OPIM 5603 : Fall 2017
Exam 1

This is a two-and-a-half hour exam worth a total of 100 points.

There are two types of problems in this exam. Problem starting with an (R) should be completed in `R` and submitted as an RNotebook. Set the heading of the RNotebook as an `html_document`, with a table of contents and without numbered sections. Add your name and a date to the header as well. The solution to each problem should be a separate section (specified by #), and each subproblem should be set as a subsection (specified as ##). For example, if Problem 2 is denoted with an (R), you should have a section titled Problem 2, specified by:

```
# Problem 2
```

in your RNotebook. Also, for subproblem b in Problem 2, you should have a subsection, specified by:

```
## Problem 2b
```

Note that R code submitted should work independent of the data that sits in the data structure. For example, suppose there was a vector `r_vec` with the values $(1, 2, 6)$ and the problem asks for you to create R code to create a vector `answer` which doubles each element of `r_vec`. The answer

$$\texttt{answer} \leftarrow \mathrm{c}(2, 4, 12)$$

would be given no credit. The answer

$$\texttt{answer} \leftarrow 2^{*}\texttt{r\_vec}$$

would be an appropriate answer. You must show all steps in your solution. For example, if a problem asks for the expected value of a random variable that is binomially distributed with $n = 10$ and $\pi = 0.3$, and you simply write

$$3,$$

this will be given no credit. However,

$$10 * 0.3$$

would be given credit.

The remaining questions are specified with a (W) to indicate that the answer must be written directly in the exam printout, in pen, in black or blue ink. All steps must be clearly described and the answers must be written legibly. The only answers that will be graded is what is written in the area provided. Anything written elsewhere will be ignored when grading. You must show all work, otherwise no credit will be given.

Good luck on the exam!

ACKNOWLEDGMENT AND AGREEMENT TO RULES

You may use any references, websites, or class documents that you would like, except for interactions with other people and/or looking at someone else's exam. This includes emails, text, tweets, message boards, facebook, snapchat, or any other medium where you can correspond regarding the problems on the exam. You can read, for example, message boards, but absolutely cannot post anything. You are NOT permitted to use your cell phones for any purpose at all. Use of a cell phone will results in an automatic 0 on the exam. You are also not permitted to use headphones. Any attempt to cheat will result in the highest degree of sanctions, as outlined in the University of Connecticut Student Code and/or other documents regarding academic dishonesty released by the University of Connecticut. By signing below, you agree that you have taken the exam individually and according to the rules/guidelines listed above and that you will NOT discuss any problems related to the exam with anyone until after the solutions are release.

You must fill this page out, sign it, and turn in your exam copy. Please print your name and your NetID on the top lines, and then sign and date the following lines. Failure to do so will result in a 0 on the exam. By signing below, you agree that you have taken the exam individually and according to the rules/guidelines listed above.

Name - Exam5 Extra	NetID - xxx55555
Signature	Date

$\boxed{\textbf{Problem 1}}$

(W)

A predictive model has been built which estimates that for three potential students that have been offered admissions, the probability that each will enroll is 1.0, 0.3, and 0.6, respectively. Each of the students report their high school GPA, which is 3.3, 3.9, and 3.8, respectively.

a. What is the probability mass function for the number of student that enroll?

b. What is the probability that two or more students enroll?

c. What is the expected number of students to enroll?

d. What is the expected value of the average of the high school GPA for those students that enroll?

e. The university is considering not offering admissions to the student with 0.6 probability to enroll and a high school GPA of 3.8 for another student that would have probability 0.8 of enrolling. If the university wants to ensure that the expected value of the high school GPA for those students that enroll is higher than that in the previous part, what is the minimum high school GPA that this student can have for the university to make the switch?

