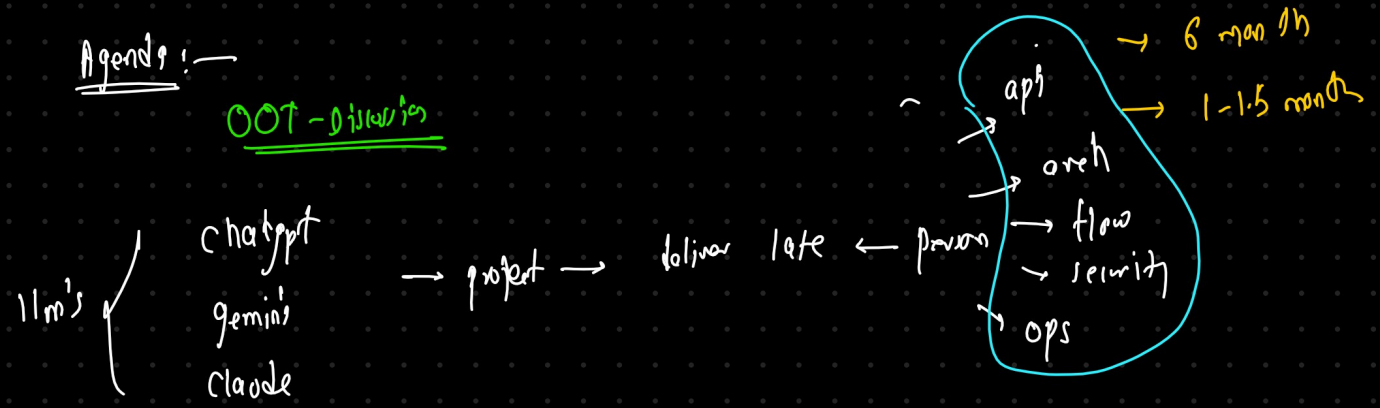
16-08-2025

Agenda :—

OOT - Discussion



Uniform Distribution:

mean outcome in a range has the same probability.

Scenario :—

perfectly fair.

Example:

Rolling a fair die (1,2,3,4,5,6)

pick a random number between 1 - 10

Example:

→ Bus arrives everytime between 10:00 & 10:30.

→ You arrive at 10:10

what's the prob the bus arrives in the next 5 minutes?

prob between 10 minutes & 15 minutes

$$P(x_1 \leq X \leq x_2) = \frac{x_2 - x_2}{b - a}$$

$a = 0$ (10:00)

$b = 30$ (10:30)

$x_2 = 15$

$x_1 = 10$

$$= \frac{15 - 10}{30} = \frac{5}{30} = \frac{1}{6}$$

10:00 ——— 10:30

↓

5 minutes

↓

6

10:00	10:05 → 1
10:05	10:10 → 2
10:10	10:15 → (3)
10:15	10:20 → 4
10:20	10:25 → 5
10:25	10:30 → 6

$$\frac{1}{6} = 0.1667 = 16.67\%$$

10:10

→

5, 5 (5)

→ bus guranteed in next 5 min

10:01 — 10:10

10:04

10:00

10:30

→ Random sampling
→ Initial assumption when no other data is available.

create random → uniform

→ D.S → job
↓ → freelancing
ML
DL
scary

→ S.D → confusion matrix
avg, mean
correlation

→ AI engineer (SD)
↓ Data, ML
↓ stats
SQL
↓

SQL →

stats → mean → ML, DL, NLP
stats

10% → 30%

part 2 - stats

Spread :-

mean = 10
SD = 5

-5 0 5 | 15 20 25

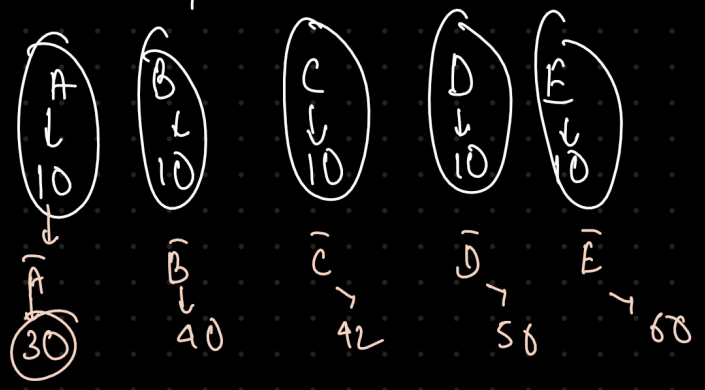
10

prob: spread of individual data points.
→ standard deviation (SD)

prob: spread of sample mean if you kept resampling

→ 100 data points → $\bar{U}$ → (50)

