

The following 54 questions are similar to the types of questions you will see on the midterm exam. The actual midterm will be 50 to 60 multiple-choice questions.

Question 1. In a statistical study, the *population* is:

- (A) The people or objects studied in the sample survey.
- (B) All people in the United States.
- (C) The group of people or objects for which conclusions are to be made.
- (D) The group of people from whom data cannot be collected.

Question 2. In a statistical study, the *sample* is:

- (A) The collection of data in sample surveys.
- (B) A subset of people in the United States.
- (C) The subset of the population on which the study collects data.
- (D) The group of people or objects for which conclusions are to be made.

Question 3. To conduct a *proper statistical study*, a statistician must:

- (A) Have a sample which is fully representative of the population.
- (B) Have a sufficiently large sample.
- (C) Decide whether the study should be an observational study or an experiment.
- (D) All of the above.

Question 4. The importance of *randomized experiments* is that they:

- (A) Require only small sample sizes.
- (B) Allow the statistician to assign units to the treatment group.
- (C) Allow the inference of causation.
- (D) None of the above.

Question 5. An advertiser of No-Pain aspirin claims it is the pain-killer most preferred by consumers. This claim was based on a consumer survey in which the choices were: Lavid, Acinna, No-Pain, and Yellnot. This is an example of:

- (A) A closed question.
- (B) An easy question.
- (C) A difficult question.
- (D) An open question.

Question 6. In a study of car ownership in Pennsylvania, the variable “Brand of car owned” is:

- (A) A discrete quantitative variable
- (B) A nominal categorical variable
- (C) A continuous quantitative variable
- (D) An ordinal categorical variable

Question 7. Which of the following measures is valid and categorical?

- (A) The sale price of a house
- (B) Weight of an individual on a scale that is sometimes 5 pounds too light, sometimes 5 pounds too heavy
- (C) Sex (male or female)
- (D) Time on a clock that is always 10 minutes fast

Question 8. When measured with extreme accuracy, the variable “height of a building” is:

- (A) A discrete quantitative variable
- (B) A nominal categorical variable
- (C) A continuous quantitative variable
- (D) An ordinal categorical variable

Question 9. The survey question, “Don’t you agree that our tax system, which is too complicated for anyone to understand, should be overhauled?” is:

- (A) Deliberately biased
- (B) Likely to lead to lying unless some form of randomized response is used
- (C) Unnecessarily complicated
- (D) Unintentionally biased

Use the following material for the next four questions: In 1982, 490,000 subjects were asked about their drinking habits. Researchers tracked subjects' death rates until 1991, and found that adults who regularly had one alcoholic drink daily had a lower death rate than those who did not drink. Most of the subjects were middle-class, married, and college-educated.

Question 10. A news headline, reporting the medical study as “Daily Drink Cuts Death!” is:

- (A) Accurate; there is fundamentally solid anecdotal evidence that one drink daily increases life length and cuts death.
- (B) Misleading; we can never infer causation from randomized experiments.
- (C) Misleading; it implies that there is a causal connection between improved health and the habit of having one drink daily; however, causation cannot be deduced from an observational study.
- (D) Accurate; stratified random samples lead to randomized experiments, from which we cannot infer causation.

Question 11. This experiment suffers from :

- (A) A lack of generalizability; most subjects were married and college educated, and so are atypical of Americans. We cannot generalize the study to all Americans.
- (B) The presence of confounding variables. For example, healthier subjects may be more likely to allow themselves one drink daily.
- (C) The presence of interacting variables; e.g., the educational level of the subjects may be interacting with whether or not they have one drink daily.
- (D) All of the above.

Question 12. This experiment is an example of:

- (A) A matched pairs design; two groups were compared, those who had one drink daily and those who did not.
- (B) A retrospective study; after the nine-year period, survivors were asked about their drinking habits.
- (C) A block design; many groups were compared, e.g., married vs. not married, middle-class vs. not middle-class.
- (D) A prospective study; after recording their drinking habits, subjects were studied into the future.

Question 13. This experiment was:

- (A) An observational study; it would not be ethical for the researchers to randomly assign subjects to drink alcohol or not.
- (B) Accurate; there is fundamentally solid anecdotal evidence that people's health will improve if they have one drink daily.
- (C) Based on a stratified random sample; subjects were stratified randomly into overlapping groups according to whether or not they had one drink daily.
- (D) A randomized experiment; subjects were assigned in a randomized manner to have one alcoholic drink each day.

