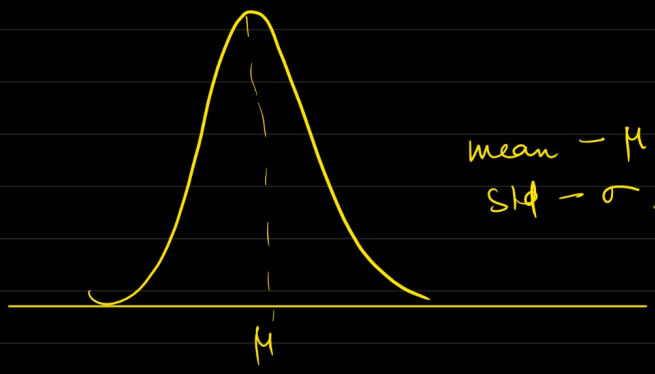
Standard Normal distribution



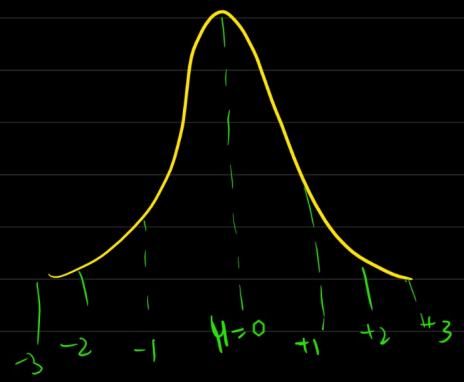
mean $= \mu$ → Can be any value

std $= \sigma$

$$pdf = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

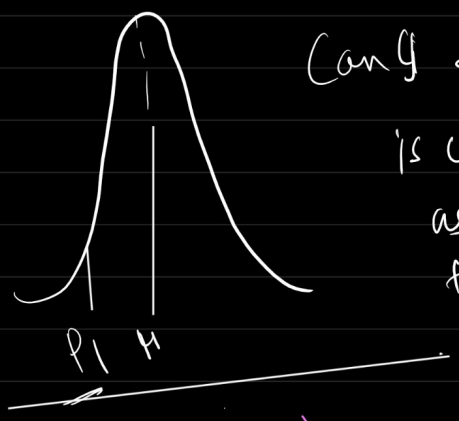
→ S.N.D is continuous prob distribution

→ A special case of N.D ⟶ $\mu = 0, \sigma = 1$



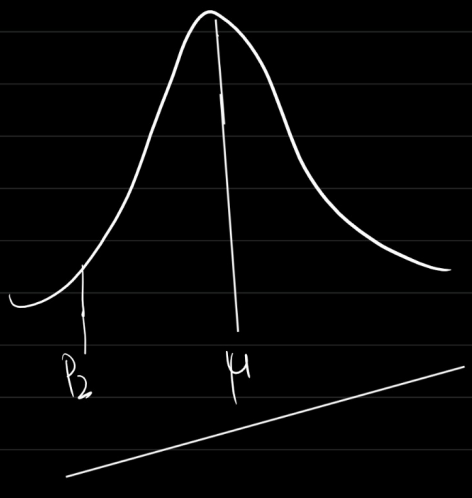
why SND if we already have N.D?

D_1



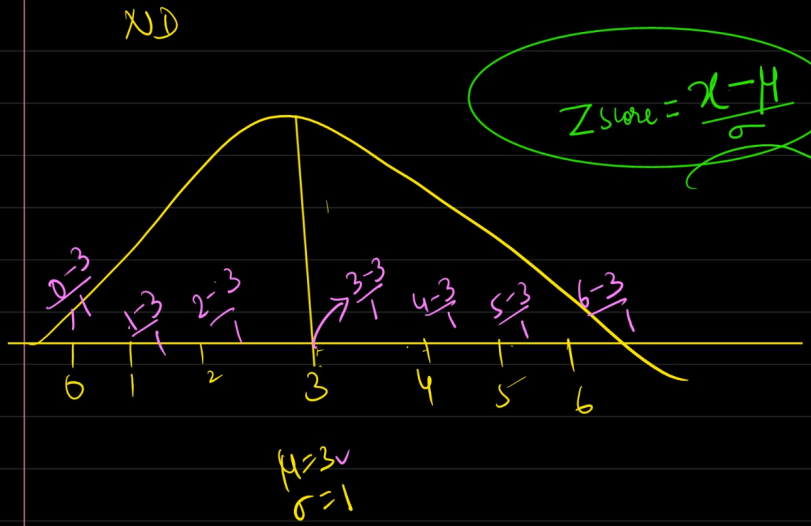
Can I say P_1 is closer to μ as compared to P_2

