Diagnostic Examination

Section I

Questions 1–40

Spend 90 minutes on this part of the exam.

Directions: The questions or incomplete statements that follow are each followed by five suggested answers or completions. Choose the response that best answers the question or completes the statement.

1. The principal of a small high school wishes to know the average GPA of all current students and asks the district IT department to run an appropriate program through the school computer files. The next day he is informed that the average GPA among the current 387 students is 2.76 (on a 4-point scale). When the principal then asks the AP Statistics teacher to calculate a 95% confidence interval of the average GPA of current students, the teacher says that this would be inappropriate. Why?

   (A) The independence assumption may have been violated.
   (B) Randomization was not used.
   (C) The 10% (of the population) condition was violated.
   (D) There is no evidence that the data comes from a normal or nearly normal population.
   (E) Some other reason.

2. In a random sample of 15 bowls of Raisin Nut Bran cereal, each with $3/4$ of a cup of cereal and a $1/2$ cup of skim milk, the mean number of calories is 220 with a standard deviation of 6. Assuming all assumptions for inference are met, what is a 95% confidence interval for the mean number of calories of such bowls of cereal?

   (A) $220 \pm 1.96 \left(\dfrac{6}{\sqrt{14}} \right)$

   (B) $220 \pm 1.96 \left(\dfrac{6}{\sqrt{15}} \right)$

   (C) $220 \pm 2.131 \left(\dfrac{6}{\sqrt{14}} \right)$

   (D) $220 \pm 2.131 \left(\dfrac{6}{\sqrt{15}} \right)$

   (E) $220 \pm 2.145 \left(\dfrac{6}{\sqrt{15}} \right)$

GO ON TO THE NEXT PAGE ➤

3. Consider two treatments for diabetes and their respective success rates. In a double blind study, Treatment A is successful on 40 out of 43 patients with mild diabetes and 96 out of 132 patients with severe disease, while Treatment B is successful on 117 out of 135 patients with mild diabetes and 27 out of 50 patients with severe disease. Which of the following is a correct statement?

(A) Treatment A has a higher success rate with patients with mild diabetes, severe diabetes, and overall than Treatment B.

(B) Treatment A has a higher success rate with patients with mild diabetes, but Treatment B has a higher success rate with patients with severe diabetes and a higher overall success rate.

(C) Treatment A has a higher success rate with patients with mild diabetes and those with severe diabetes, but Treatment B has a higher overall success rate.

(D) Treatment B has a higher success rate with patients with mild diabetes and those with severe diabetes, but Treatment A has a higher overall success rate.

(E) Treatment B has a higher success rate with patients with mild diabetes, severe diabetes, and overall than Treatment A.

4. The time required for the first stage on an assembly line is normally distributed with a mean of 32 minutes and variance of 6 minutes, while the time required for an independent second stage is normally distributed with a mean of 43 minutes and variance of 10 minutes. What is the probability that the two stages take over 80 minutes?

(A) .1056
(B) .1865
(C) .2593
(D) .3341
(E) .3773

5. Which of the following is most important in avoiding bias in surveying?

(A) Picking a sampling method in which every combination of people has the same chance of being selected

(B) Care in choosing individuals who you feel are representative of the population in terms of as many characteristics as you can list

(C) Making sure every individual in the population of interest is invited to respond

(D) Sending out a large enough number of surveys so that even if the response rate is low, you will have sufficient numbers of responses to do statistical analysis

(E) Making sure the wording is not so neutral as to bore potential respondents

6. A sample of 20 volunteers was randomly split into two groups. In one group the subjects were given caffeine pills while in the other the subjects received placebos. Heart rates were measured (in beats per minute) before and after the subjects took the pills, and the differences (before minus after) are given below.

Caffeine	−5	2	0	3	−3	−2	1	−5	0	1
Placebo	−1	0	−1	1	1	0	−2	3	2	−2

Which of the following tests should be used to determine the difference in mean heart rate changes between caffeine and placebo takers?

(A) One-sample paired *t*-test on a set of differences
(B) Two-sample *t*-test
(C) Regression *t*-test for association between caffeine and placebo
(D) Chi-square test of independence between caffeine and placebo
(E) Two-sample *z*-test on the difference between two proportions

7. A shoe manufacturer comes up with two experimental designs to test the longevity of a new material out of which to make soles. One hundred volunteers are recruited to try out new pairs of shoes for six months. Design 1 involves flipping a coin as each volunteer comes to the store, and if *heads*, giving the subject a pair of shoes with the standard soles and if *tails*, giving shoes with soles made from the new material. Design 2 involves flipping a coin as each volunteer comes to the store, and if *heads* giving the subject a pair of shoes where the right shoe has a standard sole and the left has a sole made from the new material, while if tails the subject receives a pair with a standard sole on the left shoe and a sole made of the new material on the right. Which of the following is accurate?

