

**STATISTICS
SECTION I****Time—1 hour and 30 minutes****Number of questions—40****Percent of total grade—50**

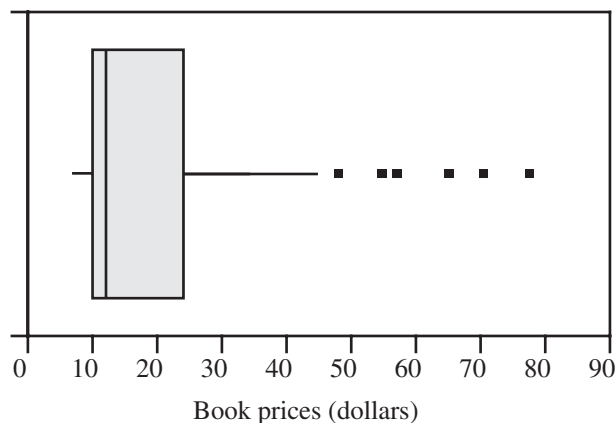
Directions: Solve each of the following problems, using the available space for scratchwork. Decide which is the best of the choices given and fill in the corresponding oval on the answer sheet. No credit will be given for anything written in the test book. Do not spend too much time on any one problem.

1. Fifty oranges of one variety were selected at random and weighed. A 95 percent confidence interval computed from the sample weights to estimate the mean weight of oranges of this variety is (7.58, 8.72) ounces. Which of the following statements is true with respect to the confidence level used?
 - (A) In repeated samplings of the same size, 95 percent of the intervals constructed will contain the true mean weight of the oranges of this variety.
 - (B) In repeated samplings of the same size, 95 percent of the intervals constructed will contain the sampled mean weight of the oranges.
 - (C) Ninety-five percent of the oranges in this sample weigh between 7.58 and 8.72 ounces.
 - (D) Ninety-five percent of the oranges of this variety weigh between 7.58 and 8.72 ounces.
 - (E) There is a 95 percent chance that the mean weight of another sample of 50 oranges from this variety will be between 7.58 and 8.72 ounces.
-
2. In the jury pool available for this week, 30 percent of potential jurors are women. A particular trial requires that, out of a jury of 12, at least three are women. If a jury of 12 is to be selected at random from the pool, what is the probability it meets the requirements of this trial?
 - (A) 0.168
 - (B) 0.843
 - (C) 0.915
 - (D) 0.949
 - (E) The answer cannot be determined without knowing the size of the jury pool.

3. The mean daily demand for bread at a popular bakery is 2,500 loaves, with a standard deviation of 225 loaves. Every morning the bakery bakes 3,000 loaves. What is the probability that today it will run out of bread? Assume that the mean daily demand for bread at this bakery is normally distributed.
- (A) 0.8333
 - (B) 0.1667
 - (C) 0.9869
 - (D) 0.0132
 - (E) 0.0900
-
4. A medicine is known to produce side effects in one in five patients taking it. Suppose a doctor prescribes the medicine to four unrelated patients. What is the probability that none of the patients will develop side effects?
- (A) 0.8000
 - (B) 0.4096
 - (C) 0.2500
 - (D) 0.2000
 - (E) 0.0016

5. An automobile service station performs oil changes and tire replacements, as well as other services. Sixty percent of its customers request an oil change, 30 percent request tire replacements, and 10 percent request both. A customer requests an oil change. What is the probability this customer does not request tire replacements?
- (A) 0.420
(B) 0.500
(C) 0.700
(D) 0.833
(E) 0.857
-

6. The following boxplot summarizes the prices of books at a bookstore.

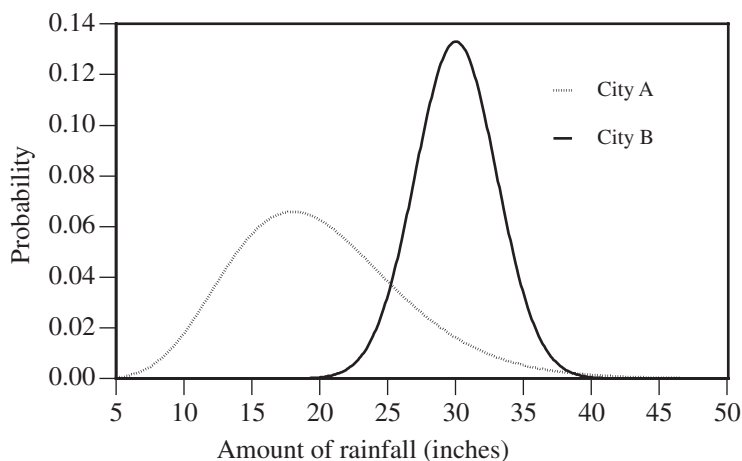


Which of the following is true about the prices of books at this store?