Question 14. The Gallup Poll, a well-known polling group, regularly surveys the public on many issues. The number of interviews on which their surveys are based is, approximately:

- (A) 60–120
- (B) The entire U.S. population.
- (C) 600,000 – 1,200,000
- (D) 600 – 1,200
- (E) 60,000 – 120,000

Use the following material for the next four questions: To estimate the unemployment rate, the Bureau of Labor Statistics contacts about 60,000 households chosen randomly from a list of all known households. The Bureau asks each adult in every sampled household whether they are in the labor force, i.e., employed or seeking employment.

Question 15. The *sampling frame* for this study is:

- (A) All households which were not selected in the sample.
- (B) All adults in the sampled households.
- (C) All sampled households.
- (D) All known households.

Question 16. The *measurements in the sample* are:

- (A) The employment statuses of the adults in the population.
- (B) The total number of adults in all sampled households.
- (C) The employment statuses of all adults who are not in the labor force.
- (D) The employment statuses of the adults in the sample.

Question 17. For this study, a *unit* is:

- (A) A unit of measurement.
- (B) An adult in any particular household.
- (C) An adult who is in the labor force.
- (D) A question on the survey answered by each adult in the sample.

Question 18. The *population* of interest is:

- (A) All known or unknown households.
- (B) All employed adults.
- (C) All adults who refused to participate in the survey.
- (D) All adults who are in the labor force.

Question 19. If a *simple random sample* of 1,600 subjects is chosen then an approximate margin of error for making inferences about a percentage of the entire population is:

- (A) $1/1,600$, or 0.06%.
- (B) Impossible to calculate without knowing the percentage observed in the sample.
- (C) Very high, because the sample size is a tiny percentage of the population size.
- (D) Approximately zero because the sample size is so large.
- (E) $1/\sqrt{1,600}$, or 2.5%.

Question 20. The term “control” means that:

- (A) We control carefully the observed outcomes of the experiment.
- (B) There must be a basis for making comparisons.
- (C) There is a need to control the subjects of the experiment.
- (D) We control carefully the cost of performing the experiment.

Question 21. The term “randomization” means that we:

- (A) Allow subjects to assign themselves randomly to the placebo or treatment group.
- (B) Assign the subjects to the random experiment in a systematic manner.
- (C) We study only a random subset of all observed outcomes of the experiment.
- (D) Use a chance mechanism to assign subjects to the treatment and control groups.

Question 22. In late 1999, a national newspaper reported that about 65% of the 78,000 voters in a Washington, D.C. election also approved the “medical” use of marijuana. This outcome was claimed to be a sign of strong support by D.C. residents for “medical marijuana.” This survey method is:

- (A) A probability sampling plan: There is a high chance that the 78,000 voters were well-scattered across the D.C., and so they constitute a representative sample.
- (B) An example of volunteer sampling: It is subject to response biases because supporters of “medical marijuana” may be more likely to have voted.
- (C) An example of stratified random sampling because all voters can be divided into two strata: those in favor of “medical marijuana” and those who are not.
- (D) A way for bureaucrats to accurately assess public opinion of “medical marijuana.”

Question 23. A *simple random sample* of ten subjects from a population is one in which:

- (A) Any group of ten subjects has the same chance of being the selected sample.
- (B) Most, but not all, groups of size ten have the same chance of being selected.
- (C) We record the data provided by the first ten subjects who respond to the survey.
- (D) A simple way was found to choose ten subjects at random from the population.

Question 24. In a study of PSU students' awareness of world issues, a television crew sampled students sitting in the Hub at 12:00 p.m. (noon). This survey is an example of:

- (A) A stratified random sample, because students are stratified by gender and they are found at the Hub in a random manner.
- (B) A haphazard (or convenience) sample, because the Hub is a convenient place to find students.
- (C) A simple random sample, because it is simple to find students at the Hub in a random manner.
- (D) A volunteer response sample because, to be seen on the evening news, students will eagerly volunteer their responses.

Question 25. A radio advertiser wishes to choose a random sample of size 100 from a population of 5,000 listeners. After observing that $5,000 \div 50 = 100$, he first selects a subject at random from the first 50 names in the sampling frame, and then he selects every 50th subject listed after that one. This method of random sampling is called:

- (A) Cluster random sampling.
- (B) Systematic random sampling.
- (C) Stratified random sampling.
- (D) Simple random sampling.

Question 26. In a second study the radio advertiser divides the population into five separate groups and then selects a simple random sample from each group. This method of random sampling is called:

- (A) Volunteer sampling.
- (B) Systematic random sampling.
- (C) Stratified random sampling.
- (D) Simple random sampling.

Question 27. In a third study, the radio advertiser divides the population into a large number of groups. He selects a simple random sample of the groups and then surveys *every* subject in each of the groups selected. This method of random sampling is called:

- (A) Random digit dialing.
- (B) Systematic random sampling.
- (C) Stratified random sampling.
- (D) Cluster sampling.

