

Practice Examination 3

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.

- Which of the following is a true statement?
 - While properly designed experiments can strongly suggest cause-and-effect relationships, a complete census is the only way of establishing such a relationship.
 - If properly designed, observational studies can establish cause-and-effect relationships just as strongly as properly designed experiments.
 - Controlled experiments are often undertaken later to establish cause-and-effect relationships first suggested by observational studies.
 - A useful approach to overcome bias in observational studies is to increase the sample size.
 - In an experiment, the control group is a self-selected group who choose not to receive a designated treatment.
- Two classes take the same exam. Suppose a certain score is at the 40th percentile for the first class and at the 80th percentile for the second class. Which of the following is the most reasonable conclusion?
 - Students in the first class generally scored higher than students in the second class.
 - Students in the second class generally scored higher than students in the first class.
 - A score at the 20th percentile for the first class is at the 40th percentile for the second class.
 - A score at the 50th percentile for the first class is at the 90th percentile for the second class.
 - One of the classes has twice the number of students as the other.

GO ON TO THE NEXT PAGE ►

3. In an experiment, the control group should receive
- treatment opposite that given the experimental group.
 - the same treatment given the experimental group without knowing they are receiving the treatment.
 - a procedure identical to that given the experimental group except for receiving the treatment under examination.
 - a procedure identical to that given the experimental group except for a random decision on receiving the treatment under examination.
 - none of the procedures given the experimental group.
4. In a random sample of Toyota car owners, 83 out of 112 said they were satisfied with the Toyota front-wheel drive, while in a similar survey of Subaru owners, 76 out of 81 said they were satisfied with the Subaru four-wheel drive. A 90% confidence interval estimate for the difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems is reported to be $-.197 \pm .081$. Which is a proper conclusion?
- The interval is invalid because probabilities cannot be negative.
 - The interval is invalid because it does not contain zero.
 - Subaru owners are approximately 19.7% more satisfied with their drive systems than are Toyota owners.
 - 90% of Subaru owners are approximately 19.7% more satisfied with their drive systems than are Toyota owners.
 - We are 90% confident that the difference in proportions between Toyota and Subaru car owners who are satisfied with their drive systems is between $-.278$ and $-.116$.
5. In a study on the effect of music on worker productivity, employees were told that a different genre of background music would be played each day and the corresponding production outputs noted. Every change in music resulted in an increase in production. This is an example of
- the effect of a treatment unit.
 - the placebo effect.
 - the control group effect.
 - sampling error.
 - voluntary response bias.
6. A computer manufacturer sets up three locations to provide technical support for its customers. Logs are kept noting whether or not calls about problems are solved successfully. Data from a sample of 1000 calls are summarized in the following table:

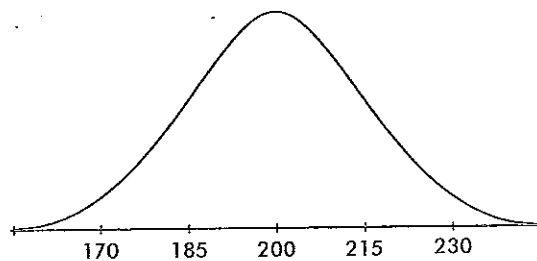
	Location			
	1	2	3	Total
Problem solved	325	225	150	700
Problem not solved	125	100	75	300
Total	450	325	225	1000

Assuming there is no association between location and whether or not a problem is resolved successfully, what is the expected number of successful calls (problem solved) from location 1?

