AP Statistics Practice Examination 4

Multiple Choice
Statistics
Section I

Time: 90 minutes

1. Which of the following statements is (are) correct?

   I. Correlation makes no distinction between explanatory and response variables.
   II. The sign of r reflects the strength of the association.
   III. r measures the strength of a *linear* relationship only.

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

2. In a random sample of 500 women, 120 are college graduates. With what confidence can we assert that between 22% and 26% of women are college graduates?
   (A) 2%
   (B) 4%
   (C) 14.75%
   (D) 24%
   (E) 70.49%

3. The median of a distribution is 150 and the interquartile range is 50. Identify the statement(s) that *must* be true.

I. 50% of the data are between 125 and 175.
II. 50% of the data are less than or equal to 150.
III. 75% of the data are greater than 125.

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

4. Twenty-five men were polled, and their annual incomes were recorded. The cumulative frequency histogram below shows the results.
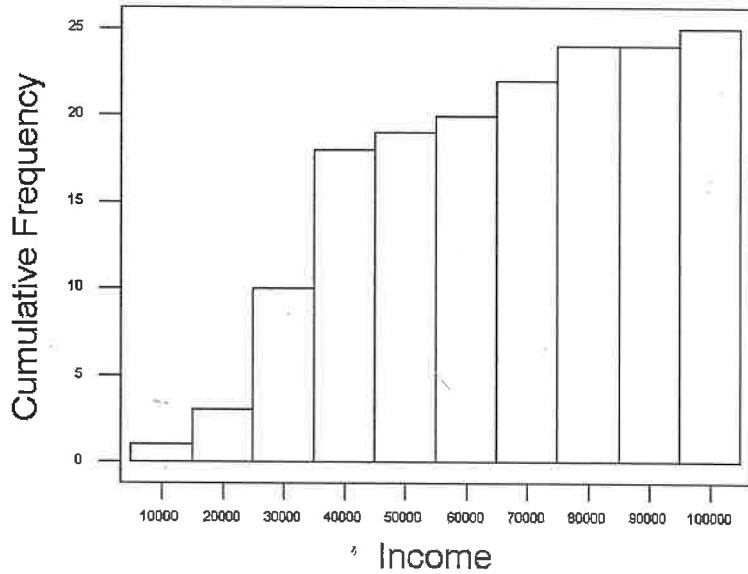
Which of the following statements can be made upon examination of the histogram?

(A) The same number of men had incomes between $75,000 and $85,000 as between $85,000 and $95,000.
(B) The median income was $55,000.
(C) The mean income was less than the median income.
(D) The modal income was between $35,000 and $45,000.
(E) No men had incomes between $75,000 and $85,000.

5. Only 6 out of every 100 people have blood type O^+. What is the probability that in a random sample of individuals, the first person with type O^+ blood will be the eighth person tested?

   (A) $C(8, 8)(0.06)^8(0.94)^0$

   (B) $C(8, 1)(0.06)^1(0.94)^7$

   (C) $C(1, 8)(0.06)^1(0.94)^7$

   (D) $(0.06)^1(0.94)^7$

   (E) $(0.06)^7(0.94)^1$

6. Mr. DeVeaux teaches two sections of AP* Physics. He has 38 seniors in one section and 24 juniors in the other section. The overall mean for both sections on the midterm exam was 87. If the junior section had a mean of 92, what was the approximate mean for the senior section on the midterm exam?

   (A) 82.6

   (B) 83.8

   (C) 89.5

   (D) 87.0

   (E) 90.4

7. Which of the following is/are acceptable ways to express your decision at the end of a hypothesis test?

   I. Fail to reject H_0; accept H_a
   II. Accept H_0; reject H_a
   III. Reject H_0; accept H_a

   (A) I only

   (B) II only

   (C) III only

   (D) All are acceptable.

   (E) None are acceptable.

