Measures of Dispersion and intro to probability

10 October 2022 06:57 PM

What is central tendency? — A representative number used to depict a data set, which can represent centrality or other measurable properties of data.

Always consider an array of numbers

eg. Mean (arithmetic average); Median; Mode;

AVERAGE { mean, median, mode } { Harmonic mean, geometric mean } $\boxed{\text{Arithmetic mean}}$

d: 6, 8, 24, 19, 5, 24, 24

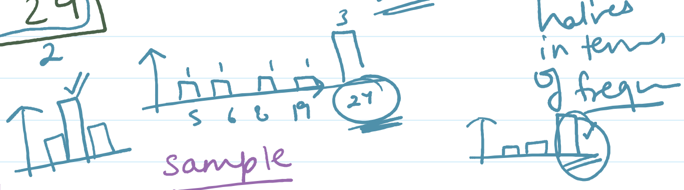
1 A.M: $\frac{\sum \text{data}}{\# \text{data}} = 12.4$

step 1 for median is always ascending your data

5, 6, 8, 19, 24

2 Median: any no. which divides my data into 2 equal halves in terms of frequency

3 Mode: 24

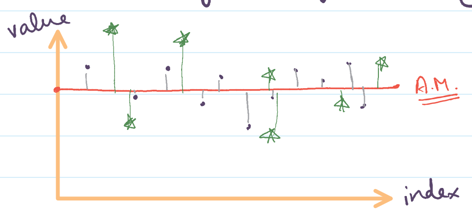


sample { n, numbers }

Study of data the 4 measures of studying any data:

1. Central tendency
2. Dispersion
3. Symmetricity } out of syll.
4. kurtosis

Deviation: moving away from something — average (A.M.)



1 6
2 3
3 2
4 4
5 ⋮
⋮
n

What is dispersion

Being stretched out or in a state of dispersion. The degree to which numerical data is ~~expected to deviate~~ deviating from an average value is known as statistical dispersion.

In other words, dispersion aids in the comprehension of data distribution.

Statistics measures of dispersion are used to assess data variability, or how homogeneous or heterogeneous the data is. Simply put, it demonstrates how constrained or dispersed the variable is.

Measures of Dispersion: measuring the degree to which data points vary or deviate from the selected average.

A few measures of dispersion are:

1. Range → Range
2. Variance
3. Standard Deviation } → one concept

Range

It is simply the difference between the maximum value and the minimum value given in a data set.

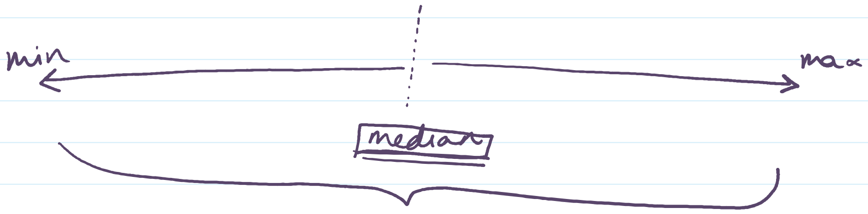
(1 to 10) natural nos. range = 10 − 1 = 9

Range = Max − Min ✓

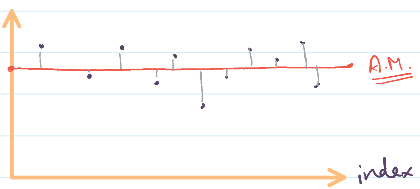
(0, 0, 0, 0, 0)

range :

−5, −6, 7, 24, 8, −10 range = 34 ✓



Variance : Variation



On an avg. how much the points deviate from A.M.

$\bar{x} = 13.33$

	x	$x - \bar{x}$	
1	20	(6.67)	+
2	10	(−3.33)	+
3	15	(1.67)	+
4	16	(2.67)	+
5	14	(0.67)	+
6	5	(−8.33)	+

Mean (Deviation from AM) = 0

ALWAYS ZERO

mean deviation from the mean is always zero.

$$\boxed{\left\{ \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \right\}}$$ ✓

VARIANCE

(Mean absolute deviation)

sample represented by small letters in english Pop^n measured by greek letters.