D_2

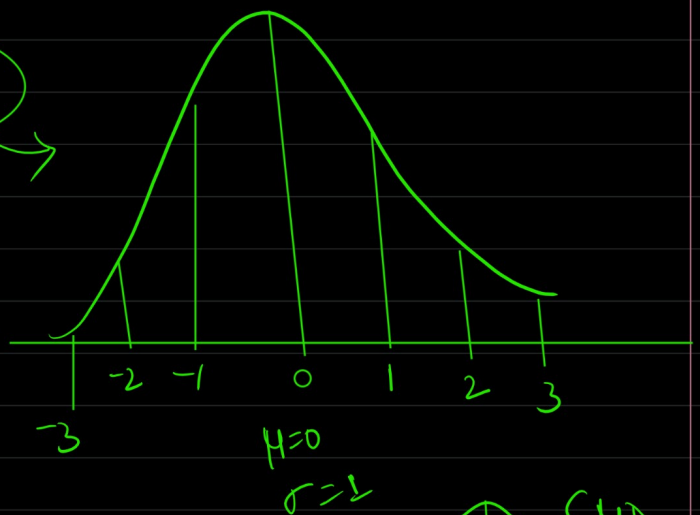


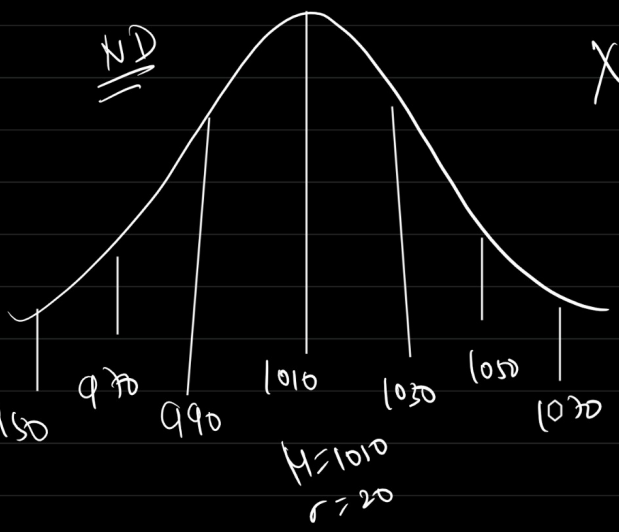
On the same scale μ & σ

N.D ⟶ Standard N.D



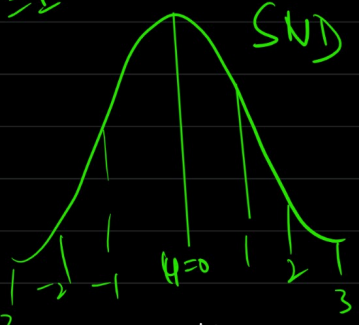
$$Z_{score} = \frac{x-\mu}{\sigma}$$





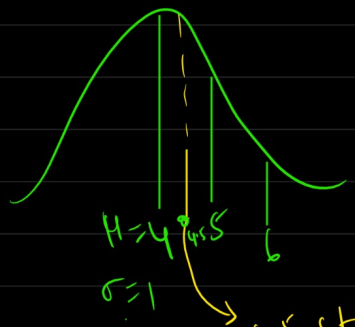
$$X \approx SND(\mu=0, \sigma=1)$$

$$Z = \frac{x-\mu}{\sigma}$$



x	$x-\mu$	$\frac{x-\mu}{\sigma}$
950	$950-1010=60$	$60/20=3$
970	$970-1010=40$	2
990	$990-1010=20$	1
1010	$1010-1010=0$	0
1030	$1030-1010=20$	1
1050	$1050-1010=40$	2
1070	$1070-1010=60$	3

Q



How many standard deviation 4.5 is away from mean?

$$Z_{score} = \frac{x-\mu}{\sigma} = \frac{4.5-4}{1} \Rightarrow 0.5$$

Q2.

