Statistics for Data Science-1

Week 8 Graded Assignment

1. Zaheer Khan has taken m five-wicket hauls in his last n matches. His match records are selected at random, one by one, and analyzed. If none of the match records is analyzed more than once, then what is the probability that the k^{th} one analyzed is his last five-wicket haul match?

a. $\dfrac{{}^{m}C_{m-1} \times {}^{n-m}C_{k-m}}{{}^{n}C_{k-1}} \times \dfrac{1}{n-k+1}$

b. $\dfrac{{}^{m}C_{m-1} \times {}^{n-m}C_{k-m}}{{}^{n}C_{k-1}}$

c. $\dfrac{1}{n-k+1}$

d. $\dfrac{{}^{n-m}C_{k-m}}{{}^{n}C_{k-1}}$

Answer: a

Solution:

Let A be the event of getting exactly $(m-1)$ five-wicket haul matches when the first $(k-1)$ matches are analyzed.

Let B be the event that k^{th} match analyzed is five-wicket haul match.

Now, P(A)= $\dfrac{{}^{m}C_{m-1} \times {}^{n-m}C_{k-m}}{{}^{n}C_{k-1}}$

$P(B|A) =$ Probability that the k^{th} match analyzed is five-wicket haul match given that the analysis of the first $k-1$ matches shows $m-1$ five-wicket haul matches.

$= \dfrac{1}{n-k+1}$

Required Probability $= P(A \cap B) = P(A) \times P(B|A) = \dfrac{{}^{m}C_{m-1} \times {}^{n-m}C_{k-m}}{{}^{n}C_{k-1}} \times \dfrac{1}{n-k+1}$

Hence, option (a) is correct.

Example: n=15,k=9, m=4

Let A be the event of getting exactly 3 five-wicket haul matches when the first 8 matches are analyzed.

Let B be the the event that 9^{th} match analyzed is five-wicket haul match.

Now, P(A)= $\dfrac{{}^{4}C_{3} \times {}^{11}C_{5}}{{}^{15}C_{8}}$

$P(B|A) =$ Probability that the 9^{th} match analyzed is a five-wicket haul match given that the analysis of the first 8 matches shows 3 five-wicket haul matches $= \dfrac{4-3}{15-8} = \dfrac{1}{7}$

Required Probability $= P(A \cap B) = P(A) \times P(B|A) = \dfrac{{}^4C_3 \times^{11} C_5}{{}^{15}C_8} \times \dfrac{1}{7} = \dfrac{8}{195}$

2. A and B predicts the outcomes of a cricket match and their chances of predicting the runs scored by a specific batsman correctly are $\frac{a}{b}$ and $\frac{c}{d}$ respectively independent of each other. If the probability of them predicting the same wrong score is $\frac{p}{q}$. Given that they predicted the same score, find the probability that their answer is correct.

   a. $\dfrac{q-p}{q}$

   b. $\dfrac{p(b-a)(d-c)}{qac+p(b-a)(d-c)}$

   c. $\dfrac{qac}{qac+p(b-a)(d-c)}$

   d. $\dfrac{p(b-a)(d-c)}{q}$

   Answer: c

   Solution:
   Let us define the following events:
   E_1: Both A and B predicted the score correctly.
   E_2: Exactly one of them predicted the score correctly.
   E_3: Neither of them predicted the score correctly.
   E: They predicted the same score.
   Now, $P(E_1) = \dfrac{a}{b} \times \dfrac{c}{d} = \dfrac{ac}{bd}$; $P(E|E_1) = 1$

   $$P(E_2) = \frac{a}{b} \times \frac{d-c}{d} + \frac{b-a}{b} \times \frac{c}{d} = \frac{a(d-c)+c(b-a)}{bd} \quad ; P(E|E_2) = 0$$

   $$P(E_3) = \frac{b-a}{b} \times \frac{d-c}{d} = \frac{(b-a)(d-c)}{bd} \quad ; P(E|E_3) = \frac{p}{q}$$

   Hence, By Bayes' Rule:

   $$P(E_1|E) = \frac{P(E_1) \times P(E|E_1)}{P(E_1) \times P(E|E_1) + P(E_2) \times P(E|E_2) + P(E_3) \times P(E|E_3)}$$

   $$= \frac{\frac{ac}{bd} \times 1}{\frac{ac}{bd} \times 1 + \frac{a(d-c)+c(b-a)}{bd} \times 0 + \frac{(b-a)(d-c)}{bd} \times \frac{p}{q}}$$

   $$= \frac{qac}{qac+p(b-a)(d-c)}$$

   Hence, option (c) is correct.