Question 28. To assess its employees' views of new student fees imposed by the state legislature, a university divided employees into two groups: those with at least 20 years, and those with fewer than 20 years, of service. A random sample was selected from each group and their views were obtained by telephone. This method of sampling is likely to be:

- (A) Unbiased, because the population has been stratified by telephone number.
- (B) Unbiased, because it is an example of stratified random sampling.
- (C) Biased, because random digit dialing should have been used to contact employees.
- (D) Biased, because the two groups are different in character.

Question 29. In any large data set, the proportion of data falling at or below Q_3 , the third quartile, is:

- (A) 68%
- (B) Dependent on the sample drawn from the population.
- (C) 99.7%
- (D) 25%
- (E) 75%

Question 30. To survey the opinions of its customers, a supermarket grouped its customers by the days when they did most of their shopping. The supermarket randomly selected two such groups, and asked all customers in those two groups to complete a survey. This method of sampling is called:

- (A) Systematic sampling.
- (B) Random digit dialing.
- (C) Cluster sampling.
- (D) Stratified random sampling.

Question 31. A supermarket manager wants to know if customers would pay slightly higher prices to have computers available throughout the store to help them locate items. An interviewer is posted at the entrance door and asked to collect a sample of 100 opinions by asking questions of the next person who came to the door each time she had completed an interview. This method of sampling is likely to be:

- (A) Unbiased, because any customer has a good chance of being asked questions.
- (B) Unbiased, because there is no reason for a customer not to answer questions.
- (C) Biased, because this is an example of a convenience sample.
- (D) Biased, because only customers will be asked questions.

Question 32. In a boxplot (oriented vertically, with the maximum at the top and the minimum at the bottom), the proportion of the data falling between the top edge and the bottom edge of the box itself is

- (A) 68%
- (B) 99.7%
- (C) Impossible to say without more information
- (D) 50%
- (E) 25%

Question 33. A pollster studying opinions on gun control divided a city into blocks, then surveyed the third house to the west of the southeast corner of each block. If that house consisted of apartments, the westernmost ground floor apartment was selected. The pollster conducted the survey during daytime, and left a phone number for those who were away when she tried to interview them. This sampling method is likely to be:

- (A) Unbiased, because the population has been stratified according to city blocks.
- (B) Biased; those at home during daytime may have sharply different views from those not at home.
- (C) Unbiased, because this is an example of systematic sampling.
- (D) Biased, because she tried to interview in person those who were away.

Question 34. When a distribution is greatly skewed to the left (i.e., has a long right tail), the median will usually be:

- (A) Smaller than the mean.
- (B) Exactly equal to the mean.
- (C) Larger than the mean.
- (D) In no relation whatsoever to the mean.

Question 35. The *Empirical Rule* states, in part, that if a data set is approximately normally distributed (or bell-shaped) then:

- (A) About 68% of all observations fall within one standard deviation of the mean.
- (B) About 37% of all observations fall within two standard deviations of the mean.
- (C) At most 90% of all observations fall within three standard deviations of the mean.
- (D) None of the above.

Question 36. For any data set, the *standard deviation* is:

- (A) A measure of the spread or variability of the data.
- (B) The average of the sample mean and quartiles.
- (C) The average of the deviations.
- (D) A measure of central tendency.

Use the following material for the next three questions: A sample of 28 temperature measurements in °F, all taken at 12:00 p.m., was collected in a coastal town in NC. The data are given in the following stemplot:

Stemplot of Temperature Readings

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

Question 37. For the data given in this stemplot, the *five-number summary* is:

- (A) 32, 68, 79, 88, 98
- (B) 32, 66, 78.5, 84.5, 98
- (C) 55, 66, 78.5, 84.5, 95
- (D) 32, 66, 78, 84, 98

Question 38. In this stemplot, the *interquartile range* is:

- (A) $84 - 66 = 18$
- (B) $84.5 - 66 = 18.5$
- (C) $88 - 68 = 20$
- (D) $95 - 55 = 40$

Question 39. A second sample of 28 temperature readings in a GA coastal town was also collected. Each GA measurement was recorded at the same time and date as the NC data value, and turned out to be exactly 5°F higher than the corresponding NC measurement. We calculate the standard deviation (S.D.) of the two data sets and conclude that:

- (A) The two data sets have the same standard deviations.
- (B) The S.D. of the NC data exceeds the S.D. of the GA data by 5°F.
- (C) The S.D. of the GA data exceeds the S.D. of the NC data by 5°F.
- (D) None of the above; there is no relationship between the two S.D.'s.

