#Basics stats - Beginner

→ Probability

1) Intro to Basic Terms
2. Variables
3. Random Variables
4) Population, Sample, Population mean, Sample mean.
5) Population Distribution, Sample Distribution & Sampling distri....
6) Mean, Median, Mode
7) Range
8) Measure of Dispersion
9) Variance
10) Std Deviation
11) Gaussian / Normal Distribution

Intermediate stats

12) Standard Normal Distribution
13) Z score
14) Probability Density function
15) Cumulative distribution fn
16) Hypothesis Testing
17) Many different plotting graphs
18) Kernel Density Estimation
19) Central Limit Theorem
20) skewness of data
21) Covariance
22) Pearson Correlation Coefficient
23) Spearman Rank Correlation
24) Hypothesis Testing.

Advanced stats

25) Q-Q plot
26) Chebyshev's inequality
27) Discrete and cts Distribution
28) Bernoulli & Binomial Distn.
29) Log Normal Distn.
30) Power law Distribution
31) Box Cox Transform
32) Poisson Distribution
33) Application of non-Gaussian Distribution

Population
sample

Sample → inference & conclusions → Population
Exit polls.

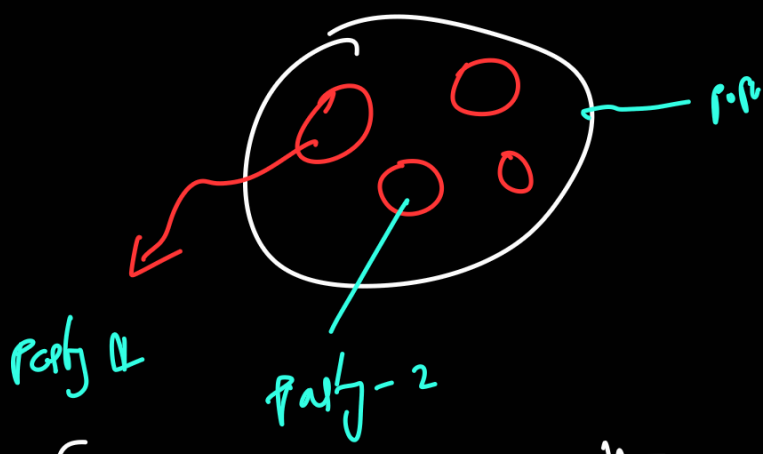
Descriptive stats

① Analyzing data, Summarizing data
organizing data, in form of
numbers, & graphs

② Bar plot, Histogram, Pie-chart,
PDF, CDFs Normal D. —

③ Measure of Central Tendency
Mean, Median, Mode

Inferential stats



① Z Test
② T test
③ Chi square Test
→ Hypothesis

Statistics

① Population → population mean
② Sample → Sample mean
③ Random variable → Discrete random variable
→ cts random variable.

mean
median
mode
} measure of central tendency

→ population mean = $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$

→ Sample mean $(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i$

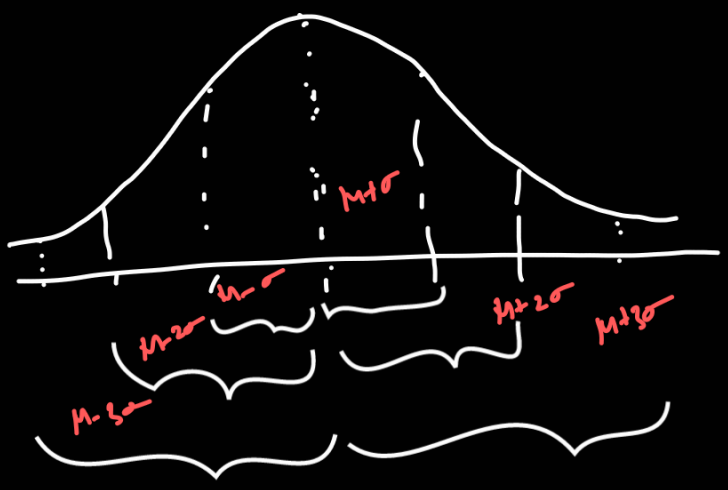
1 → Discrete → 2, 3, 4 [whole nums]
2 → cts → within a range of val.
10 — 15 [10.1, 10.2 ... 15]

③ Gaussian / Normal distribution

$x \sim G.D(\mu, \sigma)$ (mean, SD)

$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$

$Var = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$; $\sigma = \sqrt{Var}$



$P(\mu-\sigma \le x \le \mu+\sigma) \approx 68\%$

$P(\mu-2\sigma \le x \le \mu+2\sigma) \approx 95\%$

$P(\mu-3\sigma \le x \le \mu+3\sigma) \approx 99.7\%$

Empirical formulas.