Variance

variance

arithmetic mean

Variance is the ~~sum~~ of the squares of the deviations from the mean. It is employed to gauge how dispersed a set of data is about the mean. The variance is calculated by comparing each result to the mean, unlike the preceding variability measurements.
Whether you are calculating the variance for the complete population or using a sample to estimate the population variance, there are two formulas for the variance.

Population diversity

Pop^n variance

$(\sigma^2 = 1/N \, \Sigma^N_{i=1} (X_i - \mu)^2)$

Population Sample

{ } {– – – – – –}

$$\frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)^2$$ ✓

In the equation, σ^2 is the population parameter for the variance,
μ is the parameter for the population mean, and
N is the number of data points, which should include the entire population.

SUV

2.0 L
100 HP
12 km/L
5 seats

1000 avg.

Let's look at how this will be used in a practical case.
Raj wants to statistically understand the mileage he is getting from his new car - he has measured the mileage of his car and has the following readings -
13,15,14,13
N = 4
μ = mean = (13+15+14+13)/4 = 13.75
Here, this formula will become,

(13, 15, 14, 13)

$\bar{x} = \frac{1}{4}(13+15+14+13) = 13.75$

σ2 = ((13-13.75)2 + (15-13.75)2 + (14-13.75)2+ (13-13.75)2)/4
= ((-0.75) 2 + (1.25) 2 +(0.25) 2 + (-0.75) 2)/4
= (0.5625+1.5625+0.0625+0.5625)/4

$s^2 = \frac{1}{4}\left\{(13-13.75)^2 + (15-13.75)^2 + (14-13.75)^2 + (13-13.75)^2\right\}$

Here, this formula will become,

σ2 = ((13-13.75)2 + (15-13.75)2 + (14-13.75)2+ (13-13.75)2)/4
= ((-0.75) 2 + (1.25) 2 +(0.25) 2 + (-0.75) 2)/4
= (0.5625+1.5625+0.0625+0.5625)/4
= 2.75/4
= 0.6875

$$s^2 = \frac{1}{4}\left\{(13-13.75)^2 + (15-13.75)^2 + (14-13.75)^2 + (13-13.75)^2\right\}$$

$= \boxed{0.6875}$ ✓

var.s → sample , var.p → popⁿ.

Population
μ "mu" represents mean
σ^2 "sigma" represents variance
N popⁿ size

$\bar{x}$ represents mean
s^2 represents variance
n sample size

More examples in excel ✓

Why $n-1$ instead of n

advanced topic
to overcome biasedness

But why squared??

squared basically took care of -ve distance problem
but it inflated the distance!

Normalizing using std dev

What is std dev
Difference from mean deviation

Standard deviation = $\sqrt{\text{Variance}}$

$$\sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}}$$

→ normalized back / scaled back!

Standard Deviation

The standard deviation is the average distance between each data point and the mean. Simply taking the variance's square root yields the standard deviation.
You have a reduced standard deviation when the values in a dataset are clustered more closely together.
Conversely, when values are more dispersed, the standard deviation is higher because the standard deviation is higher.
The standard deviation conveniently uses the data's original units, simplifying interpretation. The standard deviation is, therefore, the most frequently employed measure of variability.

Take pizza delivery as an example; a standard deviation of 5 means that the delivery time will vary by 5 minutes from the mean. It's frequently stated alongside the average: 20 minutes (s.d. 5).

Simply taking the variance's square root yields the standard deviation. Remember that the variance is expressed in square units. The square root thereby converts the value to natural units.

Ungrouped: No categorization or grouping.
All Individual data points.