(A) Neither design uses randomization since volunteers are used rather than randomly picking people from the population of shoe wearers.
(B) Design 1 is a completely randomized design while Design 2 is a block design.
(C) Both designs use blinding but neither uses double blinding.
(D) In the second design, standard construction and new material construction are confounded.
(E) One of the two designs is actually an observational study while the other is an experiment.

8. Three herds of milk-producing cattle, one each of Guernsey, Jersey, and Holstein, are being tested with injections of bovine growth hormone, or BGH, to spur increased milk production. Two sources of BGH are being tested, one from bovine cadavers and a second from genetically engineered E. coli. In each herd the cattle are randomly sorted into two groups, with one group receiving BGH from each of the two sources. Mean daily milk production is compared between the two groups in each herd. Which of the following best describes the design of this experiment?

(A) Blocking on cattle type
(B) Blocking on BGH source
(C) Matched pairs (bovine cadavers versus genetically engineered)
(D) Two factors (cattle type and BGH source)
(E) Completely randomized

9. Everyone in a gathering of 50 high school students (equal numbers of males and females) takes his/her resting pulse rate, and the following back-to-back stemplot displays the results:

| Male		Female
	4	67
8741	5	389
99530	6	024699
8666422	7	166678999
86441	8	02377
9650	9	
SD = 13.4		SD = 11.8

How do the means and medians of the males and females compare?

(A) Both the mean and the median of the males are greater than the mean and the median of the females.

(B) The medians are the same, however because of the larger standard deviation, the males will have a larger mean.

(C) The medians are the same; however, the distributions (one roughly symmetric, the other with two low outliers) indicate that the mean of the males is greater than the mean of the females.

(D) The medians are the same; however, the distributions (one roughly symmetric, the other skewed to the lower values) indicate that the mean of the males is greater than the mean of the females.

(E) The median of the males (74) is less than the median of the females (76), but the mean of the males (74.9) is greater than the mean of the females (70.5).

10. Five new estimators are being evaluated with regard to quality control in manufacturing ball bearings of a given diameter. Each estimator is tested every day for a month on samples of sizes $n = 10$, $n = 20$, and $n = 40$. The ball bearings actually produced that month had a consistent mean diameter of 2.3 cm. The distributions given by each estimator are as follows:

GO ON TO THE NEXT PAGE ➤
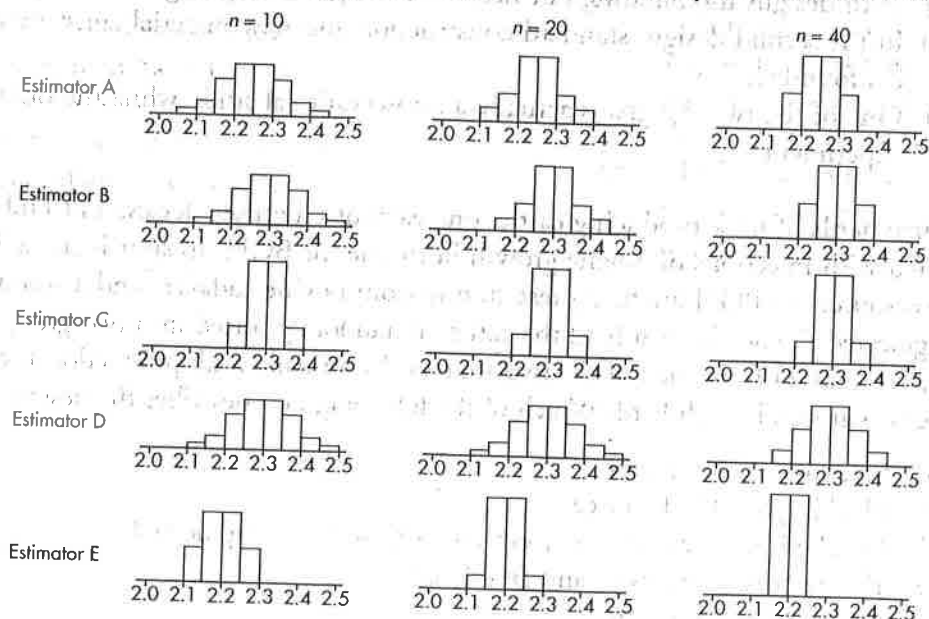
The selected estimator will eventually be used with a sample of size $n = 100$. Which estimator is best?