#Log normal Distribution

$X \sim$ Log normal Distribution

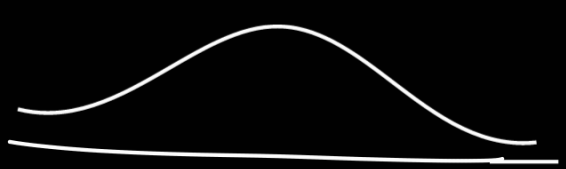
if $\ln(x)$ is normally Distributed

→ $X = \{x_1, x_2, x_3 \ldots x_n\}$

$\log(x_1), \log(x_2), \ldots, \log(x_n)$

↓

normally distributed



→ log normal

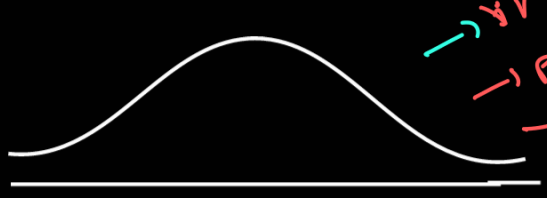
↓

$X \sim$ log normal

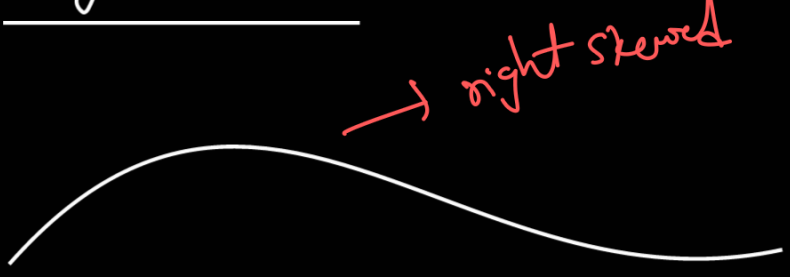
if $\log(x)$ is normally distributed.

↳ $\log(x)$ normal distribution ⇒ a log normal distribution

Gaussian distribution



Log normal



Eg: income of people.

↳ feedback of product

↳ comments length in amazon.

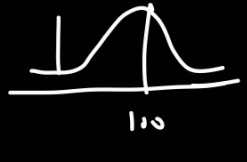
→ why distributions are in?

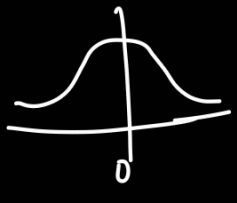
→ to convert all into same std scale.

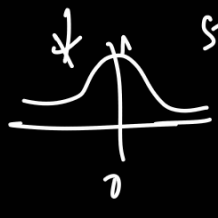
Ex:

R&D	market	stock	profit
↓	↓	↓	↓

gaussian → 

lognormal → log(....) → 

SDN $\frac{x_i - \mu}{\sigma}$ → 

SDN $\frac{x_i - \mu}{\sigma}$ → 

std scaler

Covariance

Ex:

size of house	price of house
↑	↑
↓	↓

$$Cov(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)$$

$$Var(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

$\Rightarrow Cov(x, x) = Var(x)$

→ Cov(x, y) <
- x ↑ y ↑ → +ve cov
- x ↑ y ↓ → -ve cov

→ Covariance only says +ve & -ve
↳ but don't say how much +ve & -ve, so we use pearson correlation coefficient.

---

mean, median, mode

[1, 2, 3, 4, 5] : mean = $\frac{15}{5} = 3$

$\mu = 3$

If I add an outlier 50

[1, 2, 3, 4, 5, 50] = mean = $\frac{65}{5} = 13$

$\mu_{new} = 13$

→ So, mean is highly influenced by outliers

↳ So, we use median.