- (A) This store carries more books priced above the mean than below the mean.
(B) This store carries more books priced below the mean than above the mean.
(C) This store carries about the same number of books at different prices in the entire price range.
(D) The mean price of books at this store is the same as the median price of books.
(E) The mean price of books at this store is lower than the median price of books.

7. Random variable X is normally distributed, with a mean of 25 and a standard deviation of 4. Which of the following is the approximate interquartile range for this distribution?
- (A) $25.00 - 22.30 = 2.70$
 - (B) $27.70 - 22.30 = 5.40$
 - (C) $27.70 \div 22.30 = 1.24$
 - (D) $2.00(4.00) = 8.00$
 - (E) $37.00 - 13.00 = 24.00$
-
8. An experiment was designed to test the effects of three different types of paint on the durability of wooden toys. Because boys and girls tend to play differently with toys, a randomly selected group of children was divided into two groups by sex. Which of the following statements about this experiment is true?
- (A) There are three types of paint and two sex groups, giving a total of six treatment combinations in this experiment.
 - (B) Type of paint is a blocking factor.
 - (C) Sex is a blocking factor.
 - (D) This is a completely randomized design.
 - (E) This is a matched-pairs design in which one boy and one girl are matched by age to form a pair.

9. The following graph summarizes data collected on annual rainfall in two cities for the past 150 years.



Which of the following conclusions can be made from this graph?

- (A) The cities have different mean annual rainfalls, but the range of their annual rainfalls is approximately the same.
 - (B) On average, City B gets more rain than City A, but it has a smaller range of annual rainfall.
 - (C) On average, City B gets more rain than City A, and it has a larger range of annual rainfall.
 - (D) On average, City A gets more rain than City B, but it has a smaller range of annual rainfall.
 - (E) On average, City A gets more rain than City B, and it has a larger range of annual rainfall.
-
10. According to the central limit theorem, regardless of the population distribution, the sample mean \bar{X} is approximately normally distributed with $\mu_{\bar{X}} = \mu_X$ and $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$,
- (A) provided that a large number of random samples of the same size are selected from the population
 - (B) provided that a large random sample of size n is taken from the population
 - (C) provided that the population median is equal to the population mean
 - (D) provided that the sample size does not exceed 30
 - (E) It is always true. No other condition is required.

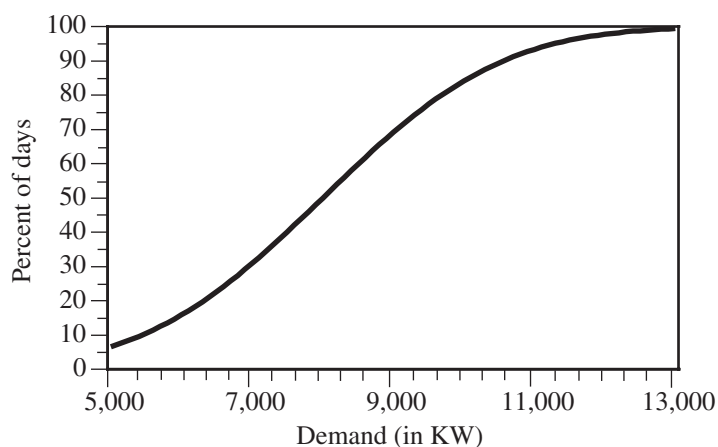
11. Which of the following is a discrete random variable?
- (A) The number of times a student guesses the answers to questions on a certain test
 - (B) The amount of gasoline purchased by a customer
 - (C) The amount of mercury found in fish caught in the Gulf of Mexico
 - (D) The height of water-oak trees
 - (E) The time elapsed until the first field goal at home football games
-
12. A random sample of 300 shoppers was selected to estimate the proportion of customers satisfied with the floor displays of merchandise throughout the store. What is the maximum margin of error if a 90 percent confidence interval is to be constructed?
- (A) 0.0033
 - (B) 0.0475
 - (C) 0.0566
 - (D) 0.0949
 - (E) 0.1132

