

Q1. Assume there are 120 students in our B565 class, and each person has a 2% of chance of carrying coronavirus. We also know that the Omicron variant dominates and let's assume that it accounts for 90% of the new Covid cases. What's the probability that the entire class is free of coronavirus and Omicron, respectively? Show your work. [10 pts]

i. The probability that the entire class is free of coronavirus:

If each student has a 2% chance of carrying coronavirus, then each student has a 98% chance of not carrying coronavirus.

$1 - 0.02 = 0.98$  is the probability that a student doesn't have the coronavirus.

Hence, if a class of 120 students doesn't have coronavirus, the probability of that happening is  **$0.98^{120} = 0.08853787272$**  which is 8.8% approximately. (We multiply the probability of each student not having coronavirus).

ii. The probability that the entire class is free of Omicron:

If each student has a 2% chance of carrying coronavirus, then each student has a  $0.02 \times 0.9 = 0.018$  probability of carrying Omicron variant. The probability of a student not having Omicron is  $1 - 0.018 = 0.982$ .

Hence, if a class of 120 students doesn't have coronavirus, the probability of that happening is  **$0.982^{120} = 0.011307810828$**  which is 11.3% approximately. (We multiply the probability of each student not having Omicron).

Q2. Let  $\Omega$  be the space of possible outcomes of a fair die (with six sides) thrown twice. What is  $\Omega$ ? Let A be the event that a 4 is observed on either individual throw or the sum of both throws is at least 5. Let B be the event that the difference between the two throws is exactly two. Are A and B independent? What is the probability of B given A? What is the probability of A given B? Show your work. [15 pts]

i.  $\Omega = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$ . Size of  $\Omega$  is 36.  
 $n(\Omega) = 36$

ii. A = Event when a 4 is observed on either individual throw or the sum of both throws is at least 5

$A = \{(1, 4), (1, 5), (1, 6), (2, 3), (2, 4), (2, 5), (2, 6), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$   
 $n(A) = 36$ .

B = The event that the difference between the two throws is exactly two.

$B = \{(1, 3), (2, 4), (3, 5), (4, 6), (3, 1), (4, 2), (5, 3), (6, 4)\}$ .  
 $n(B) = 8$

Set  $A \cap B = \{(2, 4), (3, 5), (4, 6), (4, 2), (5, 3), (6, 4)\}$ .  $n(A \cap B) = 6$ .

**A and B are not independent as  $A \cap B$  is not empty. .**

$$P(A)=n(A)/n(\Omega)=30/36=5/6$$

$$P(B)=n(B)/n(\Omega)=8/36=2/9$$

$$P(A \cap B)=n(A \cap B)/n(\Omega)=6/36=1/6$$

$$\text{iii. } P(B|A)=(P(A|B)P(B))/P(A)=P(A \cap B)/P(A)=(1/6)/(5/6)=1/5$$

$$\text{iv. } P(A|B)=(P(B|A)P(A))/P(B)=P(A \cap B)/P(B)=(1/6)/(2/9)=3/4$$

Q3. Recall that given two vectors  $x$  and  $y$  of  $n$  dimensions, their cosine similarity, Euclidean distance, and correlation can be computed as follows. [15 pts]

Cosine similarity,  $\cos(x, y) = (x \cdot y) / \|x\| \cdot \|y\|$

Euclidean distance,  $d(x, y) = (\sum_{i=1}^n (x_i - y_i)^2)^{1/2}$

Correlation,  $\text{corr}(x, y) = S_{xy} / (S_x S_y)$

$$= (\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})) / (\sum_{i=1}^n (x_i - \bar{x})^2)^{1/2} \cdot (\sum_{i=1}^n (y_i - \bar{y})^2)^{1/2}$$

Sets can also be represented as vectors of zeros and ones, so for those vectors, Jaccard similarity (intersection over union) can be used. For the following vectors  $x$  and  $y$ , calculate the indicated similarity or distance measures. Show the steps.