[1, 2, **3, 4**, 5, 50]

$\frac{3+4}{2} = \frac{7}{2} = 3.5$ ↑

→ If there are lot of outliers, use median mode
↓
otherwise mean.

Random variables

① RV
② types of RV → numerical → discrete, cts
→ categorical

#Central Limit Theorem

$X \not\sim G.D(\mu, \sigma^2)$

$n \geq 30$

S_1 x_1 x_2 ..... x_{30} $= \bar{x}_1$

S_2 ..... $= \bar{x}_2$

S_3 ..... $= \bar{x}_3$

S_{∞} ..... $= \bar{x}_{\infty}$

$\bar{X}(\bar{x}_1, \bar{x}_2, \bar{x}_3 \ldots \bar{x}_{100}) \approx GD(\mu, \frac{\sigma^2}{n})$

(mean)

$\bar{x} \approx \mu$

#CHEBYSHEV'S INEQUALITY

$X \sim GD(\mu, \sigma)$

$P(\mu - \sigma \leq x \leq \mu + \sigma) \approx 68\%$

$P_r(\mu - 2\sigma \leq x \leq \mu + 2\sigma) \approx 95\%$

$P_r(\mu - 3\sigma \leq x \leq \mu + 3\sigma) \approx 99.7\%$

But if not follow GD.

$y \not\sim GD$

$P_r(\mu - k\sigma < x < \mu + k\sigma) > 1 - \frac{1}{k^2}$

$k = 1$ $P(\mu - \sigma < x < \mu + \sigma) > 1 - \frac{1}{1} = 0$

$k = 2$ $P(\mu - 2\sigma < x < \mu + 2\sigma) > 1 - \frac{1}{4} = 75\%$

$k = 3$ $P(\mu - 3\sigma < x < \mu + 3\sigma) > 1 - \frac{1}{9} = 85\%$

#PEARSON CORRELATION COEFFICIENT

① Covariance $= Cov(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x - \mu_x)(y - \mu_y)$