- $\frac{(325)(450)}{700}$
- $\frac{(325)(700)}{450}$
- $\frac{(325)(450)}{1000}$
- $\frac{(325)(700)}{1000}$
- $\frac{(450)(700)}{1000}$

7. Which of the following statements about the correlation coefficient is true?
- (A) The correlation coefficient and the slope of the regression line may have opposite signs.
- (B) A correlation of 1 indicates a perfect cause-and-effect relationship between the variables.
- (C) Correlations of $+0.87$ and -0.87 indicate the same degree of clustering around the regression line.
- (D) Correlation applies equally well to quantitative and categorical data.
- (E) A correlation of 0 shows little or no association between two variables.
8. Suppose X and Y are random variables with $E(X) = 780$, $\text{var}(X) = 75$, $E(Y) = 430$, and $\text{var}(Y) = 25$. Given that X and Y are independent, what is the variance of the random variable $X - Y$?
- (A) $75 - 25$
- (B) $75 + 25$
- (C) $\sqrt{75 - 25}$
- (D) $\sqrt{75 + 25}$
- (E) $\sqrt{75} - \sqrt{25}$
9. What is a sample?
- (A) A measurable characteristic of a population
- (B) A set of individuals having a characteristic in common
- (C) A value calculated from raw data
- (D) A subset of a population
- (E) None of the above

10. A histogram of the cholesterol levels of all employees at a large law firm is as follows:



Which of the following is the best estimate of the standard deviation of this distribution?

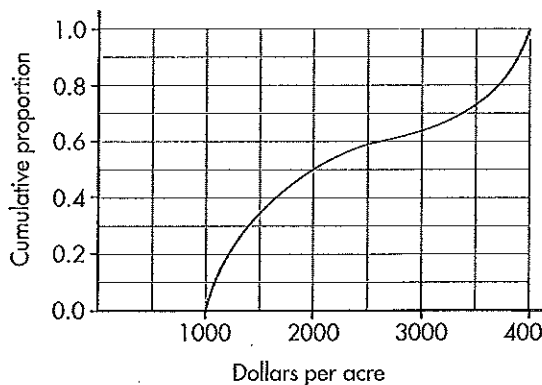
- (A) $\frac{(230-170)}{6} = 10$
- (B) 15
- (C) $\frac{200}{6} = 33.3$
- (D) $230 - 170 = 60$
- (E) $245 - 155 = 90$
11. A soft drink dispenser can be adjusted to deliver any fixed number of ounces. If the machine is operating with a standard deviation in delivery equal to 0.3 ounce, what should be the mean setting so that a 12-ounce cup will overflow less than 1% of the time? Assume a normal distribution for ounces delivered.
- (A) 11.23 ounces
- (B) 11.30 ounces
- (C) 11.70 ounces
- (D) 12.70 ounces
- (E) 12.77 ounces

12. An insurance company wishes to study the number of years drivers in a large city go between automobile accidents. They plan to obtain and analyze the data from a sample of drivers. Which of the following is a true statement?
- (A) A reasonable time-and-cost-saving procedure would be to use systematic sampling on an available list of all AAA (Automobile Association of America) members in the city.
- (B) A reasonable time-and-cost-saving procedure would be to randomly choose families and include all drivers in each of these families in the sample.
- (C) To determine the mean number of years between accidents, randomness in choosing a sample of drivers is not important as long as the sample size is very large.
- (D) The larger a simple random sample, the more likely its standard deviation will be close to the population standard deviation divided by the square root of the sample size.
- (E) None of the above are true statements.
13. The probability that a person will show a certain gene-transmitted trait is .8 if the father shows the trait and .06 if the father doesn't show the trait. Suppose that the children in a certain community come from families in 25% of which the father shows the trait. Given that a child shows the trait, what is the probability that her father shows the trait?
- (A) .245
 (B) .250
 (C) .750
 (D) .816
 (E) .860
14. Given an experiment with $H_0: \mu = 10$, $H_a: \mu > 10$, and a possible correct value of 11, which of the following increases as n increases?
- I. The probability of a Type I error.
 II. The probability of a Type II error.
 III. The power of the test.
- (A) I only
 (B) II only
 (C) III only
 (D) II and III
 (E) None will increase.
15. If all the values of a data set are the same, all of the following must equal zero except for which one?
- (A) Mean
 (B) Standard deviation
 (C) Variance
 (D) Range
 (E) Interquartile range
16. To determine the average number of minutes it takes to manufacture one unit of a new product, an assembly line manager tracks a random sample of 15 units and records the number of minutes it takes to make each unit. The assembly times are assumed to have a normal distribution. If the mean and standard deviation of the sample are 3.92 and 0.45 minutes respectively, which of the following gives a 90% confidence interval for the mean assembly time, in minutes, for units of the new product?
- (A) $3.92 \pm 1.645 \frac{0.45}{\sqrt{14}}$
 (B) $3.92 \pm 1.753 \frac{0.45}{\sqrt{14}}$
 (C) $3.92 \pm 1.761 \frac{0.45}{\sqrt{14}}$
 (D) $3.92 \pm 1.753 \frac{0.45}{\sqrt{15}}$
 (E) $3.92 \pm 1.761 \frac{0.45}{\sqrt{15}}$

