Conditional probability and introduction to statistical inference 13oct22

13 October 2022 06:45 PM

Set theory elements:

1. Union
2. Intersection
3. Difference
4. Complement
5. P(AUB) formula
6. And = x or = +

Set : is a collection of objects or elements.

$A1 = \{A, B, 1, 2, 3, \}$

$A2 = \{x : x \in \mathbb{N}\}$ $\quad \mathbb{N} \to$ natural nos.

$= \{1, 2, 3, 4, \ldots\}$ infinite elements.

Union of sets.

$X1 = \{1, 3, 5, 7, 9\}$

$X2 = \{2, 4, 6, 8, 10\}$

$X_1 \cup X_2 = \{1, 3, 5, 7, 9, 2, 4, 6, 8, 10\} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

union

Intersection of sets

$X_1 \cap X_2 = \{\ \}$ null since there are no common elements

a null set is denoted by $\varnothing$ (greek letter phi)

Difference of sets.

$X3 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ $\quad X_1 = \{1, 3, 5, 7, 9, 11\}$

$X3 - X1 = \{2, 4, 6, 8, 10\}$

$X1 - X3 = \{11\}$

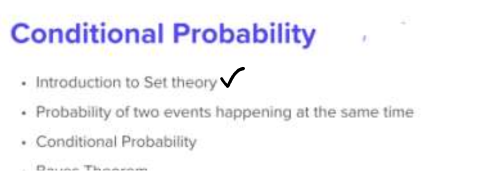
now we will use set theory operations in solving probability questions

If E_1 & E_2 are two events, then $P(E_1 \cup E_2)$ says that prob. of either E_1 occurs or E_2 occurs.

If E_1 & E_2 are 2 events then $P[E_1 \cap E_2]$ says that prob. of E_1 and E_2 both occur

What is conditional?

Conditional in this context means reducing sample space by having extra information about an event.



$E_1 \to$ getting a red card

- Introduction to Set theory ✓
- Probability of two events happening at the same time
- Conditional Probability
- Bayes Theorem

$E_1 \to$ getting a red card

$E_2 \to$ getting a face card $\{K, Q, J, A\}$

$P(E_1 \cap E_2) = \frac{8}{52}$ ✓

P[red card and face card] $= \frac{8 \text{ red face card}}{52 \text{ total}}$

$P(E_1 \cup E_2) =$ P(red card or face card) $= \frac{26+8}{52} = \frac{34}{52}$

Extra information that we have cards example

P(having a face card GIVEN THAT I HAVE A BLACK CARD)

$= 8/26$

fav. outcomes

{ 13 ♠, 13 ♣, 13 ♢, 13 ♡ }

Symbol of conditional is

$P(X|Y)$ = Prob. of event X when Y has already occured.

given extra inf.

Formula for conditional probability

DEFINITION For any two events A and B with $P(B) > 0$, the conditional probability of A given that B has occurred is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad (2.3)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

eg. P(face card | black card) $= 8/26$ from above.

FORMULA VERIFY

$P(A \cap B) =$ P(black face card) $= 8/52$

$P(B) =$ P(black card) $= 1/2$

$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{8}{52} \div \frac{1}{2} = \frac{8}{52} \times 2 = \frac{8}{26}$

4·7. Multiplication Law of Probability and Conditional Probability

Theorem 4·8. *For two events A and B*

$$P(A \cap B) = P(A) \,.\, P(B \mid A),\ P(A) > 0$$
$$= P(B) \,.\, P(A \mid B),\ P(B) > 0 \qquad \ldots(4\cdot8)$$

where $P(B \mid A)$ represents the conditional probability of occurrence of B when the event A has already happened and $P(A \mid B)$ is the conditional probability of happening of A, given that B has already happened.

$P(A \cap B) = P(A)\ P(B|A) \Rightarrow P(B|A) = \frac{P(A \cap B)}{P(A)}$

$P(A \cap B) = P(A)\, P(B|A) \Rightarrow P(B|A) = \dfrac{P(A \cap B)}{P(A)}$

$P(A \cap B) = P(B) \cdot P(A|B)$

$\Rightarrow P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

Example 2.25 Suppose that of all individuals buying a certain digital camera, 60% include an optional memory card in their purchase, 40% include an extra battery, and 30% include both a card and battery. Consider randomly selecting a buyer and let $A = \{\text{memory card purchased}\}$ and $B = \{\text{battery purchased}\}$. Then $P(A) = .60$, $P(B) = .40$, and $P(\text{both purchased}) = P(A \cap B) = .30$. Given that the selected individual purchased an extra battery, the probability that an optional card was also purchased is

0.75

extra info already occ.