② Pearson cc $= \rho_{(x,y)} = \frac{Cov(x,y)}{\sigma_x \sigma_y}$

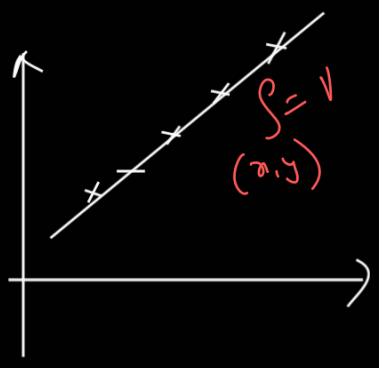
↓

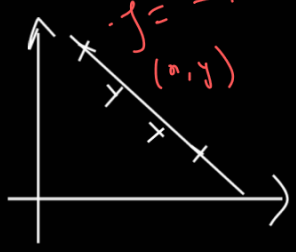
① STRENGTH

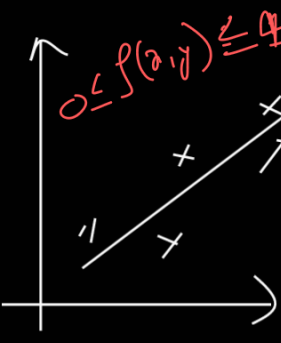
② DIRECTION OF RELATIONSHIP

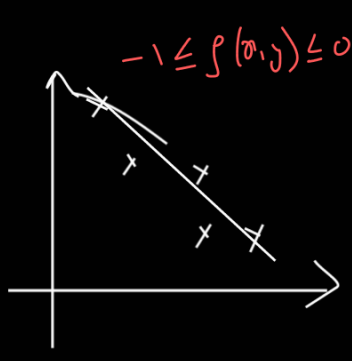
$-1 \leq \rho \leq 1$

③ Spearman Rank cc







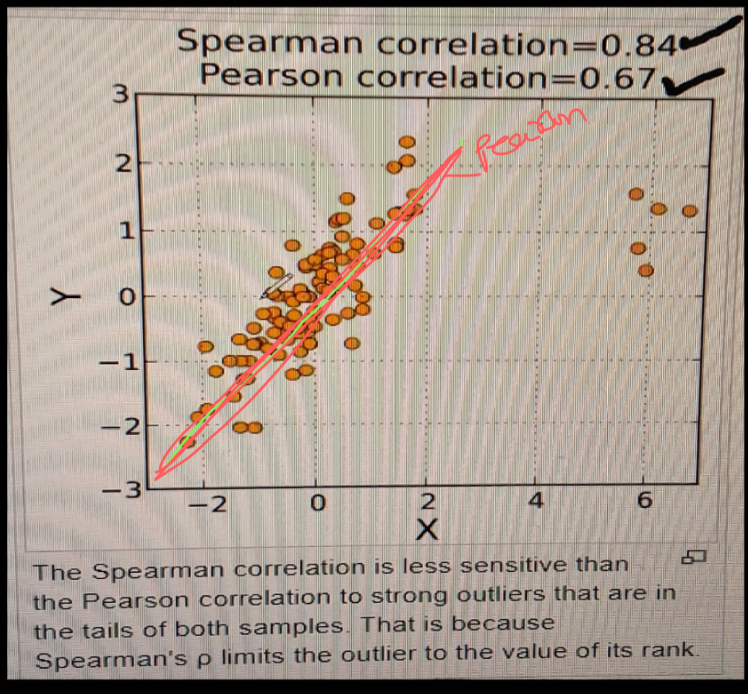


⑤ $\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$

If ρ is close to 0 → no cor

∴ 1 → highly co-

−1 → −ve cor−



The Spearman correlation is less sensitive than the Pearson correlation to strong outliers that are in the tails of both samples. That is because Spearman's ρ limits the outlier to the value of its rank.

$$\rho = \frac{cov(R_x, R_y)}{\sigma_x * \sigma_y}$$

Spearman rank Correlation

x	y	rank x	rank y	d_i	d_i^2

① sort 'x' in ascending order

② keep corresponding values in y.

③ assign 1, 2, 3 . . . rank for x.

↳ corresponding ranks to y.

④ d_i is diff of ranks(x) & ranks(y).

⑤ square d_i^2

→ Various ways to find outliers

① Using scatter plot

② Box plot

③ using z score

④ using the IQR - interquartile range.

→ 0 25 50 75 100

↑ ↑ ↑

Q_1 Q_2 Q_3

$IQR = Q_3 - Q_1$

lower bound = $Q_1 - 1.5\,IQR$

upper bound = $Q_3 + 1.5\,IQR$

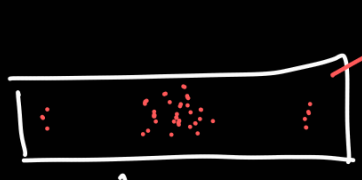
Normalization (min max normal)

↳ b/w (0,1)

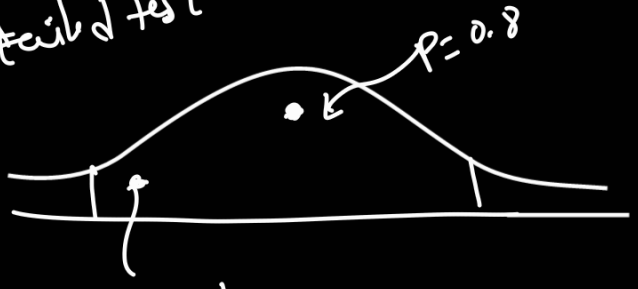
$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

---

p-value



2-tailed test



P=0.01

(1) if we repeat 100 times we will touch this place 1 time.

Penly

P=0.8 → if we repeate 100 times we will touch 80 times at the place

→ **Null hypothesis :**

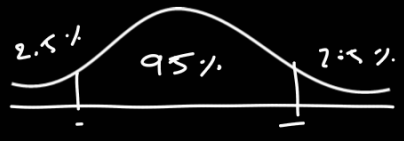
Treats everything same or equal.

→ **p-value :**

It is the probability for the "Null Hypothesis" to be true.

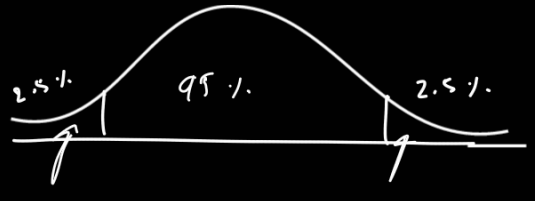
Confidence Intervals

① Point estimate : $\bar{x} \rightarrow \mu$



① Population std given ② Population std not given

① Average size of sharks in sea. with 95% CI.