Grouped: Frequency type data each value has a frequency attached to it.

x	f
2	2

[2, 2, 3, 4, 4, 4, 5, 5]

frequency attached to it.

x	f
2	2
3	1
4	3
5	2

grouped format

[2, 2, 3, 4, 4, 4, 5, 5]

ungrouped format

$n = 1000$

Class Intervals

much readable grouped format

2×2 $\quad 4\times3 \quad 3\times1 \quad 5\times2$

ll	ul	f
100	110	26
110	120	13
120	130	11
130	140	10
140	150	19
150	160	18
160	170	19
170	180	11
180	190	8
190	200	15

$(x_i - \bar{x})$

$(2-\bar{x})\ 2$

$(4-\bar{x}) + (4-\bar{x}) + (4-\bar{x})$

$(4-\bar{x})\times 3$

$$\boxed{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}}$$

for std dev take sq root.

Grouped formula

Formula for Variance And Standard Deviation Ungrouped Data

Grouped vs. Ungrouped

The formula for the variance of an entire population is the following:

$$\sigma = \sqrt{\frac{\Sigma(X-\mu)^2}{n}}$$

In the equation,

- σ2 is the population parameter for the variance
- σ is the population parameter for the standard deviation
- μ is the parameter for the population mean, and
- n is the number of data points, which should include the entire population

This is the same formula as before.

See excel question

Question-2: Calculate the variance and standard deviation of (3, 8, 6, 10, 12, 9, 11, 10, 12, 7.)

$$(\sigma) = 1/N \sqrt{N\Sigma fx^2 - (\Sigma fx)^2}$$

$$\sigma = \sqrt{\frac{1}{N}\sum\left\{f_i(x_i-\mu)^2\right\}}$$

$$\sigma^2 = \frac{1}{N}\sum\left[f_i(x_i-\mu)^2\right] = \frac{1}{N}\sum\left[f_i\left(x_i^2+\mu^2-2\mu x_i\right)\right] = \frac{1}{N}\sum\left[f_i x_i^2 + \mu^2 f_i - 2\mu f_i x_i\right]$$

$$= \frac{1}{N}\left[\sum f_i x_i^2 + \mu^2\sum f_i - 2\mu\sum f_i x_i\right] = \frac{1}{N}\left[\sum f_i x_i^2 + \mu^2\sum f_i - 2\mu\cdot\mu\cdot\sum f_i\right]$$

$$= \frac{1}{N}\left[\sum f_i x_i^2 + \mu^2\sum f_i - 2\mu^2\sum f_i\right] = \frac{1}{N}\left[\sum f_i x_i^2 - \mu^2\sum f_i\right] = \frac{1}{N}\left[\sum f_i x_i^2 - \frac{(\sum f_i x_i)^2}{(\sum f_i)^2}\cdot\sum f_i\right]$$

$$= \frac{1}{N}\left[\sum f_i x_i^2 - \frac{(\sum f_i x_i)^2}{N}\right];$$

now $\times \div N$ $\left\{\frac{1}{N^2}\left[N\sum f_i x_i^2 - (\sum f_i x_i)^2\right]\right.$

$N \Big[\quad \quad \frac{}{N} \Big] \qquad\qquad \Big) \; N^2 \Big[\qquad \Big]$

take root $\quad \sigma = \frac{1}{N}\sqrt{N\sum f_i x_i^2 - (\sum f_i x_i)^2}$

Formula for Variance And Standard Deviation Grouped Data

Difference between grouped and ungrouped

The formula for standard deviation becomes:

$(\sigma) = 1/N\sqrt{N\Sigma fx^2 - (\Sigma fx)^2}$

Question-3: Calculate the mean and standard deviation for the below data

x	f
60	2
61	1
62	12
63	29
64	25
65	12
66	10
67	4
68	5

Solution:

x	f	xf	fx^2
60	2	120	7200
61	1	61	3721
62	12	744	46128
63	29	1827	115101
64	25	1600	102400
65	12	780	50700
66	10	660	43560
67	4	268	17956
68	5	340	23120
	100	6400	409806

Mean	64
Variance	2.86
Standard Deviation	1.691153453

Variance is 2.86 and standard deviation is 1.69

PROBABILITY : chance of occurance of an event

Why do we need probability?

Knowing and comprehending the probabilities of different outcomes can be quite helpful in a world full of uncertainty. You can arrange your plans accordingly. I would bring my umbrella if it looked like it could rain. I will have myself tested if my eating habits indicate that I might have diabetes. I will notify my customer of a renewal premium if he is unlikely to do so without a reminder.
So knowing the likelihood might be very beneficial.