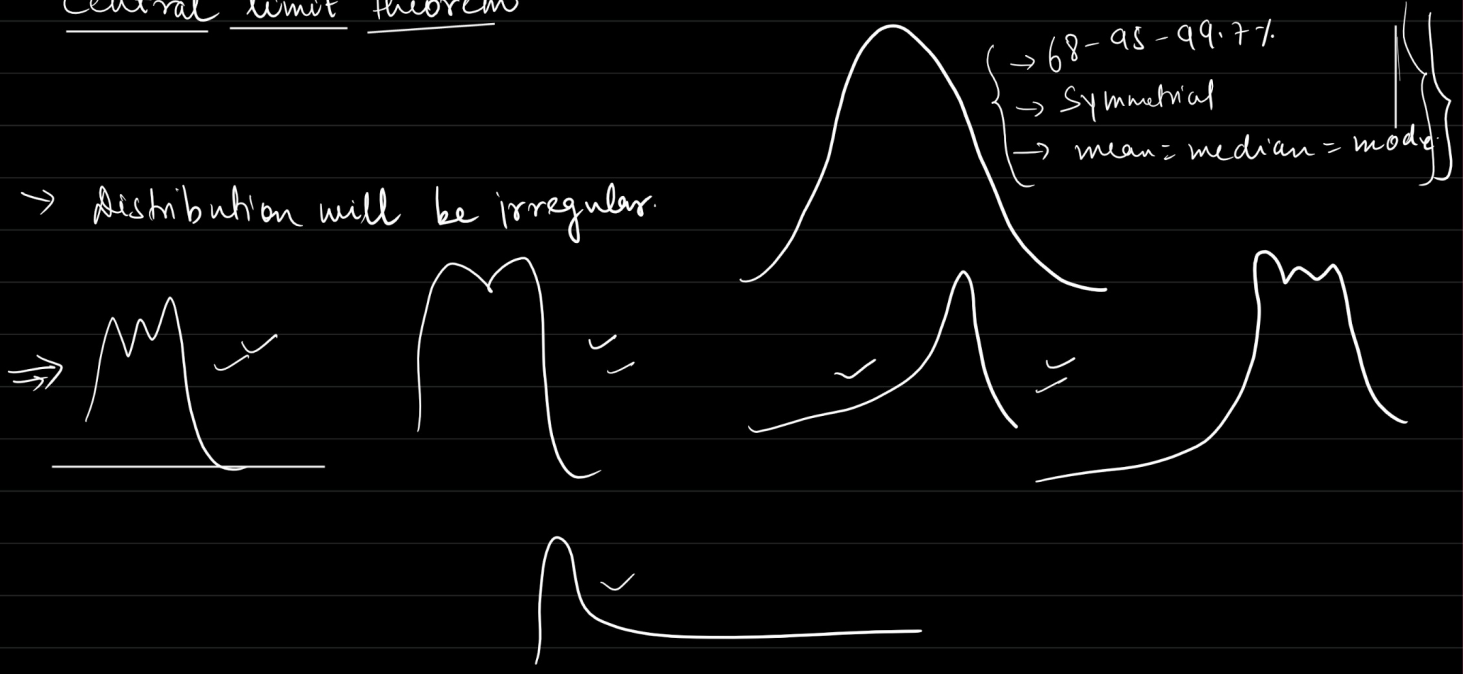
No of rooms	Area of House (sqft)	locality	distance from airport	y (price of house) (in lakhs)
1	1100	1	20 km	70
2	1200	2	30 km	80
3	1150	3	15 km	90
4	1250	1	60 km	110
5	1300	1	70 km	120
-	-	2	-	-
-	-	-	-	-
-	-	-	-	-

