Summary

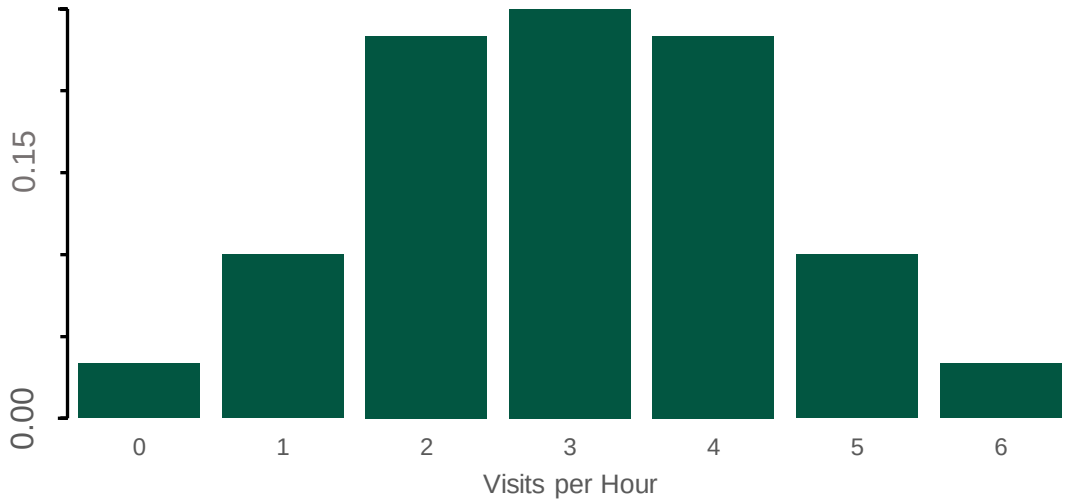
Please typeset these exercises using R-Markdown. See the appendix at the end of this document for additional information.

Discrete Probability Exercises

Web Analytics

1. A website is monitored for two months and data gathered on the number of visitors (visits) per hour. A *visit* is a collection of *hits* which may include hundreds of page and image views as the user progresses through the site. Only after a visitor leaves for 45 minutes will their next activity be considered a new visit. (e.g. There is no activity from a user between a hit at time t_i and their next hit $t_{i+1} - t_i \geq 2700$ secs). The number of new visits initiated per hour at a website is modeled by a random variable, *X.* (A random variable, as you recall, takes on a probability *p* between 0 and 1). The probability mass function (pmf) of the visits per hour is given by the following histogram:

```
barplot(c(.03,.09,.24,.27,.25,.09,.03),names.arg=c(0,1,2,3,4,5,6),xlab="Visits perHour")
```



- For a given hour, what is the probability that the number of visits will be between 4 and 6?
- Compute the mean, the variance, and the standard deviation of the number of visits per hour.

2. Your client estimates that 68% of all visitors to the site are current college students. The remaining proportion have not attended or have already graduated. Suppose the site had 11 visitors in the last two hours.
    - Let *X* be the number of visitors (among the 11) who are college students. What assumptions should be satisfied for *X* to obey a binomial distribution *B(n,p)*?
    - Modeling *X* as a binomially distributed random variable, what is *n* and what is *p* in this case?
    - What is the probability that exactly 7 of the 11 visitors are college students?
    - What is the expected number of visitors (of the 11) who are college students? What is the standard deviation of the number of visitors who are college students?

3. A website, hugeretail.com, features a huge variety of products for upscale consumers. These consumers are segmented into "mobile" and "desktop" consumers, as a function of how they complete the majority of their purchases, either via phone or some other device such as a laptop. 38% of the site's consumer base are mobile consumers. The Consumer Fraud Division of the Department of Justice has begun a preliminary investigation into the online practices of hugeretail.com after an article exposed complaints of padded charges. Their first task is to choose a random sample of 20 customers for interviews and data collection regarding their transactions with hugeretail.com
    - What is the probability that 8 of the 20 randomly chosen consumers will be mobile consumers?
    - What is the probability that between 12 and 14 of the 20 randomly chosen consumers will be desktop consumers?
    - What is the probability that 3 or fewer of the randomly chosen consumers will be desktop consumers?

A Game of Chance

4. You have made it to the final round of a game show. The announcer asks you the final multiple-choice question, which has four possible answers: (a), (b), (c), and (d). If you answer the question correctly, you win $1,000,000. After considering the question, you realize you are not sure, but you have your suspicions. You think the answer is either choice (a) with probability 60%, or choice (b) with probability 40%. You are certain the answer is not choice (c) or (d). You have three options:
    - *Choose not to answer the question*: You walk away with $500,000.
    - *Select an answer (a or b)*: If you are correct, you win $1,000,000. If you are incorrect, you win only $32,000.
    - *Phone a friend*: In this option, the announcer allows you to call a friend for help.

    After listening to your friend's response, you must then answer the question. You know that:

- Given the correct answer is (a), your friend will say “a” with probability 80%.
- Given the correct answer is (b), your friend will say “b” with probability 80%.
- What is the probability that your friend says the answer is “a”?
- What is the probability that your friend says the answer is “b”?
- Make a decision tree to find your best strategy. In *words*, what should your strategy be? How much money do you expect to win?

Healthcare Analytics

5. With much acclaim, you have written a machine learning-based classifier which diagnoses a particular disease state. Given a person who has the disease, the probability that the classifier will fail to reveal the disease is 0.06. Given a person who does not have the disease, the classifier will correctly identify that the person does not have the disease with probability 0.91.
    - If someone has the disease, what is the probability that the classifier will identify that person as having the disease?
    - Suppose three unrelated individuals who do not have the disease are classified by your algorithm. What is the probability that at least one of the three individuals will be identified by your classifier as having the disease?
    - In a particular area of a particular country, 17% of the population suffers from this disease. If a random person from this area is tested and the results indicate that the person has the disease, what is the chance that the person actually has the disease?

Airline Analytics

6. Let’s look at overbooking. JetPurple Airlines’ first-class cabins have 10 seats per plane. JetPurple’s overbooking policy is to sell up to 11 first class tickets, since cancellations and no-shows are always possible (and quite likely). For a given flight on JetPurple, there were 11 first class tickets sold. Suppose that each of the 11 persons who purchased tickets has a 20% chance of not showing up for the flight, and that the events that different persons show up for the flight are independent (e.g. the passengers are not traveling as part of a classy high school sports team).
    - What is the probability that at most 5 of the 11 persons who purchased first class tickets show up for the flight?
    - What is the probability that exactly 10 of the persons who purchased first class tickets show up for the flight?
    - Suppose that there are 10 seats in first class available and that the cost of each first-class ticket is $1,200. (This contributes entirely to profit since the variable cost associated with a passenger on a flight is close to zero.) Suppose further that any overbooked seat costs the airline $3,000, which is the cost of the free ticket plus some cost in damaged customer relations. (First class passengers do not expect to be bumped!) Thus, if 10 of the first-class passengers show up for the flight, the airline’s profit is $12,000. If 11 first class passengers show up, the profit is $9,000. What is the expected profit from first class passengers for this flight?

- Suppose that only 10 first class tickets were sold. What would be the expected profit from first class passengers for this flight?
- (Think) People often travel in groups of two or more. Does this affect the independence assumption about passenger behavior? Why or why not?

Appendix

Here is a nice overview of installing MikTek (LaTeX distribution) so that R-Studio can find it. If you can get in the habit of using R-Markdown for your assignments, it will benefit you in all of your classes, as well as next semester, and in your job in business analytics.