Problem 2

(W)

Motherboards used in the computers assembled by Bergman, Inc. are provided by three different companies. 75% are provided by company X, 20% are provided by company Y, and 5% are provided by company Z. It is known that each of the motherboard provider companies have defects with certain probabilities; the motherboards provided by company X have a 4% rate of defect, those provided by company Y have a 3% rate of defect, and those provided by company Z have a 2.4% rate of defect.

a. If a computer is randomly chosen, what is the probability that it's motherboard has a defect?

b. If a randomly selected computer is chosen and it's motherboard has a defect, what is the probability that it came from company Z?

Problem 3

(R)

The number of trout a fisherman catches each day follows a Poisson distribution with $\lambda = 18$ and the number of pompono is binomially distributed with $n = 12$ and $p = 0.7$. The probability of having to throw a fish back into the ocean because it doesn't meet weight requirements is 0.1 and 0.2, respectively, for trout and pompono. The cost to scale a trout is normally distributed with mean \$1 and standard deviation \$0.1 while the cost of scaling a pompono is uniformly distributed between \$1 and \$1.25. Also the revenue the fisherman receives for each trout varies uniformly between \$15 and \$22 and the revenue for pompono varies uniformly between \$22 and \$27. The probability that a trout will be purchased if it is caught and scaled is 0.9, and this probability for pompono is 0.85.

a. What is the expected number of fish caught each day?

b. What is the expected number of fish that are sold each day?

c. What is the expected total daily cost for scaling fish?

d. What is the probability that the daily cost of scaling exceeds \$23?

e. What is the variance of the operating daily revenue (total revenue for fish sold minus costs for scaling)?

Problem 4

(W)

A laboratory blood test is 95 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1 percent of the healthy persons tested (i.e., if a healthy person is tested, then, with probability 0.01, the test result will imply the person has the disease). Suppose 0.5 percent of the population actually has the disease.

a. What fraction of the population test positive for the disease?

b. What is the probability that a person has the disease given that the test result is positive?

c. What is the probability that the person does not have the disease given that the test result is negative?

$\boxed{\textbf{Problem 5}}$

(W)

For the following five statements, indicate whether the statement is true or false by circling **True** or **False**, and provide a reason for your answer. If you answer true, then explain why. If you answer false, then provide a counter example.

a. The probability density function of a continuous random variable can never be greater than 1. **True** or **False**

b. The probability mass function of a discrete random variable is always less than or equal to 1. **True** or **False**

c. Complement events are not collectively exhaustive and mutually exclusive. **True** or **False**

d. It is impossible for an event to be independent of it's complement. **True** or **False**

e. Suppose A is an event for which $0 < P(A) < 1$. Furthermore, suppose event B always occurs if A occurs. It is possible for A and B to be independent. **True** or **False**

Problem 6

(R)

The *coefficient of variation* is a standardized measure of dispersion, often used to normalize the standard deviation of a random variable by its mean. For a random variable X with mean μ and standard deviation σ, the coefficient of variation c_v is defined by

$$c_v = \frac{\sigma}{\mu} \cdot 100\%.$$

It is a useful statistic for comparing the degree of variation from one data series to another, even if the means are drastically different from one another.
Assuming data come from a normal distribution, an unbiased estimator for c_v is

$$\left(1 + \frac{1}{4n}\right) \cdot \frac{s}{\bar{x}},$$

where s is the sample standard deviation and $\bar{x}$ is the sample mean.

a. For a random variable Y that is binomially distribution with parameters $n = 10$ and $\pi = 0.4$, what is the coefficient of variation?

b. Consider the series of data in the file `stocks.csv`, which contains the daily stock prices for two stocks, A and B. Provide a line of code for reading the data into a data frame in `R`.

c. Assuming that the prices for each of the two stocks come from a normal distribution, what is your estimate for the population coefficient of variation for each stock price? Do not use any packages for this problem—you must use the unbiased estimator defined above.

d. In the `raster` package there is a function `cv` which takes as argument a vector and calculates the coefficient of variation. Install the package, and use the function to calculate the coefficient of variation for a 10-value vector with each component generated from a normal distribution with mean 40 and standard deviation 2. Does it appear that the function returns an unbiased estimator for the population coefficient of variation? Explain why or why not.