Linear Reg
logistic Reg
Clustering

→ Standardization → $Z\text{score} = \frac{x - \mu}{\sigma}$

$= \frac{x_i - \text{Area of house}}{\sigma}$

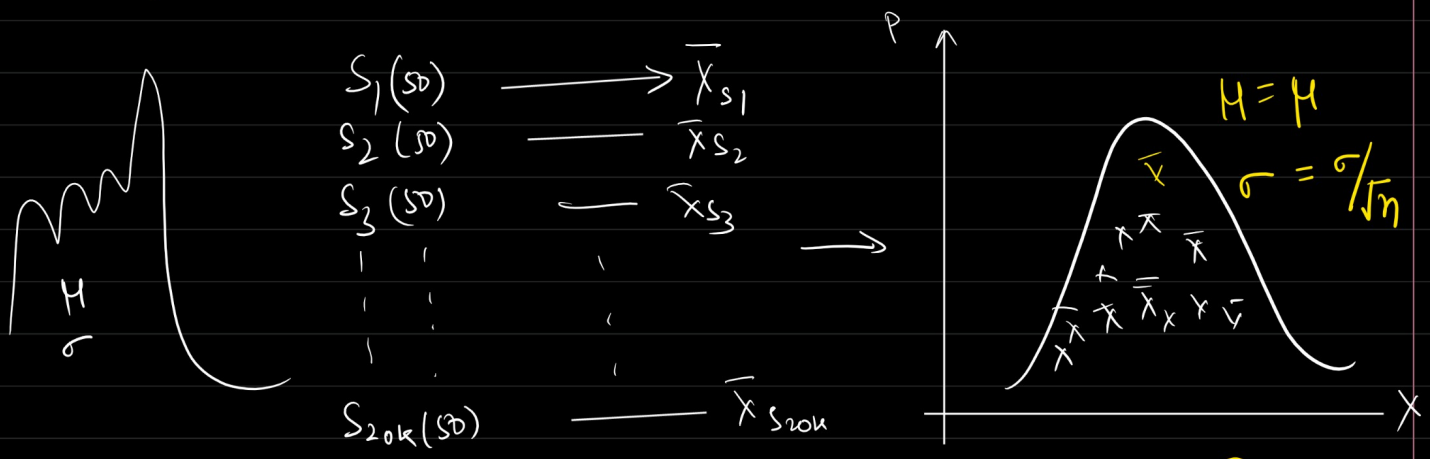
Central limit theorem



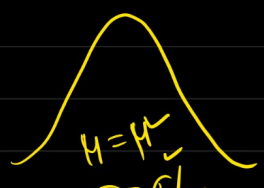
→ Distribution will be irregular.

→ The CLT states that if you have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population with replacement, then the distribution of sample means will be approximately normally distributed.

→ Sampling: mean of a population (μ, σ) will approximately be a normal distribution ⟶ $M = \mu$, $\sigma = \sigma/\sqrt{n}$



CLT → Sampling distribution of the mean ⟹

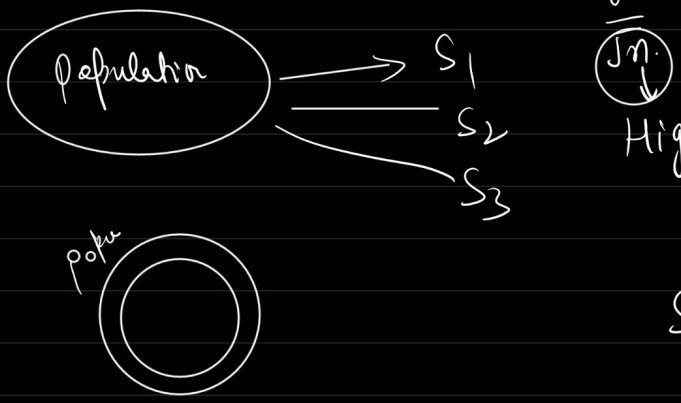


Population (μ, σ) ⟶ large No of sample ⟶ Sampling mean ⟶ plot ↳ N.D $M = \mu$, $\sigma = \sigma/\sqrt{n}$

Two conditions of CLT :-

1. The no. of samples should be large.
2. The sample size should be greater than or equals to 30 (except the population distribution which is already a Normal Dist.)

Standard error = $\sigma/\sqrt{n}$



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

Higher the sample size, SE will be low

$SE \propto \frac{1}{n}$ $\quad SE \downarrow \; n \uparrow$

Q You have a population with a $\mu = 100$ and Std dev (σ) = 20. If you take sample size 50 from this population. What is the prob that sample mean will be less than 105.

→ $\mu = 100, \quad \sigma = 20, \quad n = 50, \quad \bar{X} = 105$

$$Z\,score = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{105 - 100}{\frac{20}{\sqrt{50}}} = \frac{5 \cdot \sqrt{2}}{4}$$

$$Z\,score = \frac{5 \cdot \sqrt{2}}{4} \Rightarrow \frac{5 \times 1.414}{4} = \frac{7.07}{4} = 1.7675$$

$$= 0.8944$$

Estimate → A specified observed value used to estimate an unknown population parameter using sample.

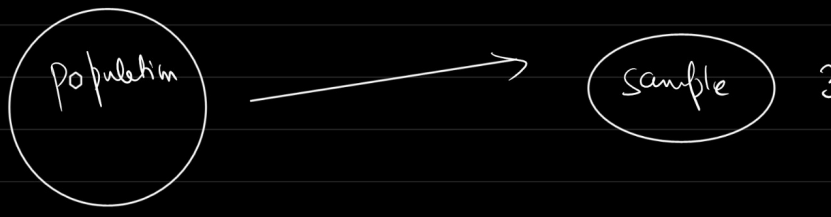
① Point estimate → A point estimate is a single value that is used to estimate the true value of a population parameter.