5 sample dataset of size 10 randomly,



Standard Error :

Example : A → 10 people → find their avg height → (avg-A) → 150

B → 10 people → find their avg height → (avg-B) → 160

C → 10 people → find their avg height → (avg-C) → 155

how much sample mean vary if i kept sampling

Z-statistics (Standard Error)

SE = sample mean vary if i kept sampling

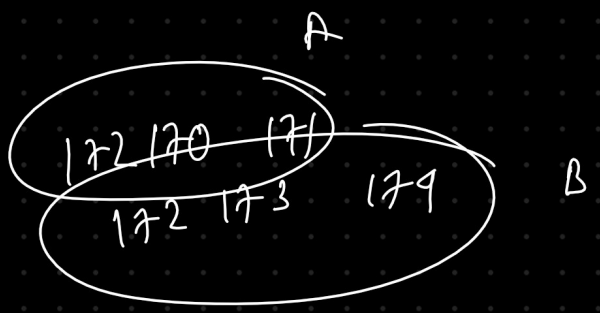
Example: A → 10 people height → 170 cm

B → 172 cm ⟷ Standard Error (change)

$$SE = \frac{\sigma}{\sqrt{n}}$$

σ → SD of population

n → sample size



SE → smaller if :

→ Data isn't too spread out

→ more samples

Z-score :—

$$z = \frac{\bar{x} - \mu}{SE}$$

$\bar{x}$ → sample mean

μ → population mean

SE → standard error.

Q: A company claims that their every drink has 200 mg caffeine on average.

→ we tested $n = 40$ cans:

sample_mean $\simeq$ 198 mg

population σ = 10 mg

$\mu = 200$ mg
$\bar{x} = 198$ mg
$\sigma = 10$ mg

Q: Is your sample mean significantly different from the claimed 200 mg?

Step by step:

(1) Find SE = $\frac{10}{\sqrt{40}} \approx 1.58$

(2) Find z = $z = \frac{198 - 200}{1.58} \approx -1.27 \approx |z| \downarrow 1.27$

$|z| = 1.27 < 1.96$ (95% confidence)
, is not statistically different

z score - range

95% confidence →	± 1.96
99% confidence →	± 2.58

$|z| < 1.96$ → Not statistically significant at 5% level
→ Difference could be due to random chance

$|z| \geq 1.96$ → statistically significant at 5% level
→ unlikely difference is due to chance. (weak evidence)

$|z| \geq 2.58$ → significant at 1% level
→ very strong evidence against the true statement.

→ 200 mg , 198 mg

$|z| \rightarrow 3.40$

$|z| \rightarrow 0.96$

$|z| \rightarrow 2.0$

A → 10% < 1.96

A → 50% ≥ 1.96

A → 80% $|z| \geq 2.58$

(A) (B) (C) → Pills

Significance levels: 5% and 1%

: 5% level ($\alpha = 0.05$)

→ we are okay with a 5% chance of being wrong when there is a difference.

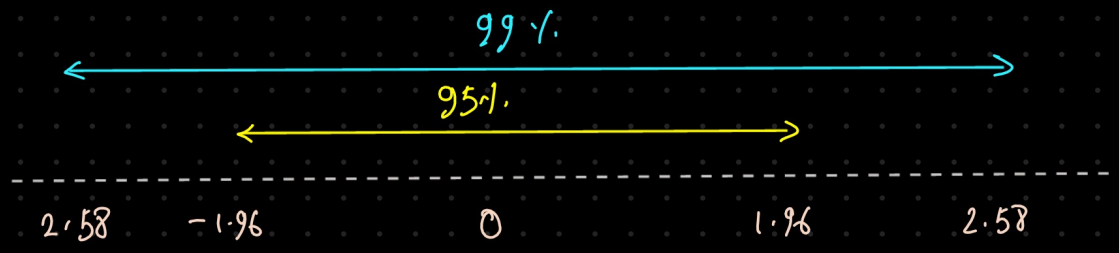
→ 95% confident

: 1% level ($\alpha = 0.01$)

