HW3 | Chandan Suri | CS4090

1. When an individual is randomly selected from the population, the events can be defined as:-

$A = \{\text{type A selected}\}$

$B = \{\text{type B selected}\}$

$C = \{\text{ethnic group 3 selected}\}$

(a)

P(A) = Probability that an individual of type A is selected (irrespective of the ethnic group)

$= 0.106 + 0.141 + 0.2$

$= 0.447 \approx \underline{0.45}$ Ans.

P(C) = Probability that an individual belonging to ethnic group 3 is selected

Now, we will add the probabilities of an individual being in ethnic group that is selected for all the blood groups.

$\Rightarrow 0.215 + 0.2 + 0.065 + 0.02$

$= \underline{0.5}$ Ans.

$P(A \cap C)$ = Probability that an individual selected is of blood group A that belongs to the ethnic group 3.

Thus, this should be the entry in the table provided where it's both blood group A and ethnic group 3 probability,

which is $\Rightarrow \underline{0.2}$. Ans

(b) $P(A|C)$ = Probability of an individual being selected is from type A blood group given that the selected individual belongs to the ethnic group 3.

$$= \frac{P(A \cap C)}{P(C)} = \frac{0.2}{0.5} = \underline{0.4} \quad \underline{\underline{\text{Ans}}}.$$

$P(C|A)$ = Probability of an individual being selected is from the ethnic group 3 given that the selected individual is from the type A blood group.

$$= \frac{P(A \cap C)}{P(A)} = \frac{0.2}{0.447} = 0.447 \approx 0.45 \quad \underline{\underline{\text{Ans}}}.$$

(c) Let D be an event as $\Rightarrow$

$D = \{\text{ethnic group 1 selected}\}$

$\because B = \{\text{type B selected}\}$

$\therefore B^c = \{\text{type B is not selected}\}$

Probability that an individual being selected is from ethnic group 1 given that the selected individual does not have type B blood.

$$P(D|B^c) = \boxed{\frac{P(D \cap B^c)}{P(B^c)}} \text{ --- } (1)$$

$$\because P(B^c) = 1 - P(B) = 1 - [0.065 + 0.018 + 0.008] = 1 - 0.091 = \underline{0.909}$$

&.

$P(D \cap B^c)$ = Probability that the selected individual belongs to the ethnic group 1 and does not have type B blood group.

$= 0.082 + 0.106 + 0.004$

$= \underline{0.192}$

Using the values computed above and putting it in the equation (1) above,

$$P(D \mid B^c) = \frac{P(D \cap B^c)}{P(B^c)} = \frac{0.192}{0.909} = \frac{192}{909}$$

$= \underline{0.21}$

<u>Ans.</u>

2. (a) $P(M \cap LS \cap Pr)$ = Probability that the next shirt sold is a medium (M), long-sleeved (LS) and print (Pr) shirt ⇒

This, we can directly get from the table for LS (long-sleeved) shirt at the entry for (M) medium and (Pr) print shirt ⇒ 0.05 Ans (because all of these should be applicable at the same time).

(b) $P(M \cap Pr)$ ⇒ Probability that the next shirt sold is a medium (M) and print (Pr) shirt ⇒

for this, the shirt can either be short-sleeved (SS) or long-sleeved (LS) ⇒

(*) The short forms in the parenthesis belong to the corresponding events.

So,

$$P(M \cap Pr) = P(M \cap LS \cap Pr) + P(M \cap SS \cap Pr)$$

$$= 0.05 + 0.07$$

$$= 0.12$$

Ans.

(c) $P(SS)$ ⇒ Probability that the next shirt sold is a short-sleeved shirt.

for this we need to add all the probabilities for different combinations of sizes and patterns for the shirts.

This would correspond to adding all the elements in the table provided for short-sleeved shirt

⇒

$\therefore P(SS) \Rightarrow P(SS \cap S \cap Pl) + P(SS \cap S \cap Pr) + P(SS \cap S \cap St)$
$+ P(SS \cap M \cap Pl) + P(SS \cap M \cap Pr) + P(SS \cap M \cap St)$
$+ P(SS \cap L \cap Pl) + P(SS \cap L \cap Pr) + P(SS \cap L \cap St)$

$\Rightarrow 0.04 + 0.02 + 0.05 + 0.08 + 0.07 + 0.12 + 0.03$
$+ 0.07 + 0.08$

$\Rightarrow \underline{\underline{0.56}}$ Ans.

$P(LS) \Rightarrow$ Probability that the next shirt sold is a long-sleeved shirt.

for this, we need to add all the probabilities for different combinations of sizes and patterns for the long-sleeved shirts.

This would correspond to adding all the elements in the table provided for ~~short~~ long-sleeved shirt $\Rightarrow$

$\therefore P(LS) \Rightarrow P(LS \cap S \cap Pl) + P(LS \cap M \cap Pl) + P(LS \cap L \cap Pl)$
$+ P(LS \cap S \cap Pr) + P(LS \cap M \cap Pr) + P(LS \cap L \cap Pr)$
$+ P(LS \cap S \cap St) + P(LS \cap M \cap St) + P(LS \cap L \cap St)$

$\Rightarrow 0.03 + 0.10 + 0.04 + 0.02 + 0.05 + 0.02$
$+ 0.03 + 0.07 + 0.08$

$= \underline{\underline{0.44}}$ Ans

(d) $P(M) \Rightarrow$ Probability that the next shirt sold is of medium size.