8. The following table gives the percentage of nursing home residents by age.

Age	under 65	65–74	75–84	85 and older
Percentage	9.8	12	31.8	46.5

In a random sample of 150 families with a relative in a nursing home, the following was the distribution of the relatives' ages: 10 under 65, 20 ages 65–74, 52 ages 75–84, and 68 ages 85 and older. If a goodness-of-fit test were performed, what would be the value of the χ^2 statistic?

(A) 0.541

(B) 0.707

(C) 2.156

(D) 2.810

(E) 28.110

9. If $P(A) = 0.7$, $P(not\ B) = 0.4$, and $P(A\ and\ B) = 0.5$, find $P(A\ or\ B)$.

(A) 1.3

(B) 0.6

(C) 0.8

(D) 0.1

(E) 1.1

10. If you wanted to find the average GPA for seniors at your school who have been accepted into college, what would be the most appropriate technique to use to gather the data?

(A) Census

(B) Simple random sample

(C) Stratified random sample

(D) Systematic random sample

(E) Controlled experiment

11. Given that a population has a standard deviation of 0, which of the following statements *must* be true?

    I. The standard deviation of a sample drawn from the population is also 0.

    II. The sample mean of a sample drawn from the population is also 0.

    III. The sample mean and median are equal for a sample drawn from the population.

    (A) I only

    (B) III only

    (C) I and II only

    (D) I and III only

    (E) I, II, and III

12. A meteorologist wants to simulate 20 days of weather in a region where the probability of precipitation during the season being simulated is 40%. She uses a random-number table as shown.

23091   05892   21007   43902   62973   29940   69630   91312

If R represents a rainy day and N represents a day without rain, which of the following could be the meteorologist's simulation result?

    (A) *NRNRR*   *RNRNR*   *NRNNR*   *NRRNN*   10 of the 20 days will be rainy.

    (B) *RRRNN*   *RRNRR*   *NNRNR*   *NNRRR*   12 of the 20 days will be rainy.

    (C) *RRRNR*   *RNNNR*   *RRRRN*   *NRNRR*   13 of the 20 days will be rainy.

    (D) *RRRNR*   *RNNNR*   *RRRRN*   *RRNRR*   14 of the 20 days will be rainy.

    (E) There is an insufficient number of random digits to conduct this simulation.

13. The American judicial system is based on the assumption that a person is innocent until proved guilty. A defendant is accused of a crime. What is the consequence of a Type II error?

    (A) The jury finds the defendant innocent; he is innocent.

    (B) The jury finds the defendant innocent; he is guilty.

    (C) The jury finds the defendant guilty; he is innocent.

    (D) The jury finds the defendant guilty; he is guilty.

    (E) The jury declares a mistrial because an error has been made.

14. A random sample of 27 individuals is selected, and the age and income of each individual is recorded. Regression analysis is performed, with the following results.

```
Dependent variable is:     Income
No Selector
R squared = 2.0%      R squared (adjusted) = -1.9%
s = 2.734e4  with  27 - 2 = 25  degrees of freedom

Source        Sum of Squares    df    Mean Square    F-ratio
Regression    3.90728e8          1    3.90728e8      0.523
Residual      1.86828e10        25    7.4731e8

Variable    Coefficient    s.e. of Coeff    t-ratio    prob
Constant    27300.4         1.576e4         1.73       0.0956
Age           244.203        337.7          0.723      0.4763
```

Is there a significant relationship between age and income?

(A)   A significant relationship exists between age and income at the $\alpha = 0.10$ level but not at the $\alpha = 0.05$ level.

(B)   A significant relationship exists between age and income at the $\alpha = 0.05$ level but not at the $\alpha = 0.01$ level.

(C)   A significant relationship exists between age and income at the $\alpha = 0.01$ level but not at the $\alpha = 0.001$ level.

(D)   A significant relationship exists between age and income at the $\alpha = 0.001$ level.

(E)   A significant relationship between age and income does not exist at any of the commonly accepted levels.

