You are expected to complete this exam during any two-hour window within the exam period, which runs from Wednesday 11:30 AM to Thursday 11:30 AM. The exam is open-book and open-note:

- You may use all of this course's resources: the textbook, your class notes, your old homework, and the course web site. You may not share any resources with any other person while you are taking the exam.
- You may cite material (definitions, theorems, examples, etc.) from class, the assigned textbook readings, and the assigned homework problems. You do not have to redevelop or reprove that material. On the other hand, you may not cite results that we have not studied.
- You may not consult any other books, papers, Internet sites, etc. You may use a computer for viewing the course Moodle/web site, running R commands, typing up your answers, and e-mailing with me. If you want to use a computer for other purposes, then check with me first.
- You may not discuss the exam in any way — spoken, written, etc. — with anyone but me, until everyone has handed in the exam. During the exam period you will inevitably see your classmates around campus. Refrain from asking even seemingly innocuous questions such as "Have you started the exam yet?" If a statement or question conveys any information about the exam, then it is not allowed. If it conveys no information, then you have no reason to make it.

Feel free to ask clarifying questions in person or over e-mail. You should certainly ask for clarification if you believe that a problem is mis-stated. If you cannot receive clarification, then explain your interpretation in your solution.

Your solutions should be thorough, self-explanatory, neat, concise, and polished. You might want to work first on scratch paper and then recopy your solutions. Alternatively, you might want to type your solutions. Always show enough work and justification so that a typical classmate could understand your solutions. If you cannot solve a problem, then write a brief summary of the approaches you've tried. Partial credit is often awarded. Present your solutions in the order assigned.

Good luck. :)

This first problem is based on a true story from a few months ago. Suppose that a contentious election is about to happen. You are organizing volunteers to assist voters at polling places. You have s sites to cover and v volunteers. You'd like to have at least one volunteer at each site, but that's not going to happen, because $v < s$. To make matters worse, your volunteers are inexperienced and undisciplined. They don't respond to messages promptly. On election day, they might just go to sites randomly.

A. Assuming that each of your volunteers uniformly randomly picks a site, what is the probability that you cover v sites?

For the second problem, a *hard disk drive* is a mechanical device that rotates thousands of times per second, often for years on end. Consequently, it lasts only a few years before breaking. Suppose that, on any given day, any given hard drive has a 0.1% chance of breaking. In your job at a technology company, you oversee a data center that contains 100,000 hard drives. Whenever one of them breaks, a worker replaces it at the end of the day, so that the next day begins anew with 100,000 functioning hard drives.

B.A. How many of your hard drives will break today? Answer in the form of a random variable X. Be sure to specify the distribution of that random variable, including its parameter values. Also mention any assumptions that you make.
B.B. What is the probability that more than 100 hard drives break today? Give your answer as a mathematical expression, showing as much detail as possible. Simplify that expression down to a decimal number, if possible.
B.C. Let Y be the number of hard drives that broke yesterday. What is the support of $X + Y$?
B.D. What is the PMF of $X + Y$? Show as much detail as possible.

This next problem is related to Classwork Problem #15, which you might want to review. You are a chemist investigating a poorly understood chemical reaction. There are three mutually exclusive explanations for how the reaction works, which you call H_1, H_2, H_3. Based on your knowledge of theory and past experiments, you assign probabilities $P(H_1) = 0.6, P(H_2) = 0.3, P(H_3) = 0.1$ to these explanations. Then you invent a new experiment and run that experiment to obtain a data set D. This new data set surprises you a bit. You calculate that the probability of seeing such data is 0.03 if H_1 is true, 0.01 if H_2 is true, and 0.11 if H_3 is true.

C. What probability do you assign to H_3 now?

Finally, at a certain fictional boarding school, students are sorted upon arrival into four groups — Heroic, Sinister, Bumbling, and Boring — that supposedly reflect their characters. (If this school sounds cruel, I haven't even described the frequent violence, both physical and emotional, that occurs there. Let's just say that the school is the target of many lawsuits.) The students are not necessarily allocated to groups evenly. For example, when 80 students arrive, the counts might end up as 20 Heroic, 20 Sinister, 20 Bumbling, and 20 Boring, but the counts might instead end up

- 25 Heroic, 15 Sinister, 30 Bumbling, 10 Boring,
- 10 Heroic, 50 Sinister, 20 Bumbling, 0 Boring,
- 20 Heroic, 10 Sinister, 0 Bumbling, 50 Boring,

or any other allocation that adds up to 80.

D. How many allocations are possible, assuming that 80 students arrive at the school? (To clarify, we are concerned about the *number* of students in each group, but not the *individual identities* of the students in each group.)