For example: a=1, b=3, c=1, d=4, p=1 and q=228

Let us define the following events:

E_1: Both A and B predicted the score correctly.

E_2: Exactly one of them predicted the score correctly.

E_3: Neither of them predicted the score correctly.

E: They predicted the same score.

Now, $P(E_1) = \frac{1}{3} \times \frac{1}{4} = \frac{1}{12}$; $P(E|E_1) = 1$

$P(E_2) = \frac{1}{3} \times \frac{3}{4} + \frac{2}{3} \times \frac{1}{4} = \frac{5}{12}$; $P(E|E_2) = 0$

$P(E_3) = \frac{2}{3} \times \frac{3}{4} = \frac{6}{12}$; $P(E|E_3) = \frac{1}{228}$

Hence, By Bayes' Rule:

$$P(E_1|E) = \frac{P(E_1) \times P(E|E_1)}{P(E_1) \times P(E|E_1) + P(E_2) \times P(E|E_2) + P(E_3) \times P(E|E_3)}$$

$$= \frac{\frac{1}{12} \times 1}{\frac{1}{12} \times 1 + \frac{5}{12} \times 0 + \frac{6}{12} \times \frac{1}{228}} = \frac{\frac{1}{12} \times \frac{38}{38}}{\frac{1}{12} \times \frac{38}{38} + 0 + \frac{1}{12} \times \frac{1}{38}} = \frac{38}{39}$$

An item is produced in three factories A, B and C. Factory A produces x times the number of items produced by factory B, and the factories B and C produces the same number of items. It is known that $p\%$, $q\%$, $r\%$ of the items produced by factories A, B and C respectively are defective. All items produced in the three factories are stocked, and an item is selected at random. On the basis of given information, answer the questions (3) and (4).

3. What is the probability that the selected item is defective?(Enter the answer correct to two decimal places)

Solution:

Let the number of items produced by each of the factories B and C be n. Then the number of items produced by factory A is xn.

Let us define the events:

A= Item produced is defective.

E_1= Item is produced by factory A.

E_2= Item is produced by factory B.

E_3= Item is produced by factory C.

Now,

$P(E_1) = \frac{xn}{xn + n + n} = \frac{x}{x+2}$ and, $P(E_2) = \frac{n}{xn + n + n} = \frac{1}{x+2} = P(E_3)$

Also it is given in the question that,

$P(A|E_1) = \frac{p}{100}$, $P(A|E_2) = \frac{q}{100}$ and, $P(A|E_3) = \frac{r}{100}$

Hence,
P(Selected item at random is defective)$= P(A) = \sum_{i=1}^{3} P(A \cap E_i)$
$= \sum_{i=1}^{3} P(A|E_i) \times P(E_i)$
$= P(E_1)P(A|E_1) + P(E_2)P(A|E_2) + P(E_3)P(A|E_3)$

$$= \left(\frac{x}{x+2} \times \frac{p}{100}\right) + \left(\frac{1}{x+2} \times \frac{q}{100}\right) + \left(\frac{1}{x+2} \times \frac{r}{100}\right)$$

For Example: $x = 5$**,** $p = 8$**,** $q = 7$ **and** $r = 4$
Let the number of items produced by each of the factories B and C be n. Then the number of items produced by factory A is $5n$.
Let us define the events:
$A=$ Item produced is defective.
$E_1=$ Item is produced by factory A.
$E_2=$ Item is produced by factory B.
$E_3=$ Item is produced by factory C.
Now,
$P(E_1) = \dfrac{5n}{5n+n+n} = \dfrac{5}{7}$ and, $P(E_2) = \dfrac{n}{5n+n+n} = \dfrac{1}{7} = P(E_3)$
Also it is given in the question that,
$P(A|E_1) = \dfrac{8}{100}$, $P(A|E_2) = \dfrac{7}{100}$ and, $P(A|E_3) = \dfrac{4}{100}$

