ST370 Exam 1 Review

The following concepts are all fair game for the first exam. Use the following check list and the review problems to help you in your studying. (Most of the review problems come from old tests.)

- Populations, samples, and processes
  - Sampling methods
    - Voluntary response sample
    - Simple random sample
    - Stratified random sample
    - Multistage random sample
  - Problems with samples
    - Undercoverage
    - Nonresponse
    - Response bias
    - Poorly worded questions
  - Observational Study vs. Designed Experiment
  - Design of Experiments
    - Experimental unit
    - Treatment
    - Factors
    - Levels
    - Response variable or outcome
  - Problems with experiments
    - Lurking variables
    - Placebo effect
    - Bias
    - Lack of realism
  - Three Principles of Experimental Design
    - Compare
    - Randomize
    - Repeat
  - Types of Experimental Design
    - Completely randomized design
    - Randomized block design
    - Matched Pairs
- Descriptive statistics
  - Qualitative vs. Quantitative
  - Graph types
    - Stem and Leaf Displays
    - Dotplots
    - Histograms
    - Boxplots
  - Creating and interpreting histograms

        - Skew, shapes, modes
    - Measures of center
        - Mean, median, mode, trimmed mean
    - Measures of spread
        - Range, interquartile range (IQR), five number summary, standard deviation
        - How these impact boxplots and modified boxplots
        - Using Q1 - 1.5IQR, Q3 + 1.5IQR to find outliers
- Probability
    - Events, sample space, Venn diagrams
    - Addition property and disjoint events
    - Complement property
    - Fundamental Counting Principle
    - Permutations
    - Combinations
    - Conditional Probability and independent events
    - Law of Total Probability
    - Bayes Rule
- Discrete random variables
    - Interpreting distributions… P(X>2)
    - Cumulative distribution function
    - Expected value (mean)
    - Variance
    - Linear transformations of mean and variance
    - Binomial Probability Distribution
        - Finding the binomial pmf for exact values (ex. P(X=3)) – know the formula and execute it
        - Cumulative probability (ex. P(X>3)) - (interpreting, but not calculating) – know the syntax of commands and be able to choose the correct from a MC list
        - Mean, variance, and standard deviation of a binomial
    - Geometric Distribution
        - Finding probability of x, given p.
    - Poisson Distribution
        - Find the Poisson distribution for exact values (ex. P(X=3)) – know the formula and execute it
        - Cumulative probability (ex. P(X>3)) - (interpreting, but not calculating) – know the syntax of commands and be able to choose the correct from a MC list
- Continuous random variables
    - There is no pmf, but we will do cdf
    - Uniform distribution
    - Continuous functions that are nonnormal
        - Finding cdf using integration
        - Using percentile to find bounds
        - Expected value and variance using integration
    - Normal distribution

        - Empirical rule
        - Z-scores (calculating z or x)
        - Normal distribution vs standard normal distribution
    - Turning z-scores into area (no tables or technology required… know the logic of the process)
        - Percentiles into z-scores (no tables of technology required… know the logic of the process)
        - Critical values (no tables or technology required… know the logic of the process)
        - Normality tests (normal quantile plots… I won't ask to create, but know the steps to create and know how to interpret)
        - Sampling distributions (know what it is and how the population distribution can turn into the sampling distribution)
        - What is the sampling distribution parameters and how are they derived - the mean and standard deviation of sample means
        - Central Limit Theorem - what is it and why does it matter?
        - Solving sampling problems of skewed (or could be normal) data
            - Turn mean and standard deviation into a z
            - Write a probability statement (in terms of the variable or in terms of the test statistic)
            - Interpreting p-value (no tables of technology required - know the logic of the process)
            - Completing a diagram of the normal curve that shows the names and values of the parameters

1. Choose the letter for the graph that is most appropriate for displaying/analyzing the data for the following questions. Each letter may be used once, more than once, or not at all. For some questions, two or more graphs may be appropriate.

A. Stem Plot	B. Side-by-side boxplot	C. Pie Chart

   a. What % of golfers use Top-Flight, Nike, Titleist, Callaway, and "all other" golf balls?
   b. How far does a Titleist golf ball go? Ten Titleist balls were hit by a mechanical golfer and the distances were recorded.
   c. Which type of ball goes farther? Ten Titleist and Ten Nike balls were hit by a mechanical golfer and the distances were recorded

2. George has a taxi company with 17 taxicabs. George keeps records of many things including the miles driven per year for each taxi. Below are the miles driven in 2004 (rounded to the nearest thousand miles) for each of his 17 taxicabs.

   28 30 42 34 29 31 52 28 32 29 28 45 39 27 36 25 29

   a. Generate a stemplot of the data.

   b. Looking at the stemplot you developed, how would you describe the shape of the data distribution?

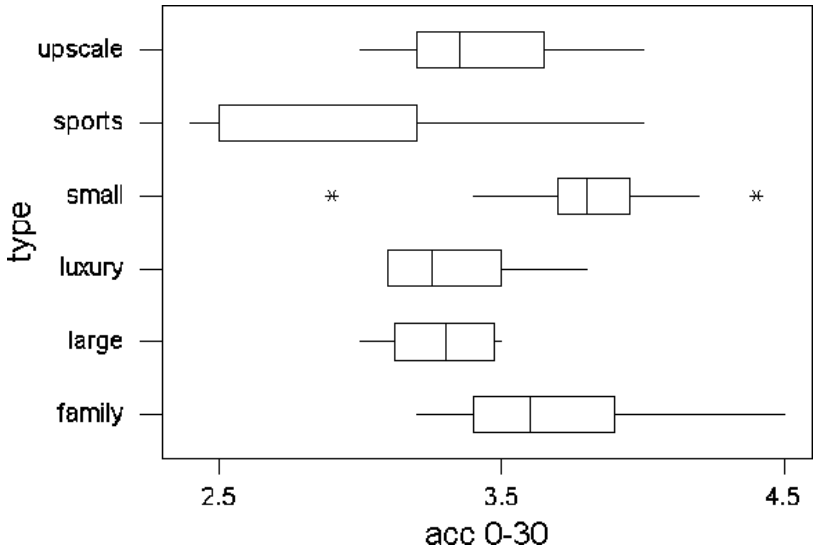
   c. The mean is 33.2 and the standard deviation is 7.2. What is the five-number summary?

   d. Are there any outliers? Justify your answer.

e. Make a modified boxplot. Be sure to label the plot.

f. What measures of center and spread are most appropriate for this data? Why?