17. A company has 1000 employees evenly distributed throughout five assembly plants. A sample of 30 employees is to be chosen as follows. Each of the five managers will be asked to place the 200 time cards of their respective employees in a bag, shake them up, and randomly draw out six names. The six names from each plant will be put together to make up the sample. Will this method result in a simple random sample of the 1000 employees?
- (A) Yes, because every employee has the same chance of being selected.
 (B) Yes, because every plant is equally represented.
 (C) Yes, because this is an example of stratified sampling, which is a special case of simple random sampling.
 (D) No, because the plants are not chosen randomly.
 (E) No, because not every group of 30 employees has the same chance of being selected.
18. Given that $P(E) = .32$, $P(F) = .15$, and $P(E \cap F) = .048$, which of the following is a correct conclusion?
- (A) The events E and F are both independent and mutually exclusive.
 (B) The events E and F are neither independent nor mutually exclusive.
 (C) The events E and F are mutually exclusive but not independent.
 (D) The events E and F are independent but not mutually exclusive.
 (E) The events E and F are independent, but there is insufficient information to determine whether or not they are mutually exclusive.
19. The number of leasable square feet of office space available in a city on any given day has a normal distribution with mean 640,000 square feet and standard deviation 18,000 square feet. What is the interquartile range for this distribution?
- (A) 652,000 – 628,000
 (B) 658,000 – 622,000
 (C) 667,000 – 613,000
 (D) 676,000 – 604,000
 (E) 694,000 – 586,000
20. Consider the following back-to-back stemplot:
- | | | |
|-------|---|-------|
| | 0 | 348 |
| | 1 | 01256 |
| 843 | 2 | 29 |
| 65210 | 3 | 2557 |
| 92 | 4 | |
| 7552 | 5 | 6 |
| | 6 | 1458 |
| 6 | 7 | 09 |
| 8541 | 8 | |
| 90 | 9 | |
- Which of the following is a correct statement?
- (A) The distributions have the same mean.
 (B) The distributions have the same median.
 (C) The interquartile range of the distribution to the left is 20 greater than the interquartile range of the distribution to the right.
 (D) The distributions have the same variance.
 (E) None of the above is correct.

