Prelim 1 for BTRY6010/ILRST6100

September 29, 2016: 7:30pm-9:30pm

Name:_____

Lab: (circle one)

Lab 402: Tues 1:25PM - 2:40PM

Lab 403: Tues 2:55PM - 4:10PM

Lab 404: Wed 2:55PM - 4:10PM

Lab 405: Tues 7:30PM - 8:45PM

Score: _____ / 75

Instructions

1. Please **do not turn to next page** until instructed to do so.
2. You have 120 minutes to complete this exam.
3. The last page of this exam has some useful formulas.
4. No textbook, calculators, phone, computer, notes, etc. allowed (please keep your phones off or do not bring them to the exam).
5. Please answer questions in the spaces provided.
6. When asked to calculate a number, it is sufficient to write out the full expression in numbers without actually calculating the value. E.g., $\frac{1+3\times\frac{4}{7}}{3+0.7}$ is a valid answer.
7. Please sign the following statement before beginning the exam.

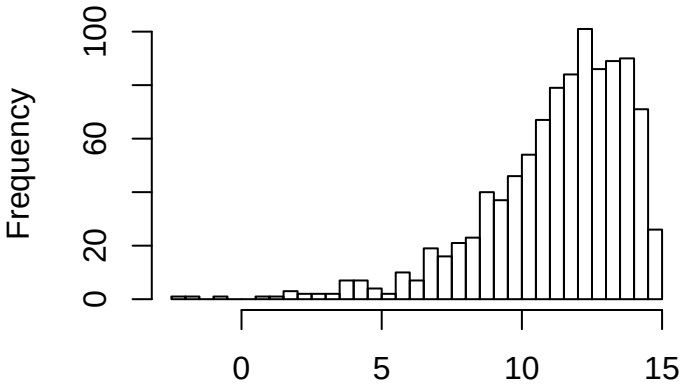
Academic Integrity

I, _____, certify that this work is entirely my own. I will not look at any of my peers' answers or communicate in any way with my peers. I will not use any resource other than a pen/pencil. I will behave honorably in all ways and in accordance with Cornell's Code of Academic Integrity.

Signature: _____ **Date:**_____

For each problem in this section, please circle one of the following answers. No justification is required.

1. [2 points]



The histogram shown above represents data that are

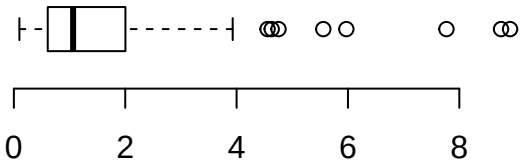
a) skewed right

b) skewed left

c) not skewed

d) not enough information provided

2. [2 points]



The box plot shown above represents data that are

a) skewed right

b) skewed left

c) not skewed

d) not enough information provided

3. [2 points]

Which of the following can be used effectively to display two numerical variables?

a) color-coded box plots

b) a histogram

c) stacked bar plots

d) none of the above

True or False? For each of the following questions, please answer either *True* or *False*. While justification is not required, it is encouraged and may allow in some cases for partial credit to be awarded.

4. [2 points]

If a distribution is bimodal, then it must be symmetric.

5. [2 points]

If events A and B are disjoint, then $P(A \cup B) = P(A) + P(B)$.

6. [2 points]

If A and B are independent, then $P(A|B) = \frac{P(A)}{P(B)}$.

7. [2 points]

The events A and A^C are disjoint.

8. [2 points]

(Recall that a deck of playing cards has 26 red cards and 26 black cards.) Suppose you shuffle a deck of cards thoroughly and then select the top card. Let X be 1 if the card is red and 0 if the card is black. The random variable X is Bernoulli(0.5).

9. [2 points]

You shuffle a deck of cards thoroughly and then select the top 5 cards. Let X be the number of red cards. The random variable X is Binomial$(5, 0.5)$.

Please answer the following question.

