Southern University of Science And Technology
School of Life Sciences
BIO210 Biostatistics
(Spring, 2022)

Assignment 3
Due on 26th Oct, 11 p.m.

1. **Travelling**: Han Meimei is taking a probability course and want to think about her commutes using probabilistic modelling. She starts promptly at 8am. Han Meimei drives and thus is at the mercy of traffic lights. When all traffic lights on her route are green, the entire trip takes 18 minutes. Han Meimei's route includes 5 traffic lights, each of which is red with probability 1/3, independent of every other light. Each red traffic light that she encounters adds 1 minute to her commute (for slowing, stopping, and returning to speed).

   1.1) **(2.5 points)** Find the PMF, expectation, and variance of the length (in minutes) of Han Meimei's commute.

   1.2) **(5 points)** Given that Han Meimei's commute took her at most 19 minutes, what is the expected number of red lights that she encountered?

   1.3) **(2.5 points)** Given that Han Meimei encountered a total of three red lights, what is the probability that exactly two out of the first three lights were red?

2. **Rolling dice**: Two fair four-sided tetrahedral dice are rolled simultaneously. Let the random variable $\boldsymbol{X}$ be the absolute difference of the two rolls.

   2.1) **(5 points)** Calculate the PMF, the expected value, and the variance of $\boldsymbol{X}$.

   2.2) **(5 points)** Calculate and plot the PMF of $\boldsymbol{X}^2$.

3. **Achieving a High Score**: Suppose that a person plays a game in which his score must be one of the 50 numbers 1, 2, 3, ..., 50 and that each of these 50 numbers is equally likely to be his score. The first time he plays the game, his score is $\boldsymbol{X}$. He then continues to play the game until he obtains another score $\boldsymbol{Y}$, such that $\boldsymbol{Y} \geqslant \boldsymbol{X}$. It may be assumed that all plays of the game are independent.

   3.1) **(2.5 points)** List all possible values that $\boldsymbol{Y}$ can take.

   3.2) **(5 points)** What is the probability that $\boldsymbol{Y}$ takes the value 50 ?

   3.3) **(5 points)** Compute the PMF of $\boldsymbol{Y}$.

4. Going To The Lab: Han Meimei lives in an old town without real-time traffic information. Her lab is located inside a city nearby. Han Meimei usually checks the traffic report in the city to decide how she is going to the lab. If the report is **"traffic jam"**, she will always ride her bike. If the report is **"no traffic jam"**, she will always drive her car. Now, we know that if the report is **"traffic jam"**, the probability of actually having a traffic jam is 0.8. On the other hand, if the report is **"no traffic jam"**, the probability of actually having a traffic jam is 0.1. On *Monday* and *Friday*, the report is **"traffic jam"** 70% of the time and on other days it is 20%.

4.1) (5 points) One day, Han Meimei missed the report and there is a traffic jam in the city. What is the probability that the report was **"traffic jam"** if it was on Monday or Friday? What is the probability that the report was **"traffic jam"** if it was on other days of the week?

4.2) (5 points) The probability of Han Meimei missing the report is equal to 0.2 on any day of the week. If she misses the report, Han Meimei will flip a fair coin to decide whether to ride her bike or drive her car. Let event A = { Han Meimei is riding her bike }, and event B = { The report is **"no traffic jam"** }. Are events A and B independent? Does your answer depend on the days of the week?

4.3) (5 points) Han Meimei is riding a bike and there is no traffic jam. What is the probability that she saw the report? Does it depend on the days of the week?

Han Meimei's university encourages low carbon travel. If she rides her bike to the lab, she gets 5 points on that day; if she drives her car to the lab, she gets 0 points on that day. Let the random variable $\boldsymbol{X}$ represent the points that Han Meimei gets on a day.

4.4) (2.5 points) Which of the following statements are correct about the random variable $\boldsymbol{X}$? Put a ✓ in front of the statements that you think are correct.

[ ] $\boldsymbol{X}$ is a discrete random variable
[ ] $\boldsymbol{X}$ is a Bernoulli random variable
[ ] $\boldsymbol{X}$ is a Binomial random variable
[ ] $\boldsymbol{X}$ is a continuous random variable

4.5) (5 points) Compute the **PMF** of $\boldsymbol{X}$.

4.6) (5 points) Compute $E[\boldsymbol{X}]$ and $var(\boldsymbol{X})$

5. **Who is correct?** In a class of 25 students, 11 of them have type **O** blood, 6 type **A**, 5 type **B** and 3 type **AB**. If we randomly select a sample of 5 students, and let the random variable $\boldsymbol{X}$ represents the number of students with type **B** within the selected sample. Compute the PMF $p_{\boldsymbol{X}}(x)$.

