

NOTE to students: This is intended to give you an idea of the type questions the instructor asks and the approximate length of the exam. It does NOT indicate the exact questions or the topics covered. Students should refer to the information posted on Moodle to determine the coverage of the material.

Instructions:

- Read each question carefully.
- Provide one answer to each question in the space provided.
- For the short answer problems:
 - Show all work or explain your reasoning to receive full credit.
 - Make all work legible.
 - If asked to draw a well-labeled picture, be sure to:
 - Label the name and parameter values of the distribution (e.g. $N(7,2)$ or $t(7)$)
 - Label the mean on the horizontal axis
 - Label the X or t value of interest
 - Shade and label the relevant area under the curve
- You may use a basic calculator (no graphing calculators, cell phones, or web-enabled devices).

Honor Pledge:

I certify that I have not received or given unauthorized aid in taking this exam.

Signed: _____

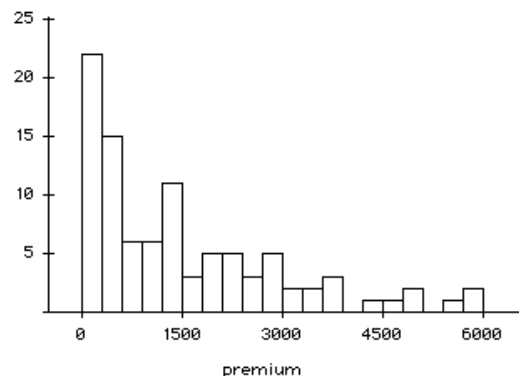
Multiple Choice - Circle the best answer (3 points each).

1. When considering a point estimator, an estimator that is *precise*
 - a. Is equal to the true value of the parameter on average.
 - b. Consistently estimates the true value of the parameter.
 - c. Has low variability in the values of its estimates.
 - d. Has high variability in the values of its estimates.
2. The standard error of a statistic describes
 - a. the variability of a sample statistic.
 - b. the variability of the individuals.
 - c. the error that occurs due to non-response and measurement errors.
 - d. all of the above.
3. For a t-distribution, as the degrees of freedom increases
 - a. The distribution becomes more skewed to the right.
 - b. The distribution becomes more skewed to the left.
 - c. The distribution becomes more like the standard normal distribution.
 - d. The distribution stays exactly the same.
4. A sociologist randomly selected 100 people from a county that had about 10,000 people in it and found that the 95% confidence interval for the true proportion of people in this county that attended religious services was $74\% \pm 8.6\%$. Which of the following would increase the size of the margin of error in this situation (e.g., make it something bigger than 8.6%)?
 - a. If they had sampled 200 people rather than 100 people.
 - b. If they had used 99% confidence rather than 95% confidence.
 - c. If there had been 50,000 people in the county instead of 10,000.
 - d. Both b. and c. but not a.
5. We should be wary of a poll that suffers from undercoverage because
 - a. those who refused to respond may be different from those who participate.
 - b. the sampling frame was not representative of the population.
 - c. the standard deviation will be very large.
 - d. the mean and median will be different because the results are skewed.
6. A rv has density function: $f(x) = 2x$ for $0 < x < 1$. What is $P(X \leq 0.4)$?
 - a. 0.40
 - b. 0.16
 - c. 0.32
 - d. 1.00
7. Which of the following graphical summaries is most appropriate for a numeric variable?
 - a. Bar chart
 - b. Histogram
 - c. Boxplot
 - d. Both b. and c. but not a.

Use the following for questions 8-10: Does "listening" to classical music make plants grow? A study was conducted in which 40 potted house plants were randomized to be either in a room while classical music was played from a radio or in a room with no music, with 20 plants being assigned to each type of room. The rooms had similar amounts of light and the plants were given the same amount of water during the study. The size of the plants was measured at the start of the study and again after two weeks, and the amount of growth for each plant was calculated. It was found that the plants in the room where classical music was played were significantly larger, on average, than the plants in the room with no music.