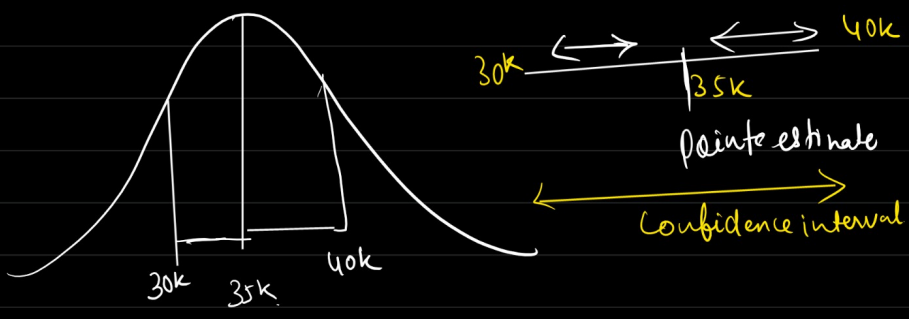
Avg salary of IT employees

Sample of 100 students → avg ht

→ mean, variance

② Interval estimate — Range of values used to estimate the unknown population parameter

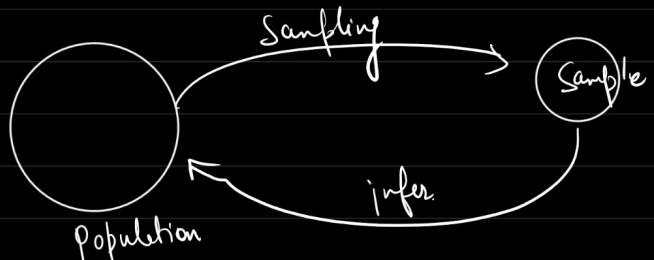




P value

→ The p-value is the probability value, calculated from a statistical test.

→ Pvalue in Hypothesis testing is used to decide whether to reject a null hypothesis or not.



* Mage of the employee is 45 years (pop)

→ You took a sample
→ Calculated avg age of that sample → 50 years.

Zscore ⇒ $\frac{\bar{x} - H}{\sigma/\sqrt{n}}$ → Corresponding to this Z score, you will be having prob. Value

How far?

Clinical trial of a medicine

↳ 1000 person

↳ 950 → got cured with this medicine.

↳ Medicine is working 95%.

→ Out of 100 times, 95% will work. → [95% confident]

→ 5% of the time medicine doesn't work. — [5% margin of error]

⇓

level of significance

⇓

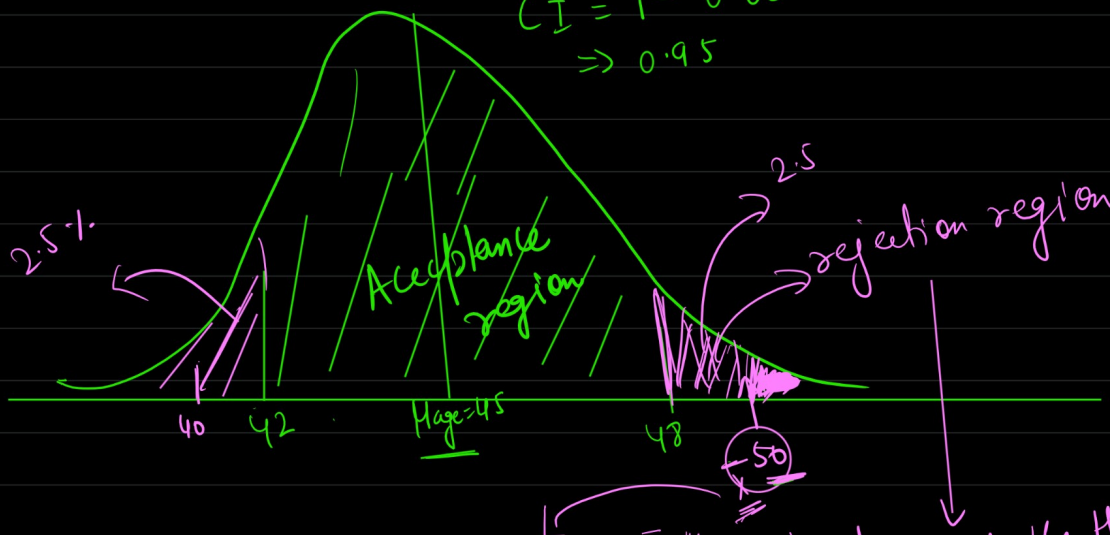
α.