21. Which of the following is a correct statement?
- (A) A study results in a 99% confidence interval estimate of (34.2, 67.3). This means that in about 99% of all samples selected by this method, the sample means will fall between 34.2 and 67.3.
- (B) A high confidence level may be obtained no matter what the sample size.
- (C) The central limit theorem is most useful when drawing samples from normally distributed populations.
- (D) The sampling distribution for a mean has standard deviation $\frac{\sigma}{\sqrt{n}}$ only when n is sufficiently large (typically one uses $n \geq 30$).
- (E) The center of any confidence interval is the population parameter.
22. The binomial distribution is an appropriate model for which of the following?
- (A) The number of minutes in an hour for which the Dow-Jones average is above its beginning average for the day.
- (B) The number of cities among the 10 largest in New York State for which the weather is cloudy for most of a given day.
- (C) The number of drivers wearing seat belts if 10 consecutive drivers are stopped at a police roadblock.
- (D) The number of A's a student receives in his/her five college classes.
- (E) None of the above.
23. Suppose two events, E and F , have nonzero probabilities p and q , respectively. Which of the following is impossible?
- (A) $p + q > 1$
- (B) $p - q < 0$
- (C) $p/q > 1$
- (D) E and F are neither independent nor mutually exclusive.
- (E) E and F are both independent and mutually exclusive.
24. An inspection procedure at a manufacturing plant involves picking four items at random and accepting the whole lot if at least three of the four items are in perfect condition. If in reality 90% of the whole lot are perfect, what is the probability that the lot will be accepted?
- (A) $(.9)^4$
- (B) $1 - (.9)^4$
- (C) $4(.9)^3(.1)$
- (D) $.1 - 4(.9)^3(.1)$
- (E) $4(.9)^3(.1) + (.9)^4$
25. A town has one high school, which buses students from urban, suburban, and rural communities. Which of the following samples is recommended in studying attitudes toward tracking of students in honors, regular, and below-grade classes?
- (A) Convenience sample
- (B) Simple random sample (SRS)
- (C) Stratified sample
- (D) Systematic sample
- (E) Voluntary response sample
26. Suppose there is a correlation of $r = 0.9$ between number of hours per day students study and GPAs. Which of the following is a reasonable conclusion?
- (A) 90% of students who study receive high grades.
- (B) 90% of students who receive high grades study a lot.
- (C) 90% of the variation in GPAs can be explained by variation in number of study hours per day.
- (D) 10% of the variation in GPAs cannot be explained by variation in number of study hours per day.
- (E) 81% of the variation in GPAs can be explained by variation in number of study hours per day.

27. To determine the average number of children living in single-family homes, a researcher picks a simple random sample of 50 such homes. However, even after one follow-up visit the interviewer is unable to make contact with anyone in 8 of these homes. Concerned about nonresponse bias, the researcher picks another simple random sample and instructs the interviewer to keep trying until contact is made with someone in a total of 50 homes. The average number of children is determined to be 1.73. Is this estimate probably too low or too high?
- (A) Too low, because of undercoverage bias.
 (B) Too low, because convenience samples overestimate average results.
 (C) Too high, because of undercoverage bias.
 (D) Too high, because convenience samples overestimate average results.
 (E) Too high, because voluntary response samples overestimate average results.
28. The graph below shows cumulative proportions plotted against land values (in dollars per acre) for farms on sale in a rural community.
29. An experiment is to be conducted to determine whether taking fish oil capsules or garlic capsules has more of an effect on cholesterol levels. In past studies it was noted that daily exercise intensity (low, moderate, high) is associated with cholesterol level, but average sleep length (< 5, 5 – 8, > 8 hours) is not associated with cholesterol level. This experiment should be done
- (A) by blocking on exercise intensity
 (B) by blocking on sleep length
 (C) by blocking on cholesterol level
 (D) by blocking on capsule type
 (E) without blocking
30. A confidence interval estimate is determined from the monthly grocery expenditures in a random sample of n families. Which of the following will result in a smaller margin of error?
- I. A smaller confidence level
 II. A smaller sample standard deviation
 III. A smaller sample size
- (A) II only
 (B) I and II
 (C) I and III
 (D) II and III
 (E) I, II, and III



What is the median land value?

- (A) \$2000
 (B) \$2250
 (C) \$2500
 (D) \$2750
 (E) \$3000

GO ON TO THE NEXT PAGE >

31. A medical research team claims that high vitamin C intake increases endurance. In particular, 1000 milligrams of vitamin C per day for a month should add an average of 4.3 minutes to the length of maximum physical effort that can be tolerated. Army training officers believe the claim is exaggerated and plan a test on an SRS of 400 soldiers in which they will reject the medical team's claim if the sample mean is less than 4.0 minutes. Suppose the standard deviation of added minutes is 3.2. If the true mean increase is only 4.2 minutes, what is the probability that the officers will fail to reject the false claim of 4.3 minutes?

