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

Assignment 4
Due on 24th Mar, 11 p.m.

1. **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 $p_{\boldsymbol{X}}(x)$.

   1.1) **(7.5 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.

   1.2) **(7.5 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.

   1.3) **(7.5 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		

1.4) **(7.5 points)** Who do you think is correct and who is wrong? Explain your answers.

1.5) **(10 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 $p_{\boldsymbol{Y}}(y)$ 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?

2. Counting E-mails (10 points): During Lecture 11, we talked about using the **Poisson** distribution to model the number of E-mails received per day. Now we know that in order for the Poisson distribution to work, there are some assumptions that need to be met (Lecture 11, slide 15). Do you think the number of E-mails received per day follows a Poisson distribution? Why or why not?

3. Checking independence of a collection of events (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?

4. Telecommunication (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?

5. Bacteria In A River (5 points): Assume that bacteria of a species called Leptospira are randomly distributed in a river according to the Poisson distribution with an average concentration of 16 per 40 ml of water. If we draw 10 ml of water from the river using a test tube, what is the approximate probability that the number of bacteria Leptospira in the sample is exactly 4?

6. **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.

   6.1) **(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?

   6.2) **(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%.

   6.3) **(5 points)** What is the probability distribution of $\boldsymbol{X}$?

   6.4) **(5 points)** What is the mean of $\boldsymbol{X}$?

   6.5) **(5 points)** What is the variance of $\boldsymbol{X}$?

7. **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.

   7.1) **(5 points)** Find the probability of getting 3 or more episodes of otitis media in the first 2 years of life.

   7.2) **(5 points)** Find the probability of not getting any episodes of otitis media in the first year of life.