ST 314 Midterm Exam

SPRING 2023

Name: Ya Zou

Student ID Number: 934-417-686

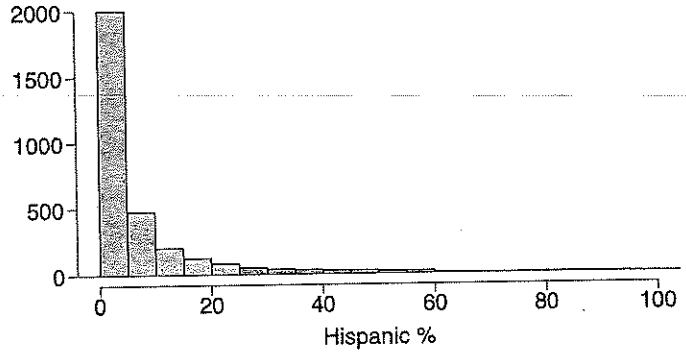
Instructions

- You will have exactly **80 minutes** to complete the exam.
  - Do not open the exam until you are told to do so.
  - You must stop working when Erin announces the exam is over.
  - If you finish early, please hand your exam to Erin or one of the TAs.
- There are 17 questions on this exam.
- You may use any hard copy/paper notes during the exam.
- You MAY NOT use any digital notes, a calculator, computer, tablet, or phone during the exam.
- You are NOT expected to work out calculations on the exam.
  - For example, if an answer to a question is found by doing $10 + \frac{85}{\sqrt{150}}$, you may leave it as the expression previously written. Do not spend time trying to work out the simplified answer to a complicated expression.
- For multiple choice questions, please fill in the bubble corresponding to the answer you wish to select.
  - For example, to select answer B, mark:

○ A
● B
○ C
○ D

Use the following information for questions 1-2

The US census collects data on the race and ethnicity of Americans, among many other variables. The histogram below shows the distribution of the percentage of the population that is Hispanic in 3,142 counties in the US in 2010.



1. (1 point) Which of the following statements are true regarding the distribution of the percentage of the population that is Hispanic in the 3,142 counties. **Select all that apply.**

   - ☒ The mean of these data is greater than the median.
   - ☐ The mean of these data is less than the median.
   - ○ To describe the spread of these data, the standard deviation is a more appropriate metric than the quartiles and interquartile range (IQR).
   - ○ The sampled distribution (displayed in the histogram above) appears to be approximately normal.
   - ◉ The sampled distribution (displayed in the histogram above) is right skewed.

2. (1 point) A histogram is not the only way the quantitative variable "percentage of county population that is Hispanic" can be visualized. Which of the following options would be an appropriate alternative way to display these data?

   - ○ Pie chart
   - ○ Bar plot
   - ◉ Box plot
   - ○ Discrete probability model

3. (1 point) A university wants to determine what fraction of its undergraduate student body supports a new $25 annual fee to improve the student union. There are 25,000 undergraduate students at the university and 5 different colleges. Within each of the five colleges, 50 students were randomly selected and asked "Do you approve a $25 annual fee to improve the student union?" What sampling design was used?

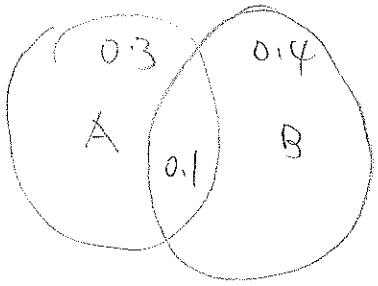
- ○ Simple random sampling
- ○ Systematic random sampling
- ○ Convenience sampling
- ◉ Stratified random sampling

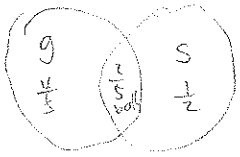
4. (1 point) Suppose you have a data set of 100 values where the mean is equal to 90 and the standard deviation is 12. Consider a transformation of this data set where every observation was decreased by 20 units. Which of the following statements is true regarding the **transformed** data set?

- ◉ The mean of the transformed data set is 70 and the standard deviation of the transformed data set is 12.
- ○ The mean of the transformed data set is 90 and the standard deviation of the transformed data set is 12.
- ○ The mean of the transformed data set is 70 and the standard deviation of the transformed data set is 1.
- ○ The mean of the transformed data set is 90 and the standard deviation of the transformed data set is 1.

5. Suppose you are a sales manager at a company that sells two types of products, A and B. The probability that a customer buys product A is 0.3. The probability that a customer buys product B is 0.4. The probability that a customer buys both product A and product B is 0.1. Which of the following expressions correctly describes the probability that a customer buys either product A or product B?

- ◉ P(A or B) = $0.3 + 0.4 - 0.1$
- ○ P(A or B) = $0.3 + 0.4$
- ○ P(A or B) = $0.3 + 0.4 + 0.1$
- ○ P(A or B) = 0.3×0.4





6. (1 point) Suppose you're out for a walk around a pond in February. The probability that you will see a Canada goose on your walk is $\frac{4}{5}$. The probability that you will see a Trumpeter swan is $\frac{1}{2}$. The probability that you will see both a Canada goose and a Trumpeter swan is $\frac{2}{5}$. Let C represent the event "seeing a Canada goose" and T represent the event "seeing a Trumpeter swan". Which of the following statements is true about the events C and T?

   - ○ The events are disjoint because it is impossible to see both a Canada goose and a Trumpeter swan on the same walk.
   - ◉ The events are independent because $P(C \text{ and } T) = P(C) \times P(T)$.
   - ○ One of the events is guaranteed to occur.
   - ○ The events are complements of each other.