13. Which of the following statements correctly describes the relation between a t -distribution and a standard normal distribution?
- (A) The standard normal distribution is centered at zero, whereas the t -distribution is centered at $(n - 1)$.
 - (B) As the sample size increases, the difference between the t -distribution and the standard normal distribution increases.
 - (C) The standard normal is just another name for the t -distribution.
 - (D) The standard normal distribution has a larger standard deviation than the t -distribution.
 - (E) The t -distribution has a larger standard deviation than the standard normal distribution.
-
14. Researchers are concerned that the mean clutch size (number of eggs per nest) of robin's eggs in a particular forest has decreased from the mean of 4.8 eggs measured 10 years ago. They plan to send out teams to locate and count the number of eggs in 50 different nests, and then perform a hypothesis test to determine if their concerns are well-founded. What is the most appropriate test to perform in this situation?
- (A) A one-tailed t -test on one population mean.
 - (B) A one-tailed t -test on the difference between two independent population means.
 - (C) A one-tailed t -test on the mean difference between ordered pairs.
 - (D) A one-tailed z -test on a population proportion.
 - (E) A chi-squared test on independence between two variables.

15. Scores on the take-home part of a test are approximately normally distributed, with a mean of 40 and a standard deviation of 3. Scores on the in-class part of the same test are also approximately normally distributed, with a mean of 34 and a standard deviation of 6. The final score on the test is the sum of the score on the take-home part and the score on the in-class part. What are the mean and standard deviation of the final test scores?
- (A) Mean 74 and standard deviation 4.50
 - (B) Mean 37 and standard deviation 6.71
 - (C) Mean 37 and standard deviation 4.50
 - (D) Mean 74 and standard deviation 6.71
 - (E) Mean 74 and standard deviation 3.00
-

16. On a boxplot, an observation is classified as an outlier if
- (A) it is smaller than $\bar{X} - 3S$ or larger than $\bar{X} + 3S$
 - (B) it is larger than $\bar{X} - 3S$ and smaller than $\bar{X} + 3S$
 - (C) it is smaller than Q_1 or larger than Q_3
 - (D) it is smaller than $Q_1 - 1.5IQR$ or larger than $Q_3 + 1.5IQR$
 - (E) it is larger than $Q_1 - 1.5IQR$ and smaller than $Q_3 + 1.5IQR$

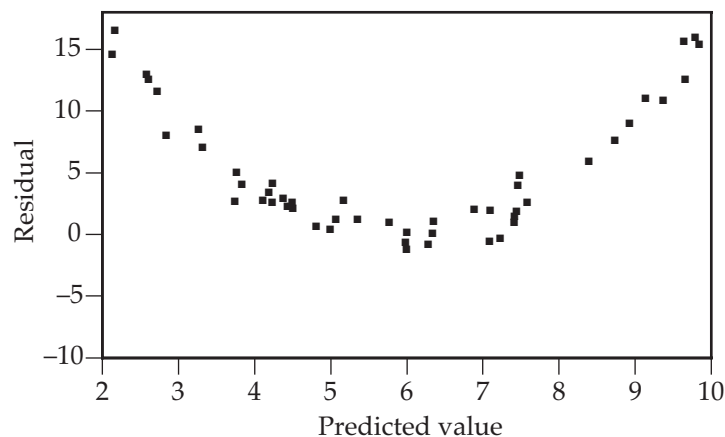
17. The following cumulative graph gives the electricity demand of a certain town.



The power plant that provides electricity to this town is capable of generating 12,000 kW daily. On what percent of days will the power plant not be able to meet the demand for electricity?

- (A) About 3 percent
- (B) About 60 percent
- (C) About 80 percent
- (D) About 92 percent
- (E) About 97 percent

18. Sixty pairs of measurements were taken at random to estimate the relation between variables X and Y . A least squares regression line was fitted to the collected data. The resulting residual plot is as follows:



Which of the following conclusions is appropriate?

- (A) A line is an appropriate model to describe the relation between X and Y .
- (B) A line is not an appropriate model to describe the relation between X and Y .
- (C) The assumption of normality of errors has been violated.
- (D) The assumption of constant sample standard deviations has been violated.
- (E) The variables X and Y are not related at all.

19. A consumer awareness group has received several complaints that the price of an asthma medicine has significantly increased in recent years. Two years ago, the mean price for this medicine was estimated to be \$78.00, and the price at different pharmacies was approximately normally distributed. The group decides to select 10 pharmacies at random and record the price of the medicine at each of the pharmacies. Assuming a 5 percent level of significance, which of the following decision rules should be used to test the hypotheses below?

H_0 : The mean price of medicine is \$78.00.

H_a : The mean price of medicine is higher than \$78.00.