3. The 1999 Consumer Reports New Car Buying Guide reported on the number of seconds required for a variety of cars to accelerate from 0 to 30 mph. The cars were also classified in six categories according to type. The following boxplots display these distributions:



a. Which type of car tends to accelerate the fastest?

b. Which type of car has the smallest IQR?

4. One in 20 people have a food allergy. Suppose I have a classroom with 90 students in it. What is the expected value of students with allergies? What is the standard deviation?

5. An education researcher wishes to study the association between GPA and whether a student takes the SAT as a sophomore for Broughton High School students. There are 800 sophomore students in the school, and the researcher randomly selects 150 of these students to record their sophomore-year cumulative GPA and whether they took the SAT.
    a. What is the unit/subject for this study?

    b. What is the population for this study?

    c. What is the sample for this study?

    d. What are the response variables for this study? Are they categorical or quantitative?

    e. What type of sampling design did the researcher use?

6. Suppose the local Childhood Lead Poisoning Prevention Council in a metropolitan area in western Tennessee undertakes the responsibility of determining the proportion of homes in their city that have unsafe lead levels. Because of the great expense involved in performing spectrometric testing they decide to test only some of the homes. The Council assumes that houses built prior to 1970 are more likely to have unsafe lead levels. Consequently, they divide their population into homes built prior to 1970 and homes built after 1970. They then take a random sample of 100 homes built prior to 1970 and 100 homes built after 1970 and record the lead levels for each of the 200 homes. This is an example of a:
    a. Randomized block design
    b. Simple random sample
    c. Stratified random sample
    d. Completely randomized design
    e. Multistage random sample

7. Suppose in the process of obtaining the population of all homes in the previous problem, several neighborhoods were mistakenly omitted. What type of bias will this study potentially suffer from?
    a. Non-response
    b. Response bias
    c. Double blind
    d. Placebo
    e. Undercoverage

8. Do piano lessons improve the spatial-temporal reasoning of preschool children? Preschool children were randomly selected to participate in this study. Researchers gave one group of 34 randomly selected children six months of piano lessons, and gave a second group of 34 randomly selected children six months of computer lessons.

   At the end of six months the children were given a test to measure improvement in spatial-temporal reasoning (values ranging from -4 to 9).

   The average improvement for the piano lesson children was 3.618 points. The average improvement for the computer lesson children was 0.386 points.

   a. What is the population?

   b. What is the sample?

   c. What is the response variable? Is it categorical or quantitative?

   d. Is this an observational study or an experiment? Why?

   e. What is the design used here? Give a detailed outline.

9. Hospital floors are usually covered by bare tiles. Carpets would cut down on noise but might be more likely to harbor germs. To study this possibility, investigators randomly assigned 8 of 16 available hospital rooms to have carpet installed. The others were left bare. Later, air from each room was pumped over a dish of agar. The dish was incubated for a fixed period, and the number of bacteria colonies were counted.
   a. Select the appropriate statistical term for the 16 hospital rooms.
      A. Experimental Units
      B. Treatments
      C. Response
      D. Control Group
   b. Select the appropriate statistical term for number of colonies in a dish.
      A. Experimental Units
      B. Treatments
      C. Response
      D. Control Group

10. The sampling distribution of a statistic is
    a. The probability that we obtain the statistic in repeated random samples
    b. The mechanism that determines whether randomization was effective
    c. The distribution of values taken by a statistic in all possible samples of the same size from the same population
    d. The extent to which the sample results differ systematically from the truth

11. Since 1976 the Nurses' health study has followed more than 100,000 nurses. Every two years, the nurses fill out a questionnaire about their habits and their health. Results from this study indicated that post-menopausal women have a reduced risk of heart disease if they take a hormone replacement drug. Which of the following is NOT true concerning this study?
    a. It is obvious from this study that the hormone-replacement drug reduces the risk of heart disease.
    b. A possible lurking variable may have caused the relationship, making it not so obvious that hormone-replacement drug reduces the risk of heart disease.
    c. A randomized experiment would be a better approach to testing the effects of the hormone-replacement drug.
    d. This study is an observational study.

12. The following table gives the distribution of the number of servings of fruits and vegetables consumed per day in a population:

Number of servings X	0	1	2	3	4	5
Probability	0.4	0.1	0.1	0.2	0.1	0.1

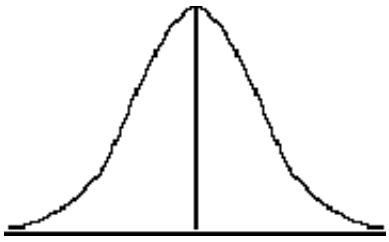
    a. Find the expected value for this random variable

    b. Find the variance for this random variable

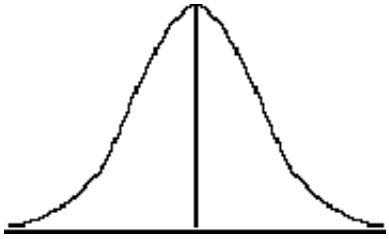
13. Which of the following is NOT a step in creating normal quantile plots?
    a. Calculate quantiles, using qi = (i - 0.5)/n
    b. Plot the data values on the x-axis and the quantiles on the y-axis
    c. Order the data from smallest to largest
    d. Calculate the z-score corresponding to each quantile

14. Weights of many dog breeds are normally distributed. West Highland Terrier (Westie) weights are normally distributed with a mean of 16.3 pounds and a standard deviation of 2.1 pounds. **For each question below, show your work, answer the question, clearly label the graph, and shade in the area that represents the percentage.**

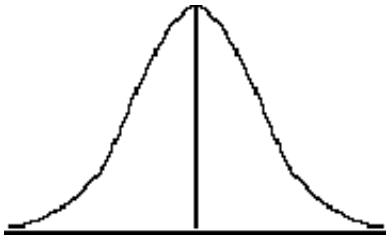
    a. What is the range for the central 95% of Westie weights? Use empirical rule.