(A) Estimator A
(B) Estimator B
(C) Estimator C
(D) Estimator D
(E) Estimator E

11. A plot of average selling price of new homes against square footage shows a strong correlation of $r = .78$. What correlation would you expect to see in a plot of selling prices of individual homes against their square footage?

(A) $r = 0$
(B) $0 < r < .78$
(C) $r = .78$
(D) $.78 < r < 1$
(E) $r = 1$

12. A medical researcher determines a linear relationship giving diastolic blood pressure in terms of systolic blood pressure with correlation $r = .65$. If the same data is used to find a linear relationship this time giving systolic blood pressure in terms of diastolic blood pressure, what will the correlation be?

(A) .35
(B) .50
(C) .65
(D) –.35
(E) –.65

13. A simple random sample is defined by

(A) the method of selection
(B) examination of the outcome
(C) whether or not a random number generator is used
(D) how representative the sample is of the population
(E) the assignment of different numbers associated with the outcomes of some chance situation

14. Research has shown that the age at which babies begin to crawl is normally distributed with a mean of 30 weeks. Consider random samples of 15 babies each. The variable $\dfrac{\bar{x} - 30}{\dfrac{s}{\sqrt{15}}}$

(A) has a normal distribution
(B) has a t-distribution with $df = 14$
(C) has a t-distribution with $df = 15$
(D) has a chi-square distribution
(E) has a distribution that cannot be specified because the sample size is too small ($n < 30$)

15. The speeds of cars passing through an automated E-Z Pass toll booth have a nearly normal distribution with a mean of 5 mph and a standard deviation of 0.6 mph. What is a speed at the 30th percentile?

(A) 4.48 mph
(B) 4.69 mph
(C) 4.82 mph
(D) 5.31 mph
(E) 5.52 mph

16. At a large university 55% of the students are women, and 30% of the students participate in intramural sports. If the probability that a student plays intramural sports given that the student is a woman is .24, what is the probability that a student is a woman given that the student plays intramural sports?

(A) .1309
(B) .44
(C) .49
(D) .61
(E) .6875

GO ON TO THE NEXT PAGE ➤

17. Three AP Statistics classes, each with the same number of students, are given the same four-point quizzes. Following are the histograms for the student scores in each class. Put the standard deviations from the three sets of scores in order from smallest to largest.

Class A Class B Class C
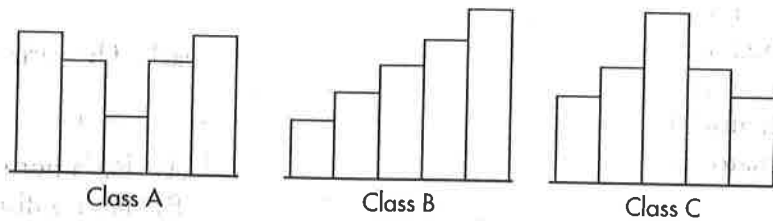
(A) SD(A) < SD(B) < SD(C)
(B) SD(B) < SD(A) < SD(C)
(C) SD(C) < SD(B) < SD(A)
(D) SD(B) < SD(C) < SD(A)
(E) SD(C) < SD(A) < SD(B)

18. Nutrient control is fundamental in water treatment plants. Two such nutrients, carbon (organic substances) and nitrogen (compounds) are ideally found in wastewater in the ratio of 20:1. In ten random samples of wastewater in one treatment plant, data was collected on carbon and nitrogen with the following regression analysis.

```
Dependent variable is: Carbon

Source          SS           MS            F
Regression      33522        33522         1.22e3
Residual        220.352      27.5439

Variable    Coef        s.e.        t         p
Constant    -1.569      7.411      -0.212    0.837
Nitrogen    20.157      0.577      34.9      0.000

R-sq = 99.3%   R-sq(adj) = 99.3%    s = 5.248
```

Which of the following gives a 99% confidence interval for the slope of the regression line?

(A) $20.157 \pm (3.250)\left(\dfrac{0.577}{\sqrt{10}}\right)$

(B) $20.157 \pm (3.250)(0.577)$

(C) $20.157 \pm (3.355)(0.577)$

(D) $20.157 \pm (3.250)(5.248)$

(E) $20.157 \pm (3.355)\left(\dfrac{5.248}{\sqrt{10}}\right)$

19. Three randomly chosen cars of a particular model are subjected to sideswipe (shallow-angle) collisions and the monetary damages are noted. Assuming that monetary damages in this type of accident are normally distributed, a 95% *t*-based confidence interval (with $t = 4.303$) for the mean is found. Which of the following is a correct statement?