8. This study is best described as
 - a. an observational study.
 - b. a matched pairs design.
 - c. a randomized block design.
 - d. a randomized experiment.
9. In this study, the response was
 - a. the 40 potted house plants.
 - b. the rooms with and without classical music.
 - c. the size of the plant at the end of the study
 - d. the amount of growth for each plant.
10. In this study, the primary purpose of the use of random assignment is
 - a. to ensure that the experimenter doesn't know whether a plant was in the room with or without classical music.
 - b. to ensure that the groups are similar in all respects except for whether or not classical music was played in the room.
 - c. to ensure that the study participants are representative of the population.
 - d. to account for the placebo effect.

11. Recently a major consumer magazine reported the annual auto insurance premiums (in dollars) of a sample of randomly selected families in Texas. This data was used to produce the histogram at the right. The mean of this distribution would be

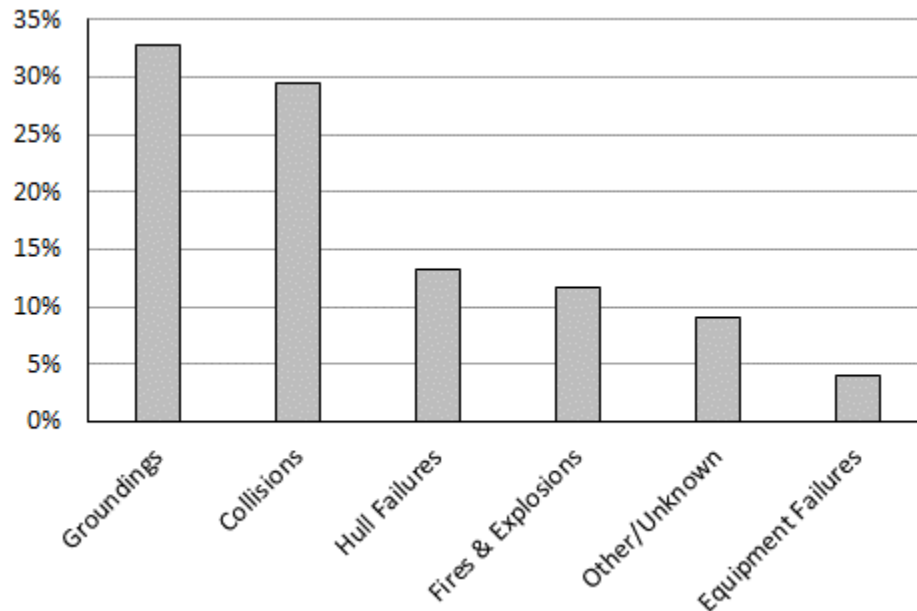


- a. Less than the median
 - b. About the same as the median
 - c. Greater than the median
 - d. Either a. or c. but not b.
12. Consider the data presented in the previous question. The sample standard deviation was 1500. Which of the following is the best interpretation of this value?
 - a. On average, the annual auto insurance premiums for each family were about \$1500 away from the mean premium, roughly speaking.
 - b. On average, the annual auto insurance premiums for each family were about \$1500 away from each other, roughly speaking.
 - c. Roughly speaking, the average distance between the individual values for annual auto insurance premiums for each family and the mean premium was \$1500.
 - d. Both a. and c. but not b.

Use the following for questions 13 and 14: A researcher was interested in utilities provided by city governments. The researcher randomly selected 20 counties from a list of all counties in the U.S. From each of these counties the researcher then contacted each city government (a total of 192) and found that 12 (6.25%) of them provided electricity to their residents.

13. In this situation the sampling frame is
- the list of all counties in the U.S.
 - the 316 cities.
 - the 20 counties.
 - 6.25%.
14. The type of sample used in this example is a
- simple random sample.
 - cluster random sample.
 - stratified random sample.
 - all of the above.
15. An instructor in a college class recently gave an exam that was worth a total of 100 points. The instructor inadvertently made the exam harder than he had intended, so he plans to curve the grades by adding 10 points to everyone's score. Which of the following summary measures would this curve affect?
- Mean
 - Median
 - Standard deviation
 - Both a. and b. but not c.
16. Which of the following variables is Nominal?
- Age (in years)
 - Age (grouped as: less than 18, 18-29, 30-59, 60 and older)
 - Car color (e.g. red, blue, grey)
 - Letter grade in a course (e.g. A+, A, A-, B+, B)
17. An administrator for a large community college knows that 82% of the students at the school receive financial aid. The administrator is going to take a sample of 100 students to complete a survey about the costs of attending that college. Which formula would be used to calculate the standard deviation of the sample proportion in this situation?
- $\frac{0.82}{100}$
 - $\frac{0.82}{\sqrt{100}}$
 - $\frac{(0.82)(0.18)}{100}$
 - $\sqrt{\frac{(0.82)(0.18)}{100}}$

Use the following for questions 18 and 19: Data from the International Tanker Owners Pollution Federation Limited give the cause of spillage for 455 large oil tanker accidents from 1970 to 2012. The data is displayed in the graph below.



