

מדעי הנתונים – פרויקט גמר – חלק א'**שאלות תיאורטיות ותכנות פשוטות**

1. **א:** לפי השאלה ישנם שלושה נתונים:

א. ההסתברות להולדת בן היא $\frac{1}{2}$ ונסמן זאת $P(A)$

(תאומים זהים בהכרח יהיו מאותו מין, תאומים שאינם זהים יכולים להיות מאותו מין או שלא מאותו מין)

ב. ההסתברות להולדת תאומים זהים היא $\frac{1}{300}$ ונסמן זאת $P(B)$

ג. ההסתברות להולדת תאומים שאינם זהים היא $\frac{1}{125}$ ונסמן זאת $P(C)$

נשתמש בנוסחת בייס:

$$P\left(\frac{\text{זהים}}{\text{בן}}\right) = \frac{P\left(\frac{\text{בן}}{\text{זהים}}\right) * P(\text{זהים})}{P\left(\frac{\text{בן}}{\text{זהים}}\right) * P(\text{זהים}) + P\left(\frac{\text{בן}}{\text{לא זהים}}\right) * P(\text{לא זהים})} = \frac{\frac{1}{2} * \frac{1}{300}}{\frac{1}{2} * \frac{1}{300} + \frac{1}{4} * \frac{1}{125}}$$

$$\frac{\frac{1}{600}}{\frac{1}{600} + \frac{1}{500}} = \frac{\frac{1}{600}}{\frac{11}{3000}} = \frac{5}{11}$$

כלומר ההסתברות שלאליס היה תאום זהה היא $\frac{5}{11}$

ב1:

א. הסבירות לבחירת קערה באקראי היא $\frac{1}{2}$

ב. ההסתברות לבחור עוגיית שוקולד בקערה 1 היא $\frac{3}{4}$, ולכן $\frac{1}{2} * \frac{3}{4} = \frac{3}{8}$

ג. ההסתברות לבחור עוגיית שוקולד בקערה 2 היא $\frac{1}{2}$, ולכן $\frac{1}{2} * \frac{1}{2} = \frac{1}{4}$

לפי בייס:

$$P\left(\frac{\text{שוקולד 1 קערה}}{\text{שוקולד}}\right) = \frac{\frac{1}{2} * \frac{3}{4}}{\frac{1}{2} * \frac{3}{4} + \frac{1}{2} * \frac{1}{2}} = \frac{\frac{3}{8}}{\frac{3}{8} + \frac{1}{4}} = \frac{3}{5}$$

2. להלן:

א. הסבירות לבחירת קערה באקראי היא $\frac{1}{2}$

ב. ההסתברות לצהובה מ-1994 היא 0.2

ג. ההסתברות לצהובה מ-1996 היא 0.14

$$P\left(\frac{1994 \text{ צהובה}}{\text{ירוקה וצהובה נבחרו}}\right) = \frac{\frac{1}{2} * \frac{1}{5} + \frac{1}{2} * \frac{1}{5}}{\left(\frac{1}{2} * 0.14 + \frac{1}{2} * 0.1\right) + \left(\frac{1}{2} * 0.2 + \frac{1}{2} * 0.2\right)} = \frac{\frac{1}{5}}{\frac{1}{5} + \frac{3}{25}} = \frac{5}{8}$$

3. להלן:

א. הסיכוי שהשפעת תפגע בפועל $\frac{1}{10,000}$

ב. הסיכוי לחיובי למרות שהוא לא חולה בפועל הוא $\frac{1}{100}$

ג. הסיכוי לחולה בפועל שיצא שלילי הוא 0.

א

א. נגדיר: האיש חולה – H, הבדיקה חיובית – E

$$i. P(E/H) = \text{הסיכוי שהוא חולה לפי הבדיקה} = 0.99$$

$$ii. P(H) = \text{הסיכוי שהוא חולה לפי שיעור המחלה באוכלוסייה} = 10^{-4}$$

$$iii. P(E) = P\left(\frac{\text{חיובי}}{\text{חולה}}\right) * P(\text{חולה}) + P\left(\frac{\text{חיובי}}{\text{בריא}}\right) * P(\text{בריא}) = 0.99 * 10^{-4} + 0.01 * (1 - 10^{-4})$$