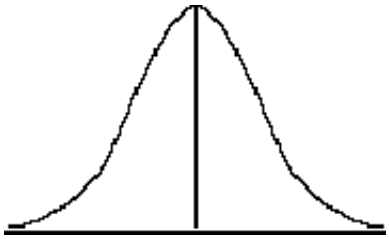
    b. A certain Westie breeder is accused of underfeeding his dogs. A sample of 25 dogs is taken from the breeder, and the dogs are weighed. What is the probability the sample mean is less than 14 pounds?



    c. Tom has a Westie that weighs 20 pounds. Tom thinks his dog might be eating too well. What percentage of Westies weigh as much as Tom's or more?



    d. How much do the highest 10% of Westies weigh?



15. Undergraduate students are classified into one of four groups: freshmen (35% of students), sophomores (28% of students), juniors (23% of students), or seniors (14% of students). Of those that are freshmen, 46% live on campus. Of those that are sophomores, 23% live on campus. Of those that are juniors, 17% live on campus. And of those that are seniors, 14% live on campus.
    a. What is the probability a randomly selected student lives on campus? Round your answer to 4 decimal places.

    b. If a student lives on campus, what is the probability they are a senior? Round your answer to 4 decimal places.

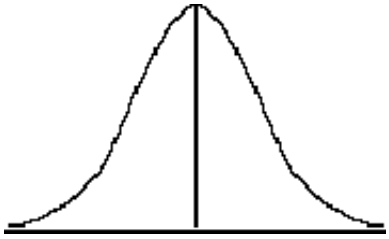
16. Roberto makes and sells two types of handmade soaps: lavender and goat's milk. In a week, he produces an average of 120 bars of lavender soap, with a standard deviation of 7 bars, and an average of 135 bars of goat's milk soap, with a standard deviation of 4 bars. The lavender bars sell for $6 per bar, while the goat's milk bars are $8 per bar. Let X be the total worth (in dollars) of soap produced per week (of both types)
    a. Find the mean of X

    b. Find the standard deviation of X. Round your answer to 2 decimal places.

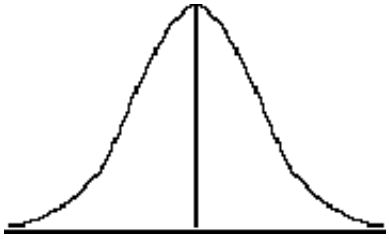
17. Suppose $P(B) = 0.7$, $P(D) = 0.4$, and $P(B \cap D) = 0.3$. Find $P(B \cup D)$.

18. Suppose $P(A) = 0.2$, $P(B) = 0.4$, and $P(A|B) = 0.25$. Find $P(A \cap B)$ and $P(B|A)$

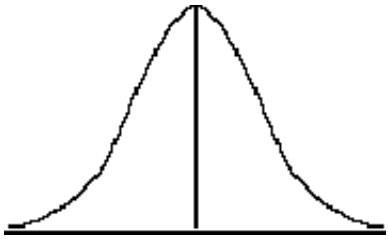
19. According to Harper's magazine, the time spent by kids in front of the television set per year can be modeled by a normal distribution with a mean of 1500 hours and a standard deviation of 250 hours.
    a. What percent of kids watch television for less than 1200 hours per year? (Answer the question, clearly label the graph, and shade in the area that represents the percentage).



    b. How many hours of TV do the highest 6 percent of kids watch? (Answer the question, clearly label the graph, and shade in the area that corresponds to the probability.)



    c. A researcher followed a random sample of 49 children for one year and calculated the mean number of hours that these 49 children watched television. Approximate the probability that the sample mean is between 1400 and 1600 hours. (Answer the question, clearly label the graph, and shade in the area that corresponds to the probability.



For each story in the questions below, choose which type of sampling or experimental design was used in the data collection process.

A. Simple Random Sample	B. Stratified Random Sample	C. Multistage Random Sample	D. Completely Randomized Design
E. Randomized Block Design	F. Voluntary Response Sample	G. Matched Pairs Design	H. Anecdotal Evidence

20. A marketing department of a major oil company wants to investigate whether cars get better mileage using their gas (Brand A) than using gas from another company (Brand B). They use 20 identical Ford Escorts for the study. They randomly assign 10 cars to Brand A and 10 to Brand B and record the mileage from cars in each group.

21. A marketing department of a major oil company wants to investigate whether cars get better mileage using their gas (Brand A) than using gas from another company (Brand B). They use 10 identical Ford Escorts and 10 identical Toyota Camrys for the study. They randomly select 5 Fords to receive Brand A and independently randomly select 5 Toyotas to receive Brand A. The remaining cars receive Brand B. They then conduct the experiment and compare the results.

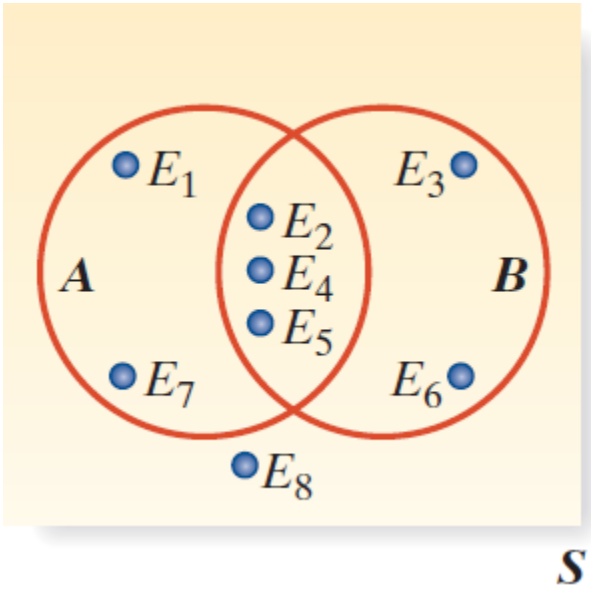
22. You plan to sample residents of registered nursing homes in your state. You randomly select 5 counties from your state. From the 5 counties you obtain a list of all nursing homes in those counties and randomly select 5 from each county. You then obtain a list of all residents in the 25 selected nursing homes and randomly select 5 individuals from each nursing home to sample.