Hence,
P(Selected item at random is defective)$= P(A) = \sum_{i=1}^{3} P(A \cap E_i)$
$= \sum_{i=1}^{3} P(A|E_i) \times P(E_i)$
$= P(E_1)P(A|E_1) + P(E_2)P(A|E_2) + P(E_3)P(A|E_3)$

$$= \left(\frac{5}{7} \times \frac{8}{100}\right) + \left(\frac{1}{7} \times \frac{7}{100}\right) + \left(\frac{1}{7} \times \frac{4}{100}\right) = \left(\frac{51}{700}\right) = 0.07$$

4. If an item selected at random is found to be defective, what is the probability that it was produced by factory B?(Enter the answer correct to two decimal places)
Solution:
P(Item is produced by factory B | Item is defective)$= P(E_2|A)$
By using the Bayes' Rule, we get:

$$P(E_2|A) = \frac{P(E_2)P(A|E_2)}{P(A)} = \frac{\frac{1}{x+2} \times \frac{q}{100}}{\left(\frac{x}{x+2} \times \frac{p}{100}\right) + \left(\frac{1}{x+2} \times \frac{q}{100}\right) + \left(\frac{1}{x+2} \times \frac{r}{100}\right)}$$

For Example: $x = 5$**,** $p = 8$**,** $q = 7$ **and** $r = 4$
P(Item is produced by factory B | Item is defective)$= P(E_2|A)$
By using the Bayes' Rule, we get:

$$P(E_2|A) = \frac{P(E_2)P(A|E_2)}{P(A)} = \frac{\frac{1}{7} \times \frac{7}{100}}{\left(\frac{5}{7} \times \frac{8}{100}\right) + \left(\frac{1}{7} \times \frac{7}{100}\right) + \left(\frac{1}{7} \times \frac{4}{100}\right)} = \frac{7}{51} = 0.14$$

5. A particular task is given to three persons, Manoj, Kalpana and Ananya whose probabilities of completing it are $\frac{a}{b}, \frac{c}{d}$ and $\frac{e}{f}$ respectively, independent of each other. What is the probability that the task will be completed? (Enter the answer correct to two decimal places)

   Solution:

   Let A,B,C denote the events that the task is completed by Manoj, Suresh and Kapil respectively. Then
   $P(A) = \frac{a}{b}$, $P(B) = \frac{c}{d}$, and $P(C) = \frac{e}{f}$
   The task will be completed if at least one of them completes the task.Thus, we have to calculate the probability of occurrence of at least one of the three events A,B,C, i.e, $P(A \cup B \cup C)$.
   Now, $P(A \cup B \cup C) = 1 - P(A \cup B \cup C)^c = 1 - P(A^c \cap B^c \cap C^c)$
   Since A,B,C are mutually independent $\implies$ A^c,B^c and C^c are mutually independent.
   Therefore, $P(A \cup B \cup C) = 1 - P(A^c)P(B^c)P(C^c)$
   $P(A \cup B \cup C) = 1 - (1 - \frac{a}{b})(1 - \frac{c}{d})(1 - \frac{e}{f})$

   For example:

   Let A,B,C denote the events that the task is completed by Manoj, Suresh and Kapil respectively. Then
   $P(A) = \frac{1}{4}$, $P(B) = \frac{2}{6}$, and $P(C) = \frac{1}{5}$
   The task will be completed if at least one of them completes the task. Thus, we have to calculate the probability of occurrence of at least one of the three events A,B,C, i.e, $P(A \cup B \cup C)$.
   Now, $P(A \cup B \cup C) = 1 - P(A \cup B \cup C)^c = 1 - P(A^c \cap B^c \cap C^c)$
   Since A,B,C are mutually independent $\implies$ A^c,B^c and C^c are mutually independent.
   Therefore, $P(A \cup B \cup C) = 1 - P(A^c)P(B^c)P(C^c)$
   $P(A \cup B \cup C) = 1 - (1 - \frac{1}{4})(1 - \frac{1}{3})(1 - \frac{1}{5}) = 1 - \frac{2}{5} = \frac{3}{5} = 0.6$

