STAT 521: Assignment 8

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

Problem 1 (Computer exercise)

A person's muscle mass is expected to decrease with age. To explore this relationship in women, a nutritionist randomly selected 15 women from each 10-year age group, beginning with age 40 and ending with age 79, and measured their muscle mass.

a) Create a scatter plot between age and muscle mass. Choose variables for x- and y-axes appropriately. Is there any relationship between age and muscle mass? If so, is the association positive or negative? Does the relationship appear to be linear? Any outliers? **Describe your findings**.

b) Obtain the Pearson correlation coefficient between age and muscle mass. Is the correlation significantly different from zero? Report the 95% confidence interval for ρ

c) Run simple regression. **Again, make sure to choose an appropriate variable for each of X and Y.** Report the estimated regression equation. Attach the ANOVA table. What is the value of R^2 and its interpretation? What is a point estimate for σ^2? Produce a scatter plot with the fitted regression line.

d) What is a point estimate of the difference in the mean muscle mass for women differing in age by one year? Report its 95% confidence interval too.

e) Suppose you wish to predict muscle mass of a woman aged 60 based on your regression model. Calculate her predicted muscle mass. Report its 95% prediction interval too. If the woman has muscle mass of 105, what is the value of the residual for her?

f) Is it appropriate to estimate muscle mass of a woman aged 20 using the regression equation you obtained above? Discuss.

g) Produce a residual plot against fitted (predicted) values, as well as a normal probability plot of residuals. Are there any outliers? Are residuals normally distributed? Is there any non-linear pattern in residuals? How about the equal variance assumption?

Problem 2

Complete the ANOVA table below for simple linear regression of $n = 27$ and answer the following questions.

Source	***df***	***SS***	***MS***	***F***
Model			840	10.5
Error				
Total				

a) Calculate R^2.

b) It is known that X and Y used in here are negatively associated. Using the information above, what is the correlation coefficient between X and Y?

c) Means and standard deviations of X and Y are given below. Using this and part (b), obtain the estimated regression equation.

	Mean	**SD**
X	110.2	20.5
Y	55.0	8.2

Problem 3 (Biostats students only)

Expected value of a function of random variables:

In Assignment #5, you learned the joint probability density (or mass) function. Suppose random variables X and Y have a joint probability density/mass function, $f_{X,Y}(x,y)$. How can we calculate the expected value of any function of X and Y, e.g., $E\left[g\left(X,Y\right)\right]$?

Definition:

For discrete random variable X and Y:

$$E\left[g\left(X,Y\right)\right] = \sum_x \sum_y g(x,y)\, f_{X,Y}(x,y)$$

For continuous random variable X and Y:

$$E\left[g\left(X,Y\right)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{X,Y}(x,y)\, dxdy$$

Example: Let X and Y, both continuous 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 $E(XY)$.

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\, dxdy \\ &= \int_0^1\int_0^2 xy\frac{1}{6}(x+4y)\, dxdy \\ &= \int_0^1\int_0^2 \frac{1}{6}(x^2y+4xy^2)\, dxdy \\ &= \int_0^1 \left(\frac{1}{18}x^3y + \frac{1}{3}x^2y^2\Big]_0^2\right) dy \\ &= \int_0^1 \left(\frac{4}{9}y + \frac{4}{3}y^2\right) dy = \frac{2}{9}y^2 + \frac{4}{9}y^3\Big]_0^1 = \frac{2}{3} \end{aligned}$$

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

$$f_{X,Y}(x,y) = \begin{cases} 4xy, & 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0, & \text{elsewhere} \end{cases}$$

a) Find $E(X)$ and $E(Y)$.

b) The covariance between X and Y can be written as:

$$cov(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right] = E(XY) - \mu_X\mu_Y$$

where $\mu_X = E(X)$ and $\mu_Y = E(Y)$. Find $E(XY)$ and $cov(X,Y)$.

c) Use the part (b) to find the correlation coefficient between X and Y: $\rho_{X,Y}$ Remember, from Lecture 10, the correlation coefficient is defined as:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$$

Problem 4 (Biostats students only)

Marginal probability distributions, independence of RVs

Suppose you have a joint probability distribution of two discrete random variables X and Y, $p_{X,Y}(x,y)$, as follows (example taken from Assignment #5):

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

How do we get the marginal probability distribution of X, that is, $p_X(x)$? This is easy because:

$$\begin{aligned} p_X(0) &= Pr(X=0) = p_{X,Y}(0,0) + p_{X,Y}(0,1) = 0.61 \\ p_X(1) &= Pr(X=1) = p_{X,Y}(1,0) + p_{X,Y}(1,1) = 0.28 \\ p_X(2) &= Pr(X=2) = p_{X,Y}(2,0) + p_{X,Y}(2,1) = 0.11 \end{aligned}$$

In general, marginal probability mass functions (for discrete cases) are given by:

$$p_X(x) = \sum_y p_{X,Y}(x,y) \qquad \text{and} \qquad p_Y(y) = \sum_x p_{X,Y}(x,y)$$

Similarly, if X and Y are continuous, marginal density functions are:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \qquad \text{and} \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

Example: Let X and Y, both continuous, 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 $f_X(x)$ and $f_Y(y)$.

For $0 \le x \le 2$,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^1 \frac{1}{6}(x+4y)\,dy = \frac{1}{6}xy + \frac{1}{3}y^2\Big]_0^1 = \frac{x+2}{6}$$

and $f_X(x) = 0$ elsewhere. For $0 \le y \le 1$,

$$f_y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^2 \frac{1}{6}(x+4y)\,dx = \frac{1}{12}x^2 + \frac{2}{3}xy\Big]_0^2 = \frac{4y+1}{3}$$

and $f_Y(y) = 0$ elsewhere.

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

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

a) Find $f_X(x)$ and $f_Y(y)$.

b) If the joint density can be written as the product of individual density functions, i.e.,

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

then X and Y are said to be independent random variables. Are X and Y independent?

c) If random variables X and Y are independent, then we have this property:

$$E(XY) = E(X)E(Y)$$

Find $E(XY)$. Is your answer same as the one you get for Problem 3 part (b)?

Additional notes: If random variables X and Y are independent, then the covariance between X and Y is zero and so is the correlation.

$$\text{If } X \text{ and } Y \text{ independent} \implies cov(X,Y) = 0$$

But the converse is not necessarily true. You can have a correlation of zero between X and Y that are not independent. See the following joint probability distribution.

X	Y			
	0	1	2	Total
0	1/3	0	1/3	2/3
1	0	1/3	0	1/3
Total	1/3	1/3	1/3	1

Verify that $cov(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{3} - \left(\frac{1}{3}\right)(1) = 0$. But X and Y are not independent because $Pr(X=0\,|\,Y=0) \neq Pr(X=0)$.