23. In Fall 1995, the BBC in Britain requested viewers to call the network and indicate their favorite poem.

---

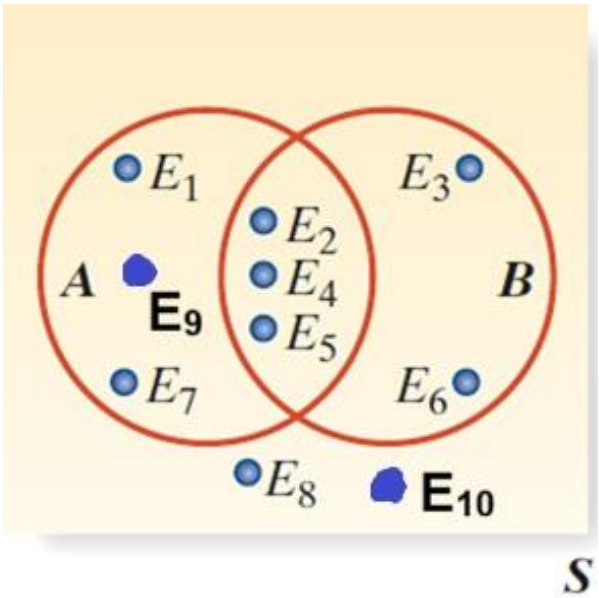
24. Which of the following will not necessarily decrease the bias?
    a. Increase the sample size.
    b. Use a probability-based sampling technique.
    c. Make sure all members of the population have the same chance of being selected into our sample.
25. Given the distribution function of a continuous variable, $f(x) = \frac{4}{27}x\left(x^2 - \frac{9}{2}x + 6\right), 0 \le x \le 3$
    a. Find P(X > 1.5)
    b. How would you set up finding the 70th percentile?

26. For the Venn diagram below, $P(E_1) = 0.11$, $P(E_2) = 0.06$, $P(E_3) = P(E_4) = 0.22$, $P(E_5) = 0.05$, $P(E_6) = 0.24$, $P(E_7) = 0.09$, $P(E_8) = 0.01$.



a. $P(A^c)$
b. $P(B^c)$
c. $P(A^c \cap B)$
d. $P(A \cup B)$
e. $P(A \cap B)$
f. $P(A^c \cap B^c)$

27. Modifying the above Venn diagram, prove whether or not $A \perp B$.



28. Assume that x is a random variable having a Poisson probability distribution with a mean of 5.8.
    a. Find P(x = 3)
    b. Which of the following commands in MATLAB would find P(x > 9)?
        A. poisscdf(9, 5.8)
        B. poisscdf(10, 5.8)
        C. 1 – poisscdf(9, 5.8)
        D. 1 – poisscdf(10, 5.8)
29. A store sells 2 different brands of smart TVs. Of its smart TV sales, 60% are brand A (less expensive) and 40% are brand B. Each manufacturer offers a 1-yr warranty on parts and labor. It is known that 25% of brand A's smart TVs require warranty repair work, whereas 10% for brand B.
    a. What is the probability that a randomly selected purchaser bought a smart TV of Brand A and will need repair while under warranty?

    b. Now if a customer returns to the store with a smart TV that needs warranty repair work, what is the probability that it is Brand A? Brand B?

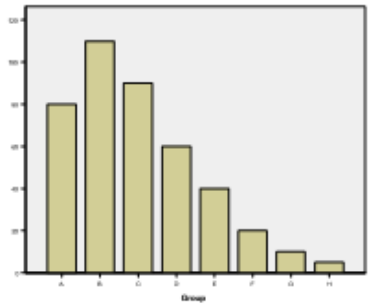
30. In a hospital emergency room, 3 patients arrive on average every 20 minutes. Let x be equal to the number of patients arriving at the emergency room in any 20-minute period. Assume that x has a ppo distribution with mean 3.
    a. What is the probability that 5 patients arrive at the emergency room in the next 20 minutes?
    b. Which MATLAB command would be used to calculate the probability of 5 or more patients coming in the next 20 minutes?
        A. poisscdf(5,3)
        B. poisscdf(4,3)
        C. 1 – poisscdf(5,3)
        D. 1 – poisscdf(4,3)

For the next 6 questions, choose A, B, C, D, E, or F from the table below. Each answer may be used once, more than once, or not at all.

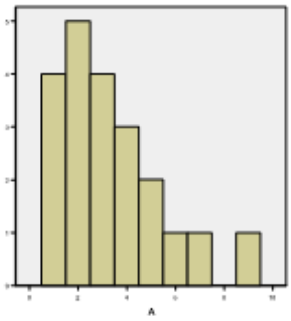
A

0	1
1	
2	
3	
4	2
5	1 4
6	0 3 7
7	1 2 3
8	4 5 9 9
9	4 5 5 5 7
10	3 4 7 8 9 9
11	0 1 1 4 5
12	1 2 3 4

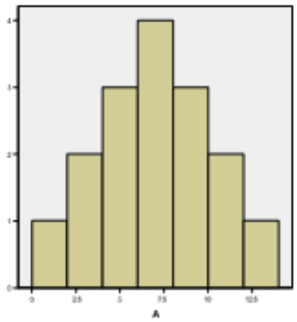
B



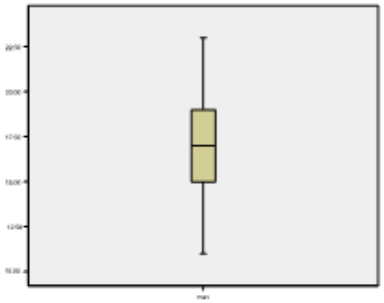
C



D



E



F

None of the Above / No graph could tell you this information

31. Which graph shows you data in which the mean is greater than the median?

32. Which graph clearly shows the 25th percentile?

33. Which graph should be used only for categorical data?

34. Which graph has a distribution that is similar to the distribution in graph D? (Do not choose D as your answer!)

35. Which graph shows a distribution that is skewed right?

36. Which graph shows you data that has undercoverage bias?

---

37. In a quality control process, there exists a known defect rate for a manufacturing process of 0.05. What is the probability that a defective part is encountered on the 8^{th} part checked? What is the expected value of the number of parts checked per defect?