This would encompass the case that the shirt is short sleeved or long- sleeved.
(SS) (LS)

$$\Rightarrow P(M) = \underbrace{P(M \cap SS)}_{\substack{\text{short-sleeved} \\ \text{and} \\ \text{medium}}} + \underbrace{P(M \cap LS)}_{\substack{\text{long-sleeved} \\ \text{and} \\ \text{medium.}}}$$

Now,

$$P(M \cap SS) = P(M \cap SS \cap Pl) + P(M \cap SS \cap Pr) + P(M \cap SS \cap St)$$

from the table, we would have to take the entire row of medium and short-sleeved.

$$P(M \cap SS) = 0.08 + 0.07 + 0.12 = \underline{\underline{0.27}}$$ Ans

And,

$$P(M \cap LS) = P(M \cap LS \cap Pl) + P(M \cap LS \cap Pr) + P(M \cap LS \cap St)$$

from the table, we would have to take the entire row of medium and long-sleeved.

$$P(M \cap LS) = 0.1 + 0.05 + 0.07 = \underline{\underline{0.22}}$$ Ans

$$\therefore P(M) = 0.27 + 0.22 = \underline{\underline{0.49}}$$ Ans.

$\Rightarrow$ $P(Pr)$ = Probability that the next shirt sold has a pattern of print type.

This would encompass the case that the shirt is short- sleeved or long- sleeved.
(SS) (LS)

$$\Rightarrow P(Pr) = \underbrace{P(Pr \cap SS)}_{\substack{\text{short-sleeved} \\ \text{and print}}} + \underbrace{P(Pr \cap LS)}_{\substack{\text{long-sleeved} \\ \text{and print pattern type.}}}$$

Now,

$P(Pr \cap SS) = P(Pr \cap SS \cap S) + P(Pr \cap SS \cap M) + P(Pr \cap SS \cap L)$

$\Rightarrow$ from the table, we would have to take the entire column of print pattern type and short-sleeved.

$P(Pr \cap SS) = 0.02 + 0.07 + 0.07 = \underline{\underline{0.16}}$ Ans

And,

$P(Pr \cap LS) = P(Pr \cap LS \cap S) + P(Pr \cap LS \cap M) + P(Pr \cap LS \cap L)$

from the table, we would have to take the entire column of print pattern type & long-sleeved.

$P(Pr \cap LS) = 0.02 + 0.05 + 0.02 = \underline{\underline{0.09}}$ Ans

$\therefore\ P(Pr) = 0.16 + 0.09 = \underline{\underline{0.25}}$ Ans.

(e) $P(M \mid SS \cap Pl)$ = Probability that the shirt just sold was of size medium given that the shirt sold was a short-sleeved plaid. (short-sleeved and plaid)

$$\Rightarrow P(M \mid SS \cap Pl) = \frac{\boxed{P(M \cap SS \cap Pl)}}{P(SS \cap Pl)}$$ this can be found in the table.

$$= \frac{0.08}{P(SS \cap Pl \cap S) + P(SS \cap Pl \cap M) + P(SS \cap Pl \cap L)}$$ of all the sizes

$= (0.08) / (0.04 + 0.08 + 0.03)$

$= 0.08 / 0.15 = \underline{0.53}$ Ans

(f) $P(SS \mid M \cap Pl) \Rightarrow$ Probability that the shirt sold was short-sleeved given that it was medium and plaid.

$$\Rightarrow \frac{\boxed{P(SS \cap M \cap Pl)}}{P(M \cap Pl)}$$ → this comes directly from the table

$$\Rightarrow (0.08) / \left(\underbrace{P(M \cap Pl \cap SS)}_{\text{medium plaid \& short-sleeved}} + \underbrace{P(M \cap Pl \cap LS)}_{\text{medium plaid \& long-sleeved}} \right)$$

$$\Rightarrow (0.08)/(0.08 + 0.1) = 0.08/0.18 = 8/18 = \underline{\underline{0.44}}$$ Ans.

$P(LS \mid M \cap Pl) \Rightarrow$ Probability that the shirt sold was ~~short~~ long-sleeved given that it was medium and plaid.

$$\Rightarrow \frac{\boxed{P(LS \cap M \cap Pl)}}{P(M \cap Pl)}$$ → this comes directly from the table

$$\Rightarrow 0.1 / \left(\underbrace{P(M \cap Pl \cap SS)}_{\text{medium plaid and short-sleeved}} + \underbrace{P(M \cap Pl \cap LS)}_{\text{medium plaid \& long-sleeved}} \right)$$

$$\Rightarrow 0.1/(0.08 + 0.1)$$

$$\Rightarrow \frac{0.1}{0.18} = \frac{10}{18} = \underline{\underline{0.56}}$$ Ans.

(3)

$P(D)$ = Probability of the selected individual having a certain disease

$P(D) = \underline{0.05}$

$\because P(D) + \underbrace{P(ND)}_{\downarrow} = 1$

probability of the selected individual having no disease.

