Name ____

Spring 2019 **STAT 587C** **Exam II**

(100 points)

Instructions:

- Full credit will be given only if you show your work.
- The questions are not necessarily ordered from easiest to hardest.
- You are allowed to use any resource except aid from another individual.
- Aid from another individual, will automatically earn you a 0.

1. SpaceX has recovered the first stage booster of its Falcon 9 rocket in 35 of 42 attempts. Assume the number of successes have a binomial distribution with an unknown true probability of success. For the following questions, assume the default prior for the true probability of success. (4 points each)

   (a) What is the posterior distribution for the true probability of success?

   (b) What is the posterior expectation for the true probability of success?

   (c) Provide an equal-tail 90% credible interval for the true probability of success.

   (d) Calculate the probability the true probability of success is greater than 0.9.

   (e) Are these data a random sample from a population? Explain why or why not.

2. During the spring, the city of Ames flushes all of its fire hydrants in order to remove sediment from the water line leading to the fire hydrant. To understand how often the fire hydrants should be flushed, the city selects a random sample of the fire hydrants in Ames and measures the amount of dry sediment removed when performing the flush. In 16 hydrants, the city found a sample average of 2 kilograms (kg) of dry sediment and a sample standard deviation of 1.5 kg. For the following questions, assume the dry sediment measurements are independent and normally distributed with mean μ and variance σ^2 and assume the default prior for the mean and variance. (4 points each)

   (a) State the marginal posterior for the population variance.

   (b) State the marginal posterior for the population mean.

   (c) Provide an 80% credible interval for the population mean.

   (d) Let $\tilde{Y}$ be the amount of dry sediment in an unmeasured fire hydrant water line. The predictive distribution for $\tilde{Y}$ is $t_{n-1}(\overline{y}, s^2[1 + 1/n])$. Determine the probability this amount will be less than 3 kg.

   (e) Explain why the normal distribution may not be a very good model for these data.

3. Iowa State University researchers have developed a method of testing soil health using tea bags buried in an agricultural field. After 30 days, the tea bag is recovered, dried, and weighed. Lower weight indicates healthier soil due to tea leaves decomposing. The file `tea.csv` contains measurements of tea bag weights in grams (g) from agricultural fields that have prairie strips, a treatment that is designed to increase soil health. For the following questions, assume the tea bag weights are independent and normally distributed with mean μ and variance σ^2. (4 points each)

   (a) Calculate the maximum likelihood estimator for the population mean of tea bag weights.

   (b) Calculate the maximum likelihood estimator for the population variance in tea bag weights.

   (c) Construct a 99% confidence interval for the mean tea bag weight.

   (d) From previous experience, the researchers know the mean weight of tea bags in standard agricultural fields is 1.2 g. Calculate a p-value for the null hypothesis $H_0 : \mu \geq 1.2$.

   (e) If researchers believe fields with prairie strips will have healthier soil than standard fields, is the null hypothesis in the previous question the appropriate null hypothesis? Explain why or why not.

4. Let $Y_i \overset{ind}{\sim} Bin(n_i, \theta_i)$ for $i = 1, 2$. For the following questions, assume the data are random. (5 points each)

(a) Calculate

$$E\left[\frac{Y_1}{n_1} - \frac{Y_2}{n_2}\right].$$

(b) Calculate

$$Var\left[\frac{Y_1}{n_1} - \frac{Y_2}{n_2}\right].$$

(c) Calculate a standard error for

$$\frac{Y_1}{n_1} - \frac{Y_2}{n_2}.$$

(d) Provide a formula for computing an approximate 100(1-a)% confidence interval for $\theta_1 - \theta_2$?

5. The US FDA is currrently overseeing a clinical trial for the drug *selonsertib* which is aimed at patients who have late-stage fatty liver disease. Patients at Marshall University Hospital in Huntington, West Virginia who have late-stage fatty liver disease can enroll in the clinical trial and will be randomly assigned either selonsertib or placebo (a sugar pill). Patients who enroll will have their *aspartate transaminase* levels measured before starting the pill regiment and again one year later. Doctors will compare how much these levels change for the selonsertib group compared with the placebo group.

   (a) Describe the population being studied. (3 points)

   (b) Describe the sample. (3 points)

   (c) Is this a random sample? Explain why or why not. (2 points)

   (d) Describe a reasonable model for these data to address the scientific question of interest. (12 points)