$$iv. P(H/E) = \sim \frac{1}{102}$$

ב

ב. נשנה את נתון 2:

$$i. P(E/H) = \text{הסיכוי שהוא חולה לפי הבדיקה} = 0.99$$

$$ii. P(H) = \text{הסיכוי שהוא חולה לפי שיעור המחלה באוכלוסייה} = \frac{1}{200}$$

$$iii. P(E) = P\left(\frac{\text{חיובי}}{\text{חולה}}\right) * P(\text{חולה}) + P\left(\frac{\text{חיובי}}{\text{בריא}}\right) * P(\text{בריא}) = 0.99 * \frac{1}{200} + 0.01 * \left(1 - \frac{1}{200}\right)$$

$$iv. P(H/E) = 0.0149$$

4. השאלה זהה ל-1 א', לכן הנתונים והתשובה הסופית זהים.

שאלות תחת Random Variables:

1. נראה:

א. יש $6 * 6 = 36$ אופציות שונות בהטלת שתי קוביות.

ב. מתוכם, 12 מתחלקות ב-3:

(1,2), (2,1), (5,1), (1,5), (2,4), (4,2), (3,3), (3,6), (6,3), (5,4), (4,5), (6,6)

ג. הסיכוי לקבל אופציה המתחלקת ב-3 היא $\frac{12}{36} = \frac{1}{3}$ ד. הסיכוי לקבל אופציה שאינה מתחלקת ב-3 היא $1 - \frac{1}{3} = \frac{2}{3}$ ה. השווי המוערך הוא: $0 = 2 - 2 = \frac{1}{3} * 6 + \frac{2}{3} * (-3)$

2. נראה:

א. יש $5 * 5 = 25$ צירופים אפשרייםב. יש 4 צירופים השווים ל-12: (8,4), (7,5), (10,2), (9,3), כלומר $\frac{4}{25}$ ג. יש 6 צירופים הגדולים מ-12: (10,5), (10,4), (10,3), (9,5), (9,4), (8,5), כלומר $\frac{6}{25}$ ד. ההסתברות ליתר הצירופים, הקטנים מ-12, היא $\frac{15}{25} = 1 - \frac{4}{25} - \frac{6}{25}$ ה. השווי המוערך הוא: $-2.8 = 0.8 + 0 - 3.6 = \frac{4}{25} * 5 + \frac{6}{25} * 0 + \frac{15}{25} * (-6)$

3. נראה:

א. ההסתברות לבחור גבר $\frac{x}{200} = 0.4$, כלומר יש 80 גברים.

ב. בהתאם ההסתברות לבחור אישה היא 0.6 ויש 120 נשים.

ג. ממוצע הגברים הוא $0.4 * 8 = 3.2$ ד. לפי הנלמד בהרצאה, סטיית התקן היא $std(x) = \sqrt{var(x)} = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N}}$ i. כאשר $\bar{X} = 3.2$ ובנוסף $x_i = i$ עבור $0 \leq i \leq 8$ ii. $std(x) = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N}} = \sqrt{\frac{(-3.2)^2 + (-2.2)^2 + \dots + (5.2)^2}{8}} = \sqrt{\frac{77.76}{8}} = 9.72 \approx 3.11$ 5. נחשב את השטח כאשר $P(x > 3)$ בזכות העובדה שמתקבל משולש ישר זווית:

$$P(x > 3) = \frac{0.4 * (5 - 3)}{2} = \frac{0.4 * 2}{2} = 0.4$$

6. נראה כי:

א. הסבירות שלעובד בחברה יש ילד היא 0.6, ושעובד אין ילד היא 0.4.

ב. נשתמש בטענת עזר: כאשר נרצה לבדוק סבירות של אחד מ-X, אזי יש X אופציות

לכך (כלומר x, (x-1), ..., 1) במקרה זה יש 4 משתתפים אזי יש 4 אופציות.

ג. $P(3/4 \text{ with Child}) = 4 * (0.4 * 0.6^3) = 4 * 0.0864 = 0.3456$

7. נחשב את הערך של X לפי סכמת הערכים:

א. $X = -10 * 0.1 - 5 * 0.35 + 0 + 5 * 0.35 + 10 * 35 = 0$