6. If A and B are two independent events such that $P(A^c) = m$ and $P(B^c) = x$ and $P(A \cup B) = n$, then calculate the value of x. (Enter the answer correct to two decimal places)

   Solution:

   We are given that $P(A^c) = m$ and $P(B^c) = x$ and $P(A \cup B) = n$.
   Now, $P(A \cup B) = 1 - P(A \cup B)^c = 1 - P(A^c \cap B^c)$
   $P(A \cup B) = 1 - P(A^c).P(B^c)$ [As Events are independent]

$n = 1 - mx$
$mx = 1 - n$
Hence, $x = \dfrac{1-n}{m}$

For example:

We are given that $P(A^c) = 0.6$ and $P(B^c) = x$ and $P(A \cup B) = 0.7$.
Now, $P(A \cup B) = 1 - P(A \cup B)^c = 1 - P(A^c \cap B^c)$
$P(A \cup B) = 1 - P(A^c).P(B^c)$ [As Events are independent]
$0.7 = 1 - 0.6x$
$0.6x = 1 - 0.7 = 0.3$
Hence, $x = \dfrac{0.3}{0.6} = 0.5$

Two researchers adopted different sampling techniques while investigating the same group of students to find the number of students falling in different intelligence levels.The results are given Table Q8.1.G: Answer the questions 7, 8 and 9.

Researcher	No. of students in each level		
	Below Avg.	Avg.	Above Avg
X	a	b	c
Y	d	e	f

Table 1: *
Table Q8.1.G: Intelligence Level

7. What is the probability that a student falls in below average level? (Enter the answer correct to 2 decimal accuracy).

   Solution:

   From the table we have
   Number of students investigated by researcher X =$a + b + c$
   Number of students investigated by researcher Y=$d + e + f$
   Total number of students = a+b+c+d+e+f
   Now, P(Student falls in below average level)=$\dfrac{(a + d)}{(a + b + c + d + e + f)}$

   For example: a = 76, b=70, c=54, d=50, e=23 and f=27

   From the table we have
   Number of students investigated by researcher X =$76 + 70 + 54 = 200$
   Number of students investigated by researcher $Y = 50 + 23 + 27 = 100$
   Total number of students = 300
   Now, P(Student falls in below average level)=$\dfrac{76 + 50}{200 + 100} = \dfrac{126}{300} = 0.42$

8. What is the probability that a student is of average level given that the investigation

is done by researcher Y?(Enter the answer correct to 2 decimal accuracy).

Solution:

Let us define events;

A: Student is of average level

B: Investigation is done by researcher Y.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{e}{a+b+c+d+e+f}}{\frac{d+e+f}{a+b+c+d+e+f}} = \frac{e}{d+e+f}$$

For example:

Let us define events;

A: Student is of average level

B: Investigation is done by researcher Y.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{23}{300}}{\frac{100}{300}} = \frac{23}{100} = 0.23$$

9. What is the probability that investigation is done by researcher X given that the student is of below average level?(Enter the answer correct to 2 decimal accuracy).

   Solution:

   Let us define events;

   C: Investigation is done by researcher X

   D: Student is of below average level

$$P(C|D) = \frac{P(C \cap D)}{P(D)} = \frac{\frac{a}{a+b+c+d+e+f}}{\frac{a+d}{a+b+c+d+e+f}} = \frac{a}{a+d}$$

   For example:

   Let us define events;

   C: Investigation is done by researcher X

   D: Student is of below average level

$$P(C|D) = \frac{P(C \cap D)}{P(D)} = \frac{\frac{76}{300}}{\frac{126}{300}} = \frac{76}{126} = 0.6032$$

10. During the monsoon, it rains one-third of the days and affects students travel to school. The probability that there will be heavy traffic on a rainy day is 0.5 and on a non-rainy day is 0.25. If it rains and there is heavy traffic, the probability of a student arriving late to school is 0.5. If it is a clear day and there is no traffic, this probability is reduced by $\frac{3}{8}$. In other possible situations, the probability of a student reaching school late 0.25. If on a randomly selected day, a student arrives late to school, then what is the

probability that it rained that day?
Hint: Consider the event as,
H= There is heavy traffic
H^c= There is no traffic.

