

Stat 532

Name: \_\_\_\_\_

Take home midterm

Due 10/23/17 at 5:00 PM

---

For the take home exam, you may use the textbook, any course materials provided on D2L, homeworks, and labs. You **may not** discuss questions or work together with classmates. You are welcome to contact the instructor with any questions related to better explanation or understanding of the questions themselves. Any relevant material from questions will be posted to D2L for the benefit of the entire class. For complete (and partial credit) please show all work, whether that be by hand or printed R code.

Note the first three questions are a follow up to Q1 on the in-class exam where the goal was to estimate whether MSU students go for more than an average of 10 hikes a year. Data was collected from 21 students in a section of STAT 408 at MSU and is available from the following line of R code:

```
hike <- c(10,25,10,7,35,15,10,10,75,10,15,6,20,50,30,10,0,6,4,0,0)
```

1. (14 points) Given that the response for number of hikes is a count, assume you decide to fit a Poisson sampling model coupled with a Gamma prior.
  - (a) (4 points) State and defend a prior distribution for  $\theta$ , the mean term, in a Poisson model.
  - (b) (2 points) What is the posterior distribution in this case?
  - (c) (4 points) Does this sampling model / prior combination look reasonable, why or why not?
  - (d) (4 points) Summarize the results from this model with respect to whether students MSU students hike more than an average of 10 days a year. This should include some probabilistic statements, qualitative discussion of students behavior, and potentially some visual graphics.
2. (20 points) Now consider an approach that uses a normal sampling model to estimate whether  $\theta$ , the mean term in the normal distribution, is greater than 10.
  - (a) (4 points) State and defend a prior distribution for  $\theta$ , the mean term in a normal model and  $\sigma^2$ , the variance.
  - (b) (4 points) Describe and implement a procedure to take samples from the joint posterior distribution  $p(\theta, \sigma^2 | y_1, \dots, y_n)$ .
  - (c) (4 points) Summarize the results from this model with respect to whether students MSU students hike more than an average of 10 days a year. This should include some probabilistic statements, qualitative discussion of students behavior, and potentially some visual graphics.
  - (d) (4 points) Does this sampling model / prior combination look reasonable, why or why not?
  - (e) (4 points) Discuss the similarities and differences between this procedure and a classical t-test.
3. (4 points) Compare the two approaches presented in Q1 and Q2 and then defend one choice or suggest an alternative approach.

4. (18 points) For this question, consider a subset of Bozeman housing prices. The goal is to model the average sales price of the home. Use the dataset at:  
<http://math.montana.edu/ahoegh/teaching/stat532/data/BozemanHousing2017exam.csv>
- (a) (4 points) Write out a sampling model along with the necessary priors to address this scenario.
  - (b) (4 points) Obtain samples from the posterior distribution to summarize your findings for the average housing price in Bozeman.
  - (c) (4 points) Plot your posterior predictive distribution for the closing price of a new home *and* verify that your model and prior are reasonable.
  - (d) (6 points) Now assume that you would like to build a better model for the average sales price by including other information from the data set such as square footage of living space, the number of bathrooms, and the number of bedrooms. Describe how you would do this in a Bayesian way and write out your updated sampling model and prior specification. You do not need to fit the model, but just set up the sampling model and state the necessary priors on the parameters in the model.