Homework 1

Your Name

Due: October 4, 2019 at 11:59pm

Homework Policies:

You are encouraged to discuss problem sets with your fellow students (and with the Course Instructor of course), but you must write your own final answers, in your own words. Solutions prepared "in committee" or by copying someone else's paper are not acceptable. This violates the Brown standards of plagiarism, and you will not have the benefit of having thought about and worked the problem when you take the examinations.

All answers must be in complete sentences and all graphs must be properly labeled.

For the PDF Version of this assignment: PDF

For the R Markdown Version of this assignment: RMarkdown

Turning the Homework in:

Please turn the homework in through canvas. You may use a pdf, html or word doc file to turn the assignment in.

PHP 1510 Assignment Link

PHP 2510 Assignment Link

Probability Distributions

1. There are n patients, each one has the same probability of being chosen for study screening with probability p (*Independently - No one patient has an effect on the other*). Each of these enters the study with probability r, independently.
   a. What is the distribution of the number of patients chosen for screening?
   b. What is the distribution of the number of patients who enter the study?
2. A company with n interns and m part time employees and is deciding which employees to promote. Now suppose that instead of having a predetermined number of promotions to give, the company decides independently for each employee, promoting the employee with probability p. Find the distributions of:
   a. the number of interns who are promoted
   b. the number of part time emplyees who are not promoted
   c. the total number of employees who are promoted.
3. Let $X \sim Bin(n,p)$ and $Y \sim Bin(m,p)$, where X and Y are independent.
   a. What is the distribution of $X + Y$?
   b. Explain why your answer in part a nmakes sense.
   c. Give a good example of why $X - Y$ is not binomial.
4. **(PHP 2510 Only)** It is important to begin to learn new distributions. The `exponential` distribution is a time distribution which looks at time until an event happens. It has one parameter λ which is the rate at which events occur. Give an example of data that connects the following distributions together:
   - Bernoulli
   - Exponential

   Please come up with a reasonable variable that we might consider studying in your field. Make sure you explain what kinds of questions asked on this data would lead to the above distributions.

Simulations

5. Rolling two dice

a. Simulate rolling two dice and adding their values. Do 100000 simulations and make a bar chart showing how many of each outcome occurred.
b. You can buy trick dice, which look (sort of) like normal dice. One die has numbers 5, 5, 5, 5, 5, 5. The other has numbers 2, 2, 2, 6, 6, 6. Simulate rolling the two trick dice and adding their values. Do 100000 simulations and make a bar chart showing how many of each outcome occurred.
c. Sicherman dice also look like normal dice, but have unusual numbers. One die has numbers 1, 2, 2, 3, 3, 4. The other has numbers 1, 3, 4, 5, 6, 8. Simulate rolling the two Sicherman dice and adding their values. Do 100000 simulations and make a bar chart showing how many of each outcome occurred. How does your answer compare to part a?

6. **(PHP 2510 Only)** Read sections 14.12 and 14.13 from Introduction to Data Science. Then complete the problems below:
a. Assume the distribution of female heights is approximated by a normal distribution with a mean of 64 inches and a standard deviation of 3 inches. If we pick a female at random, what is the probability that she is between 61 and 67 inches?
b. Repeat the exercise above, but convert everything to centimeters. That is, multiply every height, including the standard deviation, by 2.54. What is the answer now?