(A) $P\left(z < \frac{4.0 - 4.3}{3.2/\sqrt{400}}\right)$

(B) $P\left(z > \frac{4.0 - 4.3}{3.2/\sqrt{400}}\right)$

(C) $P\left(z < \frac{4.3 - 4.2}{3.2/\sqrt{400}}\right)$

(D) $P\left(z > \frac{4.3 - 4.2}{3.2/\sqrt{400}}\right)$

(E) $P\left(z > \frac{4.0 - 4.2}{3.2/\sqrt{400}}\right)$

32. Consider the two sets $X = \{10, 30, 45, 50, 55, 70, 90\}$ and $Y = \{10, 30, 35, 50, 65, 70, 90\}$. Which of the following is false?

- (A) The sets have identical medians.
 (B) The sets have identical means.
 (C) The sets have identical ranges.
 (C) The sets have identical boxplots.
 (E) None of the above are false.

33. The weight of an aspirin tablet is 300 milligrams according to the bottle label. An FDA investigator weighs a simple random sample of seven tablets, obtains weights of 299, 300, 305, 302, 299, 301, and 303, and runs a hypothesis test of the manufacturer's claim. Which of the following gives the P -value of this test?

- (A) $P(t > 1.54)$ with $df = 6$
 (B) $2P(t > 1.54)$ with $df = 6$
 (C) $P(t > 1.54)$ with $df = 7$
 (D) $2P(t > 1.54)$ with $df = 7$
 (E) $0.5P(t > 1.54)$ with $df = 7$

34. A teacher believes that giving her students a practice quiz every week will motivate them to study harder, leading to a greater overall understanding of the course material. She tries this technique for a year, and everyone in the class achieves a grade of at least C. Is this an experiment or an observational study?

- (A) An experiment, but with no reasonable conclusion possible about cause and effect
 (B) An experiment, thus making cause and effect a reasonable conclusion
 (C) An observational study, because there was no use of a control group
 (D) An observational study, but a poorly designed one because randomization was not used
 (E) An observational study, and thus a reasonable conclusion of association but not of cause and effect

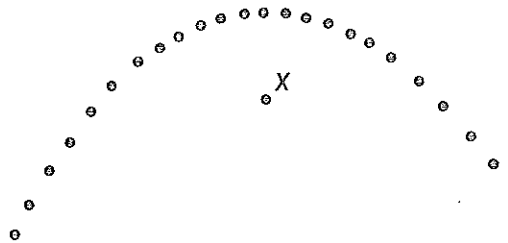
35. Which of the following is *not* true with regard to contingency tables for chi-square tests for independence?

- (A) The categories are not numerical for either variable.
- (B) Observed frequencies should be whole numbers.
- (C) Expected frequencies should be whole numbers.
- (D) Expected frequencies in each cell should be at least 5, and to achieve this, one sometimes combines categories for one or the other or both of the variables.
- (E) The expected frequency for any cell can be found by multiplying the row total by the column total and dividing by the sample size.

36. Which of the following is a correct statement?

- (A) The probability of a Type II error does not depend on the probability of a Type I error.
- (B) In conducting a hypothesis test, it is possible to simultaneously make both a Type I and a Type II error.
- (C) A Type II error will result if one incorrectly assumes the data are normally distributed.
- (D) In medical disease testing with the null hypothesis that the patient is healthy, a Type I error is associated with a *false negative*; that is, the test incorrectly indicates that the patient is disease free.
- (E) When you choose a significance level α , you're setting the probability of a Type I error to exactly α .

37.