$P(O) = 0.6 \qquad P(B) = 0.4 \qquad P(O \cap B) = 0.3$

$P(O|B) = \dfrac{P(O \cap B)}{P(B)} = \dfrac{0.3}{0.4} = \dfrac{3}{4} = 0.75$ ✓

HW

Example 2.27 Four individuals have responded to a request by a blood bank for blood donations. None of them has donated before, so their blood types are unknown. Suppose only type O+ is desired and only one of the four actually has this type. If the potential donors are selected in random order for typing, what is the probability that at least three individuals must be typed to obtain the desired type?

Making the identification $B = \{\text{first type not O+}\}$ and $A = \{\text{second type not O+}\}$, $P(B) = \frac{3}{4}$. Given that the first type is not O+, two of the three individuals left are not O+, so $P(A|B) = \frac{2}{3}$. The multiplication rule now gives

try it

$$\begin{aligned} P(\text{at least three individuals are typed}) &= P(A \cap B) \\ &= P(A|B) \cdot P(B) \\ &= \frac{2}{3} \cdot \frac{3}{4} = \frac{6}{12} \\ &= .5 \end{aligned}$$

→ answer

~~The multiplication rule is most useful when the experiment consists of several stages in succession. The conditioning event B then describes the outcome of the first stage and A the outcome of the second, so that $P(A|B)$—conditioning on what occurs first—will often be known. The rule is easily extended to experiments involving more than two stages. For example,~~

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_3 | A_1 \cap A_2) \cdot P(A_1 \cap A_2) \\ &= P(A_3 | A_1 \cap A_2) \cdot P(A_2 | A_1) \cdot P(A_1) \end{aligned} \qquad (2.4)$$

~~where A_1 occurs first, followed by A_2, and finally A_3.~~

4·7·3. Independent Events. *An event B is said to be independent (or statistically independent) of event A, if the conditional probability of B given A i.e., P (B | A) is equal to the unconditional probability of B, i.e., if*

$$P(B|A) = P(B)$$

prob. of B is unaffected by happening of event A or not
then B is indep of A, They are indep. events.
A cannot effect the occurance of B.

Independent Events

An independent event is **unrelated** to the likelihood of another event occurring (or not happening). In other words, the occurrence has no bearing on the probability that a subsequent event will occur. There is no difference between independent events in probability and independent happenings in daily life. What colour car you drive has nothing to do with where you work. No lottery ticket purchase will result in a child with blue eyes. One occurrence does not affect the likelihood of another event when the two events are independent.

Simple examples of independent events:

- Having a dog and maintaining a herb garden.
- Having enough milk and winning the lotto.
- Finding a penny on the ground after purchasing a lottery ticket (your odds of Finding a penny does not depend on you buying a lottery ticket).
- Finding your favourite movie on cable while taking a cab home.
- Getting a parking ticket and going to the casino to play craps.

no effect

$P(\text{win lott} \mid \text{more milk}) = P(\text{win lott})$ w/o condition

⇒ taking cab can't effect movies on cable

Dependent Events

When two events are reliant on one another, the likelihood of one event affects the probability of the other. An event is said to be dependent if it **needs another event to occur first**. Probabilistic dependent events are identical to dependent actual-world occurrences: If you want to go to a performance, it may rely on whether your employer will grant you time off, and if you want to visit family overseas next month, it may depend on your ability to obtain a passport in time. Formally, we can state that when two events are interdependent, one event's likelihood influences the other's likelihood.

Simple examples of dependent events:

- Going to jail after robbing a bank.
- Your power will be turned off if you don't pay your electricity payment on time.
- Seeking a decent seat before getting on a bus.
- Receiving a parking ticket for doing so. If you park illegally, your chances of receiving a ticket increase.
- Winning the lottery after purchasing ten tickets. Your chances of winning increase as you buy more tickets.
- Driving a car and being involved in a collision.

