strata so that each stratum contains a large number of individuals.
    (C) The population can be divided into strata so that the individuals in each stratum are as much alike as possible.
    (D) The population can be divided into strata so that the individuals in each stratum are as different as possible.
    (E) The population can be divided into strata of equal sizes so that each individual in the population still has the same chance of being selected.

3. All bags entering a research facility are screened. Ninety-seven percent of the bags that contain forbidden material trigger an alarm. Fifteen percent of the bags that do not contain forbidden material also trigger the alarm. If 1 out of every 1,000 bags entering the building contains forbidden material, what is the probability that a bag that triggers the alarm will actually contain forbidden material?

   (A) 0.00097
   (B) 0.00640
   (C) 0.03000
   (D) 0.14550
   (E) 0.97000

4. A candy company claims that 10 percent of its candies are blue. A random sample of 200 of these candies is taken, and 16 are found to be blue. Which of the following tests would be most appropriate for establishing whether the candy company needs to change its claim?

   (A) Matched pairs t-test
   (B) One-sample proportion z-test
   (C) Two-sample t-test
   (D) Two-sample proportion z-test
   (E) Chi-square test of association

DESCRIPTIVE STATISTICS

Variable score	N	Mean	Median	TrMean	StDev	SE Mean
	50	1045.7	1024.7	1041.9	221.9	31.4

Variable score	Minimum	Maximum	Q1	Q3	
	628.9	1577.1	877.7	1219.5	

5. Some descriptive statistics for a set of test scores are shown above. For this test, a certain student has a standardized score of $z = -1.2$. What score did this student receive on the test?

   (A) 266.28
   (B) 779.42
   (C) 1008.02
   (D) 1083.38
   (E) 1311.98

6. In a test of H_0: $\mu = 8$ *versus* H_a: $\mu \neq 8$, a sample of size 220 leads to a p-value of 0.034. Which of the following must be true?

   (A) A 95% confidence interval for μ calculated from these data will not include $\mu = 8$.
   (B) At the 5% level if H_0 is rejected, the probability of a Type II error is 0.034.
   (C) The 95% confidence interval for μ calculated from these data will be centered at $\mu = 8$.
   (D) The null hypothesis should not be rejected at the 5% level.
   (E) The sample size is insufficient to draw a conclusion with 95% confidence.

7. A summer resort rents rowboats to customers but does not allow more than four people to a boat. Each boat is designed to hold no more than 800 pounds. Suppose the distribution of adult males who rent boats, including their clothes and gear, is normal with a mean of 190 pounds and standard deviation of 10 pounds. If the weights of individual passengers are independent, what is the probability that a group of four adult male passengers will exceed the acceptable weight limit of 800 pounds?

   (A) 0.023
   (B) 0.046
   (C) 0.159
   (D) 0.317
   (E) 0.977

8. Consider a data set of positive values, at least two of which are not equal. Which of the following sample statistics will be changed when each value in this data set is multiplied by a constant whose absolute value is greater than 1?

   I. The mean
   II. The median
   III. The standard deviation

   (A) I only
   (B) II only
   (C) III only
   (D) I and II only
   (E) I, II, and III

9. Each person in a simple random sample of 2,000 received a survey, and 317 people returned their survey. How could nonresponse cause the results of the survey to be biased?

   (A) Those who did not respond reduced the sample size, and small samples have more bias than large samples.
   (B) Those who did not respond caused a violation of the assumption of independence.
   (C) Those who did not respond were indistinguishable from those who did not receive the survey.
   (D) Those who did not respond represent a stratum, changing the simple random sample into a stratified random sample.
   (E) Those who did respond may differ in some important way from those who did not respond.

10. In a certain game, a fair die is rolled and a player gains 20 points if the die shows a "6." If the die does not show a "6," the player loses 3 points. If the die were to be rolled 100 times, what would be the expected total gain or loss for the player?

    (A) A gain of about 1,700 points
    (B) A gain of about 583 points
    (C) A gain of about 83 points
    (D) A loss of about 250 points
    (E) A loss of about 300 points

11. The Attila Barbell Company makes bars for weight lifting. The weights of the bars are independent and are normally distributed with a mean of 720 ounces (45 pounds) and a standard deviation of 4 ounces. The bars are shipped 10 in a box to the retailers. The weights of the empty boxes are normally distributed with a mean of 320 ounces and a standard deviation of 8 ounces. The weights of the boxes filled with 10 bars are expected to be normally distributed with a mean of 7,520 ounces and a standard deviation of

    (A) $\sqrt{12}$ ounces
    (B) $\sqrt{80}$ ounces
    (C) $\sqrt{224}$ ounces
    (D) 48 ounces
    (E) $\sqrt{1,664}$ ounces

12. Exercise physiologists are investigating the relationship between lean body mass (in kilograms) and the resting metabolic rate (in calories per day) in sedentary males.