15. Suppose that 25% of horses live over 23.4 years, while 85% live less than 25.2 years. Assuming the ages of horses are normally distributed, what are the mean and standard deviation for the life expectancy of horses?

(A)   mean 20.14; standard deviation 4.86

(B)   mean 22.690; standard deviation 1.052

(C)   mean 22.690; standard deviation 4.97

(D)   mean 24.110; standard deviation 1.052

(E)   Not enough information is given to find the mean and standard deviation.

16. Statistics show that 7.3% of workers between the ages of 16 and 24 earn the minimum wage or less. What is the probability that if three young adults between the ages of 16 and 24 are polled, two or more will earn the minimum wage or less?

(A)   0.0004

(B)   0.0148

(C)   0.0152

(D)   0.0627

(E)   0.0677

17. To work in the word-processing department at Dewey, Cheatem, and Howe, a large center-city law firm, you must be able to type at least 80 words per minute. The director of the human resources department is revising the job description for word processors. She believes that it is possible to adjust the typing speed upward and still have a sufficient number of qualified candidates. She takes a random sample of 15 employees from the word-processing department and gives them a typing test. The mean typing speed is 93 words per minute with a standard deviation of 7 words per minute. Assume that typing speeds follow an approximately normal distribution. A 98% confidence interval for the mean number of words typed by word processors at this law firm is (88.26, 97.74). What is the t^* critical value used to compute this interval?

(A) 2.131

(B) 2.145

(C) 2.249

(D) 2.264

(E) 2.624

18. Two random samples from two independent populations are taken with the following results.

Sample 1	Sample 2
$n = 30$	$n = 40$
$\bar{x} = 26$	$\bar{x} = 31$
$s = 3.2$	$s = 3.8$

The standard error of the sampling distribution of the differences of the means is

(A) 0.594

(B) 0.838

(C) 4.968

(D) 7.000

(E) 24.646

19. Of the registered voters in a community, 58% are female. A local politician running for office has the support of 48% of the registered women and 53% of the registered men. What percentage of the vote can the politician expect to get?

(A) 49.8%

(B) 50.1%

(C) 50.5%

(D) 58.58%

(E) Not enough information is given to determine the percentage of support for the politician.

20. The percentages of children living below poverty levels for white and Asian children in the United States from 1987 to 2000 are given in the back-to-back stemplot.

```
            White         Asian
                    | 10 |
                    | 11 | 8
                9   | 12 |
                5   | 13 |
              8 5   | 14 | 4
            9 3 1   | 15 |
        9 8 3 2 1   | 16 | 4
              8 4   | 17 | 5 6
                    | 18 | 0 2 3
                    | 19 | 5 5 8
                    | 20 | 3
                    | 21 |
                    | 22 |
                    | 23 | 5
                    | 24 | 1          11 | 8 = 11.8%
```

Which of the following is a statement that can be made from an examination of the back-to-back stemplot?

(A) For every year from 1987 to 2000, the percentage of white children living in poverty is lower than the percentage of Asian children.

(B) If the outliers are removed from the distribution of Asian poverty-level percentages, the range for the distribution of white children will be larger than that for Asian children.

(C) The mean poverty-level percentage for both white and Asian children is less than the respective median value.

(D) The poverty-level percentages for both white and Asian children have increased over time.

(E) There is more variability in the poverty-level percentages for Asian children than for white children.

21. Which of the statements listed below is correct?

    I. The slope of a regression line can be calculated from the formula
    $$b = \frac{r \cdot s_y}{s_x}.$$
    II. Residual = actual value − fitted value
    III. Causation is demonstrated by the correlation coefficient.