→ only 1% chance of being wrong

5% level : bet on this if I'm 95% sure horse → A / B → win

1% level : bet on this if I'm 99% sure



95% of all possible values fall within 1.96 standard error

z-score unit → standard error

$$A = \frac{1 + CL}{2} \qquad CL = 95 \qquad A = \frac{1 + 0.95}{2} = \frac{1.95}{2} = 0.975$$

↓ z score table

CL → confidence level

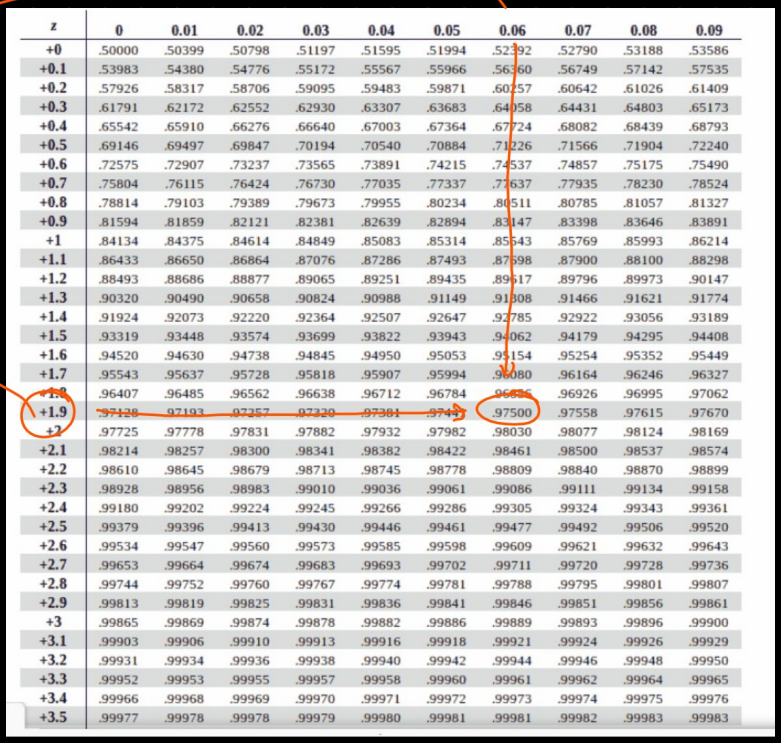
|z| → 95% → 1.96

1.9 + 0.06

= 1.96

P.S → 60% confidence

→ Z-score →



z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
+0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
+0.1	.53983	.54380	.54776	.55172	.55567	.55966	.56360	.56749	.57142	.57535
+0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
+0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
+0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
+0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
+0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
+0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
+0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
+0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
+1	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
+1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
+1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
+1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
+1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
+1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
+1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
+1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
+1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
+1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
+2	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
+2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
+2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
+2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
+2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
+2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
+2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
+2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
+2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
+2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
+3	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
+3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
+3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
+3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
+3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
+3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983

Central limit Theorem

Enough random samples & average them, those averages will follow a normal (bell-shape) distribution. even if original data is not normal,

→ when working with real-data, not every dataset looks like a nice bell-curve or close to normal distribution.

→ But CLT says if we work with sample means, we can often use ND formula: z-score, confidence interval, etc.

→ work on large sample $n \geq 30$

→ mean of sample → mean ≈ population mean

→ spread of sample mean → Standard Error $= \frac{\sigma}{\sqrt{n}}$

raw → normal distributed data
X

estimate → population parameter → population is very

Hypothesis testing :—

H_0 → Null Hypothesis
we assume it's true unless evidence says other wise.

H_1 → Alternate Hypothesis
this is the new claim we are testing for.

Rule : we never prove H_0 is true, we either reject it or fail to reject it.

P.) : Your pizza shop advertises Average delivery time = 30 min

→ A customer group claims you're slower :

$n = 36$ deliveries

sample - mean = 31 minutes

population σ = 4 minutes

$\alpha = 0.05$ (95% confidence) (1.96)

Step 1 : Hypothesis

$H_0 : \mu = 30$

$H_1 : \mu > 30$ (one - tailed test)

Step 2 : Significance level

$\alpha = 0.05$ (single tail)

(double tail)