STAT 311: Homework 5
Due: Aug 5, in class

Name:

The material covered includes CH 8 and the very beginning of CH 9 from U+H. In general, rounding to 4 digits is sufficient for this assignment.

1 Problem 1: Counting Rules

1. *Each NBA team carries 12 active players. However, only 5 players can play on the court at any given time. How many combinations of on court players (set of 5) can be made from the team of 12 players*

   Since we form a team (without replacement) and are only interested in combinations-

   $$\binom{12}{5} = 792$$

2. *Suppose the team is split into 3 positions: Guards, Forwards and Centers and an NBA team has 5 guards, 4 Forwards and 3 centers on their roster. On the court, they want 2 guards, 2 forwards and 1 center. How many combinations of on court players can be made which satisfy those constraints.*

   Now we can't put any player anywhere, but need to select from the specific positions. So we find the number of guard combinations $\binom{5}{2}$ and the number of forward combinations $\binom{4}{2}$ and center combinations $\binom{3}{1}$. Since the guards and forwards and centers are all seperate groups, to get the total number of combinations, we just multiply them together

   $$\binom{5}{2}\binom{4}{2}\binom{3}{1} = 180$$

3. *Suppose the team is split into 3 positions: Guards, Forwards and Centers and an NBA team has 5 guards, 4 Forwards and 3 centers on their roster. On the court, they want 2 guards, 2 forwards and 1 center. Suppose I need one combination of 5 players to play for the entire first half, and one combination of players to play for the entire second half. If a player played the first half, they cannot be selected again for the second half. How many different line ups can be created?*

   The problem above gave us a way to calculate the number of first half line-ups we could've produced. Now we just need to find the number of second half line-ups (given the first half line up) and multiply the numbers together to get the total number of combinations. Given that we've selected a specific first half line-up, the number of second half line-ups is

   $$\binom{3}{2}\binom{2}{2}\binom{2}{1} = 6$$

   So the total number of full game line-ups is $180 \times 6 = 1080$

2 Guessing on the SAT

Back when I took the SAT, test-takers received 1 answer point for each correct answer, were deducted 1/4 answer point for each incorrect answer and received 0 points for any unanswered questions. These answer points were then translated to a 200-800 point score. Each multiple choice question had 5 possible answers.

1. *What are the possible outcomes (in terms of answer points) and associated probabilities when guessing randomly?*

$$P(1) = 1/5$$
$$P(-1/4) = 4/5$$

2. *Suppose you come to a question where you have no idea what the right answer is. What is the expected number of answer points when guessing randomly? What is the variance?*

$$\mathcal{E}(X) = (1)P(1) + (-1/4)P(1/4) = 1(1/5) + (-1/4)(4/5) = 0$$
$$\text{var}(X) = (1-0)^2 P(1) + (-1/4-0)P(1/4) = 1(1/5) + (1/16)(4/5) = 1/4$$

3. *If you guess randomly on 5 questions, what is the probability that you get exactly 3 answers correct?*

   Since we are now just interested in the number correct (not the number of points), we can just use binomial probabilities where we have 5 trials and the probability of success is 1/5.

$$P(C = 3) = \binom{5}{3}(1/5)^3(4/5)^2 = .0512$$

4. *If you guess randomly on 5 questions, what is the probability that you get at least 3 answers correct? (Hint: you may want to use* `R` *to calculate this.)*

$$P(C \geq 3) = P(C = 3) + P(C = 4) + P(C = 5) =$$

   You could calculate that by hand, or using `R`

```
# P(C = k) for k = 3, 4, and 5
dbinom(c(3,4,5), 5, .2)

## [1] 0.05120 0.00640 0.00032

# P(C >= 3)
sum(dbinom(c(3,4,5), 5, .2))

## [1] 0.05792
```

5. *If you guess randomly on 5 questions, what is the expected number of answer points? What is the variance of the number of answer points?*

   If you guess randomly on 5 questions, then the total number of points you reicieve, T is just the sum of the points you get on each individual question. We also assume that the questions are independent.

$$\mathcal{E}(T) = \mathbb{E}(X_1 + \ldots X_5) = \mathbb{E}(X_1) + \ldots \mathbb{E}(X_5) = 5(0) = 0$$

$$\text{var}(T) = \text{var}(X_1 + \ldots X_5) = \text{var}E(X_1) + \ldots \text{var}(X_5) = 5(1/4) = 5/4$$

6. *Suppose you come to a question where you are able to (correctly) rule out one of the possible answers. So now, you are randomly guessing from 4 choices. What is the expected number of answer points now when guessing randomly from 4 random choices?*

   The possible outcomes do not change, but the probabilities for each outcome do. In particular, $P(-1/4) = 3/4$ and $P(1) = 1/4$, since there are now only 4 possible questions to randomly pick from.