    (A) The confidence interval is valid.
    (B) The confidence interval is invalid because the sample size is too small.
    (C) The confidence interval is invalid because $np = 3 < 10$.
    (D) The confidence interval is invalid because $t = 3.182$ should have been used.
    (E) The confidence interval is invalid because with the normality assumption, a *z*-based confidence interval with $z = 1.96$ should have been used.

20. According to one PC magazine, 40% of all e-mail messages are spam. Assuming this to be the case, and assuming that e-mail messages are independent, if a person receives three spam messages in a row, what is the probability the next two messages will also be spam?

    (A) $(.4)^2$
    (B) $(.4)^5$
    (C) $10(.4)^2(.6)^3$
    (D) $1 - (.4)^3$
    (E) $1 - (.6)^3$

21.
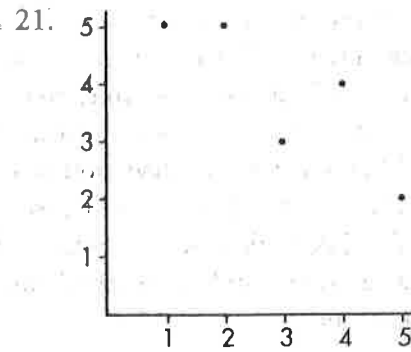
    The least squares regression line for the above scatterplot is $\hat{y} = 5.9 - 0.7x$. What is the residual for the point (3, 3)?

    (A) 0.8
    (B) 3.8
    (C) −3.8
    (D) −3.0
    (E) −0.8

22. Which of the following is *not* characteristic of cluster sampling?

    (A) Randomization plays a role.
    (B) It is often utilized for cost or convenience considerations.
    (C) The population is divided into homogeneous groups called clusters.
    (D) Each cluster is representative of the population.
    (E) Typically a census (every individual) in the chosen clusters is part of the sample.

23. A professional track star runs the mile (the only non-metric distance recognized for record purposes) in times that are normally distributed with a mean of 3:50 and a standard deviation of 0:01. Which of the following is closest to $P(X < 3:49)$, that is, to the probability that he runs a mile in under 3 minutes, 49 seconds?

    (A) $P(3:49 < X < 3:51)$
    (B) $P(3:49.3 < X < 3:50.7)$
    (C) $P(3:49.5 < X < 3:50.5)$
    (D) $P(3:49.8 < X < 3:50.2)$
    (E) $P(3:49.95 < X < 3:50.05)$

24. In a study of whether working men make more money than working women, a researcher randomly selected 500 households, interviewed all working adults in these households, and recorded their salaries. The results are summarized as follows:

	Men	Women
n	420	450
$\overline{x}$	$25,430	$23,270
s	$5,100	$4,200

    Why is it *not* appropriate to use a two-sample *t*-test for the difference between two means?

    (A) The two sample sizes are different.
    (B) It is unclear whether a simple random sample (SRS) of households was used.
    (C) The sample means are not from independent samples.
    (D) It is unknown whether the salary populations have normal distributions.
    (E) The equal variance assumption is violated.

25. The distribution of weights of a particular brand of "368 g" bags of tortilla chips is roughly normal with a mean of 370 grams and standard deviation 1.4 grams. Which of the following is the greatest?

    (A) A weight of 371 g
    (B) A weight with a z-score of 1
    (C) A weight with a percentile rank of 80%
    (D) A weight at the third quartile, Q_3
    (E) Unknown without knowledge of the *median* of the distribution.

26. In an SRS of 500 credit card customers, a computer printout of summary statistics for monthly expenditures (in dollars) was

    $x = 478$
    $Sx = 129$
    $n = 500$
    $\min X = 102$
    $Q_1 = 397$
    $Med = 469$
    $Q_3 = 603$
    $\max X = 824$

    About 250 customers had expenditures that were

    (A) above $478
    (B) between $340 and $598
    (C) between $349 and $607
    (D) between $391 and $565
    (E) between $397 and $603

27. Suppose we have a random variable X where the probability associated with the value $\binom{8}{k}(.45)^k(.55)^{8-k}$ for $k = 0, \ldots, 8$. What is the mean of X?

    (A) 0.45
    (B) 0.55
    (C) 1.98
    (D) 3.6
    (E) 4.4

28. Which of the following is the most reasonable conclusion to the sentence beginning "After several unusually cold winters in a row,"

    (A) by the law of averages a warmer-than-usual winter is due
    (B) by the law of small numbers a warmer-than-usual winter is due
    (C) by the law of large numbers it is likely that the string of unusually cold winters will continue
    (D) through the use of simulation it can be shown that the string of unusually cold winters is likely to continue
    (E) the probability of another unusually cold winter has little to do with the results of the several previous winters

29. An adjustable-elbow-angle golf swing trainer is set at the same fixed position each morning, and is subject to heavy use during the day. It is tested on a random sample of 30 days at the evening closing time. A mean of 81° with a standard deviation of 2° is noted. Which of the following is the best interpretation of a 95% confidence interval for the average angle for which this trainer is found at the daily closing?