2 tail test

lets assume : $\sigma = 100$ (if given) $n = 30$ $\bar{x} = 500$

$\mu \rightarrow 386 - 613$

CI = Point Estimate ± Margin Error

$= \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

$= 500 \pm z_{\frac{0.05}{2}} \frac{100}{\sqrt{30}}$

lower limit $= 500 - 1.96 \times \frac{100}{\sqrt{30}} = 386$

Higher limit $= 500 + 1.96 \times \frac{100}{\sqrt{30}} = 613$

α = significance value

$\alpha = 0.05$

$z_{0.025} = 1 - 0.025 = 0.975$

↳ 1.96

5 number summary

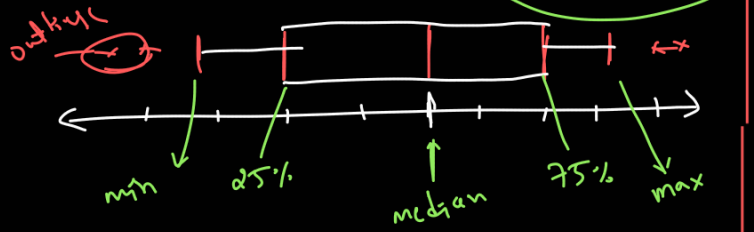
① Minimum ② 25% percentile Q_1

④ Median ③ 75% Q_3 ⑤ Max

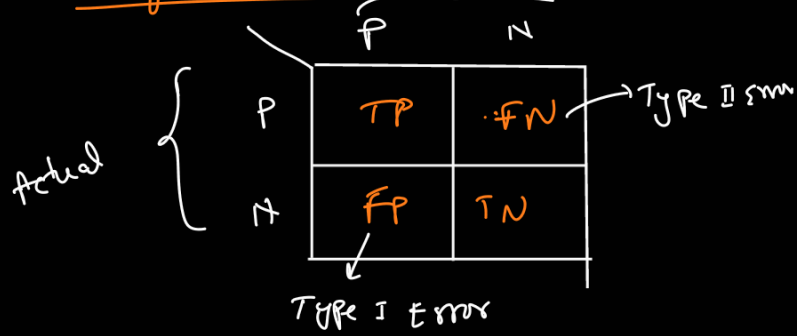
$IQR = Q_3 - Q_1$

[lower bracket ..... higher bracket]

$Q_1 - 1.5\,IQR$ $Q_3 + 1.5\,IQR$



Confusion matrix



① Whether the person has a disease or not.

→ for positive [cancer], if detects -ve, big problm.

So, we need to decrease (FN)

↳ there is no problm, if for -ve, gives +ve.

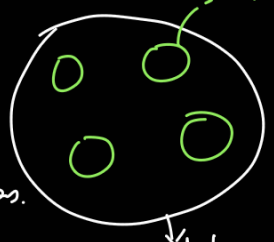
↳ so we adopt $\frac{TP}{TP+FP}$ → Precision

↳ TP ↑, TN ↑, (FN ↓↓↓), FP ↓

Q> What are Different sampling Techniques?

① Random Sampling Technique

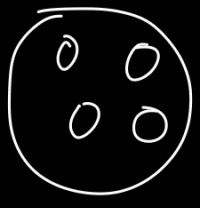
→ for exit polls, we are picking some random people from different areas. weather (boys/girls)



② stratified sampling Technique

→ if we pick random samples such that we take males : females = 1 : 1

↳ no bias



③ Systematic Sampling

⇒ if we select a particular location and if i select every "n" person of that place

④ cluster Sampling

↳ if I want to do a survey of AI ML ↳ only take survey of domain people.

② market will crash or not

↳ same, if market falls [true], & Predict -ve, so loose money.

↳ Also, if market don't fall, also loose money by selling.

↳ so, we need to focus on both FN ↓, FP ↓

$\frac{TP+TN}{TP+TN+FP+FN} = Accuracy$ & f score

↳ In this problm, both reducing & focusing on both FP & FN is imp.