* Conduct experiment with 5% level of significance

⇓

$CI = 1 - 0.05$

$\Rightarrow 0.95$



$Z\text{ score} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ → p-value → reject the H_0

2 tail

p value < Significance Value ↳ -1.2

reject the H_0

else fail to reject the H_0

Application of Z score

$$Z\text{score} = \frac{x - \mu}{\sigma}$$

Q $N(\mu = 50\text{cm}, \sigma = 20, D = 110)$

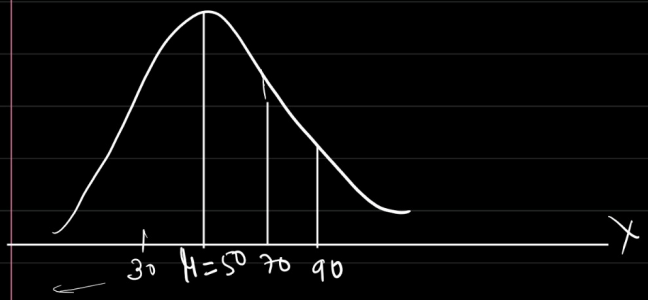
How many standard deviation D is away from mean?

$$Z\text{score} = \frac{x_i - \mu}{\sigma} = \frac{110 - 50}{20} = \frac{60}{20} = 3$$

$D = 110$ is 3σ away from mean.

$Z = 3 \rightarrow D = 110$ is 3 sd away from mean

$\sigma = 20 \rightarrow$ the dp on an avg is 20 units away from mean.



1 Sd away from mean $\Rightarrow \mu + 1\sigma$
$= 50 + 1 \times 20 = 70$

2 " " " " $= \mu + 2\sigma = 50 + 2 \times 20 \Rightarrow 90$

3 sd " " " $= \mu + 2\sigma = 50 + 3 \times 20 \Rightarrow 110$

$= \mu - 1\sigma$
$= \mu - 2\sigma$
$= \mu - 3\sigma$

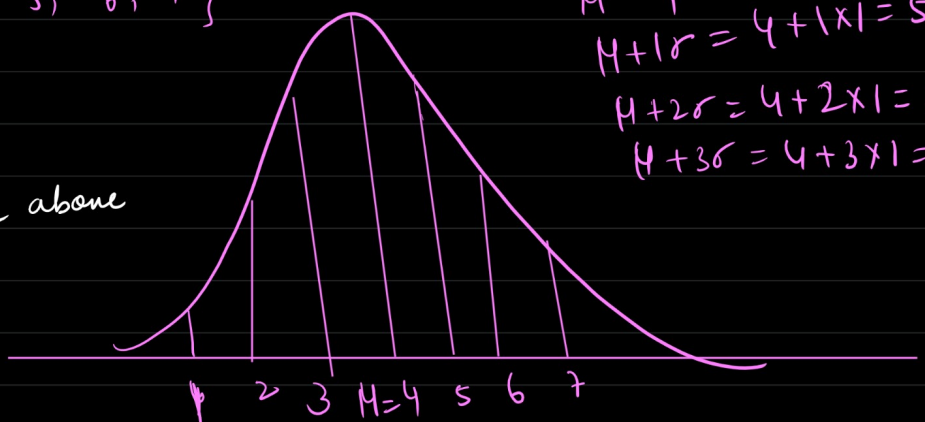
$\mu - 3\sigma = 4 - 3 \times 1 = 1$
$\mu - 2\sigma = 4 - 2 \times 1 = 2$
$\mu - 1\sigma = 4 - 1 = 3$
$\mu = 4$
$\mu + 1\sigma = 4 + 1 \times 1 = 5$
$\mu + 2\sigma = 4 + 2 \times 1 = 6$
$\mu + 3\sigma = 4 + 3 \times 1 = 7$

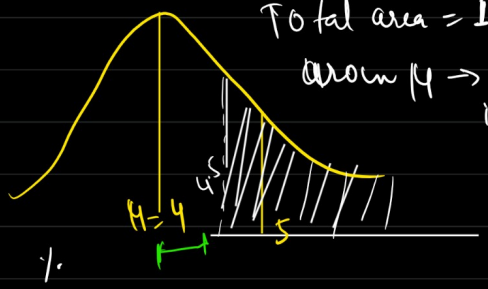
Q $X = \{1, 2, 3, 4, 5, 6, 7\}$

$\mu = 4$
$\sigma = 1$

What % of score will fall above 4.5?

