STAT 260 Fall 2023: R Assignment 2

Due: Friday March 17th BEFORE 11:59pm (PT) on Crowdmark.

Introduction to R: Before attempting this assignment, read and work through the **Introduction to R Assignment 2** file posted on Crowdmark. This file contains a list of all the R commands needed to complete this assignment.

Submission: Since you will need to copy your R code and the code output, your answers must be typed. The best way to complete this assignment is by using a word processor such as MS Word or Open Office Document. (I have a link posted on our Brightspace page in the "Useful Links" module where you can download Microsoft 365 for free.)

Complete Parts 1, 2, 3, and 4 below, copying and pasting the required R commands, R outputs, and analysis into your Word documents. Save your assignment in the PDF file format. On Crowdmark you will be asked to upload your solutions to each part separately, so the best way to save your work would be to put your solutions to each part requiring a separate submission upload on separate pages in your Word file. To convert your Word document to a PDF use the ''Save As" feature - PDF is one of the output options there. When uploading your work to Crowdmark put your submission in the first upload area, then drag and drop the pages for the other parts into the proper submission areas.

Upload your files for submission to the assignment on Crowdmark, before Friday March 17th at 11:59pm (PT). Solutions submitted within 1 hour of the Friday deadline will have a 5% late penalty applied within Crowdmark. Solutions submitted after 1 hour of the Friday deadline but before the final Sunday cutoff will have a 20% late penalty applied. Solutions submitted after the final Sunday cutoff will be graded for feedback, but marks will not be awarded.

Note: For each of the following, carry out your calculations **only** using R or RStudio. Copy and paste your command(s) and the output into your document as indicated. Format your solutions in a somewhat formal way, as if you are writing a lab report (that is, use somewhat formal language and complete sentences).

Background: The binomial distribution is used to model the number of "successes" in a fixed number of trials, where trials are independent and the probability of success remains fixed from trial to trial. The Poisson distribution is used to model the number of "successes" (or arrivals or events) in a time interval (or area interval or length interval). As seen in our lectures, the Poisson distribution is useful to approximate the binomial distribution in the case of rare events (where the probability of success is quite small and the number of trials is quite large).
Continuous distributions such as the normal distribution can be applied to many situations that arise naturally. As we have seen in Set 16 the normal distribution can also be used to approximate the binomial distribution in some situations.
The use of computer software such as R allows us to perform precise calculations of probabilities which we would normally not be able to complete by hand with our stat tables. It also allows us to make calculations in situations where our approximations would not be appropriate.

Scenario: While cdf tables are useful for calculating probabilities by hand, they can only be used when the parameter values for our problem appear in the stat tables. In situations where the parameter values of the problem are not in the tables, the use of statistical software such as R is extremely helpful. We will encounter such problems in Part 1 and Part 2 where our cdf tables cannot be used to calculate the desired probabilities by hand. Statistical software is also useful in that it can randomly generate data for us so that we can run a simulation of an experiment before collecting data from the field (which can be expensive to do). In Part 4 we will generate values to simulate a binomial experiment that is completed many times.

For each of the following, carry out your calculations **only** using R or RStudio. For each part, include a copy of your R code used and the output of your code. (i.e. Copy and paste the relevant pieces into a Word document (or other word processing document).)

Note: R can complete basic arithmetic operations such as addition (+), subtraction (-), multiplication (*), and division (/). It should be noted that parameter values can be entered as arithmetic expressions. For example, we could enter an arithmetic expression using multiplication when writing the value of lambda in the desired Poisson function.

Part 1 A radioactive object emits particles according to a Poisson process at an average rate of 5.5 particles per second. We observe the object for a total of 6.5 seconds.

(a) [**1 mark**] What is the probability that no more than 40 particles will be emitted during this interval?

(b) [**1 mark**] What is the probability that exactly 38 particles will be emitted during this interval?

(c) [**2 marks**] Suppose it is known that at least 34 particles will be emitted during this interval. What is the probability that no more than 42 particles will be emitted during this interval?

Part 2 A manufacturer of ceramic blades estimates that 0.81% of all blades produced are too brittle to use. Suppose we take a random sample of 145 blades and test them for brittleness. We want to find the probability that at least 3 blades will be too brittle to use.

(a) [**1 mark**] Find the exact probability that at least 3 blades will be too brittle to use.

(b) [**1 mark**] Use an appropriate approximation to find the approximate probability that at least 3 blades will be too brittle to use.

Part 3 The fracture toughness (in $MPa\sqrt{m}$) of a particular steel alloy is known to be normally distributed with a mean of 29.2 and a standard deviation of 2.17. We select one sample of this alloy at random and measure its fracture toughness.

(a) [**1 mark**] What is the probability that the fracture toughness will be between 24.8 and 31.5?

(b) [**1 mark**] What is the probability that the fracture toughness will be at least 28.2?

(c) [**2 marks**] Given that the fracture toughness is at least 26, what is the probability that the fracture toughness will be no more than 32.1?

Part 4 The purpose of this question is to help you visualize the normal approximation to the binomial distribution which have seen in Set 16.

(a) [**1 mark**] Let $X \sim binomial(n = 85, p = 0.32)$. Create a vector called **simulation.data** which contains a simulation for 3700 values for X. (i.e. Simulate 3700 experiments, each being binomial with $n = 85$ and $p = 0.32$.) Provide a copy of the R command which you used to create this vector. You do **not** need to copy the 3700 values you generated.

(b) [**2 marks**] Create a histogram of **simulation.data** and copy it and your line of R code into your assignment. Your histogram should have an appropriate title and an appropriate label on the x-axis. Comment on the shape of the histogram. (We are looking for a single phrase here to describe the histogram. It should be a shape we've discussed recently.)

(c) [**2 marks**] Calculate the sample mean of **simulation.data**. Copy the command used, and the output. How close is your sample mean to what you would expect? (Hint: We have discussed the expected value of the sample mean $\overline{X}$. We have also discussed the expected value of a binomial random variable X.)