STAT 521: Assignment 5

Make sure to show your computation and/or attach appropriate output.

Problem 1

Conduct the most appropriate test for Example 8 (page 13) in Lecture 7 note. Make sure to state null and alternative hypotheses. What is your conclusion?

Problem 2

Conduct the most appropriate test for Example 9 (page 14) in Lecture 7 note. Make sure to state null and alternative hypotheses. What is your conclusion?

Problem 3

The following are intraocular pressure (mmHg) values recorded for a sample of 21 elderly subjects:

14.5	12.9	14.0	16.1	12.0	17.5	14.1
12.9	17.9	12.0	16.4	24.2	12.2	14.4
17.0	10.0	18.5	20.8	16.2	14.9	19.6

a) Use SAS, R or Python to calculate the sample mean and standard deviation. Please attach your output.

b) Create a histogram and a normal probability plot. Does the data look approximately normal? Attach your output.

c) We want to test if the mean of the population from which the sample was drawn is equal to 14. State the null/alternative hypotheses.

d) Conduct the appropriate test using $\alpha = 0.05$. Do this by hand using the information obtained in part (a), and verify the result using SAS, R or Python. Report a two-sided p-value, as well as a 95% confidence interval. Attach your output.

e) Write your conclusions.

Problem 4

Jacquemyn et al. conducted a survey among gynecologists-obstetricians in the Flanders region and obtained 295 responses. Of those responding, 90 indicated that they had performed at least one cesarean section on demand every year. Does this study provide sufficient evidence for us to conclude that less than 35% of the gynecologists-obstetricians in the Flanders region perform at least one cesarean section on demand each year?

a) State the null/alternative hypotheses.

b) Conduct the appropriate test using $\alpha = 0.05$. What is your conclusion?

c) Calculate the p-value.

Problem 5

Studies in the general population suggest that the gestational length of uncomplicated pregnancies varies according to a normal distribution with $\mu = 39$ weeks with $\sigma = 2$ weeks. A sample of 22 middle-class African American women demonstrates an average gestation length of 38.5 weeks.

a) Test whether the mean gestation period in the African American women is significantly different from the population mean of 39 weeks. State the null/alternative hypotheses. Use $\alpha = 0.05$. What is your conclusion?

b) Suppose the average gestational length in African American women is actually 38.5 weeks. What was the power of the test to detect the difference from $\mu = 39$ weeks at $\alpha = 0.05$? Use SAS, R, Python, or G*Power. Attach your output.

c) How large a sample would be needed to detect the proposed difference in gestational lengths with 90% power at $\alpha = 0.05$? Do this by hand and compare the result using SAS, R, Python or G*Power. Attach your output.

Problem 6 (Biostats students only)

Joint distribution – discrete cases:

Let X and Y be discrete random variables. The joint probability function for X and Y is given by:

$$p(x, y) = P(X = x, Y = y)$$

Now define the *joint distribution function* of X and Y as:

$$F_{X,Y}(x, y) = P(X \leq x,\ Y \leq y) = \sum_{i \leq x} \sum_{j \leq y} p(i, j)$$

Example: Let X and Y be the numbers facing up when you toss two dice (one for X, one for Y). Then the probability of obtaining $(X, Y) = (1, 1)$ is:

$$p(1, 1) = P(X = 1,\ Y = 1) = \frac{1}{36}$$

Actually, since $p(x) = P(X = x) = \frac{1}{6}$ for $x = 1, 2, \cdots, 6$ and $p(y) = P(Y = y) = \frac{1}{6}$ for $y = 1, 2, \cdots, 6$,

$$p(x, y) = P(X = x,\ Y = y) = \frac{1}{36} \qquad \text{for all pairs of values of } (x, y)$$

This is a joint probability mass function of X and Y.

Now, what is $F_{X,Y}(2, 3)$ in this experiment?

$$\begin{aligned}F_{X,Y}(2,3) &= P(X \le 2,\ Y \le 3)\\ &= p(1,1)+p(1,2)+p(1,3)+p(2,1)+p(2,2)+p(2,3)\\ &= \frac{6}{36} = \frac{1}{6}\end{aligned}$$

Note that $p(x,y) \ge 0$ for all (x,y) and $\sum_i \sum_j p(i,j) = 1$. The sum of the joint PMF must be 1.

Question: Below is the joint probability function associated with data obtained in a retrospective cohort study of Ebola virus infection among healthcare workers treating Ebola patients in West Africa. Specifically, the study focused on what type of protection the healthcare worker used and whether or not he/she was infected with the virus. Define:

$$X = \begin{cases} 0, & \text{if no protection used} \\ 1, & \text{if only a mask was used} \\ 2, & \text{if a hazmat suit was used} \end{cases}, \qquad Y = \begin{cases} 0, & \text{if not infected} \\ 1, & \text{if infected} \end{cases}$$

X	Y		
	0	1	Total
0	0.05	0.56	0.61
1	0.10	0.18	0.28
2	0.09	0.02	0.11
Total	0.24	0.76	1.00

Find $F_{X,Y}(2,0)$. What is the interpretation of this value?

Problem 7 (Biostats students only)

Joint distribution – continuous cases:

Let X and Y be continuous random variables. The joint distribution function for X and Y is given by:

$$F_{X,Y}(x,y) = P(X < x,\ Y < y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t_1,t_2)\ dt_2 dt_1$$

where $f_{X,Y}(x,y)$ is the *joint probability density function* (PDF) of X and Y.

A joint PDF must satisfy:

a) $f_{X,Y}(x,y) \ge 0$ for all x, y
b) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dxdy = 1$. The total volume under surface of a joint PDF must be 1.

Example: Let X and Y have the joint probability density function:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{6}(x+4y), & 0 \le x \le 2,\ 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

We want to find $F\left(1, \frac{1}{2}\right) = P\left(X \le 1, Y \le \frac{1}{2}\right)$.

$$
\begin{aligned}
F\left(1, \frac{1}{2}\right) &= \int_0^1 \int_0^{\frac{1}{2}} \frac{1}{6}(x+4y)\ dydx \\
&= \int_0^1 \left(\frac{1}{6}xy + \frac{1}{3}y^2 \Big]_0^{\frac{1}{2}} \right) dx \\
&= \int_0^1 \left(\frac{1}{12}x + \frac{1}{12} \right) dx = \frac{1}{24}x^2 + \frac{1}{12}x \Big]_0^1 = \frac{1}{8}
\end{aligned}
$$

Question: Let X and Y have the joint probability density function given by:

$$
f_{X,Y}(x,y) = \begin{cases} kxy, & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}
$$

a) Find the value of k that makes this a joint PDF.

b) Find the joint distribution function $F_{X,Y}(x,y)$.

c) Find $P\left(X \le \frac{1}{2},\ Y \le \frac{3}{4}\right)$.