$P(\text{rain}) > 0.5$

bring umbrella

Definition of probability chances

Probability is the measurement of an event's likelihood. Probability aids in determining the chance of an event occurring because many events cannot be predicted with 100% accuracy. It is the proportion of positive events to all of the events in an experiment.

Probability(Event)=Favorable Outcomes/Total

favourable outcomes

occurring because many events cannot be predicted with 100% accuracy. It is the proportion of positive events to all of the events in an experiment.

Probability(Event)=Favorable Outcomes/Total

$$\frac{\text{favourable outcomes}}{\text{total outcomes}}$$

Numerically the probability value always lies between 0 and 1.

0≤P(E)≤1

$$P(\text{King}) = \frac{4}{52} = \frac{1}{13}$$

It is expressed in percentage, decimal, or fraction.

Prob. always lie b/w 0 & 1. 0 being an impossible event and 1 being certain event.

Sample Space : a set of all possible outcomes

SET

tossing 2 coins : { HH HT TH TT }

$$P[\text{at most 2 heads}] = \frac{4}{4} = 1$$ definite event

$$P[\text{at least 2 heads}] = \frac{1}{4} = 0.25$$ ✓

2

1

0

event : toss a coin & roll a die :

{ H-1 H-2 H-3 H-4 H-5 H-6

T-1 T-2 T-3 T-4 T-5 T-6 }

A : P(A) = 1/2

B : P(B) = 1/4

C : P(C) = 1/4

Key Terminologies

outcome

1 **An experiment** is a task for which the [redacted] are unknown. Every experiment has a mix of successful and unsuccessful results. Before Thomas Alva Edison successfully attempted to create the light bulb, he made more than a thousand unsuccessful attempts during his historical experiments.

a task to carryout eg. (roll + toss) with multiple possibilities. {sample space}

A B C : 1/3

2 **Random Experiment:** A random experiment is one for which the range of potential outcomes is known, but it is impossible to predict which specific consequence will occur on a given execution of the experiment in advance. Random experiments include tossing a coin, rolling a dice, and picking a card randomly from a deck.

3 **Trial:** The term "trials" refers to all attempts made throughout an experiment. In other terms, a trial is any specific outcome of a random experiment. Tossing a coin, for instance, is a trial.

4 **Event:** A trial with a certain conclusion qualifies as an event. An event might be something like throwing a coin and getting a tail.

5 **Random Event:** A random event is something unforeseen. You can never assign an exact value or probability because it is unexpected. For instance, because falling down a flight of stairs is purely random, it is impossible to predict the likelihood that you will do so in the next ten years.

6 **Outcome:** This is the outcome of the trial. Two obvious consequences exist when a sportsperson kicks a ball towards the goal post. He has the potential to score or fail to do so.

7 **Possible Outcomes:** A conceivable outcome is just a list of every possible result of an experiment. There are two possible results when tossing a coin: heads or tails.

8 **Equally likely outcomes:** Equally likely outcomes result from an experiment when each possible result has the same chance of happening. Any number can be rolled on a six-sided die with an equal chance of coming up.

P(any number)=1/6

(2) Probability of each outcome must be the same.

(3) One conduct of an experiment is a trial (one attempt) one outcome will be generated.

(6) one point in the sample space

(7) Sample space [Experiment → Event → Trial → Outcomes linked would be a probability]

Sample Space: This collects results from every experiment's trials. There are six possible results when tossing a die: 1, 2, 3, 4, 5, and 6. The sample space is comprised of these results. S = {1, 2, 3, 4, 5, 6} ← all possible outcomes in a set

Probable Event: A probable event can be foreseen. We can determine the likelihood of such occurrences. One can assess the likelihood that a certain child will advance to the following grade. As a result, we can describe this as a likely event. Prob. > 0

Impossible Event: An event that does not occur during the experiment or does not fit within the results' sample space is an impossible event. In a region with a temperate climate, there is no snowfall. Since there is zero chance that it would snow in this situation, it can be said that the event is impossible. Prob = 0

(**Complementary Events:**) When there are only two possible outcomes, and one of them is completely the opposite of the other, complementary events take place. The complement of an occurrence with probability P(A) is P(A'). P(A)+P(A')=1 The success and failure events in an examination are mutually supportive. P(Success)+P(Failure)=1 In a coin toss, getting heads and tails are complementary outcomes.