$P(B \mid A) \neq P(B)$

occurance of A has affected B so B is dependent on A

$P(\text{going to jail} \mid \text{robbed}) > P(\text{going})$

$P(\text{collision} \mid \text{driving}) > P(\text{collision})$

Statistical independence

2 events are Statistically independent if

$P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$ ⇒ $\frac{P(B \cap A)}{P(A)} = P(B)$

⇒ $P(A \cap B) = P(A)P(B)$

$P(A \cap B) = P(A)P(B)$

mutually independent

IMPORTANT

STATISTICAL INDEPENDENCE:

2 events X & Y are statistically independent if and only if

$P(X \cap Y) = P(X) \times P(Y)$

They are dependent if $P(X \cap Y) \neq P(X) \times P(Y)$

Dependent or Independent Event Formula

P(A|B) = P(A) ✓
P(B|A) = P(B) ✓

A: getting a diamond card

B: getting a red card

getting a diamond card   getting a red card

$P(A) = 1/4$

$P(B) = 1/2$

$P(A \cap B) = 1/4$

$P(A|B) = \frac{1/4}{1/2} = 1/2$ $P(A|B) \neq P(A)$

DEPENDENT

C

getting an ace

$P(A \cap C) = \frac{1}{52}$

$P(C) = \frac{1}{13}$

$P(C|A) = \frac{P(A \cap C)}{P(A)} = \frac{1/52}{1/4} = \frac{4}{52} = \frac{1}{13}$

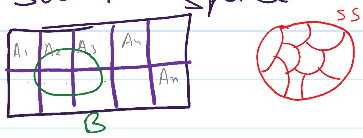
$P(C|A) = P(C)$ INDEPENDENT

The Law of Total Probability

Let $A_1, \ldots, A_k$ be mutually exclusive and exhaustive events. Then for any other event B,

$$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_k)P(A_k)$$
$$= \sum_{i=1}^{k} P(B|A_i)P(A_i) \qquad (2.5)$$

sample space   SS



$P(A_i \cap A_j) = 0$

$A_1, A_2, \cdots, A_n$ mutually exclusive events & exhaustive

→ EXCLUSIVE MEANS INTERSECTION IS NULL

→ EXHAUSTIVE MEANS TOGETHER THEY COVER FULL SAMPLE SPACE

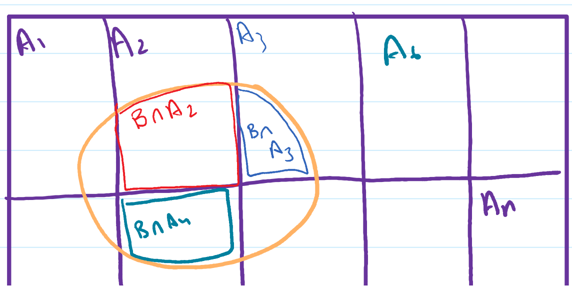
then for any other event B

$P(B) = \underbrace{P(B|A_1)}P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)$

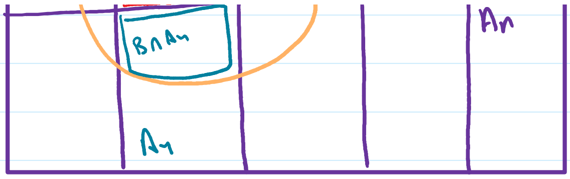
$P(B) = \frac{P(B \cap A_1)}{P(A_1)} \cdot P(A_1) + \frac{P(B \cap A_2)}{P(A_2)} \cdot P(A_2) \cdots + \frac{P(B \cap A_n)}{P(A_n)} \cdot P(A_n)$

$$\boxed{P(B) = P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_n)}$$

B



$P(B \cap A_6) = 0$



Bayes Theorem

The Bayes theorem establishes the likelihood of an event occurring given any condition. It is considered for the case of conditional probability. Also, this is known as the formula for the likelihood of "causes".

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

A,B = Events
P(A|B) = Probability of A given B is true
P(B|A) = Probability of B given A is true
P(A), P(B) = The independent probabilities of A and B

$$P(A_1|B) = \frac{P(B|A_1) \cdot P(A_1)}{P(B)} = \frac{\frac{P(B \cap A_1)}{\cancel{P(A_1)}} \cdot \cancel{P(A_1)}}{P(B)}$$

$$= \frac{P(B \cap A_1)}{P(B)}$$

$$= P(A_1|B)$$

useful Baye's thm formula.

$$\boxed{P(A_1|B) = \frac{P(B|A_1) \cdot P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)}}$$

using above law of total prob.

