Due: midnight, Sep 6 (Thursday labs); midnight, Sep 7 (Friday labs)

Lab 2: The binomial probability distribution

Objectives

- Calculate binomial probability distributions and cumulative distributions, by hand using basic R functions, and more efficiently using specialized probability functions.
- Plot binomial probability distributions and cumulative distributions.
- Calculate the mean, variance, and standard deviation of binomial random variables.
- Calculate quantiles of a binomial probability distribution.
- Generate random samples from a binomial distribution.
- Use the following R functions: `factorial`, `dbinom`, `pbinom`, `qbinom`, `rbinom`, `seq`, `:`, `par`

Background

You hypothesize that a certain lizard species has equal sex ratios. That is, the probability that any single egg produces a female offspring is 50%. To test this hypothesis, you plan to collect a sample of nine eggs, let them hatch, and determine the sex of each offspring. Before doing this, you want to calculate the probability of all possible outcomes of this experiment, assuming that your hypothesis is correct. Once you collect your sample, these calculations will help you determine whether your hypothesis is likely to be true.

Exercises

1. Consider one possible outcome: three females and six males. Write an R script to calculate and print the following (Do not use R's binomial probability functions in this script):
   - The total number of combinations in which three of the offspring are female and the other six are male.
   - The probability of each of these combinations.
   - The total probability that the sample will have three females and six males.

   Answer this question: What are five of the possible combinations with three females and six males? Pick any five you please.

2. Now write a script to calculate the full probability distribution for all possible outcomes, in terms of number of females in the sample. Again, do not use R's binomial probability functions. You should be able to do this by copying your script from exercise 1 and making a very small modification of it. Your script should calculate and print:
   - The total number of combinations for each possible outcome.
   - The probability of each of these combinations.
   - The total probability for each outcome.

3. Write a script that does the following:
   - Calculate the full probability distribution of outcomes, but this time use the appropriate R binomial probability function.
   - Make a bar plot of the probability distribution. Label each axis, each bar, and the whole plot. Make sure the y-axis is taller than the tallest bar.
   - Calculate the mean of the probability distribution, using the general definition of the mean:

     $$\mu = \sum_x x P_X(x)$$

   - Re-calculate the mean, using the shortcut formula given in lecture.
   - Calculate the variance and standard deviation of the probability distribution, using the general definition of the variance:

     $$\sigma^2 = \sum_x (x-\mu)^2 P_X(x)$$

   - Re-calculate the variance and standard deviation, using the shortcut formulas given in lecture.

   Run this script (from the start of Exercise 3) several times, each time using a different value for the probability that an offspring is female. Use the following values: 0.1, 0.3, 0.5, 0.7, 0.9. (You can also try other values, if you like.) Because you are running the same script and merely changing the inputs, you do NOT have to paste new copies of your script into your lab report or include all the output. Instead, **make a table showing the mean and variance for each value of p, and then answer the following questions:** What happens to the mean and variance of the distribution as the probability changes? At what value is the variance of the distribution highest? lowest?

4. Write an R script that makes two plots, one above the other: the probability distribution calculated in question 3, and the corresponding cumulative probability distribution. That is, for each outcome, show the probability of seeing that many or fewer females.
   Answer this question: What is the relationship between each bar of the cumulative plot and the bars of the distribution plot?

5. Write a script that uses R's binomial probability functions to answer the following questions:
   - What is the probability of exactly 4 females?
   - Exactly 7 females?
   - 7 or fewer females?
   - 4 or more females?
   - 4 or more males?
   - 7 or more offspring that are all the same sex?

6. Write a script that uses R's binomial probability functions to answer the following questions:
   - What is the 1st quartile of the number of females in a sample of nine eggs?

- What is the 5th percentile?
- What is the 95th percentile?
- What is the median?

7. You can use the function `rbinom` to simulate your planned experiment. That is, it can generate one or more values of a binomial random variable with specified values of n and p. Write a script that does the following:
   - Simulate the planned experiment 100 times. That is, create 100 iterations of the number of females in a sample of nine offspring, assuming an equal sex ratio.
   - Calculate the average of the 100 values for number of females.

   Answer these questions: Compare the average you calculated here with the mean you calculated in question 3 (specifically, the mean based on an equal sex ratio). Do you expect them to be the same? Why or why not?

Assignment: Turn in a single Microsoft Word document with a heading that includes your name, the lab number (Lab 2 for this lab), and your section (day and time). Create a subheading for each exercise. For each exercise, include (1) your annotated R script, (2) your annotated output, and (3) your answer to the final "Answer" item for each exercise (where applicable). In addition to the word document, you must also submit a separate .R file with all of your R scripts.