$\Rightarrow 0.05 + P(ND) = 1$

$P(ND) = 1 - 0.05 = \underline{0.95}$

and,

$P(P|D) \Rightarrow$ Probability that the test detects the presence of the disease.

(test was positive given that individual had disease).

$P(P|D) = 98\% = \underline{0.98}$

$P(N|D) \Rightarrow$ Probability that test came out negative given that individual had disease

$P(N|D) = 1 - 0.98 = \underline{0.02}$.

and,

$P(N|ND) \Rightarrow$ Probability that the test detects the absence of the disease correctly.

(test was negative given that individual had no disease).

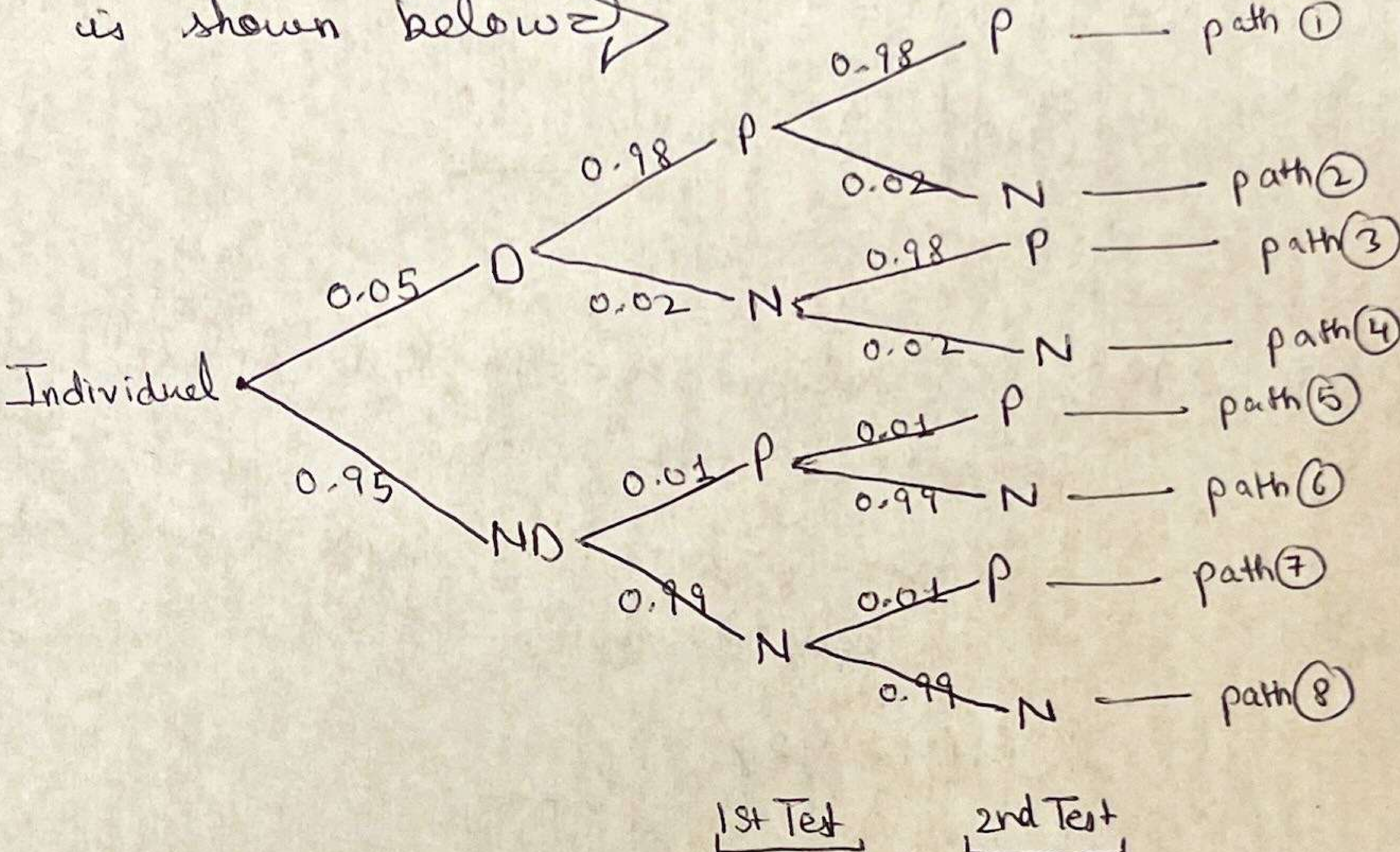
$P(N|ND) = 99\% = \underline{0.99}$

$P(P|ND) \Rightarrow$ Probability that test come out positive given that the individual had no disease

$$P(P|ND) = 1 - P(N|ND) = 1 - 0.99 = \underline{0.01}$$

These probabilities will be valid for 2nd level of tests as well.

So, the tree diagram for the full sequence of events is shown below $\Rightarrow$



$\Rightarrow$ $P(D | P \cap P) \Rightarrow$ Probability of the diseased individual given that both the tests came out positive.

$$\Rightarrow \frac{P(D \cap P \cap P)}{P(P \cap P)} = \frac{\overbrace{0.05 \times 0.98 \times 0.98}^{\text{path ①}}}{\text{path ①} + \text{path ⑤}}$$

$$= \frac{0.05 \times 0.98 \times 0.98}{0.05 \times 0.98 \times 0.98 + 0.95 \times 0.01 \times 0.01}$$

$$= (0.05 \times 0.98^2) / (0.05 \times 0.98^2 + 0.95 \times 0.01^2) = 0.998$$

$$= \underline{99.8\%} \text{ Ans}$$

4.