- (A) Reject the null hypothesis if p -value > 0.05 .
(B) Reject the null hypothesis if p -value > 0.025 .
(C) Reject the null hypothesis if $\bar{X} > 78.00$.
(D) Reject the null hypothesis if the test statistic > 1.812 .
(E) Reject the null hypothesis if the test statistic > 1.833 .
-
20. A team of engineers at the research center of a car manufacturer performs crash tests to determine the proportion of times the cars' airbags fail to operate in a crash. With the airbag system's new modified design, the team expected to reduce the failed proportion to be below last year's proportion of 0.08. They decided to test $H_0: p = 0.08$ versus $H_a: p < 0.08$, where p = the proportion of failed airbags during crash tests. If 300 crashes performed in the lab resulted in 18 failures, which of the following is the test statistic for this test?

(A) $z = \frac{0.06 - 0.08}{\sqrt{\frac{0.08(1 - 0.08)}{300}}}$

(B) $z = \frac{0.06 - 0.08}{\sqrt{\frac{0.06(1 - 0.06)}{300}}}$

(C) $z = \frac{0.06 - 0.08}{\sqrt{300(0.08)(1 - 0.08)}}$

(D) $z = \frac{0.06 - 0.08}{\sqrt{300(0.06)(1 - 0.06)}}$

(E) $z = \frac{0.06 - 0.08}{\sqrt{\frac{0.06(1 - 0.08)}{300}}}$

21. A farmer wants to know whether a new fertilizer has increased the mean weight of his apples. With the old fertilizer, the mean weight was 4.0 ounces per apple. The farmer decides to test $H_0: \mu = 4.0$ ounces versus $H_a: \mu > 4.0$ ounces, at a 5 percent level of significance, where μ = the mean weight of apples using the new fertilizer. The weights of apples are approximately normally distributed. The farmer takes a random sample of 16 apples and computes a mean of 4.3 ounces and a standard deviation of 0.6 ounces. Which of the following gives the p -value for this test?
- (A) $P(Z > 2)$
 - (B) $P(Z < 2)$
 - (C) $P(t > 2)$ with 15 degrees of freedom
 - (D) $P(t < 2)$ with 15 degrees of freedom
 - (E) $P(t > 2)$ with 16 degrees of freedom
-
22. Tina's science fair project was to estimate the mean amount of chemicals in her city's water supply. At first, she had decided to use a random sample of 15 observations. But her teacher asked her to take 35 observations. The mean and standard deviation from 35 observations turned out to be approximately the same as those from 15 observations. Is there any advantage in using 35 observations instead of 15 observations?
- (A) There is no advantage. Because the mean and the standard deviation are about the same, the confidence interval computed using 35 observations should be approximately the same as that computed using 15 observations.
 - (B) There is no advantage. In fact, the 20 extra observations will increase the likelihood of error.
 - (C) There is no advantage. Because she took 35 observations instead of 15 observations, the confidence interval using 35 observations will be wider than that using 15 observations.
 - (D) There is some advantage. Because she took 35 observations instead of 15 observations, the confidence interval using 35 observations will be narrower than that using 15 observations.
 - (E) There is some advantage. With 35 observations, she will be able to compute an exact z -confidence interval instead of an approximate t -confidence interval.

23. The Hardcore Construction Company has two offices, one in Atlanta and one in New Orleans. Fifteen engineers work in the Atlanta office, and 14 engineers work in the New Orleans office. The business manager decided to use a 2-sample t -test to compare the mean salaries of engineers in the two offices. Because there were only 15 engineers in one office and 14 engineers in the other, he used the salaries of all the engineers in the computation. Is the 2-sample t -test an appropriate inferential technique in this situation?
- (A) Yes, because he is comparing the means of two small groups.
 - (B) Yes. Both Atlanta and New Orleans are large cities, so the salaries are comparable.
 - (C) Yes. Because Atlanta and New Orleans are about 500 miles apart, the two groups of engineers can be assumed to be independent.
 - (D) No, because the number of engineers in the two offices is different.
 - (E) No, because the entire population information was used from both offices. Because no samples were taken, a t -test should not be used.
-
24. The distribution of salaries of a county school system with 4,752 employees is known to be right skewed, with the superintendent's salary an outlier on the higher side. A random sample of 20 employees was taken and their salaries recorded. A 95 percent t -confidence interval for the mean salary of the county school system employees is (\$15,360, \$32,470). The t -confidence interval is not appropriate in this situation because
- (A) the sample size is too small compared to the number of employees in the system to give a fair representation
 - (B) the skewed salary distribution tells us that assumption of normality of the sampled population will not be satisfied
 - (C) the population standard deviation of the salaries of employees of this county school system is not known
 - (D) the teacher salaries are not negotiable, but the superintendent's salary is negotiable, so the superintendent's salary should be excluded from the population sampled
 - (E) the salaries depend on the number of years of experience, which is not taken into account here