③ Vaccination side effect.

→ if +ve, predicts (-ve) big Problm, so fight

→ also, we loose Vaccination if -ve, Predict +ve.

→ So, both FN ↓, , FP ↓

⇒ FP ↓ is imp → adopt $\frac{TP}{TP+FN}$ → Recall

↳ spam mail detection.

↳ it is ok, if some (-ve) mails not detected & predict (FN); [we can remain]

↳ if shouldn't remove imp mails [FP ↓↓↓]

Z-score importance

→ it is how many std away from mean

→ helps in standardization.

→ Helps to find the best performance blw the different groups of data.

→ Ex: compare the scores blw different distri

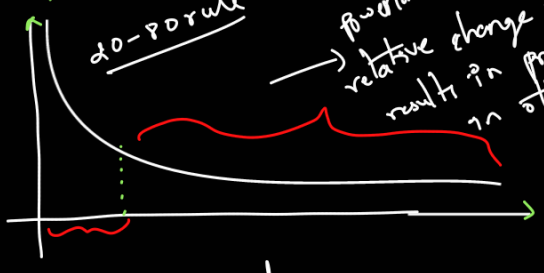
2020	2021
Avg = 181	Avg = 182
σ = 12	σ = 5
final score = 187	final = 185

$Z_{2020} = \frac{187-181}{12} = \frac{6}{12} = 0.5$

$Z_{2021} = \frac{185-182}{5} = \frac{3}{5} = 0.6$

→ So, 2021 performance is better.

Power law distribution

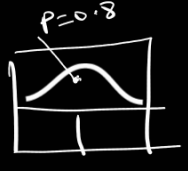


→ powerlaw is defined as a relative change in one quantity results in proportional change in other quantity.

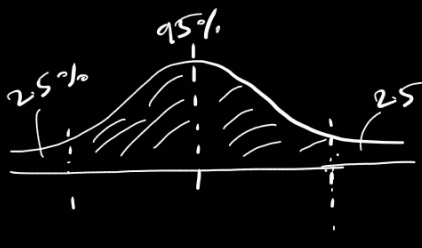
① 80% of sales are coming from 20% of products

② 80% of windows crash by 20% of bugs.

Hypothesis Testing - confidence Interval



out of 100 touches, I will touch 80 times there



→ α = Significance Value = 0.05

Hypothesis testing

→ ① Null hypothesis → Alternate hypothesis

② Perform Experiments → $\alpha = 0.05$ { z-test, t test, chi square, Anova }

③ Test value falls in extreme → Reject Null

→ Average height of people in india

↳ population (X) → samples → conduct exp → do statistical tests

Statistical test

① t test [Comparision of means] → sample (s)

② Z Test [Comparison of mean] → as npm

③ Anova Test (Analysis of Variance)

④ F Test (Comparision of Variance)

⑤ Chi square (comparision Categorical Variable)

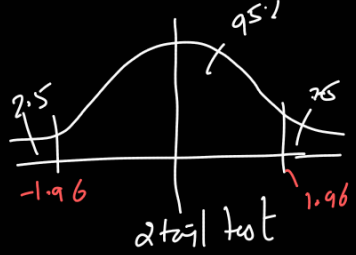
① One Sample Z test

→ In a population, the avg IQ = 100 & std is 15. The doctor tested the new medication whether it is or is the IQ. After 1 month sample of 30 people were taken, had a mean of $\bar{x} = 140$. Did medication effect intelligence? ($\alpha = 0.05$).

① Null hypothesis $H_0 \Rightarrow \mu = 100$ (didn't effect)
Alt hypothesis $\Rightarrow \mu \neq 100$ [effect].

② $\alpha = 0.05$

③ state decision rule →



if blw $\frac{-1.96 \text{ to } +1.96}{\text{accept}}$

if < -1.96 or > 1.96 } reject Null

$1 - 0.025 = 0.975$

$Z_{@0.975} = 1.96$

④ Z test : $Z = \frac{\bar{x} - M}{\frac{\sigma}{\sqrt{n}}}$

$= \frac{140-100}{\frac{15}{\sqrt{30}}} = \frac{40}{2.74} = 14.60 > 1.96$ → so reject H_0

* So, change in IQ

One Sample Z test with proportion

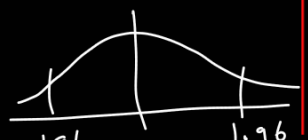
Q) A survey claims that 9 out of 10 doctors recommend aspirin for patients with headache. To test this claim, random sample of 100 doctors taken out of 82 recommends this Aspirin. Is this claim Accurate? $\alpha = 0.05$.