(a) Number of students = n

Each of those students' birthday can fall on any of the 365 days in a year.

Thus, number of ways of all the student's birthdays can be determined as =>

$$\underset{1}{\underline{365}} \quad \underset{2}{\underline{365}} \quad \underset{3}{\underline{365}} \quad \underset{4}{\underline{365}} \quad \cdots \quad \underset{n}{\underline{365}}$$

Students →

∴ Number of possible outcomes => $(365)^n$

Ans.

(b) Event A => Atleast 2 students have the same birthday.

∴ This event A would encompass all the cases that =>

{2 students have the same birthday} ∪

{3 students have the same birthday} ∪

⋮

{all (n) students have the same birthday}

Thus,

A' is the complement of A which means everything in the sample space but the outcomes possible as event A.

So,

A' = { None ~~two~~ of the students have birthday on the same date in the year }

Ans.

(c.) for choosing the ways where none of the student's birthdays fall on the same date,

~~it would mean to have/choose birthdays in an unordered fashion as the order doesn't matter~~

it would mean to have/choose birthdays in an ordered fashion as each of the students are distinct.

Thus, we would follow the rule for ordered sampling without replacement $\Rightarrow$

Number of ways the student's (n) birthdays can fall in a year when all the birthdates should be different $\Rightarrow$ $\boxed{\dfrac{365!}{(365-n)!}}$

Total number of outcomes $\Rightarrow$ $\underline{365^n}$

$\therefore$ $P(A') = \dfrac{\text{Number of ways of student's birthdays in a year when all birthdates should be different}}{\text{Total number of outcomes}}$

$$= \left(\frac{365!}{(365-n)!}\right) \Big/ 365^n$$

$$= \boxed{\frac{365!}{(365-n)!\,365^n}}$$

Ans.

(d.)

$P(A) \Rightarrow$ Probability that atleast two students have the same birthday.

$$\therefore P(A) = 1 - P(A')$$

from previous part (c),

$$P(A') = \frac{365!}{(365-n)!\,365^n}$$

$$\therefore \boxed{P(A) = 1 - \frac{365!}{(365-n)!\,365^n}}$$

Ans.

(e) Reasons for top 10:-

(1.) Considering different seasons across the United States, the month of September seems to be the most conducive one for a newborn as the weather in september is mostly pleasant and not very harsh. This kind of planning by the couples would be one of the reasons for most of the births happenning in the month of september.

(2.) According to some research on mammals, the winter season (december) has an effect on the human body and the bodies of most mammals such that more hormones are secreted. This secretion of more hormones leads to a higher chance of conception. following this fact, 9 months after december would align with top number of births happening in the month of september.

(3.) According to some research in the behavioral psychology, couples have an easier time conceiving around the Christmas time thus, making the month of september most probable for high number of births.

(4.) Lastly, couples probably plan the childbirth according to the holiday seasons around the year. Thus, the parents won't want their children to be born during the holiday breaks since this may lead to shortage of resources like staffing in the hospitals. Planning according to this makes the month of september to be the most apt one to have higher number of births.

The bottom 10 occurs in the months of January, July, November & december on some specific dates probably because of the following reasons:—

(1.) Jan (1, 2 and 3) — Happy New Year's.
July 4th → Independence Day in the USA.
November (26, 27 and 28)— Thanksgiving week.
December (24, 25 and 26)— Christmas

As we can see above, most of the big festivities in the USA happen on/or near the dates above and coincide very well with the lower number of births. This could be the reason for less reports of births in the U.S.A around those festivities.

(2.) Also, holidays such as these would means lack of staffing and other resources in the hospitals so, the couples would plan accordingly so, that the childbirth is less probable around these holidays which coincides well with the data here.

(3.) Lastly, as some of the reasons specified for the top 10 are quite significant thus, making other months /times around the year more conducive to child birth thus, making those dates less of an ideal for the child birth.

— Ans.

(5.) wine supply includes $\Rightarrow$

Zinfandel (Z) = 8
merlot (M) = 10
cabernet (C) = 12

All are from different wineries.

(a.) We want to choose 3 bottles of zinfandel out of 8 bottles in order to serve.

And, as the serving order is important.

Thus, we want to choose three as well as arrange them, (different permutations).

Thus, this case of the type "~~thus~~ Ordered without replacement".

Number of ways to do this $\Rightarrow$

$$= \binom{8}{3} \cdot 3!$$

$$= \frac{8!}{\cancel{3!}\,5!} \cdot \cancel{3!}$$

$$= \frac{8!}{5!} \qquad \left(\frac{n!}{(n-k)!} \text{ form}\right)$$

$\Rightarrow$ Ans.

(b.) Total number of wine bottles $\Rightarrow$ 30

If 6 bottles need to be randomly selected from those, so, we just need to choose 6 bottles.