    (A) On 95% of the daily closings, the angle will be between 79° and 83°.
    (B) On 95% of the daily closings, the angle will be between 80.3° and 81.7°.

    (C) On 95% of the daily closings, the mean angle will be between 79° and 83°.
    (D) We are 95% confident that the angle at any given closing will be between 80.3° and 81.7°.
   (E) We are 95% confident that the mean angle at closing times is between 80.3° and 81.7°.

30. A sample of test cars from each of two new models was run on highways and mileages (mpg) were noted.

 | | Model 1 | Model 2 |
    |--------|---------|---------|
    | n | 10 | 15 |
    | Mean | 23.4 | 25.6 |
    | Median | 22 | 26 |

   Given the above, what can be said about the mean and median of the combined 25 values?

    (A) Mean = 24.50; nothing can be said about the median
    (B) Mean = 24.50; median is between 22 and 26
    (C) Mean = 24.72; nothing can be said about the median
    (D) Mean = 24.72; median is between 22 and 26
    (E) Mean = 24.72; median = 24

31. Skewed-left distributions are most likely to be indicated by which of these boxplot features?

   (A) Medians to the left in the boxes together with outliers to the left
   (B) Medians to the left in the boxes together with outliers to the right
   (C) Medians to the right in the boxes together with outliers to the left
   (D) Medians to the right in the boxes together with outliers to the right
   (E) Very large ranges with small interquartile ranges

32. Following is the graph of the sampling distribution of the sample mean of a particular population using a large sample size n.
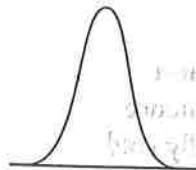
   Which of the following most likely shows the shape of the original population?

   (A)

   Exponential

   (C)

   U-shaped

   (B)

   Uniform

   (D)

   Bell-shaped
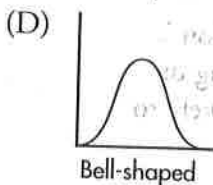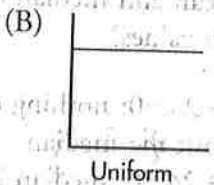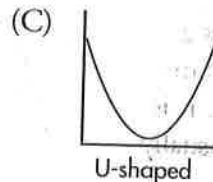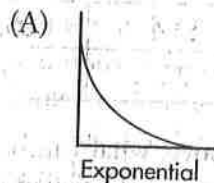
   (E) By the central limit theorem, all of the above are reasonable answers.

33. Polyphenols have been shown to have antioxidant characteristics with potential health benefits in reducing the risk of cardiovascular disease. In a major study to determine whether red or white wine increases the level of polyphenols in the blood, 500 healthy adult volunteers were randomly split into two groups, one of which drank a glass of red wine every evening while the other did the same with white wine. The level of polyphenols in the blood was measured at the beginning and end of the month-long test period, and the percent change in level for each subject was noted. The resulting summary statistics were as follows:

|	Mean percent change	SD in percent change
Red wine drinkers	5.6	2.4
White wine drinkers	5.2	2.8

   A two-sided t-test results in $t = 1.715$. Which of the following is an acceptable conclusion?

   (A) Because the P-value is less than $\alpha = .05$, there is evidence that the mean percent change in level of polyphenols in the blood from drinking red wine is greater than that from drinking white wine.

(B) Because the *P*-value is more than $\alpha = .05$, there is not sufficient evidence that the mean percent change in level of polyphenols in the blood from drinking red wine is greater than that from drinking white wine.

(C) Because the *P*-value is less than $\alpha = .05$, there is evidence that the mean percent change in level of polyphenols in the blood from drinking red wine is different from that from drinking white wine.

(D) Because the *P*-value is more than $\alpha = .05$, there is not sufficient evidence that the mean percent change in level of polyphenols in the blood from drinking red wine is different from that from drinking white wine.

(E) Because the standard deviations are large as compared to the observed difference in mean percent changes, no conclusion will be statistically significant.

34. A fortified corn-soy blend (CSB) is evaluated for vitamin C content. The standard is 25 mg vitamin C per 100 g of CSB. Consider a hypothesis test with $H_0 : \mu = 25$ and $H_a : \mu < 25$. Which of the following choices of significance level and sample size results in the greatest power of the test when $\mu = 20$?