Predictor	Coef	StDev	T	P
Constant	264.0	276.9	0.95	0.363
Mass	22.563	6.360	3.55	0.005

    S = 144.9   R-Sq = 55.7%   R-Sq(adj) = 51.3%

Based on the computer output above, which of the following is the best interpretation of the value of the slope of the regression line?

    (A) For each additional kilogram of lean body mass, the resting metabolic rate increases on average by 22.563 calories per day.
    (B) For each additional kilogram of lean body mass, the resting metabolic rate increases on average by 264.0 calories per day.
    (C) For each additional kilogram of lean body mass, the resting metabolic rate increases on average by 144.9 calories per day.
    (D) For each additional calorie per day for the resting metabolic rate, the lean body mass increases on average by 22.563 kilograms.
    (E) For each additional calorie per day for the resting metabolic rate, the lean body mass increases on average by 264.0 kilograms.

13. An investigator was studying a territorial species of Central American termites, *Nasutitermes corniger*. Forty-nine termite pairs were randomly selected; both members of each of these pairs were from the same colony. Fifty-five additional termite pairs were randomly selected; the two members in each of these pairs were from different colonies. The pairs were placed in petri dishes and observed to see whether they exhibited aggressive behavior. The results are shown in the table below.

	Aggressive	Nonaggressive	Total
Same colony	40 (33.5)	9 (15.5)	49
Different colonies	31 (37.5)	24 (17.5)	55
Total	71	33	104

A Chi-square test for homogeneity was conducted, resulting in $\chi^2 = 7.638$. The expected counts are shown in parentheses in the table. Which of the following sets of statements follows from these results?

(A) χ^2 is not significant at the 0.05 level.

(B) χ^2 is significant, $0.01 < p < 0.05$; the counts in the table suggest that termite pairs from the same colony are less likely to be aggressive than termite pairs from different colonies.

(C) χ^2 is significant, $0.01 < p < 0.05$; the counts in the table suggest that termite pairs from different colonies are less likely to be aggressive than termite pairs from the same colony.

(D) χ^2 is significant, $p < 0.01$; the counts in the table suggest that termite pairs from the same colony are less likely to be aggressive than termite pairs from different colonies.

(E) χ^2 is significant, $p < 0.01$; the counts in the table suggest that termite pairs from different colonies are less likely to be aggressive than termite pairs from the same colony.

(handwritten annotations:)
A) P-value = .076
B) P-value = .51
C) P-value = .66
D) P-value = 104
E) P-value = .006

14. Consider n pairs of numbers (x_1, y_1), (x_2, y_2), . . . , and (x_n, y_n). The mean and standard deviation of the x-values are $\bar{x} = 5$ and $s_x = 4$, respectively. The mean and standard deviation of the y-values are $\bar{y} = 10$ and $s_y = 10$, respectively. Of the following, which could be the least squares regression line?

(A) $\hat{y} = -5.0 + 3.0x$

(B) $\hat{y} = 3.0x$

(C) $\hat{y} = 5.0 + 2.5x$

(D) $\hat{y} = 8.5 + 0.3x$

(E) $\hat{y} = 10.0 - 0.4x$

15] Suppose test scores are normally distributed and 90% of people who take the test get at most a 92 and 5% get at least a 95. What are the mean and S.D. of the distribution?
(A) 81.4 and 8.3 (B) 86.1 and 7.6 (C) 80.3 and 8.5

16. George and Michelle each claimed to have the better recipe for chocolate chip cookies. They decided to conduct a study to determine whose cookies were really better. They each baked a batch of cookies using their own recipe. George asked a random sample of his friends to taste his cookies and to complete a questionnaire on their quality. Michelle asked a random sample of her friends to complete the same questionnaire for her cookies. They then compared the results. Which of the following statements about this study is false?

(A) Because George and Michelle have a different population of friends, their sampling procedure makes it difficult to compare the recipes.
(B) Because George and Michelle each used only their own respective recipes, their cooking ability is confounded with the recipe quality.
(c) Because George and Michelle each used only the ovens in their houses, the recipe quality is confounded with the characteristics of the oven.
(D) Because George and Michelle used the same questionnaire, their results will generalize to the combined population of their friends.
(E) Because George and Michelle each baked one batch, there is no replication of the cookie recipes.

17. A large company is considering opening a franchise in St. Louis and wants to estimate the mean household income for the area using a simple random sample of households. Based on information from a pilot study, the company assumes that the standard deviation of household incomes is $\sigma = \$7,200$. Of the following, which is the least number of households that should be surveyed to obtain an estimate that is within $200 of the true mean household income with 95 percent confidence?

(A)   75
(B) 1,300
(c) 5,200
(D) 5,500
(E) 7,700