Thus, this case would be unordered sampling without replacement $\Rightarrow$

$$\boxed{\binom{30}{6}} \Rightarrow \frac{30!}{6!\,(30-6)!} \Rightarrow \boxed{\frac{30!}{6!\,24!}}$$

Ans Ans.

(c) If we need to select 6 bottles such that 2 bottles should be of the same variety,

then, basically we need to select 2 bottles from each of three groups (three varieties of wine).

for choosing 2 bottles out of a variety, the order doesn't matter so, this case also of the form, "Unordered sampling without replacement".

For choosing 2 bottles of zinfandel $\Rightarrow \binom{8}{2}$

for choosing 2 bottles of merlot $\Rightarrow \binom{10}{2}$

for choosing 2 bottles of cabernet $\Rightarrow \binom{12}{2}$

$\therefore$ Total number of ways to select 2 bottles of each variety (simultaneously) $\Rightarrow$

$$\boxed{\binom{8}{2}\cdot\binom{10}{2}\cdot\binom{12}{2}} \Rightarrow \frac{\cancel{8!}}{2!\,6!}\cdot\frac{\cancel{10!}}{2!\,\cancel{8!}}\cdot\frac{12!}{2!\,\cancel{10!}}$$

Ans $\Rightarrow \boxed{(12!)/(2!\,6!\,2!\,2!)}$ Ans (d.) from part (b), the total number of ways 6 bottles can be randomly selected =>

$$\binom{30}{6} = \boxed{\frac{30!}{6!\,24!}}$$

from part (c), the total number of ways to obtain two bottles of each variety when 6 bottles here to be selected

$$\Rightarrow \binom{8}{2}\cdot\binom{10}{2}\cdot\binom{12}{2} \Rightarrow \boxed{\frac{12!}{2!\,2!\,2!\,6!}}$$

∴ Probability that randomly selecting 6 bottles of wine results in two bottles of each variety =>

$$\Rightarrow \frac{(30!)/(6!\,24!)}{(12!)/(2!\,2!\,2!\,6!)}$$

$$\Rightarrow \frac{30!}{\cancel{6!}\,24!}\cdot\frac{2!\,2!\,2!\,\cancel{6!}}{12!}$$

$$\Rightarrow \boxed{\frac{30!\,2!\,2!\,2!}{24!\,12!}}$$

Ans.

(e.) If we need to choose 6 bottles randomly, such that all of them are of the same variety, so, we can do this in 3 independent cases =>
(3 varieties)

If the 6 bottles chosen are of the variety of zinfandel, the number of ways 6 bottles of zinfandel can be chosen $\Rightarrow$

$$\binom{8}{6}$$

and, (similarly),

the number of ways 6 bottles of merlot can be chosen $\Rightarrow$

$$\binom{10}{6}$$

and,

the number of ways 6 bottles of cabernet can be chosen

$$\Rightarrow \binom{12}{6}$$

As all these cannot occur at the same time,

so,

total number of ways 6 bottles can be selected such that all of them are of the same variety

$$\Rightarrow \boxed{\binom{8}{6} + \binom{10}{6} + \binom{12}{6}} \longrightarrow 1$$

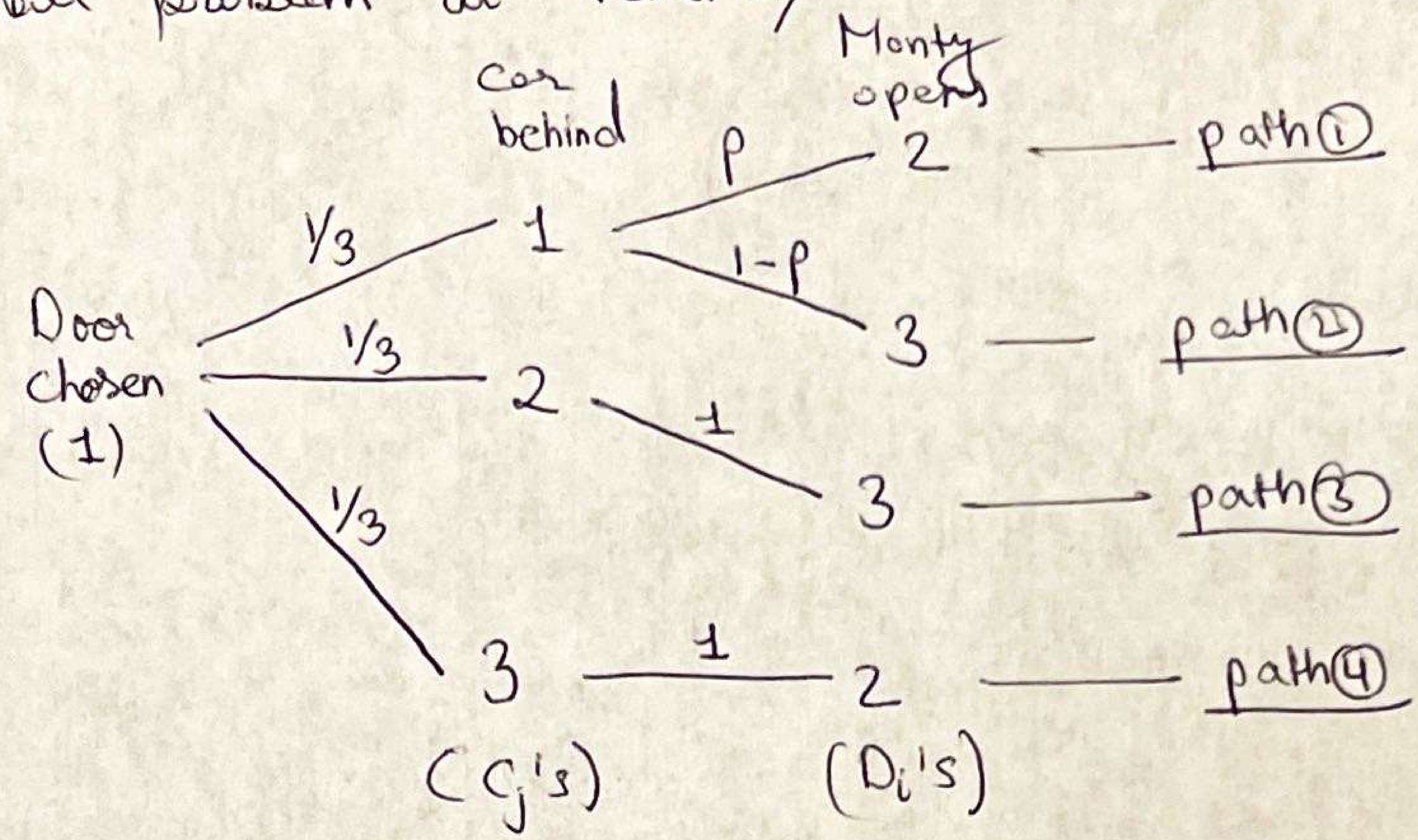
Total number of ways of randomly selected 6 bottles of wine out of all the bottles $\Rightarrow$