18. Is it appropriate to say this distribution is right skewed?
- a. Yes, since the tail trails off to the right.
 - b. No, since the distribution is clearly left skewed.
 - c. Yes, since there is a lower limit of 0%.
 - d. No, since this is a categorical variable.
19. Does this graphic accurately display the story of the data?
- a. Yes, since this is an appropriate and clearly-labeled graphic for a quantitative variable.
 - b. Yes, since this is an appropriate and clearly-labeled graphic for a categorical variable.
 - c. No, since this graphic has the baseline omitted.
 - d. No, since the percentages do not add up to 100%.

Short answer: Give one answer in the space provided. Show your work.

20. (6 points) The distribution of lifetimes (measured in 1000s of hours) for a certain electrical component is skewed to the right with a mean of 16 and a standard deviation of 3. We select a random sample of 50 components from this population. The probability that the sample mean lifetime is greater than 16.05 is equal to 0.41. Draw a well-labeled picture of this value.

21. (6 points) Consider the electrical component described in the previous question. Suppose the company producing these components knows that 95% of the components should meet the industry standard for an acceptable lifetime. We select a random sample of 50 components from this population. Describe the sampling distribution of \hat{p} in this situation; this includes commenting on its shape, center, and variability.

22. A sociologist was interested in estimating the amount of lawn space home owners in a large county in Pennsylvania allow for their pets. She randomly selected 75 homes from this county, questioned the home owner, and recorded the amount space (in square feet) they allow for their pets. From the results she found that the average amount of lawn space allowed for pets was 30 square feet with a standard deviation of 5 feet.

- a. (6 points) Is it appropriate to use this data to calculate a 90% confidence interval? If so, write the formula—with values plugged in—that would be used to calculate the interval (you may use the generic t_{n-1} notation for the multiplier as long as you plug in the appropriate degrees of freedom). In not, explain why not.

- b. (6 points) Interpret the confidence level used to calculate the interval in the previous part (i.e. what does it mean to say we have 90% confidence?)

23. (6 points) Let $Y = \max(x_1, x_2, \dots, x_n)$, where X_1, \dots, X_n are a random sample from the Uniform distribution defined over the range 0 to θ . The pdf of Y is $f_Y(y) = \frac{ny^{n-1}}{\theta^n}$ for $0 \leq y \leq \theta$ and the mean of Y is $E(Y) = \frac{n}{n+1}\theta$. Is $\hat{\Theta}_2 = \frac{n+1}{n}Y$ an unbiased estimator of θ ?

24. (7 points) A health class at a large university requires students to carry out a class project in small groups. One group in the class is going to conduct a survey of the students at their university. They are going to ask the question “Have you ever been told by a health professional that you have a sexually transmitted infection (STI)?” They will also ask for age, gender, and ethnicity. They have two possible ways of conducting their survey.
- Method 1. Stand outside of the student center for 5 consecutive days. Stop students as they pass by and ask them to fill out the survey. Have at least one member of the group standing there during each hour that the student center is open. Before having the students fill out the survey ask them if they have previously filled out the survey and only accept one survey per person.
 - Method 2. Use the list of all class offerings to randomly select 20 classes. Contact the professor for each class and arrange to administer the survey to every student in those classes. Compare the rosters for each of the 20 classes and make sure that if there are any duplicate students that they only fill out the survey one time.

Which method should the group use to carry out their survey? Explain your reasoning.

25. (6 points) 6% of the U.S. population has Type O-negative blood. Suppose that there are 20 people volunteering to donate blood (and these people can be thought of as a random sample). Write out the formula—with values plugged in—that would be used to calculate the probability that at least 2 of these people have type O-negative blood.