7. (2 points) Oregon's timber industry makes up about 16% of all timber harvested in the United States. That is, the proportion of trees harvested in the U.S. that come from Oregon is 0.16. The proportion of trees that are exported internationally given they were harvested in Oregon is 0.38. The proportion of trees that are exported internationally given they were harvested in the United States but not in Oregon is 0.22.

   Using the numerical values in the description above, write an expression for the probability that a randomly selected tree harvested in the United States is exported internationally.

   You do not need to compute the final answer. Your answer can be a mathematical expression. Please write your answer ***in the box below.***

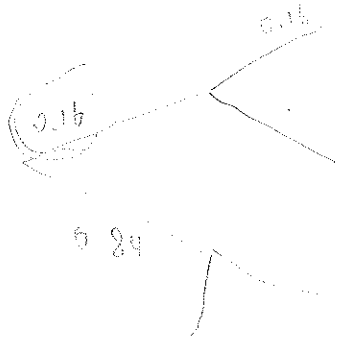
   0.38+0.16+0.22+(1−0.16)

   (0.38)(0.16)+(1-0.16)(0.22

8. A continuous random variable X has the following **cumulative density function**: $F(x) = 1 - e^{-6x}$ for $x \geq 0$.

   Write an expression for the **probability density function** of X **in the box below**.

   $$\frac{d\,1-e^{-6x}}{dx} = \frac{d\,F(x)}{dx}$$



Use the following information for questions 9-11

The discrete random variable X is defined by the following probability mass function, $p(x)$.

x	0	1	2	3	4	5
$p(x)$	0.03	0.23	0.32	0.11	0.08	0.23

9. (1 point) Which of the following correctly describes the function, $p(x)$, displayed in the table?

   - ○ The function $p(x)$ gives the probability that the random variable is greater than or equal to x, for each value of x given in the first row of the table.
   - ◉ The function $p(x)$ gives the probability that the random variable is equal to x, for each value of x given in the first row of the table.
   - ○ The function $p(x)$ gives the probability that the random variable is less than or equal to x, for each value of x given in the first row of the table.

10. (1 point) Which of the following expressions correctly represents the probability that X is greater than 2?

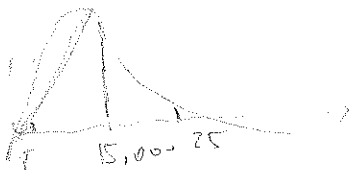
    - ○ $P(X > 2) = 0.32$
    - ○ $P(X > 2) = 1 - 0.32$
    - ○ $P(X > 2) = 0.03 + 0.23 + 0.32$
    - ◉ $P(X > 2) = 0.11 + 0.08 + 0.23$

11. (1 point) Let $F(x)$ represent the cumulative distribution function for the random variable X. Which of the following statements is FALSE?

    - ○ $F(0) = 0.03$
    - ○ $F(1) = 0.03 + 0.23$
    - ◉ $F(3) = 0.11$
    - ○ $F(5) = 1$

Use the following information for questions 12-13

In 2015 the average credit card debt per US household was $15,000. Assume the distribution of credit card debt for all households is right-skewed with a mean of $15,000 and a standard deviation of $10,000.



12. (1 point) Consider the sampling distribution of sample means from random samples of 200 households. Which of the following statements is true?

- ○ The distribution of sample means is right-skewed.
- ○ The distribution of sample means is approximately binomial.
- ◉ The distribution of sample means is approximately normal.
- ○ The distribution of sample means cannot be determined because the sample size of 200 is too small.

13. (1 point) Consider the sampling distribution of sample means from random samples of 200 households. What is the standard error of the sample mean?

- ○ 15000
- ○ 10000
- ◉ $\frac{10000}{\sqrt{200}}$
- ○ $\sqrt{\frac{10000(1-10000)}{200}}$

14. A customer support facility keeps track of the number of calls it receives and determines that, on average, 49 calls are received each hour. Which of the following distributions would be appropriate to use in modeling the random variable "number of calls received in an hour?"

- ○ Normal Distribution
- ○ Continuous Distribution
- ○ Binomial Distribution
- ◉ Poisson Distribution

Use the following information for questions 15-17

Polypropylene is a plastic material commonly used to make containers of many household items, such as plastic yogurt containers. A nonprofit organization dedicated to waste reduction is interested in estimating the proportion of recycling facilities in the United States that accept polypropylene. In a simple random sample of 50 recycling centers in the United States, 36 accept polypropylene.

15. (1 point) The critical value needed to construct a 99% confidence interval for the proportion of recycling centers in the U.S. that accept polypropylene is

- ○ The 90th percentile of a standard normal distribution.
- ◉ The 99.5th percentile of a standard normal distribution.
- ○ 1.96
- ○ impossible to determine from the information provided.

16. Consider constructing both a 90% and 99% confidence interval for the proportion of recycling centers in the U.S. that accept polypropylene. Which of the following statements are correct regarding the two confidence intervals? **Select all that apply**.

- ○ The 90% confidence interval is wider than the 99% confidence interval.
- ◉ The 99% confidence interval is wider than the 90% confidence interval.
- ◉ The critical value used to construct the 90% confidence interval is less than the critical value used to construct the 99% confidence interval.
- ○ The standard error value used to construct the 90% confidence interval is less than the standard error value used to construct the 99% confidence interval.
- ○ The intervals will be centered at different values.

17. (2 points) The 99% confidence interval is (0.556, 0.884). Interpret this confidence interval in the context of the problem **in the box below**.

The 99% confidence interval estimats the true proportion of recycly centers in the U.S that accept Poly propylene is between 0.556 and 0.884 with a point estimate of $\frac{36}{50}$