10.

a) [5 points] Suppose you want to know the average length of leaves grown in a forest. In each case below, give the name of the type of sampling described:

i. You take a leaf from the lowest branch of some trees close to your lab.

ii. You get a simple random sample of trees; then, on each of these trees, you select a simple random sample of branches; then, on each branch, you select a simple random sample of leaves.

iii. You walk through the forest with the park ranger who points out which leaves seem to be the most typical/representative of the overall population of leaves in this forest.

iv. For each species of tree, you get a simple random sample of leaves.

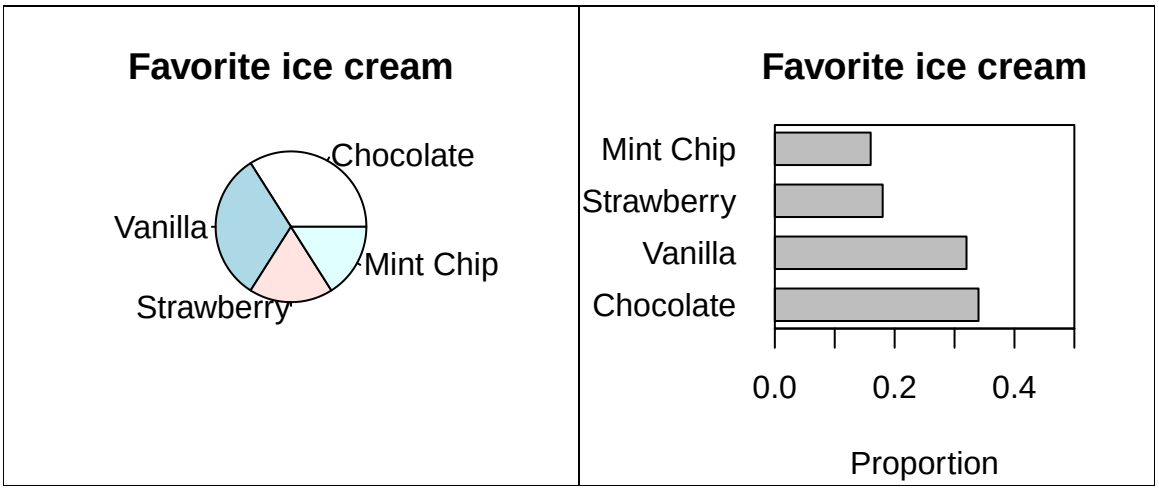
v. You walk through the forest gathering leaves until you have 50 from each species of tree.

b) [2 points] Describe why it would be difficult to get a simple random sample of leaves in the forest.

Please answer the following questions. An answer without justification will not receive full credit.

11. [2 points] Draw a histogram in which the mean is higher than the median.

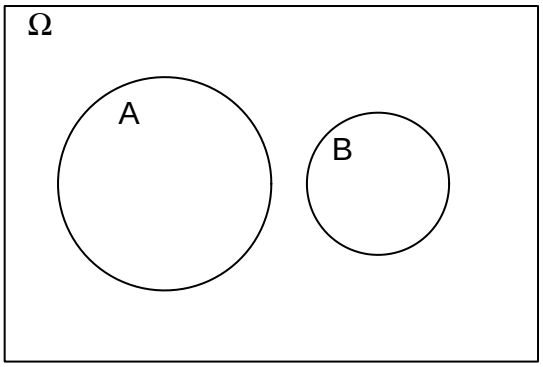
12. [3 points] Here are two ways to display the same data. Which is the more effective graphic? The more justification you give, the better.



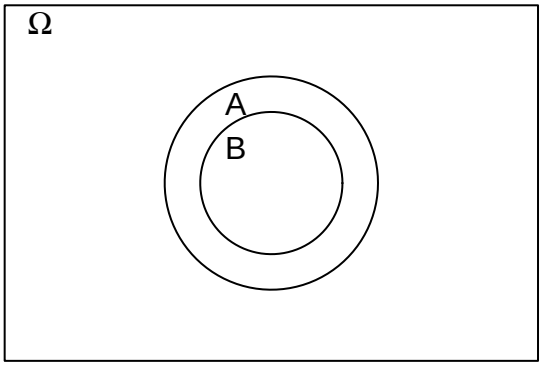
13.