→ What is probability that score is more than 4.5?





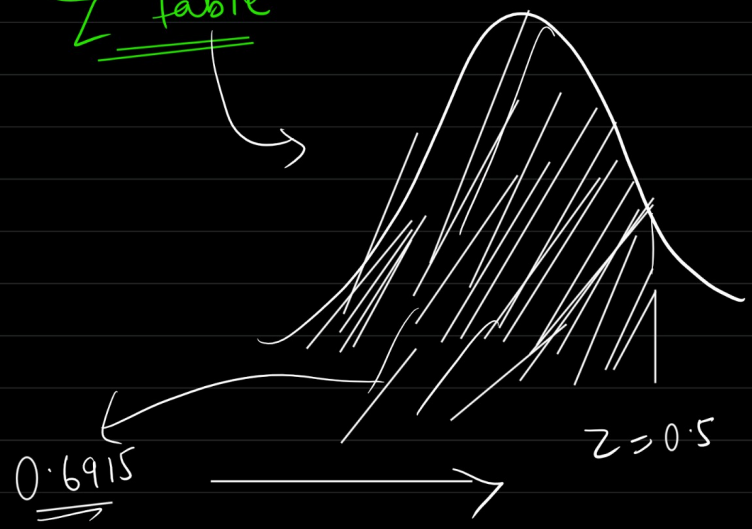
Total area = 1

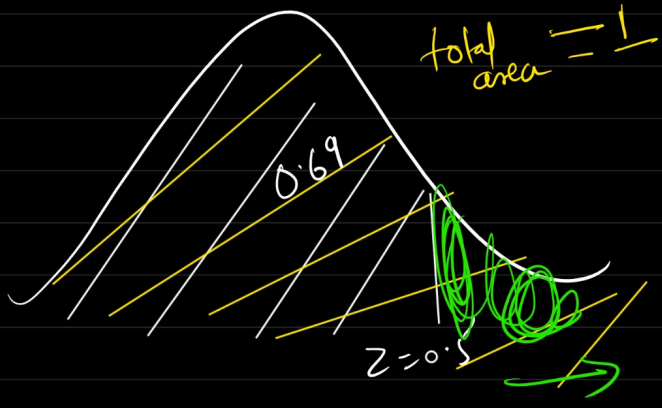
Area on μ → Area is 50% on left 50% on right

$$Z_{score} = \frac{x-\mu}{\sigma} = \frac{4.5-4}{1} = 0.5$$

4.5 is 0.5 σ away from mean

Z table





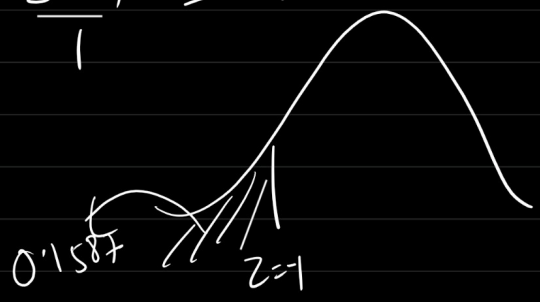
$= 1 - 0.69$

$= 0.31$

Q what is percentage of marks below 3 ?

$$Z = \frac{x-\mu}{\sigma} = \frac{3-4}{1} = -1$$

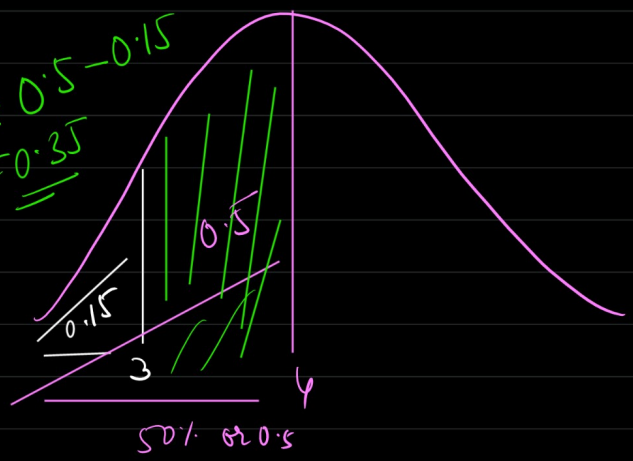
$Z = -1$



Q What is % age of marks between 4 & 3

$P(3 \le X \le 4) = 0.5 - 0.15$

$= 0.35$



Q The score follows a SND $\mu = 75$, and $\sigma = 10$

Find prob that a randomly selected student will score below 80.