Answer the next two questions based on this story. Assume that the development time for a particular type of photographic printing paper when it is exposed to a light source for five seconds has a normal distribution with mean, μ = 202 seconds, and standard deviation, σ = 4 seconds.

38. What is the probability that a particular print will require more than 205 seconds to develop?
    a. 0.0329
    b. 0.2266
    c. 0.75
    d. 1.84
    e. 0.7734
39. Suppose you wanted to find the probability that the mean of a random sample of 20 trials is more than 205 seconds. Which distribution below would you use to standardize your probability statement?
    a. $\bar{X} \sim N(202,4)$
    b. $\bar{X} \sim N(202,4/\sqrt{20})$
    c. $\bar{X} \sim N(205,4)$
    d. $\bar{X} \sim N\left(205,4/\sqrt{20}\right)$
    e. None of the above

---

40. The magazine called Literary Digest conducted a poll to predict the results of the 1936 Presidential election between Franklin Roosevelt (Democrat and incumbent) and Alf Landon (Republican). They mailed questionnaires to 10 million people and asked how they planned to vote. The list that they sampled from was constructed from telephone directories, country club memberships, and automobile registrations. At the time, the United States was in the Great Depression; those who had cars, phones, and country club memberships and thus received questionnaires tended to be relatively wealthy. Consequently, they had sampling bias due to
    a. Response bias.
    b. Nonresponse.
    c. Undercoverage.
    d. None of the above.

---

41. Suppose that 46% of adults get enough sleep, 40% get enough exercise, and 27% do both. Find the probabilities of the following events, writing the applicable probability statement that shows how to solve:

    a. Enough sleep and not enough exercise
    b. Not enough sleep and enough exercise
    c. Not enough sleep and not enough exercise

Match each of the following with the SINGLE choice that best describes the given situation. Each answer choice may be used once, more than once, or not at all.

A. Simple Random Sample	B. Stratified Random Sample	C. Multistage Random Sample
D. Voluntary Response sample	E. Capture-recapture sample	F. Completely Randomized Design
G. Randomized Block Design	H. Matched Pairs Design	I. Anecdotal Evidence

42. A study is aimed at estimating the average number of bilingual students per middle school in NC. 45 counties in North Carolina were selected randomly, from that 20 middle schools were randomly selected, and 35 randomly selected classrooms.
43. A second-grade teacher wanted to teach his students about the importance of water for plant growth. Twenty-seven pumpkin seeds were randomly selected from a pack of seeds and were planted in pots. Nine randomly selected seeds were watered only once a week. Another nine randomly selected seeds were watered twice a week. The remaining nine seeds were watered once every two weeks. Students compared how much the pumpkin plants grew after a month under the three different scenarios.
44. A study wanted to check whether milk production was affected by the type of grain mix cows are fed. Two manufacturing brands for grain mix were selected for the study. Eighty dairy cows were randomly selected out of different dairy farms. It turned out that there were four different breeds among all these cows. Before assigning the cows to be fed either of the two different brands of grain mix, they were separated into the four breeds. At random half the cows of each breed were fed one of the grain mix brands and the other half were fed the other grain mix brand.
45. Forty randomly selected women participated in a study aimed at comparing the results of two different hand creams. For a month, half of the women were asked to apply one of the creams morning and night. They were asked to not use hand cream for two weeks and then they tried the other hand cream morning and night for a month. The other group of twenty did the same, except they switched the order. Their satisfaction with the results of a cream was recorded as a number on a scale from 0 to 5 (0 – not good, 5- very good) and the difference was compared.

---

46. According to a Gallup poll, 60% of American adults prefer saving over spending. Let X = the number of American adults out of a random sample of 50 who prefer saving to spending.
    a. Find the probability that 25 adults in the sample prefer saving over spending. (Simply write the formula with numbers filled in, but solving is not required).

    b. Find the mean and standard deviation of X.

Below are the dollar amounts corresponding to receipts for 10 tables in a restaurant:

11 22 23 31 32 33 35 40 41 50

47. Determine the 5-number summary.

_____, _____, _____, _____, _____

48. Create a stem-and-leaf plot of the data.

49. The two statistics that best describe this data set are:
    a. Mean and standard deviation
    b. Median and standard deviation
    c. Mean and IQR
    d. Median and IQR

50. Calculate the interquartile range (IQR).

---

51. What statistics will be most influenced by the presence of outliers?
    a. Mean and standard deviation
    b. Median and Quartiles

The price of a gallon of fresh, whole milk (let's use X to denote this variable) follows a Normal distribution with a mean of 3.5 dollars and a standard deviation of 0.02 dollars.

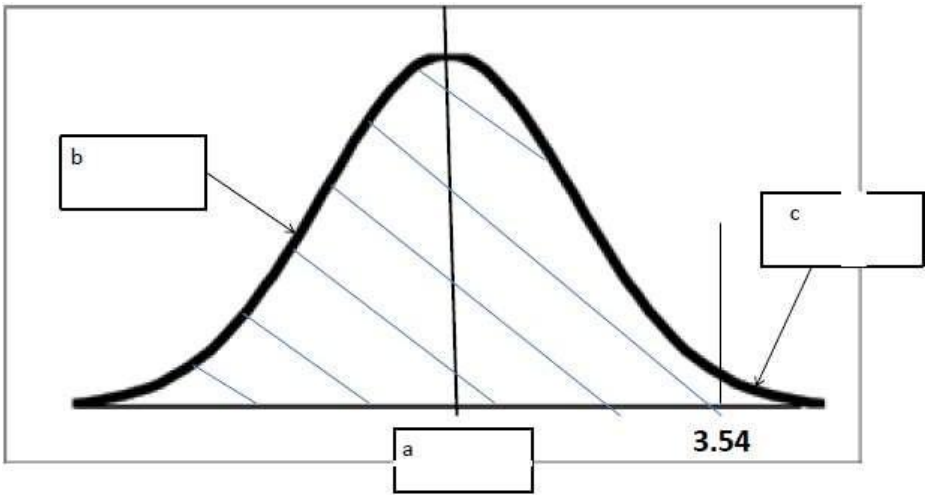
You are interested in knowing the probability that the gallon of whole fresh milk that you buy at the local grocery store costs more than 3.54 dollars.