Are the events A and B shown in the picture independent? Carefully justify your answer in each case.

a) [2 points]



b) [2 points]



14. Here is a box plot of ages of students in a certain class.

a) [2 points] Approximately 75% of students are at least _____ years old.

b) [2 points] Approximately _____% of students are at least 25 years old.

b) [2 points] Explain why the lower whisker does not extend to the 20 year old student.

b) [2 points] Suppose the 20 year old drops the class but a 21.5 year old joins the class. Draw what the boxplot would look after these changes.

15. Suppose you record the year, brand, and price of 200 cars.

a) [2 points] Describe a single kind of visual display that would effectively show all this information.

b) [2 points] Suppose there are only two brands of car considered (say, Ford and Honda). Suppose that (i) the price of cars has increased over time on average, (ii) Fords tend to be on average more expensive than Hondas in any given year, and (iii) the price of Hondas in each year has greater variance than the price of Fords. Draw the visual display from part (a) to illustrate the data in this case.

16. [4 points]

Suppose that $P(A|B) = P(A)$. Does this imply that $P(B|A) = P(B)$?

17. The fire alarm in your apartment building will go off (nearly) 100% of the time when there is a fire. However, the alarm is not perfect, especially during the summer. On a very hot and humid day, it has a 5% chance of going off even if there is no fire. The fire department knows that there have been only 2 real fires in the last 1000 hot and humid days and has no reason to think that there has been a rise or drop in the rate of fires.

a) [1 point] Based on this last sentence, what is the approximate probability that there is an actual fire on a given hot and humid day?

b) [2 points] On what fraction of hot and humid days do we expect the alarm to go off?

c) [2 points] Suppose the alarm in your building goes off on a hot and humid day. What is the probability that there is a fire?

d) [4 points] What is the sensitivity and specificity of the alarm? (Hint: Think of a fire as having a disease, and think of the alarm going off as a test giving a positive.)

18. You work at the dairy bar, serving ice cream, on Friday afternoons. You know that on average 4 people buy an ice cream every minute on Friday afternoons. Let X be the number of people who will buy an ice cream between 2:00pm and 2:01pm next Friday afternoon.

a) [3 points] What sort of probability distribution would be appropriate for modeling X? Justify your answer. In addition to naming the type of distribution, also specify the values of all parameters.

b) [2 points] What is the chance that no one buys an ice cream in that minute?

c) [2 points] What is the chance that more than 2 people buy ice cream in that minute?

d) [3 points] What is the expected value and standard deviation of the number of people buying ice cream in that minute?

19. Suppose you are taking a multiple choice exam with 10 questions and 4 choices per question. Suppose you are random guessing (you don't read the question, you just select a random choice on each question). Let X be the number of correct answers you get.

a) [2 points] What is the distribution of the random variable X? (Be sure to give the name of the distribution and the values of its parameters.)

b) [2 points] What is the probability of getting a 90% or higher on the exam?

c) [4 points] Suppose instead that you have studied for this exam, so that on any given problem you have a 60% chance of getting the right answer. Your friend has studied less and has a 40% chance of getting the right answer on any given problem. We would like to know the probability that you get a higher score on the exam than your friend. Describe (in words – no need to write any R code!) the method you could use to answer this question using a computer. Give

   i. the general name of the technique you would use and also
   ii. describe the basic idea behind what you are doing (imagine you are describing this to someone in your research group who is not in our class and does not know R). You should include in your explanation the definition of probability.

Formula Sheet

The law of total probability:

$$P(B) = P(A)P(B|A) + P(A^C)P(B|A^C)$$

Bayes' theorem:

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^C)P(B|A^C)}.$$

Binomial distribution: $X \sim \text{Binomial}(n, p)$

- probability mass function (pmf):

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0, 1, 2, \ldots, n$$

- expected value $E(X) = np$
- variance $\text{Var}(X) = np(1-p)$

Poisson distribution: $X \sim \text{Poisson}(\lambda)$

- probability mass function (pmf):

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \text{ for } x = 0, 1, 2, \ldots$$

- expected value $E(X) = \lambda$
- variance $\text{Var}(X) = \lambda$