(A) $\alpha = 0.01$, $n = 15$
(B) $\alpha = 0.05$, $n = 15$
(C) $\alpha = 0.01$, $n = 30$
(D) $\alpha = 0.05$, $n = 30$
(E) Unknown without knowledge of the strength of the given power.

35. A state inspector believes that more than 35% of cars on the state roads fail to meet pollution control requirements. Which of the following statements is appropriate in investigating this belief?

(A) The null hypothesis is that less than 35% of cars on the state roads fail to meet requirements.
(B) The null hypothesis is that more than 35% of cars on the state roads fail to meet requirements.
(C) The alternative hypothesis is that 35% of cars on the state roads fail to meet requirements.
(D) The alternative hypothesis is that less than 35% of cars on the state roads fail to meet requirements.
(E) The alternative hypothesis is that more than 35% of cars on the state roads fail to meet requirements.

36. To study water hardness, the calcium concentration (in parts per million) is measured in water collected from an SRS of 100 wells in a large farming community, and a 95% confidence interval is obtained. If a second study, with an SRS of 400 wells, shows a sample mean 10% higher and a sample standard deviation half that of the first sample, which of the following best describes the width of the new 95% confidence interval?

(A) The new width will be about the same as the old width.
(B) The new width will be about one-tenth of the old width.
(C) The new width will be about one-half the old width.
(D) The new width will be about one-quarter the old width.
(E) The new width will be about one-eighth the old width.

GO ON TO THE NEXT PAGE ➤

37. To which of the histograms could the following boxplot correspond?

(A)

(B)

(C)

(D)

(E)
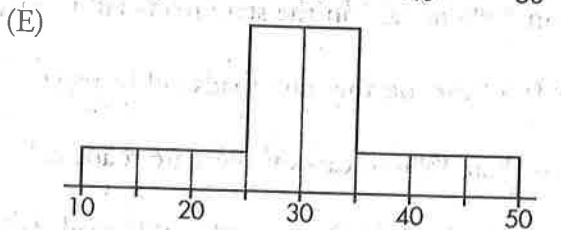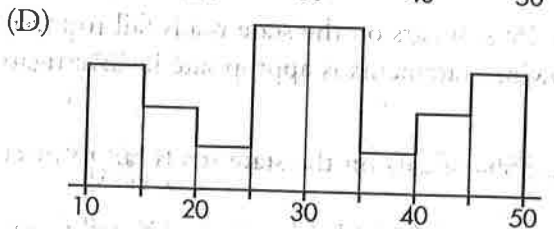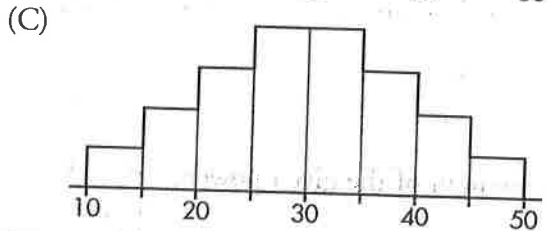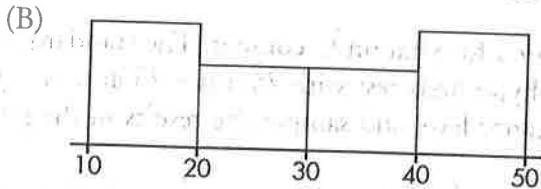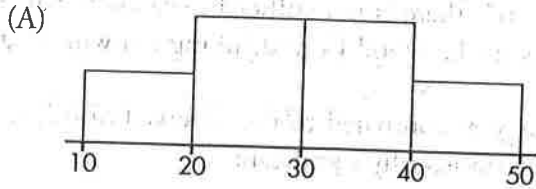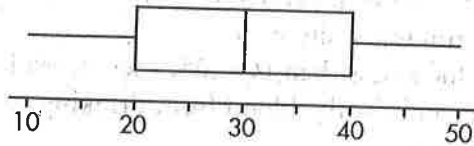
38. According to one NCAA report, only 47% of university student-athletes graduate within six years. Every week for one year, a reporter for a national sports magazine picks an SRS of 25 student-athletes who started college over six years ago and records the number who graduated within six years. What theoretical distribution describes this procedure?