   5.1) **(2.5 points)** Han Meimei approaches the problem in this way: the selection is random, and all outcomes are equally likely. Therefore, she can use the discrete uniform law to calculate probabilities. The total number of outcomes in the sample space is $|\Omega| = \binom{25}{5}$. To figure out the the number of outcomes of having k students with type **B** blood in the sample, she divides the process into two stages. The first stage is to choose k people from the 5 students with blood type **B** in the class, and the second stage is to choose $5 - k$ people from the other 20 students. The total number of outcomes is the simple multiplication of the number of choices in each stage. Now, write the PMF constructed by Han Meimei.

   5.2) **(2.5 points)** Li Lei thinks in a different way: the probability is kind of a relative frequency. Therefore, the probability of having a random student with blood type **B** is 5 out of 25. That is 0.2. If 5 students are chosen, the process of choosing one student can be treated as a Bernoulli trial, and there are a total of 5 Bernoulli trials. Therefore, the probability of observing k students with blood type **B** in the sample can be simply calculated using a binomial distribution. Now, write the PMF constructed by Li Lei.

   5.3) **(5 points)** Based on the previous two PMFs you just computed, finish the following table to see if they are different or not:

k	$P(\boldsymbol{X} = k)$ by Han Meimei	$P(\boldsymbol{X} = k)$ by Li Lei
0		
1		
2		
3		
4		
5		

   5.4) **(2.5 points)** Who do you think is correct and who is wrong? Explain your answers.

5.5) **(5 points)** In a different class of 256 students, of which 110 of them have type **O** blood, 60 type **A**, 50 type **B** and 36 type **AB**. If we randomly select a sample of 5 students, and let the random variable $\boldsymbol{Y}$ represents the number of students with type **B** within the selected sample. Repeat the analysis in **1.1)**, **1.2)** and **1.3)** (*i.e.* compute the PMF $p_{\mathbf{Y}}(y)$ using Han Meimei's and Li Lei's methods, respectively, and compare them in a table). What do you notice about the difference between probabilities calculated by Han Meimei and Li Lei?

6. Checking independence of a collection of events (2.5 points): During the lecture, we made a definition on the independence of a collection of events by using a multiplication equation (Lecture 8 Slide 6). Now suppose we have a collection of $\boldsymbol{n}$ events, how many times do you need to use the equation in order to check if they are independent or not?

7. Telecommunication (2.5 points): In a terrible environment, the probability of success in sending a character by wireless is $\frac{3}{7}$. What is the probability that 22 characters out of 44 are sent successfully, assuming the results of sending each character are independent?

8. Renal Disease: The presence of bacteria in a urine sample (bacteriuria) is sometimes associated with symptoms of kidney disease in women. Suppose a determination of bacteriuria has been made over a large population of women at one point in time and 5% of those sampled are positive for bacteriuria.

8.1) **(2.5 points)** If a sample size of 5 is selected from this population, what is the probability that 1 or more women are positive for bacteriuria?

8.2) **(2.5 points)** Suppose 100 women from this population are sampled. What is the probability that 3 or more of them are positive for bacteriuria?

One interesting phenomenon of bacteriuria is that there is a turnover; that is, if bacteriuria is measured on the same woman at two different time points, the results are not necessarily the same. Assume that 20% of all women who are bacteriuric at time 0 are again bacteriuric at time 1 (1 year later), whereas only 4.2% of women who were not bacteriuric at time 0 are bacteriuric at time 1. Let

$\boldsymbol{X}$ be the random variable representing the number of bacteriuric events over the two time periods for 1 woman and still assume that the probability that a woman will be positive for bacteriuria at any one exam is 5%.

8.3) **(2.5 points)** What is the probability distribution of $\boldsymbol{X}$?

8.4) **(2.5 points)** What is the mean of $\boldsymbol{X}$?

8.5) **(2.5 points)** What is the variance of $\boldsymbol{X}$?

9. **Otolaryngology**: Assume the number of episodes per year of otitis media, a rare disease of the middle ear in early childhood, follows a Poisson distribution with parameter $\lambda = 1.6$ episodes per year.

9.1) **(2.5 points)** Find the probability of getting 3 or more episodes of otitis media in the first 2 years of life.

9.2) **(2.5 points)** Find the probability of not getting any episodes of otitis media in the first year of life.