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

22. A set of numbers contains five values. The largest value is 500, and the range is 100. Which of the following statements is *not* true?
    (A) The largest possible value for the mean is 480.
    (B) The smallest possible value for the mean is 420.
    (C) The largest possible value for the median is 500.
    (D) The smallest possible value for the median is 400.
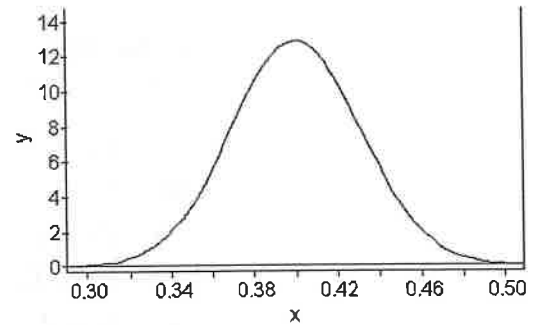   (E) The smallest possible value for the standard deviation is 0.

23. Which of the following could represent the sampling distribution of sample proportions if $p = 0.40$ and $n = 16$?
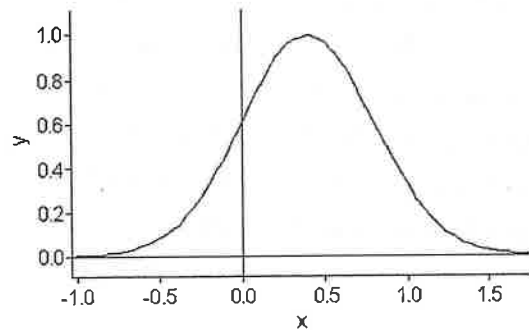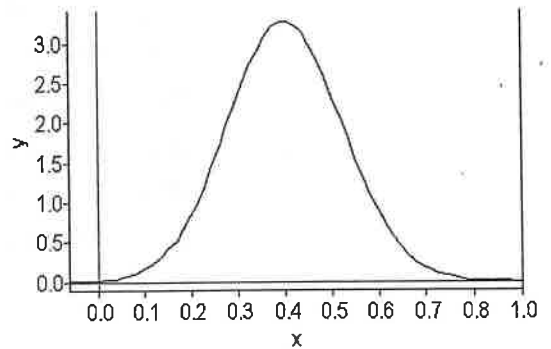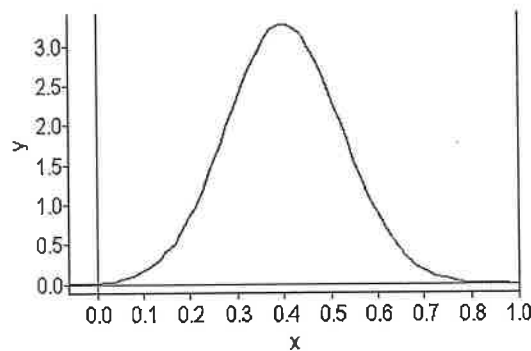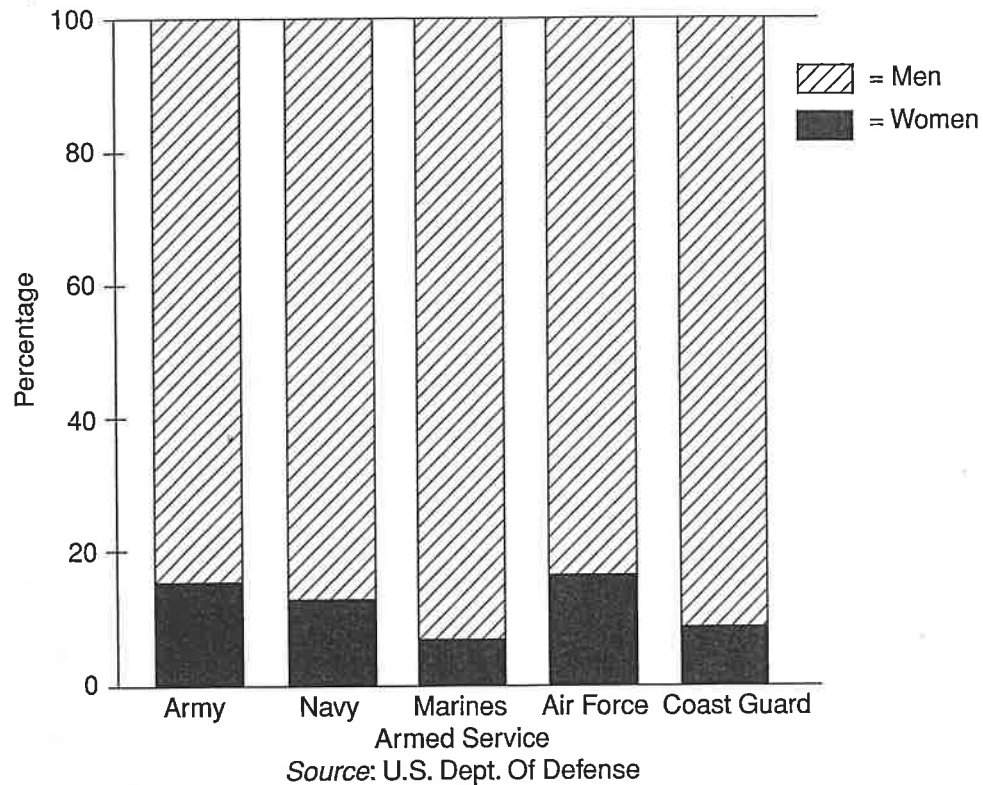
(A)
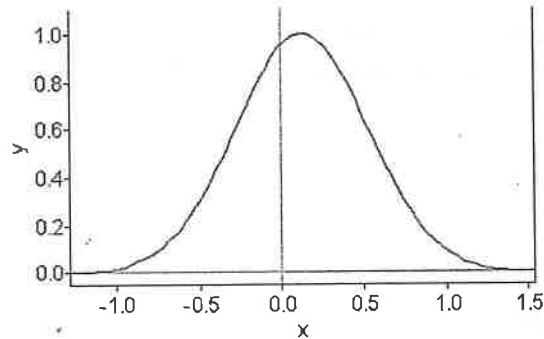
(B)