(A) A binomial distribution with $n = 25$ and $p = .47$
(B) A binomial distribution with $n = 52$ and $p = .47$
(C) The sampling distribution of sample proportions with $n = 25$ and $p = .47$
(D) The sampling distribution of sample proportions with $n = 52$ and $p = .47$
(E) The sampling distribution of sample means with $\mu = (25)(.47)$ and $\sigma_{\bar{x}} = \sqrt{(25)(.47)(.53)}$

39. In a test on whether or not atmospheric nitrogen content was once lower than the current 78.08% level, the bubbles in several samples of ancient amber are tested. A t-test of $H_0 : \mu = 78.08$ against $H_a : \mu < 78.08$ gives a P-value of 0.072. Using this data, among the following, which is the largest level of confidence for a two-sided confidence interval that does not contain 78.08?

(A) 85%
(B) 90%
(C) 92%
(D) 93%
(E) 96%

40. A long-term study of 243 soccer players who had sustained anterior cruciate ligament tears during their teenage years noted gender and whether or not 12 years later they had X-ray findings consistent with osteoarthritis. The question was whether or not developing early radiographic osteoarthritis is independent of gender among teenage soccer players with the given injury. Partial computer output is as follows:

	Arthritis	Healthy	Total
Male	43	102	145
	45.95	99.05	
Female	34	64	98
	31.05	66.95	
Total	77	166	243

What is the resulting chi-square value?

(A) $\dfrac{(43-45.95)^2}{43}+\dfrac{(102-99.05)^2}{102}+\dfrac{(34-31.05)^2}{34}+\dfrac{(64-66.95)^2}{64}$

(B) $\dfrac{(43-45.95)^2}{45.95}+\dfrac{(102-99.05)^2}{99.05}+\dfrac{(34-31.05)^2}{31.05}+\dfrac{(64-66.95)^2}{66.95}$

(C) $\dfrac{(77)(145)}{243}+\dfrac{(166)(145)}{243}+\dfrac{(77)(98)}{243}+\dfrac{(166)(98)}{243}$

(D) $\left(\dfrac{43}{45.95}\right)^2+\left(\dfrac{102}{99.05}\right)^2+\left(\dfrac{34}{31.05}\right)^2+\left(\dfrac{64}{66.95}\right)^2$

(E) $\sqrt{(43-45.95)^2+(102-99.05)^2+(34-31.05)^2+(64-66.95)^2}$

STOP

If there is still time remaining, you may review your answers.

SECTION II

PART A

Questions 1–5

Spend about 65 minutes on this part of the exam.
Percentage of Section II grade—75

> You must show all work and indicate the methods you use. You will be graded on the correctness of your methods and on the accuracy of your results and explanations.

1. A horticulturist plans a study on the use of compost tea for plant disease management. She obtains 16 identical beds, each containing a random selection of five mini-pink rose plants. She plans to use two different composting times (two and five days), two different compost preparations (aerobic and anaerobic), and two different spraying techniques (with and without adjuvants). Midway into the growing season she will check all plants for rose powdery mildew disease.

   (a) List the complete set of treatments.
   (b) Describe a completely randomized design for the treatments above.
   (c) Explain the advantage of using only mini-pink roses in this experiment.
   (d) Explain a disadvantage of using only mini-pink roses in this experiment.

2. A top-100, 7.0-rated tennis pro wishes to compare a new Wilson N1 racquet against his current model. He strings the new racquet with the same Luxilon strings at 60 pounds tension that he uses on his old racquet. From past testing he knows that the average forehand cross court volley with his old racquet is 82 miles per hour (mph). On an indoor court, using a ball machine set at 70 mph, the same speed he had his old racquet tested against, he takes 47 swings with the new racquet. An associate with a speed gun records an average of 83.5 mph with a standard deviation of 3.4 mph. Assuming that the 47 swings represents a random sample of his swings, is there statistical evidence that his speed with the new racquet is an improvement over the old? Justify your answer.

3. In October, 2008, a comprehensive residential college in upstate New York reported undergraduate enrollment by ethnic/racial categories as follows: 2.7% non-Hispanic Black, 3.7% Asian or Pacific Islander, 4.0% Hispanic, 80.0% non-Hispanic White, and the rest other/unknown. While racial/ethnic status is not considered in the admissions process, an admissions counselor is interested in whether or not the makeup of the new freshman class will change, and plans to do a statistical analysis on an appropriately drawn simple random sample.

 (a) What statistical test/procedure should be used?
 (b) State the null and alternative hypotheses. Is the test appropriate for an intended sample size of 200?
 (c) If the admissions counselor performs the indicated test on the following data, is there statistical evidence of a change in ethnic/racial composition? Explain.

	Non-Hispanic Black	Asian or Pacific Islander	Hispanic	Non-Hispanic White	Other/Unknown
Number of Students	3	4	14	150	29

 (d) Suppose the data was obtained by noting the racial/ethnic status of a simple random sample of 200 potential new students visiting the campus during fall 2008. Did the test/procedure target the intended population? Explain.

