R Tutorial 9

Instructions:

- Answer all questions.
- Ensure that your findings and results are clearly stated and thoroughly discussed. Please support your arguments using suitable R code with the relevant outputs, interpretations, plots and graphs whenever possible. You should support your argument using appropriate theory that is appropriately referenced.
- The R commands that you use in obtaining your results for all questions must be documented in a R script file. These scripts must be clearly commented. Ensure that any output is clearly stated and interpreted separately from the code as additional comments.
- Include the task name, your name and surname, and your student number in your R script file.
- You MUST label each answer by question number and, where a question has multiple parts, label each part of the question CLEARLY.
- On completion of your assignment, please submit onto RUconnected. If there are any issues uploading onto RUconnected, you may email your submission to: s.izally@ru.ac.za. Please submit your R script file and any other saved data files and plots mentioned in the questions below. Your student number should be included in the name of each file that you submit.
- Each student must complete an individual assignment. You will be assessed based on the quality and/or correctness of the R code, its outputs, and your explanations and interpretations. Acknowledge any help you may have received. Feel free to note any help you may have given to other students in the course.
- This assignment must be submitted by Tuesday, 25 September 2024 by 17:00. Late submissions will be penalized.
- Please note the Rhodes University and the Rhodes University Department of Statistics plagiarism policies.

Questions:

1. Let $Y \sim geo\,(p = 0.28)$, where Y is the number of trials before the first success. Use R to determine the probability that:
   (a) $P\,(Y = 3)$.
   (b) $P\,(Y = 2)$.
   (c) $P\,(Y = 1)$.
   (d) $P\,(Y \leq 3)$.
   (e) $P\,(Y > 3)$.
   (f) Use the rgeom command to simulate 100 000 values from the geometric random variable X, the number of failures in a sequence of Bernoulli trials before a success occurs. **Do not copy the output in your Word document when you submit.**
   (g) How would you simulate in R, a geometric random variable Y, where Y is the number of trials before the first success? **Do not copy the output in your Word document when you submit.**
   (h) What is the theoretical mean of X and the theoretical mean of Y? Compute this in R using the formulas derived in lectures for the mean of a geometric distribution.
   (i) Use the mean function in R to determine the mean of X and the mean of Y that was simulated in parts (f) and (g). How do the values obtained in parts (h) and (i) compare? If there are differences, why is this the case?
   (j) Use the barplot(table()) command to plot the distribution of X and Y side by side. **HINT:** Use par(mfrow=c(1,2)).
2. A water bore hole driller will drill a succession of holes in a given area to find a productive well. The probability that he is not successful on any hole (or trial) is 0.8. Let us assume that the probability of a successful hole is independent of any other wells drilled. Use R to answer the following questions.

(a) What is the probability that the third hole drilled will be the first to yield a productive well?

(b) What is the probability that the second hole drilled will be the first to yield a productive well?

(c) How many holes does the driller expect to drill until the first productive well is found?

(d) If the prospector can afford to drill at most ten wells,

   i. What is the probability that he has three successful wells?

   ii. What is the probability that he will fail to find a productive well?