- $x = (1, 1, 1, 1)$ ,  $y = (2, 2, 2, 2)$  cosine, correlation and Euclidean.
- $x = (0, 1, 0, 1)$ ,  $y = (1, 0, 1, 0)$ , cosine, correlation, Euclidean, Jaccard
- $x = (0, -1, 0, 1)$ ,  $y = (1, 0, -1, 0)$ , cosine, correlation, Euclidean
- $x = (1, 1, 0, 1, 0, 1)$ ,  $y = (1, 1, 1, 0, 0, 1)$  cosine, correlation, Jaccard
- $x = (2, -1, 0, 2, 0, -3)$ ,  $y = (-1, 1, -1, 0, 0, -1)$  cosine, correlation

$$\text{i. } x = (1, 1, 1, 1), y = (2, 2, 2, 2)$$

a. Cosine Similarity:

$$x \cdot y = 1 \cdot 2 + 1 \cdot 2 + 1 \cdot 2 + 1 \cdot 2 = 8$$

$$\|x\| = (1^2 + 1^2 + 1^2 + 1^2)^{1/2} = 2$$

$$\|y\| = (2^2 + 2^2 + 2^2 + 2^2)^{1/2} = 4$$

$$\cos(x, y) = (x \cdot y) / \|x\| \cdot \|y\| = 8 / (2 \cdot 4) = 1$$

b. Correlation:

$$\bar{x} = (\sum_{i=1}^n (x_i)) / n = (1+1+1+1) / 4 = 1$$

$$\bar{y} = (\sum_{i=1}^n (y_i)) / n = (2+2+2+2) / 4 = 2$$

$$(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})) = (1-1) \cdot (2-2) + (1-1) \cdot (2-2) + (1-1) \cdot (2-2) + (1-1) \cdot (2-2) = 0$$

$$(\sum_{i=1}^n (x_i - \bar{x})^2) = (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 = 0$$

$$(\sum_{i=1}^n (x_i - \bar{x})^2)^{1/2} = 0$$

$$(\sum_{i=1}^n (y_i - \bar{y})^2) = (2-2)^2 + (2-2)^2 + (2-2)^2 + (2-2)^2 = 0$$

$$(\sum_{i=1}^n (y_i - \bar{y})^2)^{1/2} = 0$$

Since, both the numerator and the denominator are 0, the correlation for vectors **x** and **y** are not defined.

c. Euclidean distance:

$$\sum_{i=1}^n (x_i - y_i)^2 = (1-2)^2 + (1-2)^2 + (1-2)^2 + (1-2)^2 = 4 \cdot (-1)^2 = 4$$

$$(\sum_{i=1}^n (x_i - y_i)^2)^{1/2} = 4^{1/2} = 2$$

ii.  $x = (0, 1, 0, 1)$ ,  $y = (1, 0, 1, 0)$

a. Cosine Similarity:

$$x \cdot y = 0 \cdot 1 + 1 \cdot 0 + 0 \cdot 1 + 1 \cdot 0 = 0$$

$$\|x\| = (0^2 + 1^2 + 0^2 + 1^2)^{1/2} = 2^{1/2}$$

$$\|y\| = (1^2 + 0^2 + 1^2 + 0^2)^{1/2} = 2^{1/2}$$

$$\cos(x, y) = (x \cdot y) / (\|x\| \cdot \|y\|) = 0$$

b. Correlation:

$$\bar{x} = (\sum_{i=1}^n x_i) / n = (0 + 1 + 0 + 1) / 4 = 1/2$$

$$\bar{y} = (\sum_{i=1}^n y_i) / n = (1 + 0 + 1 + 0) / 4 = 1/2$$

$$(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}))$$

$$= (0 - 1/2)(1 - 1/2) + (1 - 1/2)(0 - 1/2) + (0 - 1/2)(1 - 1/2) + (1 - 1/2)(0 - 1/2) = (-1/4) \cdot 4 = -1$$

$$(\sum_{i=1}^n (x_i - \bar{x})^2) = (0 - 1/2)^2 + (1 - 1/2)^2 + (0 - 1/2)^2 + (1 - 1/2)^2 = 4 \cdot 1/4 = 1$$