4. Concrete is made by mixing sand and pebbles with water and cement and then hardening through hydration. Different densities result from different proportions of the aggregates. Assume that concrete densities are normally distributed with mean 2317 kilograms per cubic meter and standard deviation 128 kilograms per cubic meter.

 (a) What is the probability that a given concrete density is over 2400 kg/m^3?
 (b) In a random sample of five independent concrete densities, what is the probability that a majority have densities over 2400 kg/m^3?
 (c) What is the probability that the mean of the five independent concrete densities is over 2400 kg/m^3?

5. A small art gallery in Laguna Beach has the choice of stocking either oil paintings or finger paintings for a given tourist season. The oil paintings require a substantial investment, but the potential returns are also greater. The return (profit or loss) depends on whether or not the tourists that season are primarily serious art collectors or more casual buyers. A sales analysis gives the following expectations.

Season return ($1000)	Type of tourists	
Stock decision	Art collectors	Casual buyers
Oil paintings	135	−35
Finger paintings	−5	25

Let p be the probability that the type of tourist is primarily art collectors, so $(1 - p)$ is the probability of primarily casual buyers.

(a) As a function of p, what is the expected return for stocking oil paintings?

(b) As a function of p, what is the expected return for stocking finger paintings?

(c) For what value of p are the two expected returns the same, and what does it mean in context for p to be greater or less than this value?

(d) In a random sample of similar establishments in similar tourist regions, 33 out of 150 reported seasons with tourists who were primarily art collectors. Construct a 95% confidence interval for the proportion of similar establishments with tourists who were primarily art collectors.

(e) Use the above results to justify a decision to stock finger paintings.

SECTION II

PART B

Question 6

Spend about 25 minutes on this part of the exam.
Percentage of Section II grade—25

6. A national retail chain classifies cashiers as "entry level" for the first
    ten years. The series of boxplots below shows the relationship between
    yearly wages (in $1000) and years of experience (YOE) for a random
    sample of these employees. Below the boxplots are computer regression
    outputs.
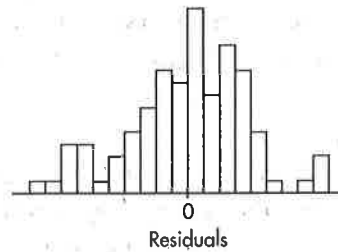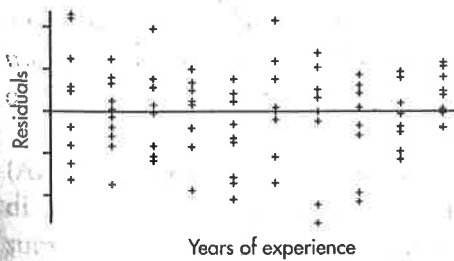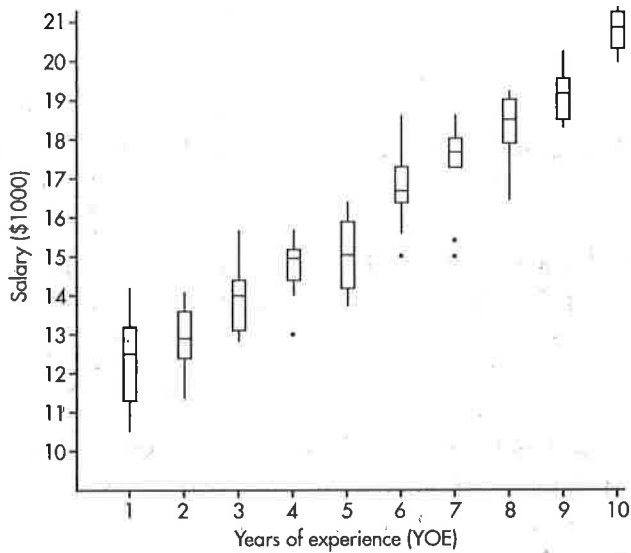

```
Dependent variable is: WAGES
Var          Coef      s.e. Coef      t          p
Constant     11.1113     0.2031      54.7     ≤ 0.0001
YOE          0.910485    0.03273     27.8     ≤ 0.0001
R-sq = 88.8%   R-sq(adj) = 88.6%
s = 0.9402 with 100 - 2 = 98 degrees of freedom
```

(a) Discuss the relationship between salary and experience based on the boxplots.

(b) Discuss how conditions for regression inference are met.

(c) Determine a 95% confidence interval for the regression slope, and interpret in context.

(d) Using only the given information, give a rough estimate of the probability that a salary is at least $1000 over what is predicted by the regression line.

STOP