25. A random sample of families was taken in order to estimate the relation between fertility and level of education (measured in number of years). A confidence interval needs to be constructed for the slope of the regression line. The social worker in charge of the project is debating whether to use a 90 percent or a 95 percent confidence interval. Which of the following statements about the length of these intervals is true?
- (A) The 95 percent confidence interval will be wider than the 90 percent confidence interval.
 - (B) The 95 percent confidence interval will be narrower than the 90 percent confidence interval.
 - (C) Both intervals will be of the same length, because they are computed from the same sample.
 - (D) The length of the confidence interval will depend on the sample size, not on the confidence level.
 - (E) The length of the confidence interval will depend on the sample standard deviation, not on the confidence level.
-
26. An insurance agent is successful in selling a life insurance policy to 20 percent of the customers he contacts. He decides to construct a simulation to estimate the mean number of customers he needs to contact before being able to sell a policy. Which of the following schemes should he use to do the simulation?
- (A) Assign numbers 0, 1 to successfully selling a policy to a customer and numbers 2, 3, 4, 5, 6, 7, 8, 9 to failing to sell a policy to a customer.
 - (B) Assign numbers 0, 1 to successfully selling a policy to a customer and numbers 2, 3, 4 to failing to sell a policy to a customer.
 - (C) Assign number 0 to successfully selling a policy to a customer and number 1 to failing to sell a policy to a customer.
 - (D) Assign numbers 0, 1, 2, 3, 4 to successfully selling a policy to a customer and numbers 5, 6, 7, 8, 9 to failing to sell a policy to a customer.
 - (E) Assign number 20 to successfully selling a policy to a customer and numbers 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 to failing to sell a policy to a customer.

27. The mean height of adult men is 70 inches, with a standard deviation of four inches. The mean height of adult women is 66 inches, with a standard deviation of three inches. Between a man with a height of 74 inches and a woman with a height of 70 inches, who is more unusually tall within his or her respective sex?
- (A) It cannot be determined, because the mean heights for the two groups are different.
 - (B) It cannot be determined, because the standard deviations of the heights for the two groups are different.
 - (C) The man, because he is 74 inches tall and the woman is only 70 inches tall.
 - (D) Both should be considered equally tall, because both of them are four inches taller than the average height of their respective sex groups.
 - (E) The woman, because her height is 1.33 standard deviations above the mean height of all women, whereas the man's height is only one standard deviation above the mean height of all men.
-
28. The amount of rainfall per month in a certain city is approximately normally distributed, with a mean of six inches and a standard deviation of 1.6 inches. Which of the following is the highest amount of rainfall, in inches, this city could have this month for the month to be among the 10 percent driest months the city has seen?
- (A) 8.05
 - (B) 7.60
 - (C) 3.95
 - (D) 3.37
 - (E) 2.28

29. A large city was interested in annexing part of the surrounding county. In a survey conducted by the local newspaper, 58 percent of respondents said they were against the annexation. During the actual vote, not all eligible voters voted, but 56 percent of the respondents voted against the annexation. Which of the following best describes the difference in the percentages obtained from the newspaper poll and the vote itself?
- (A) It is an example of nonresponse bias, the systematic tendency of individuals with particular characteristics to refuse to answer a survey question.
 - (B) It is the systematic difference between a statistic and parameter caused by the nonrandom selection of surveyed persons.
 - (C) It is the difference between the same statistics computed from two different samples.
 - (D) It is the difference between the statistic and the truth due to use of a random sample.
 - (E) It was caused by the large number of newspaper subscribers in the county.
-
30. In a clinical trial, 30 sickle cell anemia patients are randomly assigned to two groups. One group receives the currently marketed medicine, and the other group receives an experimental medicine. Each week, patients report to the clinic where blood tests are conducted. The lab technician is unaware of the kind of medicine the patient is taking. This design can be described as
- (A) a completely randomized design, with the currently marketed medicine and the experimental medicine as two treatments
 - (B) a matched-pairs design, with the currently marketed medicine and the experimental medicine forming a pair
 - (C) a randomized block design, with the currently marketed medicine and the experimental medicine as two blocks
 - (D) a randomized block design, with the currently marketed medicine and the experimental medicine as two treatments
 - (E) a stratified design with two strata, patients with sickle cell disease forming one stratum and those without sickle cell disease forming the other stratum