Bayes Theorem: Example-2

Question: Think about the three machines. Each machine can produce one thousand pins at once. A defective pin will be produced at a rate of 10% from Machine 1, 20% from Machine 2, and 5% from Machine 3. What is the likelihood that ~~a pin produced from the first machine would be defective~~?

when a defective pin comes it's from M_1

Solution:

Let us take the probability of choosing a faulty pin randomly be represented by P(A);
Pin choose from the first machine be represented by M1;
Pin choose from the second machine be represented by M2;
Pin choose from the third machine be represented by M3;
Chance of choosing pin any one of the three machines = P(M1) = P(M2) = P(M3) = 1/3
The probability of choosing a faulty pin from 1st machine is

$$P(M_1|A) = \frac{P(M_1)P(A|M_1)}{P(M_1)P(A|M_1)+P(M_2)P(A|M_2)+P(M_3)P(A|M_3)}$$

$$= \frac{(1/3) \times 0.1}{(1/3) \times 0.2 + (1/3) \times 0.1 + (1/3) \times 0.05}$$

$$= 2/7$$

let
$D \rightarrow$ defective
$P(M_1) = 1/3 = P(M_2) = P(M_3)$
$P(D|M_1) = 0.1$
$P(D|M_2) = 0.2$
$P(D|M_3) = 0.05$
$P(M_1|D) = ?$

$$P(M_1|D) = \frac{P(D|M_1)P(M_1)}{P(D|M_1)P(M_1) + P(D|M_2)P(M_2) + P(D|M_3)P(M_3)}$$

$$= \frac{0.1 \times 1/3}{\quad + 0.05 \times \frac{1}{3}}$$

2

$$= \frac{0.1 \times 1/3}{0.1 \times \frac{1}{3} + 0.2 \times \frac{1}{3} + 0.05 \times \frac{1}{3}}$$

$$= \frac{0.1}{0.1 + 0.2 + 0.05} = \frac{0.1}{0.35} = \frac{10}{35} = \frac{2}{7}$$

Example 2.30 An individual has 3 different email accounts. Most of her messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3. Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively. What is the probability that a randomly selected message is spam?

HW

To answer this question, let's first establish some notation:

$A_i = \{\text{message is from account } \# \, i\}$ for $i = 1, 2, 3,$ $\quad B = \{\text{message is spam}\}$

Then the given percentages imply that

$$P(A_1) = .70, \; P(A_2) = .20, \; P(A_3) = .10$$
$$P(B \mid A_1) = .01, \; P(B \mid A_2) = .02, \; P(B \mid A_3) = .05$$

Now it is simply a matter of substituting into the equation for the law of total probability:

$$P(B) = (.01)(.70) + (.02)(.20) + (.05)(.10) = .016$$

In the long run, 1.6% of this individual's messages will be spam. ■

INTRO TO STATISTICAL INFERENCE

We want to infer or gather some information about our population.

\# Issue is the availibility of data for the entire population
eg. I want to test whether my new flavour of chips will be well received or not.

I can launch my chips if it will score more than 7 in the market. {population}
* since I can't give one pack to each person on planet, I collect a sample of people and give them trial and note their score.

Based upon this score I can infer something about my population's probable score.

ESSENTIAL

population's probable score.

ESSENTIAL

Good Sample ✓

eg. If I take lays lovers. my score is biased.

Big Sample ✓

big sample size

· Sample should be a good representative of popⁿ.

Bigger the sample more it will represent population.

Inference keeps getting better when these qualities improve.

Sample Vs Population

A **population** in research is the complete group that you are interested in examining. This can refer to a group of individuals (such as all US adults or all employees of a corporation), but it can also refer to a group that includes other kinds of elements, such as things, occasions, organisations, nations, species, and organisms.

Sample is a subset of Population

Why do we need a sample?

Population

Let's imagine you wanted to know which movie genres Delhi residents preferred. How do you manage that?

