

Project 6

ALY 6000

Project Instructions

For your final project, you will use R to solve problems about probability distributions. Specifically, you will make use of the `d`, `p`, `q`, and `r` functions built into R for working with probability distributions. In most cases, you will need to determine the type of probability distribution that is described and use R to determine a numerical answer.

Setting up Your Project

When working in RStudio, be careful with cloud drives. RStudio does **not** always play well with cloud drives' longer file path names. You will be best served by using a local drive whenever possible until you are comfortable troubleshooting technical issues.

1. Create a new project called "Lastname-Project6".
2. Create an R Script file within your project called "Lastname-Project6.R".
3. Include your name, the date, and the class in a comment as the first line of the script.
4. Include any script initialization code and library loading.

Testing Your Solution

You can evaluate your project using the `project5_tests.R` test file. This will only work if you store results in the exact variable names specified in bold and in parentheses at the end of the problem.

Problems

Analyzing a baseball probability distribution

In the next group of problems, consider the Boston Red Sox playing a stretch of seven games, where the probability of winning a game is .65 and an outcome is the number of wins during those seven games.

1. What is the probability that the Red Sox will win exactly 5 games (**prob1_result**)?

2. Create a data.frame or tibble with each possible outcome and the probability of that outcome. Name your columns **wins** and **probability (prob2_result)**.
 3. What is the probability that the Red Sox will win fewer than 5 games (**prob3_result**)?
 4. What is the probability that the Red Sox will win between 3 and 5 games inclusively (**prob4_result**)?
 5. What is the probability of the Red Sox winning more than 4 games (**prob5_result**)?
 6. What is the theoretical expected value of the number of wins for the Red Sox in a 7-game series (**prob6_result**)?
 7. What is the theoretical variance of the number of wins for the Red Sox in a 7- game series (**prob7_result**)?
 8. Generate 1,000 random values for the number of wins by the Red Sox in a 7- game series. Use **set.seed(10)** before generating the random values.
 9. Compute the sample mean of the 1,000 random values (**prob9_result**).
 10. Compute the sample variance of the 1,000 random values (**prob10_result**).
-

Analyzing calls in a call center

The number of calls received each hour at a call center follows a Poisson distribution averaging seven calls per employee per hour.

11. What is the probability that an employee will receive exactly 6 calls in the next hour (**prob11_result**)?
12. What is the probability that an employee will receive 40 or fewer calls in the next 8 hours (**prob12_result**)?
13. Assuming that there are 5 employees working eight-hour shifts, what is the probability that they will meet the quota of 275 or more calls during the shift (**prob13_result**)?
14. If one employee is sick, what is the probability that the remaining team will still meet the quota of 275 or more calls during their shift (**prob14_result**)?
15. For a single employee working an 8-hour shift, how many calls are necessary for the day to be considered in the top 10% of days volume-wise (**prob15_result**)?
16. Generate 1,000 random values for the number of calls for a single employee during an 8-hour shift. Use a **set.seed(15)** before creating values.
17. Compute the sample mean of the 1,000 random values (**prob17_result**).

18. Compute the sample variance of the 1,000 random values (**prob18_result**).
-

Analyzing the lifespans of light bulbs

The life spans of light bulbs at a certain manufacturing company follow a normal distribution, with a mean life span of 2,000 hours and a standard deviation of 100 hours.

19. What is the percentage of light bulbs with a lifespan of between 1,800 and 2,200 hours (**prob19_result**)?
 20. What is the percentage of light bulbs with a life span of more than 2,500 hours (**prob20_result**)?
 21. Light bulbs that fall in the bottom 10% of life spans are considered defective and can be returned for a full refund. What is the maximum number of hours in a light bulb's life span for it to fall into the defective category? Round your result up to the nearest integer value (**prob21_result**)?
 22. Generate 10,000 random values for the life spans of manufactured light bulbs. Use **set.seed(25)** before generating the values. For the remaining problems, consider this the population of light bulbs.
 23. Compute the population mean for the random values (**prob23_result**).
 24. Compute the population standard deviation for the random values (**prob24_result**).
 25. Take 1,000 different samples from the random values, where each sample contains 100 values. For each of the 1,000 different samples, compute the sample mean and store all 1,000 results in a vector. Use **set.seed(1)** before computing the samples (**prob25_result**).
 26. With the result of the prior problem, create a histogram.
 27. Compute the mean of the of the values from problem 25 (**prob27_result**).
-

Analyzing the flipper length of penguins

From the **palmerpenguins** package, utilize the existing **penguins** data set to work on each of the problems below.

28. Explore the distribution of flipper length of the Adélie penguin. What distribution does it most likely follow? Justify your decision with evidence from techniques learned in this course.
29. Explore the relationship between the flipper length and beak depth of the gentoo penguin. Justify any relationship you identify with evidence from techniques learned in this course.

30. Write up your results about the penguins in an executive summary called **LastName_Project6_Report.pdf**.
-

When you are satisfied with your solution, submit two (2) files in Canvas:

1. Your R script, named **Lastname_Project6.R**.
2. A PDF of your four-page report titled **Lastname_Project6_Report.pdf**.

Your report should contain the following information, formatted as specified:

Title Page

Include your name, assignment title, and submission date.

Introduction and Key Findings

Include an overview of the assignment and any findings.

Conclusion/Recommendations

Include evidence-based recommendations and visualizations or direct presentation of tabular data.

Works Cited

Include all sources, including YouTube videos, instruction materials, Google search results, and texts that informed your study of statistics and R.

Your report should be as concise as possible while maintaining fluency. Your key findings will be strongest if supported by visualizations or direct presentation of tabular data.

Your summary must adhere to APA guidelines, including page numbers on each page (including the title page) in the upper right corner. See the following examples for [title pages](#), [citations](#), and [general APA formatting](#).

Congratulations on completing Project 6!