$$\boxed{\binom{30}{6}} \longrightarrow 2$$

$\therefore$ The probability that all the 6 bottles randomly selected are of the same variety $\Rightarrow$

$$1 / 2 \Rightarrow \left[\binom{8}{6} + \binom{10}{6} + \binom{12}{6}\right] / \binom{30}{6}$$ Ans.

6. (a) Let's first draw the tree diagram for the monty-hall problem at hand ⇒



Let C_j ⇒ event that the car is hidden behind door j.

W ⇒ event that win using the switching strategy.

So,

using the law of total probability and the tree diagram above ⇒

$$P(W) = P(W|C_1).P(C_1) + P(W|C_2).P(C_2) + P(W|C_3).P(C_3)$$

$$= 0.\tfrac{1}{3} + 1.\tfrac{1}{3} + 1.\tfrac{1}{3}$$

$$= \boxed{2/3}$$

Ans

∵ switching should always succeed, so, $P(W|C_1)=0$

(b) Using the tree diagram above,

P(switching succeeds | Door 2 opened by Monty)

$$= \frac{P(\text{switching succeeds} \cap \text{Door 2 opened by Monty})}{P(\text{Door 2 opened by Monty})}$$

The numerator in the formule before can be computed as from the tree diagram because switching succeeds at the time when monty opens door2 which leaves us with the bottom most path in the tree.
(path-4)

$$\Rightarrow \frac{1}{3} \cdot 1 = \frac{1}{3}$$

The denominator in the formule can be computed by taking the paths 1 and 4 into consideration as those are the paths when monty opens door 2.

So,

$$\Rightarrow \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot p = \left(\frac{1+p}{3}\right)$$

$\therefore$ P(switching succeeds | Door 2 opened by Monty)

$$\Rightarrow \frac{\cancel{1/3}}{(1+p)/\cancel{3}} \Rightarrow \boxed{\frac{1}{1+p}}$$ Ans.

c.

P(switching succeeds | Door 3 opened by Monty)

$$= \frac{P(\text{switching succeeds} \cap \text{Door 3 opened by Monty})}{P(\text{Door 3 opened by Monty})}$$

$\Rightarrow$ Similarly following the method as for part b, the numerator can be computed by just taking the path 3. and denominator can be computed by ~~just~~ taking the paths 2 and 3

$\Rightarrow$

$$\Rightarrow \frac{\frac{1}{3}\cdot 1}{\frac{1}{3}\cdot 1+\frac{1}{3}\cdot(1-p)} = \frac{\frac{1}{3}}{\frac{1}{3}+\frac{1-p}{3}} = \frac{1/\cancel{3}}{\frac{2-p}{\cancel{3}}}$$

$$= \boxed{\frac{1}{2-p}}$$ Ans

(d.) Given that Monty opens door 2,

probability that switching succeeds $= \frac{1}{1+p}$

probability that switching does not succeed = probability that staying with the initial choice succeeds

$$= 1-\frac{1}{1+p}$$

$$= \frac{1+p-1}{1+p}$$

$$= p/_{1+p}$$

To find a value of p such that,

$$\frac{1}{1+p} < \frac{p}{1+p}$$

$(\because p \neq -1) \rightarrow$ as it's a probability, $(0 \leq p \leq 1)$ we can multiply both sides of the inequality by $(1+p)$,

$$\frac{1}{\cancel{1+p}}\cdot\cancel{(1+p)} < \frac{p}{\cancel{1+p}}\cdot\cancel{(1+p)}$$ [no inequality sign change]