31. The registrar's office at a university has noticed that a large number of students fail to report a change of address. The registrar decides to take a random sample of 150 students from the current directory of students and determine the number of students with the correct addresses on record. He then uses this information to construct a 95 percent confidence interval. Which of the following statements must be true?
- (A) The true proportion of students at this university with the correct address on record is within the confidence interval constructed.
 - (B) Ninety-five percent of students at this university have the correct address on record.
 - (C) The true proportion of students at this university with the correct addresses on record is within the confidence interval 95 percent of the time.
 - (D) The sample proportion of students at this university with the correct address on record is within the confidence interval 95 percent of the time.
 - (E) The sample proportion of students at this university with the correct address on record is within the confidence interval.
-
32. In which of the following situations is a binomial model not an appropriate model to describe the outcome?
- (A) The number of heads in three tosses of a coin
 - (B) The number of rainy days in a given week
 - (C) The number of girls in a family of five children
 - (D) The number of students present in a class of 22
 - (E) The number of defective computer monitors out of seven purchased

33. Which of the following statements about any two events A and B is true?
- (A) $P(A \cup B)$ implies events A and B are independent.
 - (B) $P(A \cup B) = 1$ implies events A and B are mutually exclusive.
 - (C) $P(A \cap B) = 0$ implies events A and B are independent.
 - (D) $P(A \cap B) = 0$ implies events A and B are mutually exclusive.
 - (E) $P(A \cap B) = P(A) - P(B)$ implies A and B are equally likely events.
-
34. A newspaper reporter examined police reports of accidents during the past 12 months to collect data about the speed of a car and its stopping distance. The reporter then constructed a scatterplot and computed a correlation coefficient to show the relation between a car's speed and its stopping distance. This is an example of
- (A) a double-blind study
 - (B) a single-blind study
 - (C) a study involving no blinding at all
 - (D) an observational study
 - (E) a well-designed experiment

35. A company with offices in five different countries is interested in estimating the proportion of its employees in favor of banning smoking on the office premises. It is known that the views of people from different countries on issues like smoking tend to vary due to the influence of different local social structures. Which of the following is an appropriate sampling technique to use in this situation, and why?
- (A) Stratified sample, because the population is divided into five strata—namely, five offices in five different countries
 - (B) Cluster sample, because the population is divided into five clusters—namely, five offices in five different countries
 - (C) Simple random sample, because this is the simplest sampling method
 - (D) Simple random sample, because this is the only method that gives unbiased results
 - (E) Simple random sample, because this is the only method for which inferential techniques are available
-
36. In a clinic, 50 patients with sleep disorders are randomly assigned to one of two different groups. Patients in one group are given medication before bedtime. Patients in the other group are given blindfolds and played soft music at bedtime. Each patient is attached to a machine that records breathing patterns. From the patterns, it is possible to determine if the patient is awake or asleep. The data will be used to decide which method is more effective in helping patients with sleep disorders. Which of the following statements is correct in the context of this experiment?
- (A) It is not possible for this experiment to be single-blind or double-blind, because human subjects are involved.
 - (B) This is only a single-blind experiment because the experimenters did not include a control group that received neither medication nor a blindfold and soft music.
 - (C) This is only a single-blind experiment because patients will know which treatment they are receiving, although the examining doctor might not.
 - (D) This is a double-blind experiment because the experimenters could prevent both groups of patients from knowing what sort of treatment the other group is receiving.
 - (E) This is a double-blind experiment because the patients were divided into the two treatment groups randomly.

37. Data were collected on two variables X and Y and a least squares regression line was fitted to the data. The estimated equation for this data is $y = -2.29 + 1.70x$. One point has $x = 5$, $y = 6$. What is the residual for this point?
- (A) 7.91
 - (B) 6.21
 - (C) 0.21
 - (D) -0.21
 - (E) -2.91
-

38. The relation between studying time (in hours) and grade on a final exam (0–100) in a random sample of students in math class was found to be

$$\text{Grade} = 50.25 + 10.2(h)$$

How will a student's grade be affected if she studies for two hours?

- (A) Her grade will go up by 40.05 points.
- (B) Her grade will go up by 20.4 points.
- (C) Her grade will remain unchanged.
- (D) Her grade will go down by 10.2 points.
- (E) It cannot be determined from the information given.