→ 9 out 10 → proportion.

→ ① H_0 Null Hypothesis $= p = 0.90$ [9/10]

$H_1 \neq 0.90$.

② Significance $\alpha = 0.05$



2 tail test.

③ Decision Rule.

④ Z-statistics

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$\hat{p}$ = probability wrt sample.

$\hat{p} = 0.82$

$p_0 = 0.9$ → [9 out of 10]

$n = 100$ → sample.

$$Z_0 = \frac{0.82 - 0.90}{\sqrt{\frac{0.90(0.10)}{100}}}$$

$= \boxed{-2.667}$ — reject H_0

⑤ Make conclusion:

$-2.667 \neq [-1.96, 1.96]$

Reject the Null hypothesis
Accept the Alt hypothesis

* CLAIM IS inaccurate

Probability

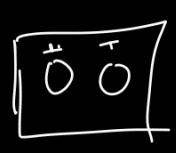
It is the measure of likelyhood of an event

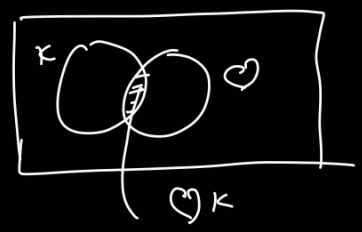
→ Eg: Tossing a coin

→ Probability = $\frac{\#\text{ of ways an event can occur}}{\#\text{ of possible outcomes.}}$

① Mutual Exclusive → tossing a coin H/T

② Not mutual Exclusive → Taking out a card from deck (K ♡)





→ Probability of (H or T) $= \frac{1}{2} + \frac{1}{2} = 1$

↳ Additive rule for mutual exclusive

→ $P(K \text{ or } ♡) = P(K) + P(♡) - P(K \cap ♡)$

$= \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52}$

Additive

ME = $P(A \text{ or } B) = P(A) + P(B)$

NME = $P(A \text{ or } B) = P(A) + P(B) - P(A \cap B)$

Multiplicative Rule

① Dependent Event

H T H

$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$

② Independent event

deck of cards

$\frac{1}{52} \times \frac{1}{51} \times \frac{1}{50}$

(conditional prob)

$P(K \cap ♡) \Rightarrow P(K) \times P(♡|K)$

$= \frac{1}{52} \times \frac{1}{51}$

Permutation & combination

6 people.

$6 \times 5 \times 4 = 120$

Roc Raj Aru Kum Chin Cut

↳ order matters

permutation

$$^nP_r = \frac{n!}{(n-r)!} = \frac{6!}{(6-3)!} = \frac{6 \times 5 \times 4 \times 3!}{3!} = 120$$

~~Combinations~~ → order don't matters. (Same combination)

$$^nC_r = \frac{n!}{(n-r)!\,r!} = \frac{6!}{(6-3)!\,3!} = \frac{6 \times 5 \times 4 \times 3!}{3! \times 3!}$$

$= 20$

Why Sample Variance is divided by n-1

Population (N)	Sample (n)
$\mu = \sum_{i=1}^{N} \frac{x_i}{N}$	$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$
Variance $\sigma^2 = \sum_{i=1}^{N} \frac{(x_i - \mu_x)^2}{N}$	$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$

→ If we select only sample which is extra so, our (s^2) & $(\bar{x})$ will vary much.

So, scientists tested $n-1$, $n-2$, ... & got good result with $(n-1)$ → $\sigma^2 \approx s^2$

↳ skewed data → unbiased estimation.