$$Z = \frac{x - \mu}{\sigma} = \frac{80 - 75}{10} = \frac{1}{2} \Rightarrow 0.5$$

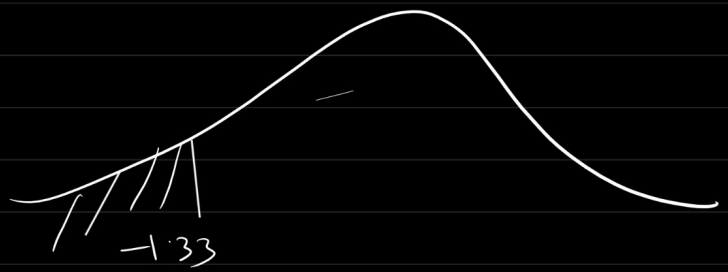
for $Z = 0.5$, AUC/cumulative prob = 0.6915

Q The avg IQ is 100 with $\sigma = 15$

% age of People lower than IQ 80

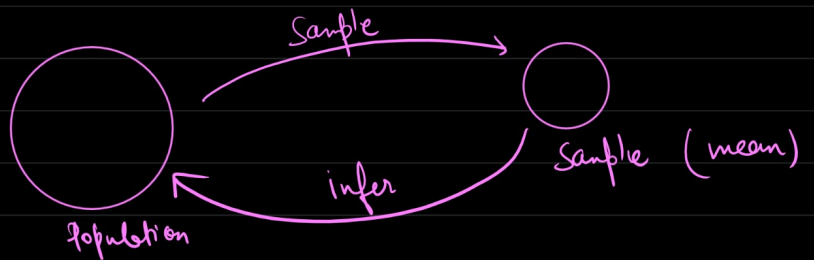
→

$$Z\text{score} = \frac{80 - 100}{15} = \frac{-20}{15} = -1.33$$



$\Rightarrow 0.9082$

Hypothesis and mechanism



→ Hypothesis testing.

* Hypothesis → It is a claim or a statement or an assumption about a population parameter that can be tested using statistical methods.

eg Avg Salary of IT employee 50K.
eg Consumption of Ice-cream is more in Summer.
eg The person is not guilty if accused of any crime.

① Null hypothesis → The initial or default assumption.

eg Person is not guilty

② Alternate Hypothesis :- opposite of Null hypothesis

eg Person is guilty.

Mechanism

① Frame the hypothesis
- H_0 (Null hypothesis)
- H_A (Alternate hypothesis)

Claim/ hypothesis :- Avg age of people in Pwskills is 45 years.

$H_0 : \mu_{age} = 45$
$H_A : \mu_{age} \neq 45$.

H_0 → will have equality sign.

Hypothesis → Avg age of Employee in ABC organisation is atleast 45 years?

→ H_0 : $H_{age} \geq 45$

→ H_A : $H_{age} < 45$

↳ $H_{age} \geq 45$

↳ $H_{age} < 45$

→ Avg age of Employees in pwskills is greater than 45 years?

$H_{age} > 45$ $H_{age} \leq 45$

H_0 : $H_{age} \leq 45$

H_A : $H_{age} > 45$

→ Avg age is greater than equals 45 years

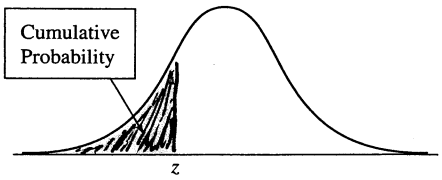
H_0 : $H_{age} \geq 45$

H_A : $H_{age} < 45$

② Experiment / statistical analysis (pvalue, significance level)

③ reject H_0 or fail to reject H_0

APPENDIX A

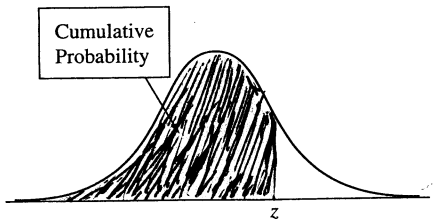


Cumulative probability for z is the area under the standard normal curve to the left of z

z	.00
−5.0	.000000287
−4.5	.00000340
−4.0	.0000317
−3.5	.000233

TABLE A Standard Normal Cumulative Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641



Cumulative probability for z is the area under the standard normal curve to the left of z

TABLE A Standard Normal Cumulative Probabilities *(continued)*

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

z	.00
3.5	.999767
4.0	.9999683
4.5	.9999966
5.0	.999999713