$\boxed{1<p} \rightarrow$ As this is not possible, so, no value of p exists for this case. Ans

(e) Given that Monty opens door 3,

probability that switching succeeds $= \frac{1}{2-p}$

probability that switching does not succeed = probability that staying with initial choice succeeds

$$= 1 - \frac{1}{2-p} = \frac{2-p-1}{2-p} = \frac{1-p}{2-p}$$

To find a value of p such that,

$$\frac{1}{2-p} < \frac{1-p}{2-p}$$

we can multiply both sides of the inequality above by $2-p$ ($\because 0 \le p \le 1$) without any sign changes & as $p \neq 2$,

$$\frac{1}{\cancel{2-p}} \cdot \cancel{(2-p)} < \frac{1-p}{\cancel{(2-p)}} \cdot \cancel{(2-p)}$$

$$1 < 1 - p$$ [subtract 1 from both sides]

$$1 - 1 < 1 - p - 1$$

$$0 < -p$$ [multiply both sides by -1 and change the sign]

$$(-1) \cdot 0 > -(-p)$$

$\boxed{p < 0}$ → As this is not possible, so, no value of p exists for this case too!

Ans

⑦ Preference is as:—

Car > Computer > Goat

ⓐ It's given that Monty will always reveal the less preferred choice.

That means that Monty will always reveal the goat unless you picked the goat.

Thus, all we learn is that the contestant did not pick the goat.

As both the doors left now will either have a car or a computer,

The chance that we win the car = $1/2$

(no matter if we switch or not).

Ans.

ⓑ Monty reveals your less preferred prize with probability $= p$

Monty reveals your more preferred prize with probability $= q = 1-p$.

We know that monty opened the door revealing the computer.

Thus, there are 2 cases now!

① If you picked the goat, Monty will reveal the computer with probability q. (as computer is more preferred than goat).

(2) If you picked the car, Monty will reveal the computer with probability p. (as computer is less preferred then car).

for the (2)nd case,

$\because$ $P(\text{car picked} \cap \text{computer revealed}) =$

$P(\text{car picked} \mid \text{computer revealed}) \cdot P(\text{computer revealed})$

$= P(\text{computer revealed} \mid \text{car picked}) \cdot P(\text{car picked})$

$= p \cdot \frac{1}{3}$ → $\because$ total outcomes $= 3$ and getting any choice is equally likely.

$= p/3$.

$\therefore$ $P(\text{car picked} \mid \text{computer revealed}) = \dfrac{p}{3\,P(\text{computer revealed})}$ — (1)

Similarly, for the (1)st case,

$P(\text{goat picked} \mid \text{computer revealed}) = \dfrac{q}{3\,P(\text{computer revealed})}$ — (2)

$\because$

$P(\text{car picked} \mid \text{computer revealed}) +$

$P(\text{goat picked} \mid \text{computer revealed}) = 1$

(because there are only 2 valid possibilities)

Solving,

$$\frac{p}{3\,P(\text{Computer revealed})} + \frac{q}{3\,P(\text{Computer revealed})} = 1$$

$\because\ q = 1 - p,$

$$\cancel{\frac{p}{3P(\text{computer revealed})}} + \frac{1}{3P(\text{computer revealed})} - \cancel{\frac{p}{3P(\text{computer revealed})}} = 1$$

$$\therefore\ \underline{P(\text{computer revealed}) = 1/3}$$

Putting this in equation ① above,

$$P(\text{car picked} \mid \text{computer revealed}) = \frac{p}{\cancel{3} \cdot \frac{1}{\cancel{3}}}$$

$$= \underline{p}$$

Ans.

Therefore,

you should switch if $\boxed{p < \frac{1}{2}}$

and stay with the initial choice if $\boxed{p > \frac{1}{2}}$.

Ans.

8. Data in the vaccine example is shown below:-

	Population		Severe cases/100k		Vaccine efficacy against severe cases
	Not Vaccinated	Vaccinated	Not Vaccinated	Vaccinated	
<50	1.12 M	3.5 M	3.8	0.3	$1-(0.3/3.8) = 92.1\%$
>50	0.18 M	2.13 M	95	13.6	$1-(13.6/95) = 85.7\%$
All ages	1.3 M	5.63 M	16.5	5.3	$1-(5.3/16.5) = 67.9\%$

Paradox here:-

$$\left.\begin{cases} 92.1\% > 67.9\% \\ \& \\ 85.7\% > 67.9\% \end{cases}\right\}$$ → the cause here seems to be the age stratification (Age groups)

The table for severe cases above includes the conditional ratios.

To get the actual numbers ⇒

{ NV → not vaccinated
{ V → vaccinated

	Age groups	Population		Severe cases/100K	
		NV	V	NV	V
1.12M	<50	~~$\frac{3.8\times1.12\times10^6}{100K} = 42.56 \approx 43$~~	3.5M	$\frac{3.8\times1.12\times10^6}{100K}$ $=42.56\simeq43$	$\frac{0.3\times3.5\times10^6}{100K}$ $=10.5\simeq11$
	>50	0.18 M	2.13M	$\frac{95\times0.18\times10^6}{100K}$ $=171$	$\frac{13.6\times2.13\times10^6}{100K}$ $=289.68$ ≈290
	All ages	1.3M	5.63 M	214	301

The paradox is related to the severe cases which in turn affects the vaccine efficacy.