You could go to every home in Delhi, ask everyone what they enjoy, and then evaluate their responses. Who would act in such a manner? That would be extremely tedious and possibly impossible.

needs too much time & money

What then, if not that? You can poll a few residents from various parts of Delhi and then use calculations to determine what most people *might* like.

good sample.

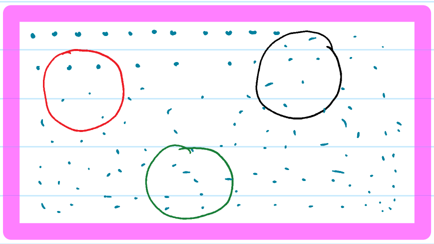
You now have it. You gathered information from a populace in the first tedious scenario because of **EVERYONE**! In the second case, you gathered information from a subset (a select number of individuals); you gathered information from a sample.

Statistic vs. parameter

$\mu \rightarrow$ mean $\sigma^2 \rightarrow$ variance

1) Parameter: fixed constant which exists for a population and states a fact about the population. It is usually denoted by greek letters.

eg. Population mean, Population variance, Population skewness Population sum or any other function of population.



POPULATION of size 'N'

$\mu \rightarrow$ popⁿ mean

$\sigma^2 \rightarrow$ popⁿ variance

---

2) Statistic : It is not a fixed constant as it changes with different samples and tells me a fact about the sample which I have chosen. usually denoted by english small letters.

statistic is a function of sample

sample is a subset of population

$\bar{x}_1 \to$ mean of sample red $\qquad$ $s_1^2 \to$ sample variance of red

$\bar{x}_2 \to$ sample mean of green $\qquad$ $s_2^2 \to$ sample va. of green

search Phases of clinical trials.

Sample Statistic Vs Population Parameter

function of samp $\qquad$ func. of popn.

Characteristic	Sample Statistic	Population Parameter
Mean	x̄	μ
Standard deviation	s	σ
Proportion	p̂	p
Sample Size	n	N → popn size

I ask if you are vegetarian or not. out of N I will get some proportion of vegetarians. the popn proportion is denoted by p. and sample proportion denoted by $\hat{p}$ $\qquad$ ^ → estimate of

$$\sum_{i=1}^{n}(x_i) \to \text{sum} = x_1 + x_2 + x_3 \cdots + x_n$$

↳ indicator going from 1 to n $\qquad$ $i = \{1, 2, 3, 4, \ldots, n\}$

$$\prod_{i=1}^{n} x_i = x_1 \cdot x_2 \cdot x_3 \cdots x_n$$

$\Pi \to$ product

$\Sigma \to$ sum

Sample Mean and Sample Standard Deviation

Calculating the sample mean is the same as calculating the population mean. We multiply the total number of values in the data set by the sum of all the data values. The sample mean is as a result;

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

The population parameter is underestimated by the standard deviation calculated using the formula $\sqrt{(\Sigma(xi-\bar{x})^2/n)}$ to calculate the standard deviation of a small sample.
The n in the numerator is changed to (n - 1) to obtain a fair estimation of the population standard deviation.
Therefore,

Sample Standard Deviation = $\sqrt{(\Sigma(xi-\bar{x})^2/n-1)}$ $\qquad s = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}}$

$$\sigma = \sqrt{\frac{\Sigma(x_i-\mu)^2}{N}}$$

when calculating sample variance we take (n-1) in denominator instead of 'n' because that gives us an unbiased estimate of σ

Question: The following are GPA scores of 30 High School students. Find the sample mean and standard deviation.
3.1, 2.9, 2.8, 2.9, 3.8, 4.8, 4.2, 3.9, 3.4, 2.5, 4.2, 3.7, 3.3, 2.1, 3.8, 3.0, 3.7, 4.0, 2.7, 3.8, 3.2, 3.5, 3.5, 3.6, 2.2, 3.1, 3.5, 4.0, 2.7, 4.5.