39. A group of scientists wanted to estimate the proportion of geese returning to the same site for the next breeding season. Suppose they decided to increase the sample size from 200 to 2,000. How will this affect the distribution of the sample proportion?
- (A) The distribution of the sample proportion will be more spread out.
 - (B) The distribution of the sample proportion will be less spread out.
 - (C) The spread of the distribution of the sample proportion will remain unaffected.
 - (D) The distribution of the sample proportion will more closely resemble the binomial distribution.
 - (E) The distribution of the sample proportion will more closely resemble the chi-square distribution.
-

40. A company has 400 employees. Their mean income is \$20,500, and the standard deviation of their incomes is \$3,750. The distribution of incomes is normally distributed. How many of the 400 employees do you expect to have an income of between \$13,000 and \$28,000?
- (A) 50
 - (B) 100
 - (C) 200
 - (D) 390
 - (E) 400

STATISTICS**SECTION II****Time—1 hour and 30 minutes****Number of questions—6****Percent of total grade—50****Part A****Questions 1–5****Spend about 60 minutes on this part of the exam.****Percent of Section II grade—75**

Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your method as well as on the accuracy of your results and explanation.

1. A short-term parking facility allows cars to be parked for a maximum of four hours. For any fraction of an hour a car is parked, the facility charges for the full hour. For example, if a person parks at this facility for one hour and 25 minutes, then the person is charged for two full hours. The probability distribution of the parking time at this short-term parking facility, as measured in full hours, is as follows:

Hours	1	2	3	4
Probability	0.15	0.25	0.45	0.15

- (a) What is the mean parking time charged for in this facility?

Section II

- (b) The facility charges \$3.00 for the first hour of parking and \$2.50 per hour thereafter. Let X be the parking fee charged for a car parked in this facility. In the table below, give the probability distribution function for X .

x	$P(X = x)$

- (c) Determine the mean and standard deviation of the parking fee charged for a car parked at this facility.
- (d) The parking facility offers a monthly parking permit, which allows a car to park without paying the hourly parking fee for an entire month. This permit costs \$100. The driver of a cab parks his cab for one hour or less in the facility 80 percent of times he parks there, and he never parks in the facility for over two hours. What is the minimum number of times in a month he would need to park in this facility in order for the monthly parking permit to cost less than his expected cost for paying the hourly rate?

2. Scientists have suspected that animals, when deficient in certain chemicals, tend to ingest natural resources that have high concentrations of those chemicals to offset the deficiency. In a study, scientists drained saliva from the parotid gland of sheep in order to make them sodium deficient. Then they offered these sheep a solution of sodium bicarbonate and measured the sheep's sodium intake. The sodium deficiency and the sodium intake, both measured in millimoles, are recorded as follows:

Sodium Deficit (in millimoles)	Sodium Intake (in millimoles)
100	110
200	180
570	610
850	790
700	750
425	390
375	420
325	380
450	300
850	790

The summary statistics and the regression output for this data are as follows:

Variable	N	Mean	Median	StDev	Q1	Q3
Deficit	10	484.5	437.5	256.4	293.7	737.5
Intake	10	472.0	405.0	250.0	270.0	760.0

Predictor	Coef	StDev	T	P
Constant	15.55	47.94	0.32	0.754
Deficit	0.94211	0.08843	10.65	0.000

Section II

(a) Does a line appear to be a reasonable model for this data? Explain your answer.

(b) State and interpret the slope in terms of the problem.

GO ON TO THE NEXT PAGE.

(c) Estimate the correlation between sodium deficit and sodium intake. Interpret your answer in the context of this problem.

(d) Estimate the amount of a sheep's sodium intake if the sheep is found to be deficient in sodium by 800 millimoles.

3. A university offers degrees in the following four areas:

- Arts and Sciences (A&S)
- Engineering (Eng)
- Business and Management (B&M)
- Computer Science (Comp)

Incoming freshmen apply for majors in A&S, Eng, B&M, and Comp in a ratio of approximately 5:6:6:8, but many students change their major during the course of study. The university officials are interested in determining whether the distribution of majors among graduating students differs from that of incoming students. They took a random sample of 200 graduating students and classified them according to their majors. The distribution of graduating students in this sample is given in the table below.

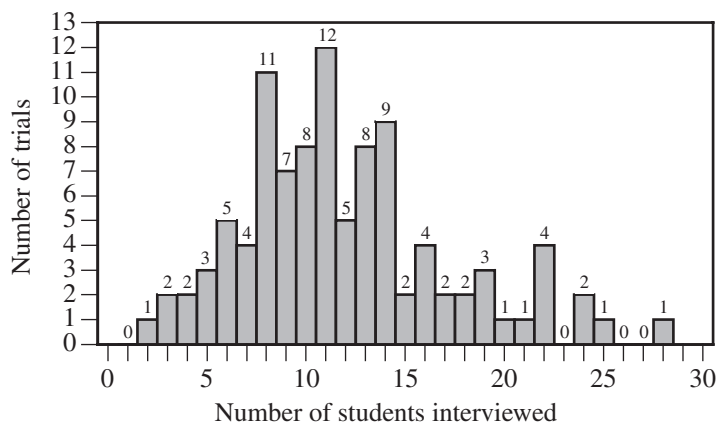
	Major			
	A&S	Eng	B&M	Comp
Number of Students	60	30	60	50

Is the distribution of majors among graduating students different from that of incoming students? Provide statistical justification for your answer.