52. This question refers to
    a. The sampling distribution of the mean
    b. The distribution of X.

53. Write down the probability statement.

54. What is the value of this probability? Use the empirical rule only. No calculator. Show all work.

55. Below is the density curve corresponding to the distribution of variable X (price of a gallon of fresh, whole milk). Fill out the empty boxes: a – value of the test statistic at the center? , b – value of shaded area?, c – value of the non-shaded area? (Round the values for b and c to 4 decimal places).



You are now interested in knowing what is the probability that the average price of 20 gallons of fresh whole milk is more than 3.52 dollars.

56. To denote this 3.52 dollars, you will use:
    a. x
    b. $\bar{x}$
57. The question refers to
    a. The sampling distribution of the mean
    b. The distribution of X

58. Write down the probability statement.

59. What is the value of this probability? (Round to 2 decimal places.) Show all work.

60. The diameter of the trunk of 200 sequoia trees was measured for a study to estimate the average size of sequoia trees. Which of the following is true?
    a. The average of the maximum diameters of the 200 sequoia trees is a statistic (or estimate).
    b. The average of the maximum diameters of the 200 sequoia trees is a parameter.
    c. The average of the maximum diameters of all sequoia trees is a statistic.

Match each of the following with the choice that best describes the given situation. Each answer choice may be used once, more than once, or not at all.

A. Simple Random Sample	B. Stratified Random Sample	C. Multistage Random Sample
D. Voluntary Response sample	E. Capture-recapture sample	F. Completely Randomized Design
G. Randomized Block Design	H. Matched Pairs Design	I. Anecdotal Evidence

61. A horticulturist wishes to estimate the mean growth of seedlings in a large timber plot last year. A random sample of n = 30 seedlings is selected and the one-year growth for each is measured.

62. There are seven sections of an introductory psychology class. A random sample of three sections is chosen, and then random samples of 8 students (4 males, 4 females) from each selected section are chosen to participate in a survey.

63. Does air pollution induce DNA mutations in mice? Starting with 40 male and 40 female mice, 20 randomly selected mice of each sex were housed in a polluted industrial area downwind from a steel mill. The other 20 of each sex were housed at an unpolluted rural location 30 km away.

---

64. Using the same horizontal scale, how would the graph of the Normal density curve showing the distribution of the sample mean for a sample of size 50 be affected if the sample size was changed to 100 (samples out of the same population)?

    a. The center of the distribution would be shifted to the right.
    b. The height of the curve would increase.
    c. The area under the curve would increase.
    d. The curve would be skewed to the right instead of symmetric.
    e. The width of the curve would increase.

65. A study showed that people who watch more television also have higher cholesterol levels. Does watching more television cause cholesterol numbers to increase? Identify a lurking variable that might help explain this relationship

After a year of business, a store owner is concerned that the age of his clientele is not well distributed to the younger generation. So, he decides to collect the ages of 13 randomly selected customers. The following data were collected:

22 36 40 42 43 47 47 48 51 53 55 60 72

66. Determine the 5-number summary.

_____, _____, _____, _____, _____

67. Create a stem-and-leaf plot of the data.

68. Identify another type of graph that you could use to visually display this data, and tell why that type of graph would be appropriate.

69. Are there any outliers in the data set? Show all work. Identify any points that are outliers.

---

70. When you donate blood, they draw samples to test for various diseases. Pooling samples is a way to shorten screening time and reduce the number of tests needed and was used during the height of the COVID pandemic. Let's say the incidence of COVID in the population is 5%. If we pool samples in groups of 8, what is the probability that the pooled sample would test positive?

71. We know $P(A) = 0.5$ and $P(B) = 0.1$. What else do we know?

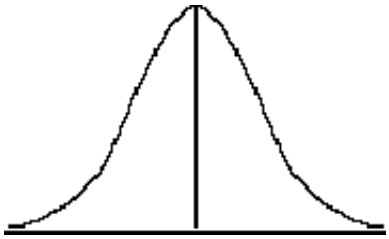
    a. P(A ∩ B) = 0.05 only if A and B are disjoint
    b. P(A ∩ B) = 0.05 only if A and B are independent
    c. P(A U B) = 0.05 only if A and B are disjoint
    d. P(A U B) = 0.05 only if A and B are independent

The amount of points the NC State women's basketball team scores per game is normally distributed with a mean of 72 and a standard deviation of 6 points.

72. How many points scored gives you the lowest 9% of all scores. Write your probability statement, show your work, and give your answer to 2 decimal places.

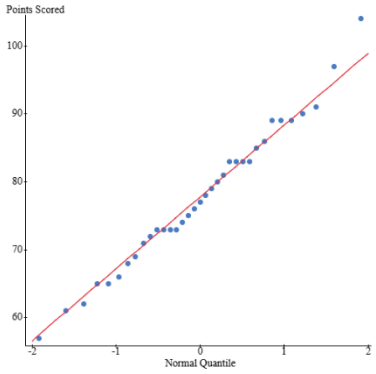
73. What is the probability that the team scores higher than 85 points in a game? Write your probability statement, show your work, and give your answer to 4 decimal places.

74. Use the Normal curve below to display the information to scale for question 71. Identify the mean, the selected score of 85, and shade in and label the correct area associated with the probability calculated in part b.



75. If you take a sample of 5 games, what is the probability that the average score of these 5 games is less than 68 points? Give your answer to 2 decimal places.