Solution:
Sample mean:
$\bar{x} = \Sigma x/n = 102.4/30 = 3.41$ ✓

calculate this

Sample standard deviation:
$S = \sqrt{(\Sigma(xi-\bar{x})^2/n-1)} = 0.65$

HW

Theory of multiple samples and sampling distribution of the mean - lite version : we can have a whole lot of samples from a given population.
we fix our sample size at 'n'

$\boxed{{}^{N}C_n \text{ samples}}$

Sample no.	Sample	sample $\bar{x}$	sample s
1	$\{x_1, x_2, x_3, \dots, x_n\}$	$\bar{x}$	S_x
2	$\{y_1, y_2, y_3, \dots, y_n\}$	$\bar{y}$	S_y
3	$\{z_1, z_2, z_3, \dots, z_n\}$	$\bar{z}$	S_z
⋮		⋮	⋮
${}^{N}C_n$	$\{k_1, k_2, k_3, \dots, k_n\}$	$\bar{k}$	S_k

→ average = μ

as analyst I will have only one sample

now if we make this table for all possible samples of size 'n' in the population we get a set of sample means $\{\bar{x}, \bar{y}, \bar{z}, \dots, \bar{k}\}$
now treating this ⤴ as a population of it's own this will also have a Pop^n mean and pop^n std. dev. and this becomes the sampling

distribution of mean.

How to choose a sample which is a good representative of the Population?

Sampling and Types of sampling

N Z

Sampling is the practice of choosing a portion of a larger population (a predetermined number of observations). It's a typical strategy where we do trials and make generalisations about the population without looking at the whole population. In this lesson, we'll examine two different sampling techniques:

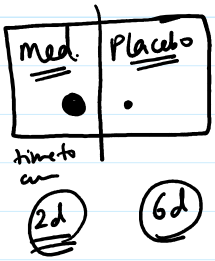
most frequently

Probability Sampling: *In this case, probability theory selects a sample.* The most important prerequisite for probability sampling is that each member of your population has an equal chance of being chosen. RANDOM SAMPLE

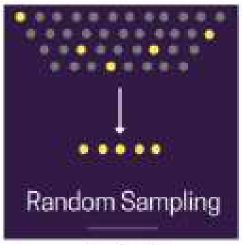
Non-Probability Sampling: *In this method, we select a sample based on factors other than chance; thus, not every member of the population has an equal chance of being included.*

Probability Sampling – Random Sampling

Pop^n → infected only
fever



Every component of the population has an equal chance of being chosen in a random sampling. *The below fig. depicts a pictorial representation of the same; all the points represent the population as a whole, with each point having an equal probability of being chosen.*



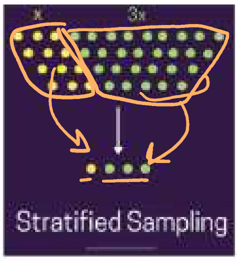
Random Sampling

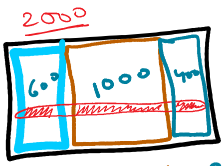
1
2
3
:
N

generate n random nos. and pick them.

2) Probability Sampling – Stratified Sampling

We group the total population into subpopulations via stratified sampling according to shared characteristics. We next take a random sample from each of those groups so that the **groups remain in the same proportion** as they were in the overall population. A graphical representation of the same is shown below in fig. We randomly choose from the yellow and green sets, which have **two groups with count ratios of x and 4x**, respectively, and represent the final set in the same ratio as these groups.





30% → 30% sample
50% → 50% samp
20% → 20% sample

eg. I need to choose sample of size 100 from pop^n of size 2000.

so i will choose $\underbrace{30} + \underbrace{50} + \underbrace{20} = 100$

India → J&K | UK | HR | DL | .... | KL | TN

Pop^n dist^n → 3% | 1% | 2% | 3% | .... | 5% | 4%

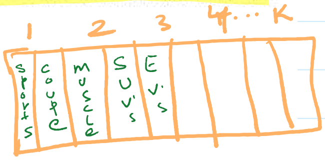
states become my strata.

to choose sample of size 'n' i will get $(3\% \times n)$ J&K
$+ (1\% \times n)$ UK + ...

Probability Sampling – Cluster Sampling

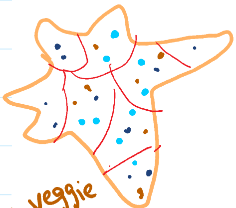
In cluster sampling, we break down the overall population into smaller groups, each of which shares the features of the population as a whole. We also choose the entire subgroups randomly rather than merely picking people.