(C)

(D)

(E)

24. A population is normally distributed with mean 42.3 and standard deviation 2.8. The variance for the sampling distribution of sample means for samples of size 10 is

(A)   0.028

(B)   0.280

(C)   0.784

(D)   0.885

(E)   2.479

25. The breakdown of percentages of men and women in the armed services in 2000 is given in the following segmented bar graph.

Source: U.S. Dept. Of Defense

Which of the following can be stated from an observation of the chart?

(A) The number of women in the Marines is less than the number of women in any other armed service.

(B) The number of men in the Air Force is less than the number of men in any other armed service.

(C) The percentage of women in the Marines is less than the percentage of women in any other armed service.

(D) The proportion of men in each of the services is the same.

(E) The percentage of women in the armed services is changing over time.

26. If $P(B) = 0.4$ and $P(A \cap B) = 0.21$, then find $P(A)$ if A and B are independent.

(A) 0.084

(B) 0.475

(C) 0.525

(D) 0.600

(E) Not possible

27. For Hospital A, the average waiting time (time between walking in the door and seeing a doctor) in the emergency room is 135 minutes with a standard deviation of 45 minutes. For Hospital B, the average waiting time in the emergency room is 90 minutes with a standard deviation of 22.5 minutes. In which hospital are you more likely to wait less than 45 minutes? Assume the distributions of waiting times are normally distributed.

   (A) Hospital B, because the average wait-time is only 90 minutes, rather than 135 minutes in Hospital A.

   (B) Hospital A, because with twice the standard deviation of Hospital B, it has twice the spread.

   (C) Neither, because for both hospitals, the probability of waiting less than 45 minutes is 2.275%.

   (D) Neither, because for both hospitals, the probability of waiting less than 45 minutes is 97.725%.

   (E) It is impossible for any patient to wait less than 45 minutes in either hospital unless the patient is in critical condition.

28. Suppose the adult unemployment rate of a city is 4.8%. If you had taken a survey of 100 adults and constructed a 95% confidence interval for the proportion of unemployed adults, which of the following would have been true?

