CS 3130 / ECE 3530: Probability and Statistics for Engineers **Due Tu 11/7**

Homework 6: Joint Probability, Covariance, Correlation

Instructions: Submit a single R Markdown file (.Rmd) of your work on Canvas by 11:59pm on the due date. You may also submit diagrams, drawings, etc. as image files (.png, .jpg, .gif)—they must be formatted into your .Rmd document (we won't look at them separately). **Be sure to show all the work involved in deriving your answers! If you just give a final answer without explanation, you may not receive credit for that question.**

You may discuss the concepts with your classmates, but write up the answers entirely on your own. Do not look at another student's answers, and do not show your answers to anyone.

1. The following data comes from the 2011 Behavioral Risk Factor Surveillance System (BRFSS) survey, which is run by the Centers for Disease Control (CDC). This is a joint probability table for the proportions of survey respondents who smoke and who have had heart attacks.

	Smoker	Non-Smoker
Heart Attack	0.03	0.03
No Heart Attack	0.44	0.50

   Answer the following questions:

   (a) What is the proportion of people who smoke?

   (b) What is the proportion of people who have had a heart attack?

   (c) If a person is a smoker, are they more likely to have had a heart attack than someone who is not a smoker? **Hint: Be careful what this question is asking!**

   (d) If someone tells you that they have had a heart attack, what is the probability that they are also a smoker?

   (e) Is smoking independent of having a heart attack?

   (f) Define Bernoulli random variables, $S = 0$ (non-smoker), $S = 1$ (smoker), and $H = 0$ (no heart attack), $H = 1$ (heart attack). What is $\text{Cov}(S, H)$?

   (g) What is $\rho(S, H)$?

2. Using R, simulate the S and H random variables from Problem 1 by sampling from the joint pmf. **Hint:** First, use the `sample` function with appropriate probabilities to draw from the four possible joint outcomes. Then, convert these into two lists of 0's and 1's for S and H.

   From your simulation, approximate the following values (corresponding to (d,f,g) above):

   (a) $P(S = 1 \mid H = 1)$

   (b) $\text{Cov}(S, H)$ **Hint:** R has a `cov` function!

   (c) $\rho(S, H)$ **Hint:** R has a `cor` function!

3. Let k be some constant number, and consider continuous random variables X and Y with joint pdf

$$f(x,y) = \begin{cases} k(x+y) & \text{for } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

   (a) Find k.

   (b) What is the joint probability $P\left(\left\{X \ge \frac{1}{4}\right\} \cap \left\{Y \le \frac{1}{2}\right\}\right)$?

   (c) What is the marginal pdf $f_X(x)$?

   (d) What is the marginal pdf $f_Y(y)$?

   (e) What is the conditional probability $P\left(X \le \frac{1}{4} \,\middle|\, Y = \frac{1}{2}\right)$?

   (f) Are X and Y independent? Explain why or why not.

   (g) What is $\mathrm{Cov}(X,Y)$?

   (h) What is $\rho(X,Y)$?

4. A polling station in Salt Lake City has 50 registered Republicans and 50 registered Democrats. After 4 people show up to vote, let R be the number of Republicans and D be the number of Democrats that have voted.

   (a) Compute and list all the entries of the joint probability table $P(R = a, D = b)$ that are non-zero.

   (b) What is $\mathrm{Cov}(R,D)$?

   (c) What is $\rho(R,D)$?

   Hint: These computations may seem daunting, but the final answers should come out to be simple numbers. You can use a calculator of course, but don't just give the final answer – show your work!