76. Below is a normal quantile plot of the points scored by the NC State women's team in the 2024-2025 season. Based on this graph, we can conclude:
    a. The scores are approximately normally distributed
    b. The scores are strongly skewed left
    c. The scores are strongly skewed right
    d. The scores are uniformly distributed.



A fertilizer company wants to know the average number of acres of soybeans grown per farm in North Carolina. They divide all farms into four classes depending on their size. From each class, they select a sample of 15 farms and count and record the number of acres of soybeans on each selected farm.

77. Which of the following is the unit for this study?
    a. an acre of ground on a farm in North Carolina.
    b. a farm in North Carolina.
    c. a single soybean plant
    d. the average number of acres of corn grown per farm in North Carolina.
    e. You are unable to determine the unit from the information given.
78. Which of the following represents the sample for this study?

    a. All farms in North Carolina.
    b. The 60 selected farms.
    c. The 4 classes of farms.
    d. The 15 selected farms.
    e. You are unable to determine the sample from the information given.

79. The response variable for this problem is

    a. A farm in North Carolina.
    b. The average size of farms in North Carolina.
    c. The size of a selected farm in North Carolina.
    d. The average number of acres of soybeans grown per farm in North Carolina.
    e. The number of acres of soybeans grown on a selected farm in North Carolina.

80. What type of sampling design was used in this study?

    a. A simple random sample.
    b. A stratified random sample.
    c. A multistage random sample.
    d. A voluntary response sample.
    e. A randomized block design.

81. Suppose some of the smaller farms were mistakenly left out of the study. What potential source of sampling bias will occur?
    a. Undercoverage
    b. Non-response
    c. Response bias
    d. No bias will occur
82. Suppose some of the farmers from the selected farms refused to give out any information. What potential source of sampling bias will occur?
    a. Undercoverage
    b. Non-response
    c. Response bias
    d. No bias will occur

83. Use the following information to determine which family makes a higher income relative to the other families in their own neighborhood.

	Family's Income	Average Income for Neighborhood	Standard Deviation for Neighborhood
Brown Family	$36,000	$28,000	$5,500
Andrews Family	$35,000	$30,000	$4,000

---

84. Data show that men who are widowed earn more than men that have not married.
    a. This is a clear relationship of causation.
    b. Lurking variables could be behind the observed relationship between marriage status and earnings.

Match each of the following with the choice that best describes the given situation. Each answer choice may be used once, more than once, or not at all.

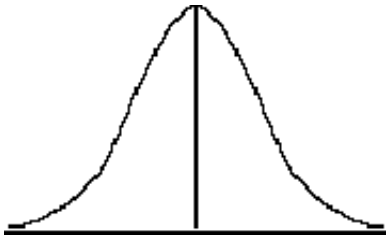
A. Simple Random Sample	B. Stratified Random Sample	C. Multistage Random Sample
D. Voluntary Response sample	E. Capture-recapture sample	F. Completely Randomized Design
G. Randomized Block Design	H. Matched Pairs Design	I. Anecdotal Evidence

85. A farmer is interested in studying the milk yield from his dairy cows. He has three types of dairy cows: Jersey, Brown Swiss and Guernsey. He takes a random sample of cows from each type and measures their milk production over a period of 10 weeks.
86. How long does the average commercial break last on the network television stations? On eight randomly selected days of the month, 5 time slots are randomly chosen and one show is randomly selected from each time slot. The length of each commercial during this time is recorded, and the average determined.
87. A Stat 370 TA is interested in how the Stat 370 labs are going this semester for her classes. She asks students to send her feedback on their labs by e-mail. Some of the students respond, and she records the number of positive and negative responses from these e-mails.
88. A consultant randomly selects 24 weekdays out of a period of 3 months and records the total amount in fees collected from neighboring parking structures in order to determine the feasibility of a new parking garage.
89. Do children pick up foreign language vocabulary more quickly if it is spoken or written? A group of 40 students is randomly selected. Twenty students are randomly selected from the group of 40 and given a list of 30 words orally in French followed by the oral English translation. The other 20 are given their words in written format in both languages. After going through the list twice, students are given a quiz (orally or written) on the words. The number of words correctly identified is recorded.

Credit card applicants have an average credit rating score of 667. Assume the distribution of credit scores is Normal with a standard deviation of 65. Use this information to answer the problems below. **Write probability statements and show all of your work.**

90. What is the probability that a single applicant for a credit card will have a credit rating score above 700?

91. Use the Normal curve below to display the information for the credit scores. Show and label the following items to scale on the plot: the mean of 667, the selected score of 700 and the probability calculated in the question above.



92. What is the probability that 10 applicants will have an average credit rating below 655?

93. Above what credit score would you find the top 10% of all individual applicants?

---

94. Suppose that X is a random variable with mean 20 and standard deviation 2. Also suppose that Y is a random variable with mean 40 and standard deviation 7. Assume that the correlation between X and Y is zero. Find the mean, standard deviation, and variance of the random variable Z for each of the following cases. Be sure to show your work.
    a. $Z = 25 – 12X$
    b. $Z = 13X – 8$
    c. $Z = X + Y$
    d. $Z = X – Y$
    e. $Z = -3X + 3Y$

95. A random sample of 262 subjects was found and the smoking habits and caffeine consumption of each subject was recorded. The data is summarized in the table below.

	No Caffeine	Occasional Caffeine	Regular Caffeine	TOTAL
Non-smoker	41	36	78	155
Smoker	12	34	61	107
TOTAL	53	70	139	262

a. Of those randomly selected subjects that do not smoke, what proportion occasionally consume caffeine?

b. What is the probability that a randomly selected subject is a smoker or occasionally consumes caffeine?

c. Show that being a smoker is dependent on regularly consuming caffeine.

96. You've scheduled a 3-day weekend at the beach. There is a 20% chance of rain each day. If rain on any day is independent of the other days, what is the probability that it will rain at least once during your trip?
   a. 0.80
   b. 0.512
   c. 0.488
   d. 0.16
   e. 0.128

97. A discrete random variable X has a probability distribution as given below.

x	-1	0	1	2
P(x)	0.1	0.3	0.1	

a. What is the expected value of X?

b. What is the standard deviation of X?

