Math 422 - Midterm 2 **Due Wednesday, Apr 18**

You may consult the textbooks and your notes for this exam, but you should not use the internet and all work must be your own. You may use software such as R, Sage, Matlab, etc. If you have any questions, please ask me. Solutions can be hand written or typed with LaTeX (or both).

1. (30 points) Sixteen student volunteers at Ohio State University each drank a randomly assigned number of cans of beer. These students were evenly divided between men and women, and they differed in weight and drinking habits. Thirty minutes later, a police officer measured their blood alcohol content (BAC) in grams of alcohol per deciliter of blood. The results are shown in the table below.

Student	Beers	BAC		Student	Beers	BAC
1	5	0.1		9	3	0.02
2	2	0.03		10	5	0.05
3	9	0.19		11	4	0.07
4	8	0.12		12	6	0.1
5	3	0.04		13	5	0.085
6	7	0.095		14	7	0.09
7	3	0.07		15	1	0.01
8	5	0.06		16	4	0.05

   (a) What is the correlation between the number of beers a student drinks and their BAC?

   (b) Find the matrices X, $\hat{\beta}$, and $\hat{y}$ for the least squares regression model $\hat{y} = X\hat{\beta}$.

   (c) In least squares regression, we assume that the entries of the matrix X are fixed, but the entries of the vector y are random, and have the form $y = X\beta + \epsilon$ where the entries of ϵ are i.i.d. Normal$(0, \sigma^2)$, and the entries of β are the parameters representing the slope and intercept for the linear relationship in the population. Use the fact that $\hat{\beta} = (X^T X)^{-1} X^T y$ to prove that $\hat{\beta}$ has a multivariate normal distribution, and find the covariance matrix for $\hat{\beta}$ in this example.

   (d) Show that the matrix $X(X^T X)^{-1} X^T$ is an orthogonal projection. What is the dimension of the subspace that it projects onto?

2. (10 points) Suppose that you book a room at a hotel in a large city. The room you get is on the 11th floor, but you don't know how tall the hotel is. Let k denote the height of the hotel (number of floors). Using a Poisson distribution with $\lambda = 20$ as a prior for k, find the posterior distribution. It is okay to use summation notation to express your final answer. (Note: the tallest hotel in the world is over 160 stories tall, but you don't need to work that information into this problem).

3. (10 points) Suppose that $X_1, \ldots, X_n$ is an i.i.d. random sample from the uniform distribution on the interval $[\theta, \theta + 3]$. Use the values of $X_1, \ldots, X_n$ to find an unbiased estimator for θ, and prove that your estimator is unbiased. Hint: *this problem has many possible solutions, some of them are fairly simple!*

4. (25 points) Suppose that when I play mini-golf, the number of strokes before I get the ball into the hole has a geometric distribution with unknown parameter θ. On a 9-hole course, the number of strokes for each hole is listed below:

$$4 \quad 5 \quad 2 \quad 3 \quad 3 \quad 6 \quad 3 \quad 2 \quad 4$$

   (a) Find the maximum likely estimate for θ.

   (b) What if I got a hole in one on each of the nine holes? What would the MLE for θ be then?

   (c) Suppose that I have no prior knowledge of what θ is, so I start with a uniform prior. Find a 95% Highest Posterior Density credible interval for θ if I get a hole in one on each of the 9 holes.

5. (25 points) Alice and Bob are using Bayesian statistics to predict whether candidate R will get more than 50% of the vote. Both are updating their priors based on the latest poll of 740 likely voters. In the poll, 378 voters plan to vote for candidate R, while the rest plan to vote for someone else. Before the poll, Alice had a prior distribution Beta(100,120) for the percent of voters who will support candidate R. Bob had a Uniform($[0, 1]$) prior.

   (a) What are Alice and Bob's posterior distributions after the poll?

   (b) What would Alice and Bob each say are the ***odds*** of candidate R winning?

   (c) What are the respective Bayes factors for Alice and for Bob?