Use the table of standard normal scores and the following material for the next four questions: As measured by the Stanford-Binet test, IQ scores are approximately normally distributed with mean 100 and standard deviation 16.

Question 40. Because of the Empirical Rule (68-95-99.7 Rule), we can conclude that approximately 68% of all IQ scores will fall between:

- (A) 84 and 116.
- (B) 32 and 168.
- (C) 68 and 132.
- (D) 0 and 200.

Question 41. *Mensa* is an organization whose members have IQ scores in the top 2% of the population. To be admitted to *Mensa*, a person's Stanford-Binet IQ score is at least:

- (A) $100 - (16 \times 0.98) = 84.32$.
- (B) $100 - (16 \times 2.05) = 67.20$.
- (C) $100 + (16 \times 2.05) = 132.80$.
- (D) $100 + (16 \times 0.98) = 115.68$.

Question 42. A score of 108 on the Stanford-Binet test falls in the:

- (A) 92nd percentile.
- (B) 69th percentile.
- (C) 50th percentile.
- (D) 80th percentile.

Question 43. A student who scores 108 on the Stanford-Binet test has a *standardized score* of:

- (A) $(108 - 100)/\sqrt{100} = 0.80$.
- (B) $(108 - 16)/100 = 0.92$.
- (C) $(108 - 100)/16 = 0.50$.
- (D) Cannot be calculated because we are given insufficient information.

Question 44. If a data set is normally distributed then:

- (A) The mean is smaller than the median or the mode.
- (B) The mean is larger than the median but not larger than the mode.
- (C) The mean, median and mode are the same.
- (D) None of the above.

Use the following material for the next four questions: A researcher asks 1,600 randomly chosen doctors whether or not they take aspirin regularly. She also asks them to estimate the number of headaches they have had in the past six months, and compares the number of headaches reported by those who take aspirin regularly to the number of headaches reported by those who do not take aspirin regularly.

Question 45. In this study, the number of headaches is a:

- (A) Response variable.
- (B) Explanatory variable.
- (C) Confounding variable.
- (D) None of the above.

Question 46. With regard to all people in the country, this study is likely to:

- (A) Provide accurate information because the subjects were chosen randomly.
- (B) Provide accurate information because a random sample of size 1,600 has a margin of error of 2.5%.
- (C) Be somewhat unrealistic because doctors are not representative of the population.
- (D) Provide inaccurate inferences because 1,600 doctors is a tiny proportion of all people in the country.

Question 47. This study may be highly unreliable because:

- (A) Subjects are unlikely to accurately recall past incidences of headaches.
- (B) Subjects should have randomly assigned themselves to the two treatments.
- (C) The researcher should have randomly assigned subjects to the two treatments.
- (D) Subjects cannot both take aspirin regularly and not take aspirin regularly.

Question 48. This type of study design is a:

- (A) Prospective study.
- (B) Census.
- (C) Retrospective observational study.
- (D) Randomized experiment.

Question 49. To test the effects of sleepiness on driving performance, twenty volunteers took a simulated driving test under each of three conditions: Well-rested, Sleepy, and Exhausted. The order in which each volunteer took the three tests was randomized, and an evaluator rated their driving accuracy without knowing the condition of the volunteer. This type of experiment is a:

- (A) Single-blind, matched-pair experiment.
- (B) Retrospective observational study.
- (C) Double-blind, matched-pair experiment.
- (D) Single-blind, block design experiment.

Question 50. In the previous question on a study of sleepiness and its effects on driving performance, the explanatory variable is:

- (A) The condition of sleepiness.
- (B) The evaluator's rating of driving accuracy.
- (C) The evaluator's rating of the driver's condition.
- (D) The order in which the volunteers took the tests.

Question 51. It has been observed that participants in a statistical experiment sometimes respond positively to a placebo, a substance which has no active ingredients. This phenomenon is called the:

- (A) Placebo effect.
- (B) Confounding effect.
- (C) Hawthorne effect.
- (D) Interacting effect.

Question 52. It has been observed that participants in a statistical experiment sometimes respond differently than they otherwise would because they know that they are in an experiment. This phenomenon is called the:

- (A) Hawthorne effect.
- (B) Interacting effect.
- (C) Confounding effect.
- (D) Placebo effect.

Question 53. For any data set, the largest number in the five-number summary is the:

- (A) Interquartile range.
- (B) Maximum.
- (C) Standard deviation.
- (D) Outlier.
- (E) Third quartile.

Question 54. For any data set, the proportion of the data that falls at or above the median is:

- (A) 95%
- (B) 25%
- (C) Dependent on the sample which was drawn from the population.
- (D) 50%
- (E) 99.7%