If sample space contains only 2 outcomes

Mutually Exclusive Occurrences: Two events are said to be mutually exclusive if one event prohibits the occurrence of the other. In other words, if two occurrences cannot happen simultaneously, they are said to be mutually exclusive. For instance, flipping a coin has a chance of producing either heads or tails. Both cannot be seen simultaneously.

✓ Total prob. must always be unity.
sum of probabilities of all outcomes in a sample space.

Question: Mary had a jar containing 8 red balls, 5 blue balls, and 7 green balls. She called one of her friends and asked them to pick a ball from the jar. What is the probability that the ball which is picked is either a red or a blue ball?

13/20

PERMUTATION AND COMBINATION

Permutation is Arranging Objects where order of the objects matter.

[A, B, C, D] Permute:

✓ABCD ✓BACD
✓ABDC ✓BADC
✓ACBD ✓BCAD
✓ACDB ✓BCDA
✓ADBC ✓BDAC
✓ADCB ✓BDCA

Combination | select or combine objects where their selection order does not matter

A B C : I need to choose 2 out of these 3.

BA
AC
BC

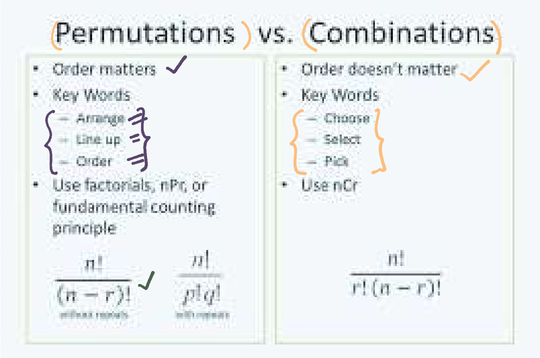
I'll choose 2 & give gold coins to both

Permutation

✓AB ✓BA
✓AC ✓CA
✓BC ✓CB

1st → gold coin
2nd silver coin

I'll choose 2 & give gold coins to both



$1^{st} \to$ gold coin

$2^{nd} \to$ silver coin

PERMUTATIONS : ^nP_r

$n \to$ total objects

$r \to$ objects to be arranged.

4 letters ABCD Permute any 2

$$^4P_2 = \frac{4!}{2!} = \frac{4 \times 3 \times \cancel{2 \times 1}}{\cancel{2 \times 1}} = \underline{\underline{12}}$$

$$^nP_r = \frac{n!}{(n-r)!}$$

$^4P_3$ $3! = \underline{\underline{3 \times 2 \times 1}}$

ABC

$\underbrace{\ }_{3}\underbrace{\ }_{\times 2}\underbrace{\ }_{\times 1} = 6 = 3!$

ABCD

AB BA
AC CA
AD DA
BC CB
BD DB
CD DC

12

ABCD

ABC BCA CBA CAB BAC ACB
ABD ADB BAD BDA DAB DBA
ACD ADC CDA CAD DCA DAC
BCD

24

$$^4P_3 = \frac{4!}{(4-3)!} = \frac{4!}{1!} = 4!$$

$$= \underbrace{4 \times 3 \times 2 \times 1} = \underline{24}$$

COMBINATION : take or pick certain elements from a set of elements.

^nC_r

$n \to$ total elements

$r \to$ elements to collect

$$^nC_r = \frac{n!}{(n-r)!\,r!} = \frac{^nP_r}{r!}$$

ABCD choose 3 out of 4 letters. order doesn't matter

ABC
ACD
BCD
ABD

ACD

$$^4C_3 = \frac{4!}{(4-3)!\,3!} = \frac{4!}{1!\,3!}$$

$$= \frac{4 \times \cancel{3 \times 2 \times 1}}{1 \times \cancel{3 \times 2 \times 1}} = 4$$

ABD

$1 \times \cancel{3 \times 2 \times 1}$

ABCD | BACD
ABDC | BADC
ACBD | B _ _ _
ACDB | B _ _ _
ADBC | B _ _ _
ADCB | B _ _ _

REP. ALLOWED!