a. $\frac{1}{8}$
b. $\frac{6}{11}$
c. $\frac{11}{48}$
d. None of the above

Answer: b
Solution:
Let us define the following events:
H: There is heavy traffic.
R: Rainy-day.
E: Student is late for school.
Now, we are given that, $P(R) = \frac{1}{3}$

$P(H|R) = 0.5$ and, $P(H|R^c) = 0.25$
$P(E|R \cap H) = 0.5$, $P(E|R^c \cap H^c) = 0.5 - \frac{3}{8} = \frac{1}{8}$ and $P(E|R^c \cap H) = P(E|R \cap H^c) = 0.25$
$P(R \cap H \cap E) = P(R)P(H|R)P(E|R \cap H) = \frac{1}{3} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{12}$

Similarly,
$P(R^c \cap H \cap E) = P(R^c)P(H|R^c)P(E|R^c \cap H) = \frac{2}{3} \times \frac{1}{4} \times \frac{1}{4} = \frac{2}{48}$

$P(R \cap H^c \cap E) = P(R)P(H^c|R)P(E|R \cap H^c) = \frac{1}{3} \times \left(1 - \frac{1}{2}\right) \times \frac{1}{4} = \frac{1}{24}$

$P(R^c \cap H^c \cap E) = P(R^c)P(H^c|R^c)P(E|R^c \cap H^c) = \frac{2}{3} \times \frac{3}{4} \times \frac{1}{8} = \frac{3}{48}$

Therefore, $P(E) = P(R \cap H \cap E) + P(R^c \cap H \cap E) + P(R \cap H^c \cap E) + P(R^c \cap H^c \cap E)$
$= \frac{1}{12} + \frac{2}{48} + \frac{1}{24} + \frac{3}{48} = \frac{11}{48}$
Also, $P(R \cap E) = P(R \cap H \cap E) + P(R \cap H^c \cap E) = \frac{1}{12} + \frac{1}{24} = \frac{3}{24}$

Now, Required Probability; $P(R|E) = \frac{P(R \cap E)}{P(E)} = \frac{\frac{3}{24}}{\frac{11}{48}} = \frac{6}{11}$

Hence, option (b) is correct.

11. There are two shops, A and B, selling t-shirts in the market. Shop A has stock of n red and 2 black t-shirts and Shop B has a stock of 2 red and n black t-shirts. One of the shops is selected at random and two t-shirts are purchased from it. If both the t-shirts purchased are red and the probability that it was purchased from a shop A is $\frac{6}{7}$, find the value of n.

   a. 3

   b. 4

   c. 5

   d. 6

   Answer:b

   Solution:

   Let us define the following events:
   $E1$: Shop A is selected.
   $E2$: Shop B is selected.
   E: Two t-shirts purchased are of red color.
   Now, $P(E1) = P(E2) = \frac{1}{2}$
   $P(E|E1) = \frac{{}^nC_2}{{}^{n+2}C_2} = \frac{n(n-1)}{(n+2)(n+1)}$
   and, $P(E|E2) = \frac{{}^2C_2}{{}^{n+2}C_2} = \frac{2}{(n+2)(n+1)}$

   Using Bayes' Theorem, we get:

$$P(E1|E) = \frac{P(E1)P(E|E1)}{P(E1)P(E|E1) + P(E2)P(E|E2)} = \frac{6}{7} \qquad \text{(Given)}$$

$$\implies \frac{\frac{1}{2} \times \frac{n(n-1)}{(n+2)(n+1)}}{\frac{1}{2} \times \frac{n(n-1)}{(n+2)(n+1)} + \frac{1}{2} \times \frac{2}{(n+2)(n+1)}} = \frac{6}{7}$$

$$\implies \frac{n(n-1)}{n(n-1)+2} = \frac{6}{7}$$

   $7n(n-1) = 6n(n-1) + 12 \implies n^2 - n - 12 = 0$
   Therefore, $n = 4, -3$. Since, n cannot be negative. So, $n = 4$
   Hence, option (b) is correct.