Prices for a woman's haircut (let's use X to denote this variable) in **all** the hair salons in a certain big city are recorded. These prices follow a Normal distribution with a mean of 30 dollars and a standard deviation of 5 dollars.

We are interested in knowing the probability that there is a hair salon in that city that will give a woman's haircut for less than 15 dollars. **FOR ALL PARTS, SHOW ALL WORK.**

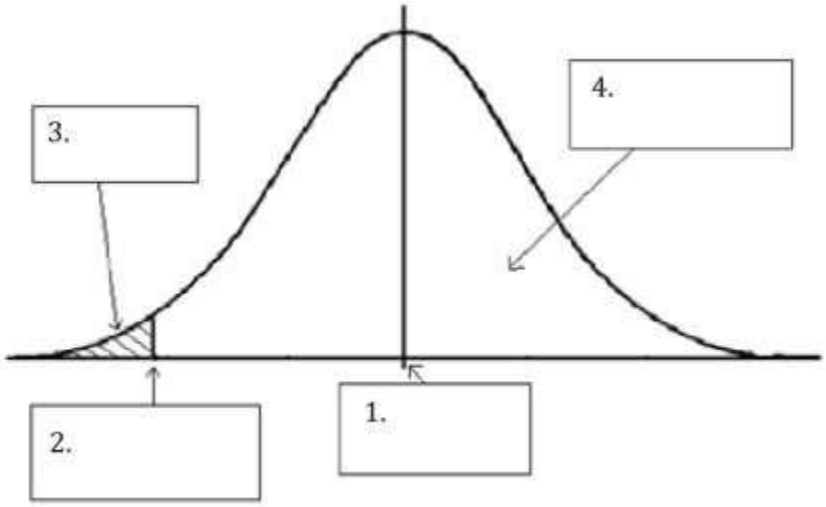
98. Write down the probability statement.

99. Calculate the z-score for X=15 (Round to two decimal places).

100. What is the probability of interest? Use the empirical rule.

101. Below is the density curve corresponding to the Standard Normal distribution. Fill out the empty boxes (Round the value of 2. to two decimal places and the values for 3. and 4. to four decimal places):

a. the value at the center.
b. the value of the z-score for X=15.
c. the value of the shaded area to the left of the z-score.
d. the value of the area to the right of the z-score.



102. At a small liberal arts college, students can register for one to six courses. Let X be the number of courses taken in the fall by a randomly selected student from this college. In a typical fall semester, 6% take one course, 6% take two courses, 12% take three courses, 20% take four courses, 41% take five courses, and 15% take six courses. Let X be the number of courses taken in the fall by a randomly selected student from this college. Describe the probability distribution of this random variable.

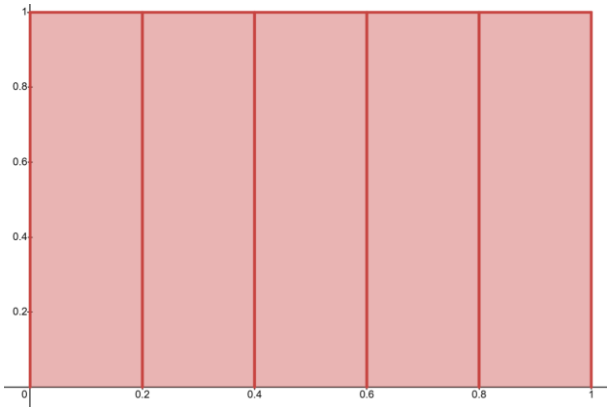
Assuming that each course is 3 credit hours:

a. Find the probability that a randomly selected student earns more than 18 credits.

b. Find the probability that a randomly selected student earns 6 or fewer credits.

c. Find the probability that a randomly selected student earns 15 credits or more.

103. Let the random variable X be a random number with the uniform density curve below.



a. $P(X \geq 0.40)$
b. $P(X = 0.40)$
c. $P(0.40 < X < 1.40)$
d. $P(0.22 \leq X \leq 0.25 \; or \; 0.42 \leq X \leq 0.45)$
e. X is not in the interval 0.5 to 0.8

104. In the population of young children eligible to participate in a study of whether or not their calcium intake is adequate, 52% are 5 to 10 years of age and 48% are 11 to 13 years of age. For those who are 5 to 10 years of age, 18% have inadequate calcium intake. For those who are 11 to 13 years of age, 57% have inadequate calcium intake. Use Bayes' rule to find the probability that a child from this population who has inadequate intake is 5 to 10 years old.

105. Suppose $P(A)$=0.2 and $P(B)$=0.6.
   a. Explain what it means for A and B to be disjoint. Assuming that they are disjoint, find the probability that A or B occurs

   b. Explain in your own words the meaning of the rule P(S)=1

   c. Consider an event A. What is the name for the event that A does not occur? If P(A)=0.7, what is the probability that A does not occur?

   d. Suppose that A and B are independent and that P(A)=0.8 and P(B)=0.3. Explain the meaning of the event {A and B}, and find its probability.

   e. Suppose P(A)=0.7 and P(B)=0.6. Can A and B be disjoint? Explain your answer.

106. Suppose the distribution of a continuous variable is given by:

$$f(x) = \frac{1}{32}(x^3 - 9x^2 + 20x)$$

And the function is defined over the interval [0,4].

   a. Find P(1 ≤ x < 3)
   b. Find the expected value of f(x)
   c. Find the standard deviation of f(x)
   d. Using technology, find the 40th percentile

107. Subhash forgot his bike lock code. It consists of 4 letters. He thinks it has a W in it. What is the probability that the bike lock code has at least one W?
   a. 0.0385
   b. 0.1538
   c. 0.1452
   d. 0.8548
   e. 0.9615

108. For a continuous random variable X, the probability density function is:

$$f(x) = \frac{1}{8}\ for\ 0 \le x \le 8\ and\ 0\ otherwise$$

   a. What is P(0 ≤ X ≤ 2)?

   b. What is P(X = 6)?

   c. What is the expected value of X?