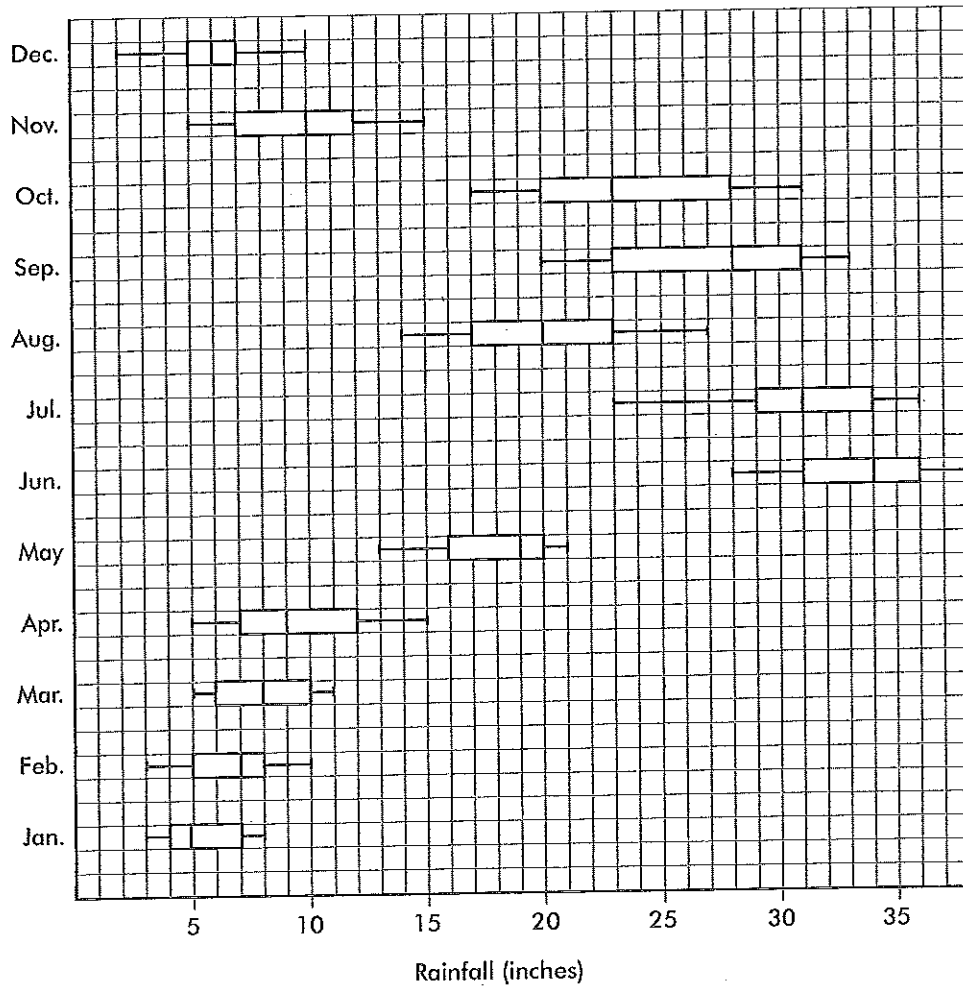
Above is a scatterplot with one point labeled X . Suppose you find the least squares regression line. Which of the following is a correct statement?

- (A) X has the largest residual, in absolute value, of any point on the scatterplot.
- (B) X is an influential point.
- (C) The residual plot will show a curved pattern.
- (D) The association between the x and y variables is very weak.
- (E) If the point X were removed, the correlation would be 1.

38. A banking corporation advertises that 90% of the loan applications it receives are approved within 24 hours. In a random sample of 50 applications, what is the expected number of loan applications that will be turned down?

- (A) $50(.90)$
- (B) $50(.10)$
- (C) $50(.90)(.10)$
- (D) $\sqrt{50(.90)(.10)}$
- (E) $\sqrt{\frac{(.90)(.10)}{50}}$

39. The parallel boxplots below show monthly rainfall summaries for Liberia, West Africa.



Which of the following months has the least variability as measured by *interquartile range*?

- (A) January
- (B) February
- (C) March
- (D) May
- (E) December

40. In comparing the life expectancies of two models of refrigerators, the average years before complete breakdown of 10 model A refrigerators is compared with that of 15 model B refrigerators. The 90% confidence interval estimate of the difference is (6, 12). Which of the following is the most reasonable conclusion?
- (A) The mean life expectancy of one model is twice that of the other.
 - (B) The mean life expectancy of one model is 6 years, while the mean life expectancy of the other is 12 years.
 - (C) The probability that the life expectancies are different is .90.
 - (D) The probability that the difference in life expectancies is greater than 6 years is .90.
 - (E) We should be 90% confident that the difference in life expectancies is between 6 and 12 years.



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