Thus, event A is the "severe cases",
$A^c \rightarrow$ "not a severe case".

The event, when changes causes the event A to change is "vaccinated" or "not vaccinated".

Thus,
$B \rightarrow$ "vaccinated"
$B^c \rightarrow$ "not vaccinated"

The confounding variable that causes the paradox is the age group.

$C \rightarrow$ ">50 age"
$C^c \rightarrow$ "<50 age"
age groups

To compute the probabilities used:-

$$\Rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\text{severe case} \cap \text{vaccinated})}{P(\text{vaccinated})}$$

$$= \frac{301 \times 10^5}{5.63 \times 10^6} = \underline{5.3}$$

$$\Rightarrow P(A|B') = \frac{P(A \cap B')}{P(B')} = \frac{P(\text{severe case} \cap \text{not vaccinated})}{P(\text{not vaccinated})}$$

$$= \frac{214 \times 10^5}{1.3 \times 10^6} = \underline{16.5}$$

$$\Rightarrow P(A|B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{P(\text{severe case} \cap \text{vaccinated} \cap >50 \text{ age})}{P(\text{vaccinated} \cap >50 \text{ age})}$$

$$= \frac{290 \times 10^5}{2.13 \times 10^6} = \underline{13.6}$$

$$\Rightarrow P(A|B' \cap C) = \frac{P(A \cap B' \cap C)}{P(B' \cap C)} = \frac{P(\text{severe case} \cap \text{not vaccinated} \cap >50 \text{ age})}{P(\text{not vaccinated} \cap >50 \text{ age})}$$

$$= \frac{171 \times 10^5}{0.18 \times 10^6} = \underline{95}$$

$$\Rightarrow P(A|B \cap C') = \frac{P(A \cap B \cap C')}{P(B \cap C')} = \frac{P(\text{severe case} \cap \text{vaccinated} \cap <50 \text{ age})}{P(\text{vaccinated} \cap <50 \text{ age})}$$

$$= \frac{11 \times 10^5}{3.5 \times 10^6} = \underline{0.3}$$

$$\Rightarrow P(A|B' \cap C') = \frac{P(A \cap B' \cap C')}{P(B' \cap C')} = \frac{P(\text{severe case} \cap \text{not vaccinated} \cap <50 \text{ age})}{P(\text{not vaccinated} \cap <50 \text{ age})}$$

$$= \frac{43 \times 10^5}{1.12 \times 10^6} = \underline{3.8}$$

(a) Using all the probabilities above,
weaker version of Simpson's paradox includes
the conditions below:-

(Let's put in the values and try and see if those
are satisfied).

$$\Rightarrow \boxed{\frac{P(A|B)}{P(A|B')} > \frac{P(A|B \cap C)}{P(A|B' \cap C)}} \text{ —— (1)}$$

$$\frac{5.3}{16.5} > \frac{13.6}{95} \Rightarrow \boxed{0.32 > 0.14}$$

true, thus,
(1) is satisfied.

$$\boxed{\frac{P(A|B)}{P(A|B')} > \frac{P(A|B \cap C')}{P(A|B' \cap C')}} \text{ --- } (2)$$

$$\frac{5.3}{16.5} > \frac{0.3}{3.8} \Rightarrow \boxed{0.32 > 0.08}$$

true, thus, (2) is also satisfied.

As both the conditions are satisfied by the vaccine data, thus, weaker version of Simpson's paradox is verified!

Ans.

(b). Stronger definition of Simpson's paradox includes:-

$$P(A|B \cap C) < P(A|B' \cap C) \text{ --- } (1)$$

Using the values computed,

$13.6 < 95 \rightarrow$ this is valid, thus, (1) is satisfied.

$$P(A|B \cap C') < P(A|B' \cap C') \text{ --- } (2)$$

Using the values computed,

$0.3 < 3.8 \rightarrow$ this is valid, thus, (2) is satisfied

$$P(A|B) > P(A|B') \text{ --- } (3)$$

$5.3 > 16.5 \longrightarrow$ this is not valid, thus, (3) is not satisfied.

As all the condns (1), (2) and (3) are not satisfied.

Ans. So, the events in vaccine data does not satisfy the stronger defn.

9. (a) As we need to distribute $K=30$ cookies among $n=4$ kids,

this will be a case of unordered sampling with replacement (as cookies are identical)

$$\Rightarrow \binom{n+K-1}{K} = \binom{30+4-1}{30} = \binom{33}{30}$$

Ans

(b) If we need to make sure that each kid gets atleast one cookie then we will first give 1 cookie to each of 4 kids leaving behind 26 cookies to be distributed among those 4 kids.

Similar to the (a) part,

the number of possibilities will be $\Rightarrow$

$$\binom{26+4-1}{26} = \binom{29}{26}$$

Ans

(c) Since there are 40 kids now basically we need to choose which 30 kids would get the cookie as each kid can only get one (at the most).

So,

$$\binom{40}{30}$$

Ans