   (A) The interval would have contained 4.8%.

   (B) The center of the interval would have been 4.8%.

   (C) You would have had a 95% probability that the interval contained 4.8%.

   (D) Increasing the sample size would have ensured capturing 4.8%.

   (E) Approximately 95% of similarly constructed intervals would have captured 4.8%.

29. A recent news program reported that the presidential approval rate was 51% with a margin of error of ±4%.

   What is meant by ±4%?

   (A) 4% of the respondents were undecided.

   (B) The proportion of Americans who approve of the president is between 49% and 53%.

   (C) The president's approval rating from those sampled was between 47% and 55%.

   (D) The proportion of Americans who approve of the president is between 47% and 55%.

   (E) Unless the true proportion of Americans who approve of the president is between 47% and 55%, it is unlikely we could have obtained these sample results.

30. An aspirin maker claims that 4 out of 5 doctors recommend its product. A consumer advocacy group believes the proportion is lower. To test the claim, a random sample of 50 doctors is selected, and 35 recommend this manufacturer's product. An appropriate test outcome is

(A) $z = -1.768$ and $p = 0.039$

(B) $z = -1.768$ and $p = 0.077$

(C) $z = -1.768$ and $p = 0.961$

(D) $t = -1.768$ and $p = 0.042$

(E) $t = -1.758$ and $p = 0.083$

31. Ruth plans to sell the jewelry she makes at an outdoor craft festival this coming Saturday. Based on her experience from past years, she can expect to make a profit of $400 if it is a sunny day, $275 if the weather is overcast, and $100 if it is raining. The weather forecaster (based on historical records) has estimated the chance of a sunny day for the day of the craft festival to be 0.65, the chance of an overcast day to be 0.15, and the chance of a rainy day to be 0.20. What is Ruth's expected profit from the sale of her jewelry?

(A) $400.00

(B) $321.25

(C) $275.00

(D) $258.33

(E) $100.00

32. Suppose we have a random variable X with probability p. The probability of exactly 3 successes in 8 trials is given by

$$P(X = 3) = C\binom{8}{3}(p)^3(0.45)^5.$$

What is the mean and standard deviation of X?

(A) mean $= 4.4$; standard deviation $= 0.2475$

(B) mean $= 3.6$; standard deviation $= 0.2475$

(C) mean $= 4.4$; standard deviation $= 1.4071$

(D) mean $= 3.6$; standard deviation $= 1.4071$

(E) There is not enough information to find the mean and standard deviation.

33. A regression equation is given as $\log \hat{y} = 0.214 - 1.28x$. What is the (approximate) predicted value for y when $x = 2$?

(A) -2.346

(B) -0.171

(C) 0.005

(D) 0.167

(E) Cannot be determined

34. We are given two sample proportions, $\hat{p}_1 = 0.38$ and $\hat{p}_2 = 0.42$. A 90% confidence interval for the true difference in population proportions if $n_1 = 40$ and $n_2 = 50$ is

(A) $-0.04 \pm 1.645\sqrt{\dfrac{(0.38)(0.62)}{40} + \dfrac{(0.42)(0.58)}{50}}$

(B) $0.04 \pm 1.697\sqrt{\dfrac{(0.38)(0.62)}{40} + \dfrac{(0.42)(0.58)}{50}}$

(C) $-0.04 \pm 1.729\sqrt{\dfrac{(0.38)(0.62)}{40} + \dfrac{(0.42)(0.58)}{50}}$

(D) $0.04 \pm 1.96\sqrt{\dfrac{(0.38)(0.62)}{40} + \dfrac{(0.42)(0.58)}{50}}$

(E) not able to be constructed.

35. The residuals for a complete data set are shown below, and r^2 for the least squares regression line that resulted in these residuals is 88.6%.

x	25	30	35	40	45	50	55	60	65	70
Residual	0	1.2	2.0	0.97	0.9	-1.1	-4.6	-0.6	0.25	0.98

Which of the following is/are true?