Permutation – Example: 1

Question: Let us take 10 numbers: 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9. What is the number of 4-digit-passcode which can be formed using these 10 numbers?

0000
0001
⋮
9999

$\boxed{^{10}P_4}$

$\frac{10!}{6!} = 9 \times \cancel{8 \times 7 \times 6} \times 10$

Solution:

These are the easiest to calculate.

When a thing has **n** different types ... we have **n** choices each time!

Here we have 10 different types. Hence, we have 10 choices

$10 \times 9 \times 8 \times 7$ ~~72 × 12 × 10~~

repetition is allowed

Choosing **r** of something that has **n** different types, the permutations are:

n × n × ... (r times)

Here, we are choosing 4 numbers, so we have 10 x 10 x 10 x10 = 10^4 times

$10 \times 10 \times 10 \times 10$

10^4

Here, the reason we have not used the formula is because the numbers can be repeated - 0000, 1111, 1122, are all possible

ABC → ABC, BCA, CAB

10,000

$10 \times 9 \times 8 \times 7$

$10 \times 10 \times 10 \times 10$

$\cancel{^{10}P_4}$

Combination – Example: 1

Question: A team of 2 is formed from 5 students (Raj, Jagdish, Naveen, Lokesh, and Ojas). Find the possible combination of teams.

R J N L O

$^5C_2 = \frac{5!}{2!\,3!} = \frac{5 \times 4}{2} = 10$

Solution:

The combinations can happen in the following 10 ways by which the team of 2 could be formed.

- ✓ Raj Jagdish
- ✓ Raj Naveen
- ✓ Raj Lokesh
- ✓ Raj Ojas
- ✓ Jagdish Naveen
- ✓ Jagdish Lokesh
- ✓ Jagdish Ojas
- ✓ Lokesh Naveen
- ✓ Lokesh Ojas
- ✓ Ojas Naveen

This is a simple example of combinations. C(5,2) = 10.

Rajesh / Salani

R: 792 S: 792

1) **Question:** Praveen has to choose 5 marbles from 12 marbles. In how many ways can she choose them?

$^{12}C_5 \quad \frac{12!}{7!\,5!}$

Rahul / Nikita

40, 14

2) **Question:** A committee of 3 members will be formed with two males and one female. Find the number of ways this committee can be formed from 5 male and four female members.

5M → 2M, 4F → 1F

$^5C_2 \times {}^4C_1$ ✓

Diksha / Diby

3) **Question:** There are 10 marbles in a rucksack, numbered from 0 to 9. How many ways of 3 different digits could be formed by picking them up from the rucksack, without replacement?

$^{10}C_3$

0 0 0 $10 \times 10 \times 10 = 10^3$ w/ repl.

A standard deck of 52 playing cards includes 4 aces, 4 kings, and 44 other cards.

Suppose that Luis randomly draws 4 cards *without replacement*.

What is the probability that Luis gets 2 aces and 2 kings (in any order)?

$\frac{\text{fav. out.}}{\text{tot. out.}}$

$$\frac{\boxed{{}^4C_2} \times \boxed{{}^4C_2}}{{}^{52}C_4} \checkmark$$

In Prob.

AND → ×

OR → +

Kyra works on a team of 13 total people. Her manager is randomly selecting 3 members from her team to represent the company at a conference.

What is the probability that Kyra is chosen for the conference?

Samara is setting up an olive oil tasting competition for a festival. From 15 distinct varieties, Samara will choose 3 different olive oils and blend them together. A contestant will taste the blend and try to identify which 3 of the 15 varieties were used to make it.

Assume that a contestant can't taste any difference and is randomly guessing.

What is the probability that a contestant correctly guesses which 3 varieties were used?