4. An education professor is interested in getting students involved in a tutoring program for area schools. The participating students are paid through a grant from the city. This year, the city has agreed to provide funds to support three students in this program. The professor has found that about four out of five students he interviews are not willing to participate in the program due to other commitments.
- (a) Describe how you would use a random number table to carry out simulations to determine the number of students the professor needs to interview to fill the three spots in the tutoring program. Describe what each random number represents in your simulation.
- (b) Run two trials of your simulation. Use the random number table provided to illustrate your simulation.

25211	75049	70678	24646	96329	63547	37255	51013	25211	75049
97077	82384	33078	59574	34916	09422	85700	74202	97077	82384
82641	66179	30341	40674	51778	97680	84707	88808	82641	66179
60675	60254	16308	70130	29610	27658	94288	88752	60675	60254
53860	97861	34625	85190	38477	60503	34561	04360	53860	97861

- (c) Suppose the professor did 100 simulations to estimate the number of students that he needs to interview to find three students willing to work for the tutoring program. The results of the simulations are shown below.



Estimate the probability that the professor has to interview more than 20 students before he finds three willing to work for the tutoring program.

5. A resident at a medical college was interested in showing that exercising regularly helps reduce cholesterol levels. Over a period of one month, he selected a random sample of 50 patients who came to his clinic for their annual physical examination. He asked the patients if they exercised regularly, and noted down the answers as “Exercises regularly” or “Does not exercise regularly.” He also recorded their current cholesterol levels and their cholesterol levels from the previous year’s physical examination, as noted in their medical records. He summarized the data on their cholesterol levels as shown in the table below.

	Exercises Regularly		Does Not Exercise Regularly	
	Last Year	This Year	Last Year	This Year
Number of Patients	18	18	32	32
Mean	223.80	189.75	268.35	265.53
Standard Deviation	15.88	18.77	32.81	28.17
Difference in Mean Cholesterol Levels (Last Year – This Year)	34.05		2.82	

- (a) The resident included the following in his report:

“On the average, the group that exercises regularly has experienced a larger decrease in cholesterol level as compared to the group that does not exercise regularly. Because the observed difference is considerably large, we can conclude that exercising regularly reduces cholesterol level.”

Why is such a causal conclusion not warranted by this data? Explain your answer.

Section II

- (b) Design an experiment that will allow the resident to draw such a conclusion.

END OF PART A

**STATISTICS
SECTION II****Part B****Question 6**

Spend about 30 minutes on this part of the exam.

Percent of Section II grade—25

6. Pam and Dave decided to buy a small house and move out of their apartment. After looking around, they decided that they both liked two neighborhoods, Sunshine Estates and Pinewoods Estates, and they wanted to buy a house in one of those two neighborhoods if they could afford it. Their real estate agent selected a random sample of houses sold in the last 12 months from each neighborhood and found out how much the houses sold for. The summary statistics for the selling prices (in \$1,000) of houses in both neighborhoods are as follows:

	Sunshine Estates	Pinewoods Estates
Sample Size	14	15
Mean Selling Price	113.357	124.333
Standard Deviation of Selling Price	3.296	14.181
Minimum Selling Price	107.000	97.000
1st Quartile Selling Price	111.000	112.000
Median Selling Price	113.500	121.000
3rd Quartile Selling Price	116.000	133.000
Maximum Selling Price	119.000	149.000

- (a) Make parallel boxplots to compare the selling prices of houses in Sunshine Estates and Pinewoods Estates. Write a few sentences describing the selling prices in these two neighborhoods.

Section II

- (b) Assume the selling prices of houses in these two neighborhoods are normally distributed. Construct a 95 percent confidence interval to compare the mean prices in the two neighborhoods. Interpret your interval.
- (c) Pam and Dave decide that they could afford to spend \$120,000 on their new house. What percent of houses in Sunshine Estates are within their budget?

- (d) In which neighborhood are they more likely to find a house within their budget? Justify your answer.
- (e) Using the information available, can we support the assumption that the prices of the houses are distributed normally? Explain your answer.

STOP

END OF EXAM