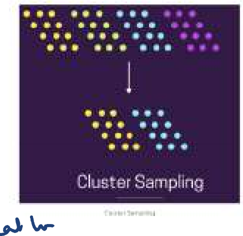
Rich people

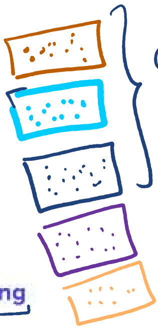


→ out of K choose some cluster and all points of the cluster are in the sample.

We had four clusters with comparable characteristics (size and form), as shown in the below figure. We chose two clusters at random and treated them as samples.







Cluster

- veggie
- fish lover
- meat lo

choose full clusters & not subsample clusters.

Probability Sampling – Systematic Sampling

Systematic sampling aims to collect samples from the population at the set, recurring intervals (basically fixed and periodic intervals). Every fifth element, the twenty-first element, and so on. Generally, this sampling technique has a higher success rate than the conventional random sampling technique. A graphical representation of the same is shown below in fig. We sample each element's ninth and seventh occurrences before repeating this process.

take samples at regular intervals

every 12^{th} value is chosen

N
1
2
3
⋮
2000

eg. $n = 50$

1, 41, 81, 121 ....

N = 60 need sample of size 3 give systematic sample starting from 3. →

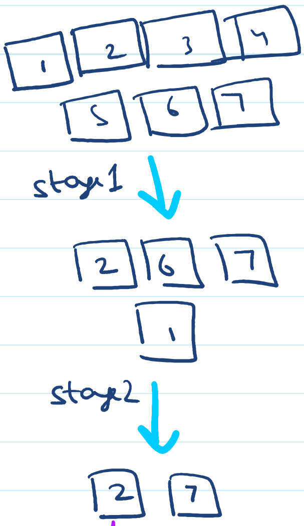
60 [1, 2, ..., 20 — 3
[21, 22, ..., 40 — 23
[41 42 .... 60 — 43

$3 \xrightarrow{9} 12 \xrightarrow{9} 21 \xrightarrow{9} 30 \xrightarrow{9} 39 \xrightarrow{9} 48 \xrightarrow{9} 57$ ✓

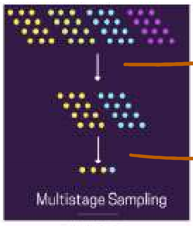
Probability Sampling – Multi-stage Sampling

Multiple sample techniques are stacked one on top of the other in multistage sampling. For instance, cluster sampling can be used to select clusters from the population in the first stage, and then random sampling can be used to select components from each cluster to create the final set. A graphical representation of the same is shown below in fig.

eg.2 stage 1 → cluster
stage 2 → cluster also possible



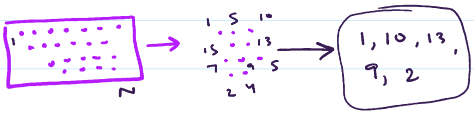
eg 1



→ stage 1 cluster

→ stage 2 random

eg 3. stage 1 → random
Stage 2 → systematic

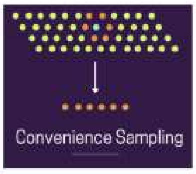


Non-Probability Sampling – Convenience Sampling

Only those people who are most accessible and available to participate in the study are included in convenience sampling. The pictorial representation of the same is shown in the figure below. The blue dot represents the researcher, and the orange dots represent the group of persons closest to orange.

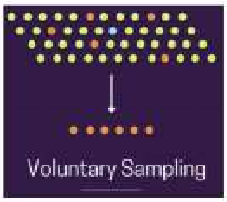
Choose on your own

do anything.



Non-Probability Sampling – Voluntary Sampling

Regarding voluntary sampling, **interested parties typically participate independently** by completing various survey forms. A current example is the frequently shown YouTube survey asking, "Have you seen any of these ads?" Here, **the survey's researcher has no authority to select a participant**. The same is depicted in pictorial form in the figure below. The researcher is shown as a blue dot, while study participants are shown as orange dots.



Non-Probability Sampling – Snowball Sampling

The final group is selected through Snowball sampling, which involves the researcher asking other known contacts to identify potential study participants. In the figure below, the researcher is depicted as the blue dot. At the same time, the orange and yellow circles represent the researcher's known contacts and the other participants in the study, respectively.

Popⁿ of class XII math students