$$\mathcal{E}(Y) = (1)P(1) + (-1/4)P(1/4) = 1(1/4) + (-1/4)(3/4) = 1/16$$

7. *What is the variance for answer points when guessing randomly from 4 random choices?*

$$\text{var}(Y) = (1 - 1/16)^2 P(1) + (-1/4 - 1/16)P(1/4) = (15/16)^2(1/4) + (5/16)^2(3/4) = .293$$

8. *Suppose in a group of 10 questions, you are able to rule out 1 incorrect answer on 4 of those questions, but for the remaining 6 questions you cannot rule out any answers. What is the expected number of answer points you will get on those 10 questions? What is the variance of the number of answer points you will receive?*

   Similar to problem The total number of points you will recieve is the sum of each of the individual questions. However, in this case, it is the sum of random variables which are not all identical. However, conceptually it is the same. Let X_i denote a variable which you are not able to rule out any of the answers and let Y denote a variable where you are able to rule out one of the potential answers.

$$\mathcal{E}(T) = \mathbb{E}(X_1 + \dots X_6 + Y_1 + \dots Y_4) = \mathbb{E}(X_1) + \dots \mathbb{E}(X_6) + \mathbb{E}(Y_1) + \dots \mathbb{E}(Y_4) = 6(0) + 4 * 1/16 = 1/4$$

$$\text{var}(T) = \text{var}(X_1 + \dots X_6 + Y_1 \dots Y_4) = \text{var}E(X_1) + \dots \text{var}(X_6) + \text{var}(Y_1) + \dots \text{var}(Y_4) = 6(1/4) + 4(.293) = 2.672$$

9. *If you leave a question unanswered, what is the expectation and variance of how many points you will receive? If you are able to rule out at least one of the answers, is it a better idea to guess randomly or leave the question unanswered?*

   If you skip a question, you automatically recieve a 0, so the expectation and variance are both 0. However, if you are able to rule out one of the answers, then your expectation is slightly positive (1/16), so you should try to guess.

3 Tight Connection

Suppose I am going to a conference in Montreal, and I have a connection in Chicago. My flight from Seattle to Chicago is supposed to arrive at 5:30 AM and my flight from Chicago to Montreal leaves at 6:30 AM. However, there is variability in when I actually arrive in Chicago, so assume that the time I arrive in Chicago is a normal distribution with a mean of 5:30 AM and a standard deviation of 20 minutes. Note that for this problem, you will need the fact that the sum of two normally distributed random variables is also a normal distribution.

1. *Assuming that I can go instantly from gate to gate in the Chicago airport, what is the probability that I will catch my connecting flight to Montreal? (ie, what is the probability that I will arrive in Chicago before 6:30 AM?)*

   There are many ways to do this question, but I find it easiest to think about the minutes to spare (ie the minutes you arrive at your gate before the flight for Montreal leaves). Let S denote the minutes to spare, then

   $$S \sim \mathcal{N}(60, 20^2)$$

   If the minutes to spare is positive, then I will catch my flight to Montreal. Thus, I am intersted in finding $P(S > 0)$. I can find the z-score $\frac{x-\mu}{\sigma} = \frac{0-60}{20} = -3$ and look that up in the table, or I can use `R`. Notice that since we want the are to the right of 0, we need to take the complement.

   ```
   1- pnorm(-3)
   ```

   ```
   ## [1] 0.9986501
   ```

2. *Slightly more realistically, I need to arrive in Chicago with enough time to travel from my arrival gate to my departure gate. Assume that the time it takes to travel by foot between the gates is also a normal distribution with a mean of 30 minutes and a standard deviation of 10 minutes. What is the distribution of how much time I will have to spare? (ie What is the distribution of the number of minutes between when I arrive at my departure gate and when the flight takes off?)*

   Now I need to account for not only my arrival time, but also the time it takes to get between gates. Let W denote the time to walk from gate to gate. So now, my total minutes to spare is $T = S - W$ where

   $$T \sim \mathcal{N}(0, 20^2 + 10^2)$$

   Notice that although we are subtracting the random variables S and W, we actually still add the variances. Now we can find $P(T > 0)$. Calculating the Z-score yields $\frac{0-30}{\sqrt{500}} = -1.34$. Using R to get the area to the right of -1.34 in the standard normal yields

   ```
   1 - pnorm(-1.341)
   ```

   ```
   ## [1] 0.9100398
   ```

3. *What is the probability that I will have between 5 and 15 minutes to spare?)*

   Here, if we are trying to get the area between 5 and 15 minutes, we can get the area to the left of the z-score for 15 and then subtract the area to the left of the z-score for 5.

