The first midterm exam will be a take-home exam which will be made available in the Midterm 1 folder under the documents section of PWeb at 5pm on Friday, March 3rd and due at 11:59pm (uploaded to Gradescope) on Monday, March 6th.

Content. The exam will cover Chapter 7 (7.1 - 7.6) and parts of Chapter 8 (8.1 - 8.3, 8.6) of DeGroot and Schervish's *Probability and Statistics.* There will be some questions that ask you to use R.

Format. The exam is intended to take 2.5 hours to complete, although you may take up to 4 hours to complete it. These 4 hours do not need to be consecutive. You should monitor your own time, and record on the test your estimate for the total amount of time you actively worked on the exam.

Your solutions to the exam should be neatly neatly written or typed. If you scan a handwritten assignment, be sure to review the legibility of your scan on Gradescope after you submit.

Resources. You may use any notes you've taken for this class, your work on any previous homework or daily assignments, lecture notes I've posted on the course website, the recorded lecture video from Monday 2/20 and DeGroot and Schervish's *Probability and Statistics* textbook, as well as Blitzstein's *Introduction to Probability* textbook.

For problems asking you to do analysis or perform computations using R, you may use either a local installation of R or the Grinnell R Studio server, and you may reference any of the R help files (available by typing `?functionname` in the console).

You may not use any other resources other than those listed above. If you have questions about whether a resource can be used, you are welcome to message me.

Preparation. The best preparation you can do for the exam is to organize your notes and/or homework to make finding information and examples as quick and efficient as possible. Beyond that, you should attempt to accurately assess what topics you have mastered and which you need to practice more. A good starting point is to review the list of objectives on each daily assignment. Another way to prepare is to create your own study guide with summaries of the important concepts, along with example problems you've designed and solved. Exam problems will be comparable in difficulty to those exhibited in class and assigned for homework. Some exam questions may be similar to problems you have seen before, while others will require you to synthesize your knowledge in new ways.

On the exam, you may be asked to do the following:

- Rephrase a key definition and/or theorem in your own words.
- Determine whether a given statement is true or false.
- Interpret or explain a statistics concept in everyday language.
- Sketch the proof of an important result discussed in class.
- Perform calculations using relevant techniques from the course.
- Provide a short, rigorous proof of a novel statement or result.
- Create and analyze a statistical model for a particular phenomenon.
- Use R to simulate a random phenomenon.

For extra practice, several additional review problems are printed below. Solutions to these problems can be found on the exams page of the course website. While these questions are representative of the typical scope and difficulty of individual exam questions, this review is not comprehensive, nor does it necessarily represent the total amount of time available for the exam.

Practice Problems.

(1) The method of *randomized response* is sometimes used to conduct surveys on sensitive topics. A simple version of the method can be described as follows:

A random sample of n people are drawn from a large population. For each person in the sample, there is probability 1/2 that the person will be asked a standard question and probability 1/2 that the person will be asked a sensitive question. Furthermore, this selection of the standard or sensitive question is made independently from person to person. If a person is asked the standard question, then there is probability 1/2 that the person will give a positive response; however, the person is asked the sensitive question, then there is an unknown probability p that they will give a positive response. The statistician can observe only the total number X of positive responses that were given by the n persons in the sample, but cannot observe which of these persons were asked the sensitive question or how many person in the sample were asked the sensitive question.

Determine the MLE of p based on the observation X.

(2) Suppose that an observation X is from from a distribution with pdf

$$f(x|\theta) = \begin{cases} \frac{1}{\theta}, & \text{for } 0 < x < \theta, \\ 0, & \text{otherwise.} \end{cases}$$

Also, suppose that the prior pdf of θ is

$$\xi(\theta) = \begin{cases} \theta e^{-\theta}, & \text{for } \theta > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Determine the Bayes estimator of θ with respect to:

(a) The mean squared error loss function;
(b) the absolute error loss function.

(3) Suppose that the random variable X has a binomial distribution with unknown value n and a known value of p for $0 < p < 1$. Determine the MLE of n based on the observation X. *Hint:* consider the ratio

$$\frac{f(x|n+1,p)}{f(x|n,p)}.$$

(4) Suppose $X_1, \ldots, X_m$ form a random sample from the Normal distribution with mean μ_1 and variance σ^2 and that $Y_1, \ldots, Y_m$ form an independent sample from the Normal distribution with mean μ_2 and variance σ^2. Let

$$S_X^2 = \sum_{i=1}^{m} (X_i - \bar{X})^2 \qquad S_Y^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2.$$

(a) For what values of (α, β) is $S^2 = \alpha S_X^2 + \beta S_Y^2$ an **unbiased** estimator of σ^2?
(b) Determine the values of α and β for which $\alpha S_X^2 + \beta S_Y^2$ will be an unbiased estimator with minimum variance.

(5) Let $X_1, \ldots, X_n$ be iid random variables with density function

$$f(x|\sigma) = \frac{1}{2\sigma} e^{-|x|/\sigma}$$

where σ is an unknown parameter with $\sigma > 0$.

(a) Find a formula for the MLE of σ.
(b) Find a formula for the Method of Moments estimator of σ.

(c) Use R to simulate the sampling distribution for the MLE estimator and for the Method of Moments estimator when $\sigma = 2$ and $n = 10$. Use your simulation to estimate the mean and variance of each estimator.

(6) The *Pareto* distribution with shape θ and minimum value 1 has PDF:

$$f(x) = \frac{\theta}{x^{\theta+1}} \qquad x > 1$$

Show that the family of Gamma distributions Gamma(α, β) is a conjugate family of prior distributions for θ, when samples are taken from a Pareto distribution with shape θ and minimum value 1.