I. The linear model is a good model for the data.
II. The sum of the squares of the residuals is zero.
III. The correlation is either $\pm\sqrt{0.886}$.

(A) I only

(B) II only

(C) III only

(D) I and III only

(E) I, II, and III

36. Two students went to their local shopping mall to conduct a survey. They wanted to know how the local population felt about boys coloring their hair. Both students had neat haircuts but one had dyed hair and one did not. What type of bias could occur in their survey?
    (A) Undercoverage
    (B) Nonresponse bias
    (C) Response bias
    (D) None of the above
    (E) A, B, and C may produce bias in this setting.

37. For the dotplot shown, which of the statistical values listed would be least affected by the rightmost value?
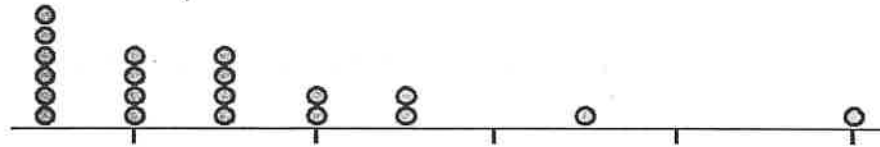
    (A) Standard deviation
    (B) Range
    (C) Variance
    (D) Mean
    (E) Median

38. A χ^2-distribution with 14 degrees of freedom is a correct model for
    (A) a comparison of the production percentage distribution of 7 car model colors with the statistically determined national preferences for those colors.
    (B) testing the question of whether 14 genetic traits are equally distributed in a population.
    (C) testing whether choice of color is independent of age among 3 age groups and 5 color choices.
    (D) testing whether the choice to smoke cigarettes or not to smoke cigarettes is independent of ethnicity among 7 different ethnic groups.
    (E) a comparison of the equality of proportions of 8 sports activities for 3 high school grade levels.

39. Starting time for hourly wage employees at a large manufacturing plant is 7 A.M. If an employee clocks in before 7:15 A.M., he is not marked as being late for work and his pay is not reduced. A random sample of 15 daily time sheets from the past two years showed that the average number of employees who arrived at work between 7 A.M. and 7:15 A.M. each day was 23 with a standard deviation of 6. Assume that the assumptions for inference have been met. Construct a 90% confidence interval for the mean number of employees who arrive at work during this time frame each day.

(A) 23 ± 2.728

(B) 23 ± 3.315

(C) 23 ± 2.630

(D) 23 ± 3.301

(E) 23 ± 2.894

40. For the given probability distribution, find the standard deviation of X.

X	1	3	5	7	9
P(x)	0.13	0.17	0.25	0.24	0.21

(A) 0.050

(B) 2.621

(C) 3.162

(D) 5.460

(E) 6.868

SECTION II: FREE RESPONSE
Part A

90 minutes
Suggested time: 65 minutes

1. Twins often show an uncanny similarity in preferences even after years of separation. At the National Twins Convention, a researcher secures the cooperation of the participants for a study. The researcher randomly selects ten pairs of twins and asks the pairs to record the number of times they have eaten dinner out in the past three months. The recorded data follow.

Twin Pair	Activity Frequency	
1	8	5
2	7	12
3	12	13
4	5	7
5	6	4
6	9	8
7	12	10
8	3	4
9	4	11
10	15	12

a. Are the preferences of twins the same? Perform a statistical test.

b. Do you have any reservations about your analysis? If so, explain.

2. The owner of a synthetic motor oil company claims that his oil maintains viscosity and protects engines up to ten times as long as petroleum motor oil in gas-powered vehicles.

a. Describe an experimental design that would test the effectiveness of the synthetic motor oil, using viscosity as a measure of effectiveness.

b. Describe any possible confounding variables.

c. Viscosity is measured numerically in units called centipoise (cP). Explain what hypothesis test you would use to test whether there is a difference in average viscosity between the synthetic motor oil and the petroleum motor oil. List both your null and alternative hypotheses.