   ```
   # Area to the left of z-score of 15
   pnorm((15 - 30)/sqrt(500))
   ```

   ```
   ## [1] 0.2511675
   ```

   ```
   # Area to the left of z-score 5
   pnorm((5 - 30)/sqrt(500))
   ```

   ```
   ## [1] 0.1317762
   ```

```r
# Area between
pnorm((15 - 30)/sqrt(500)) - pnorm((5 - 30)/sqrt(500))

## [1] 0.1193912
```

4. *What is the probability that I will have exactly 3 minutes to spare?)*

   Since this is a continuous distribution, the probability of any exact point (not an interval) is 0.

5. *Now suppose I can grab a ride on a cart instead of walking. The time it takes for the cart to get from gate to gate is a normal distribution with a mean of 20 minutes and a standard deviation of 2 minutes. What is the distribution of how much time I will have to spare? (ie What is the distribution of the number of minutes between when I arrive at my departure gate and when the flight takes off?)*

   Now the total time to spare is $S - C$ where C is the time to travel from gate to gate using a cart.

$$T \sim \mathcal{N}(40, 20^2 + 2^2)$$

6. *What is the probability that I will catch my connection if I ride the cart?*

   To calculate-

$$P(T > 0)$$

   I use the corresponding z-score of $\frac{0-40}{\sqrt{404}} = -1.99$

```r
1 - pnorm(-1.99)

## [1] 0.9767045
```

7. *Now assume that my flight touches down in Chicago at 6:15 AM. What is the probability that I will catch my connection if I walk?)*

   Since I know when I have arrived, I am no longer need to account for arrival time, but I still need to get to the gate in less than 15 minutes when walking. So to calculate $P(W < 15)$, I calculate the z score of 15, $\frac{15-30}{20} = -.75$

```r
pnorm(-.75)

## [1] 0.2266274
```

8. *Now assume that my flight touches down in Chicago at 6:15 AM. What is the probability that I will catch my connection if I ride the cart?)*

   Since I know when I have arrived, I am no longer need to account for arrival time, but I still need to get to the gate in less than 15 minutes when riding the cart. So to calculate $P(C < 15)$, I calculate the z score of 15, $\frac{15-20}{2} = -2.5$

```r
pnorm(-2.5)

## [1] 0.006209665
```

9. *Which mode of transportation should I choose now? Explain why this happens.)*

   Walking gives a greater probability of arriving on time. Although it takes longer on average, because there is a lot more variability in the time, the probability of arriving on time (since it is a firm cut-off at 15 minutes) is still higher.

4 Running for mayor

Suppose I am hired by a local politician who is running for mayor. She wants me to conduct a survey to measure what proportion of the population will vote for her. Suppose (unbeknownst to me) the true proportion of the population that will vote for her is .52.

1. *Suppose I select a random individual and record a 1 if the individual plans to vote for the candidate, and record a 0 if the individual does not plan to vote for the candidate. What type of distribution is this? What is the expectation and variance?*

   Since this is a Yes/No or 0/1 random variable, it is a Bernoulli random variable. For a Bernoulli, the expectation is $p = .52$ and the variance is $p(1-p) = .52(1-.52) = .2496$

2. *Now suppose I take a random sample of 200 individuals and count how many of them would vote for the candidate. What type of distribution is this? What is the expectation and variance?*

   Since we are now adding up 200 (independent) Bernoulli random variables, it is now a binomial distribution. For a binomial, the expectation is $np = 104$ and the variance is $np(1-p) = 49.92$.

3. *Now I take a sample of 200 individuals and instead calculate the proportion of the individuals in my sample who would vote for the candidate. Denote this statistic as $\hat{p}$. Using the central limit theorem and assuming that n is large enough so that $\hat{p}$ is approximately normal, what is the mean and variance of $\hat{p}$*

   By the central limit theorem, we know that $\hat{p} \sim \mathcal{N}(p, p(1-p)/n) = \mathcal{N}(.52, 0.0012)$.

4. *Using the distribution from above, what is the probability that a random sample of 200 individuals results in a $\hat{p}$ that is less than .5?*

   To evaluate $P(\hat{p} < .5)$, we first get the z-score $\frac{.5-.52}{\sqrt{.52(1-.52)/200}} = -.566$

```r
pnorm(-.566)

## [1] 0.2856969
```

5. *Using the distribution from above, if I want to ensure that the probability that $\hat{p}$ is less than .5 happens in less than .05 of the time, how many individuals would I need to include in my sample?*

   This is similar to the previous question, but now we need to work backwards. First, let's find the a z-score, $z^\star$ such that there is .05 to the left of $z^\star$. In particular, we can use the `qnorm` function

```r
qnorm(.05)

## [1] -1.644854
```

   So now, we know that we need .5 to have a z-score of -1.645. So we have

$$\frac{.5 - .52}{\sqrt{.52(1-.52)/n}} = -1.645$$

   Solving for n yields a value of 1688.56, so we can round up to 1689.