$$(\sum_{i=1}^n (x_i - \bar{x})^2)^{1/2} = 1$$

$$(\sum_{i=1}^n (y_i - \bar{y})^2) = (1 - 1/2)^2 + (0 - 1/2)^2 + (1 - 1/2)^2 + (0 - 1/2)^2 = 4 \cdot 1/4 = 1$$

$$(\sum_{i=1}^n (y_i - \bar{y})^2)^{1/2} = 1$$

Hence,

$$\text{corr}(x, y) = S_{xy} / (S_x S_y) = -1/1 \cdot 1 = -1$$

c. Euclidean distance:

$$\sum_{i=1}^n (x_i - y_i)^2 = (0-1)^2 + (1-0)^2 + (0-1)^2 + (1-0)^2 = 4$$

$$(\sum_{i=1}^n (x_i - y_i)^2)^{1/2} = 4^{1/2} = 2$$

d. Jaccard Coefficient:

$$JC = (f_{11}) / (f_{01} + f_{10} + f_{11}) = 0 / (2 + 2 + 0) = 0$$

iii.  $x = (0, -1, 0, 1)$ ,  $y = (1, 0, -1, 0)$

a. Cosine Similarity:

$$x \cdot y = 0 \cdot 1 + (-1) \cdot 0 + 0 \cdot (-1) + 1 \cdot 0 = 0$$

$$\|x\| = (0^2 + (-1)^2 + 0^2 + 1^2)^{1/2} = 2^{1/2}$$

$$\|y\| = (1^2 + 0^2 + (-1)^2 + 0^2)^{1/2} = 2^{1/2}$$

$$\cos(x, y) = (x \cdot y) / (\|x\| \cdot \|y\|) = 0$$

b. Correlation:

$$\bar{x} = (\sum_{i=1}^n x_i) / n = (0 + -1 + 0 + 1) / 4 = 0$$

$$\bar{y} = (\sum_{i=1}^n y_i) / n = (1 + 0 + -1 + 0) / 4 = 0$$

$$(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}))$$

$$= (0)(1) + (-1)(0) + (0)(-1) + (1)(0) = 0$$

$$(\sum_{i=1}^n (x_i - \bar{x})^2) = (0)^2 + (-1)^2 + (0)^2 + (1)^2 = 2$$

$$(\sum_{i=1}^n (x_i - \bar{x})^2)^{1/2} = 2^{1/2}$$

$$(\sum_{i=1}^n (y_i - \bar{y})^2) = (1)^2 + (0)^2 + (-1)^2 + (0)^2 = 2$$

$$(\sum_{i=1}^n (y_i - \bar{y})^2)^{1/2} = 2^{1/2}$$

Hence,

$$\text{corr}(x, y) = S_{xy} / (S_x S_y) = 0$$

c. Euclidean distance:

$$\sum_{i=1}^n (x_i - y_i)^2 = (0-1)^2 + (-1-0)^2 + (0-(-1))^2 + (1-0)^2 = 4$$

$$(\sum_{i=1}^n (x_i - y_i)^2)^{1/2} = 4^{1/2} = 2$$

iv.  $x = (1, 1, 0, 1, 0, 1)$ ,  $y = (1, 1, 1, 0, 0, 1)$

a. Cosine Similarity:

$$x \cdot y = 1 \cdot 1 + 1 \cdot 1 + 0 \cdot 1 + 1 \cdot 0 + 0 \cdot 0 + 1 \cdot 1 = 3$$

$$\|x\| = (1^2 + 1^2 + 0^2 + 1^2 + 0^2 + 1^2)^{1/2} = 4^{1/2} = 2$$

$$\|y\| = (1^2 + 1^2 + 1^2 + 0^2 + 0^2 + 1^2)^{1/2} = 4^{1/2} = 2$$

$$\cos(x, y) = (x \cdot y) / (\|x\| \cdot \|y\|) = 3 / (2 \cdot 2) = 0.75$$

b. Correlation:

$$\bar{x} = (\sum_{i=1}^n x_i) / n = (1 + 1 + 0 + 1 + 0 + 1) / 6 = 4/6 = 2/3$$

$$\bar{y} = (\sum_{i=1}^n y_i) / n = (1 + 1 + 1 + 0 + 0 + 1) / 6 = 4/6 = 2/3$$

$$(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}))$$

$$= (1-2/3)(1-2/3) + (1-2/3)(1-2/3) + (0-2/3)(1-2/3) + (1-2/3)(0-2/3) + (0-2/3)(0-2/3) + (1-2/3)(1-2/3)$$

$$= 1/9 + 1/9 - 2/9 - 2/9 + 4/9 + 1/9 = 3/9 = 1/3$$

$$(\sum_{i=1}^n (x_i - \bar{x})^2)$$

$$= (1-2/3)^2 + (1-2/3)^2 + (0-2/3)^2 + (1-2/3)^2 + (0-2/3)^2 + (1-2/3)^2 = 1/9 + 1/9 + 4/9 + 1/9 + 4/9 + 1/9 = 4/3$$

$$(\sum_{i=1}^n (x_i - \bar{x})^2)^{1/2} = (4/3)^{1/2}$$

$$(\sum_{i=1}^n (y_i - \bar{y})^2)$$

$$= (1-2/3)^2 + (1-2/3)^2 + (1-2/3)^2 + (0-1/2)^2 + (0-1/2)^2 + (1-2/3)^2 = 1/9 + 1/9 + 1/9 + 4/9 + 4/9 + 1/9 = 4/3$$

$$(\sum_{i=1}^n (y_i - \bar{y})^2)^{1/2} = (4/3)^{1/2}$$

Hence,

$$\text{corr}(x, y) = S_{xy}/(S_x S_y) = 1/3/(4/3)^{1/2} * (4/3)^{1/2} = 0.25$$

c. Jaccard Coefficient:

$$\text{JC} = (f_{11})/(f_{01}+f_{10}+f_{11}) = 3/(1+1+3) = 0.6$$

v.  $x = (2, -1, 0, 2, 0, -3), y = (-1, 1, -1, 0, 0, -1)$

a. Cosine Similarity:

$$x \cdot y = 2 \cdot -1 + -1 \cdot 1 + 0 \cdot -1 + 2 \cdot 0 + 0 \cdot 0 + -3 \cdot -1 = -2 - 1 + 3 = 0$$

$$\|x\| = (2^2 + (-1)^2 + 0^2 + 2^2 + 0^2 + (-3)^2)^{1/2} = 18^{1/2}$$

$$\|y\| = (-1)^2 + 1^2 + (-1)^2 + 0^2 + 0^2 + (-1)^2)^{1/2} = 4^{1/2} = 2$$

$$\cos(x, y) = (x \cdot y) / (\|x\| \cdot \|y\|) = 0 / (2 * 18^{1/2}) = 0.0$$

b. Correlation:

$$\bar{x} = (\sum_{i=1}^n x_i) / n = (2 - 1 + 0 + 2 + 0 - 3) / 6 = 0$$

$$\bar{y} = (\sum_{i=1}^n y_i) / n = (-1 + 1 - 1 + 0 + 0 - 1) / 6 = -2/6$$

$$(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}))$$

$$= (2)(-1 + 2/6) + (-1)(1 + 2/6) + (0)(1 + 2/6) + (2)(-2/6) + (0)(+2/6) + (-3)(-1 + 2/6) = 0$$

$$(\sum_{i=1}^n (x_i - \bar{x})^2) = (2)^2 + (-1)^2 + (0)^2 + (2)^2 + 0^2 + (-3)^2 = 18$$

$$(\sum_{i=1}^n (x_i - \bar{x})^2)^{1/2} = 18^{1/2}$$

$$(\sum_{i=1}^n (y_i - \bar{y})^2) = (-1 + 2/6)^2 + (1 + 2/6)^2 + (1 + 2/6)^2 + (-2/6)^2 + (+2/6)^2 + (-1 + 2/6)^2 = 10/3$$

$$(\sum_{i=1}^n (y_i - \bar{y})^2)^{1/2} = (10/3)^{1/2}$$

Hence,

$$\text{corr}(x, y) = S_{xy} / (S_x S_y) = 0 / ((18^{1/2}) * ((10/3)^{1/2})) = 0$$