3. A triathlon is a sports competition with three distinct components. Some triathlons consist of a 2.4-mile swim, followed by a 112-mile bike ride, and end with a 26.2-mile run. The mean times (in minutes) and standard deviations for these portions of the race are recorded below.

Component	Mean	Standard Deviation
Swim	75.465	12.378
Bike	403.506	45.023
Run	287.497	49.894

a. In a recent event, the fastest swimming time recorded was 49 minutes. If the swimming times are normally distributed, what is the probability that a randomly chosen triathlete will record a time of 49 minutes or less?

b. At the same event, the slowest time for the bike portion of the race was 8 hours and 43 minutes. If the bike times are normally distributed, what is the probability that a randomly chosen triathlete will record a time of 8 hours and 43 minutes or more?

c. What is the probability that the average run time for four randomly selected triathletes will be 5 hours or less?

d. The winning time for a recent triathlon was 8 hours and 46 minutes. If we assume the times for the three legs of the race are independent, what is the probability that a person will finish the triathlon in 8 hours and 46 minutes or less?

4. A random sample of U.S. doctors was asked the question, "In general, should a surgeon diagnosed with dementia be allowed to perform surgery?" Of the 921 sampled, 83 indicated: "Yes, with close supervision"; 838 indicated: "No, not under any circumstance."

   a. Use this information to construct and interpret a 98% confidence interval for the number of doctors who think that a surgeon diagnosed with dementia should be allowed to perform surgery with close supervision.

   b. The same question was asked of a random sample of nurses. Of the 768 sampled, 23 indicated that the surgeons should be allowed to perform surgery with close supervision. Is there significant evidence to suggest that the proportion of nurses who feel that surgeons diagnosed with dementia should be allowed to perform surgery under close supervision is less than the proportion of doctors? Give statistical evidence to support your decision.

5. A manufacturer of potato chips produces large bags of chips. The production weight is approximately normally distributed with a mean of 14.8 oz and a standard deviation of 1.2 oz.

   a. What is the probability of obtaining a bag of this company's chips that weighs less than 14 oz?

b. Round the value you obtained in part (a) to the nearest hundredth. Explain how you would conduct a simulation using the random-number table displayed below to estimate the number of bags of chips you would have to buy before obtaining one that weighs less than 14 oz. Perform your simulation seven times. Start at the leftmost digit in the first row and move across. By marking directly on or above the table, make your procedure clear enough that someone can follow what you did. Record the number of bags of chips for each of your trials.

9 2 1 9 5 6 6 8 0 3 3 1 8 2 5 6 4 6 5 1 0 5 8 2 0 9 3 9

2 3 9 8 9 2 4 8 8 7 9 3 3 2 0 3 6 2 5 0 9 4 4 3 1 1 7 6

8 4 1 6 1 4 0 3 9 7 0 1 9 8 3 7 6 7 9 3 2 6 8 3 2 8 2 3

c. Based on your results in part (b), what is the expected number of bags of potato chips you will have to buy before finding one that weighs less than 14 oz?

Part B

Suggested time: 25 minutes

6. A store that sells cellular telephones opened for business twelve years ago. Total sales of cellular telephones for each year of business are given.

Year	1	2	3	4	5	6	7	8	9	10	11	12
Units Sold	375	566	671	1106	1311	1283	2136	2967	4094	4572	5157	6621

a. Record and interpret the slope of the least squares regression line that describes the relationship between the number of cellular telephones sold and the business year.

b. Record and interpret the value of the correlation coefficient for the year of business and the number of cell phones sold.

c. Use this model to predict the number of cell phones the store will sell in the next business year.

d. Describe any shortcomings you see in this model.

e. Find a better model for predicting the number of units sold for a given business year. Justify your choice, and use three decimal places for your slope and intercept coefficients.

f. Use your model from part (e) to predict the number of cell phones the store will sell in the next business year.