#Variables

Random variables
- Numerical
  - Discrete
  - cts
- Categorical

(*) Variable Measurement Scales :

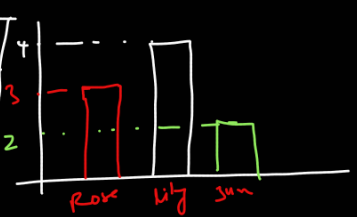
4 types of measurement variable

1. Nominal → {categorical data} colors, gender, type of flowers.
2. Ordinal → order of data matters, value doesn't
3. Interval → order & value matters, '0' not present
4. Ratio →

#Frequency distribution

→ flowers {Rose, lily, Rose, lily, sunflower. .....}

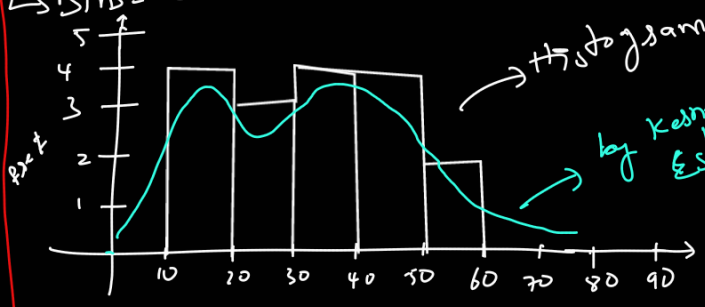
flower	freq	cumulative
Rose	3	3
lily	4	7
sunfl.	2	9



#for cts data → Histogram.

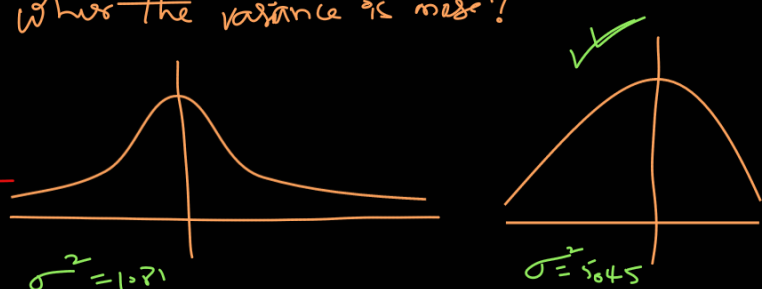
Ages = {10, 12, 14, 18, 24, 26, 30, 35, 36, 37, 40, 41 42, 43, 60, 512}

↳ Bins = 10



→ pdf → smoothing of histogram.

Where the variance is more?



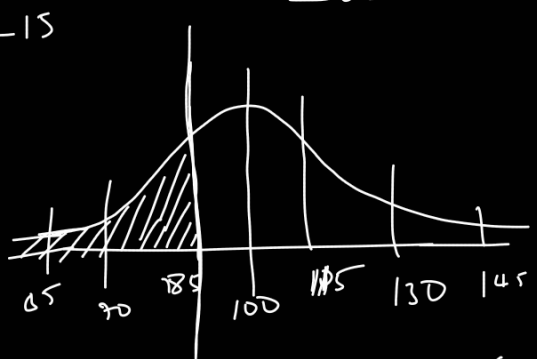
Variance → spread of data.

how much %.

$\mu = 100$

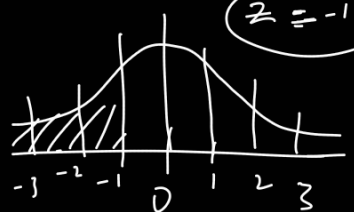
$\sigma = 15$

< 85



$$z = \frac{85 - 100}{15} = \frac{-15}{15} = -1$$

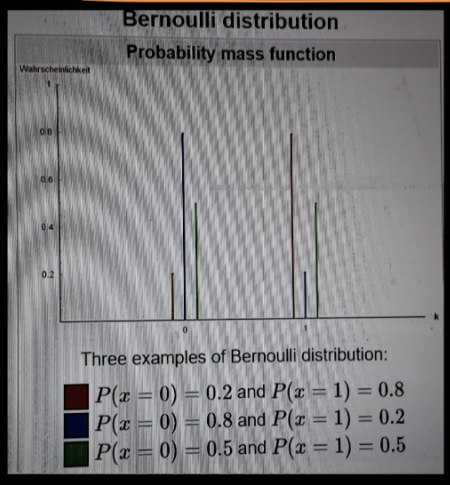
$z = -1$



$z_{@-1} = 0.15$ %

15.86 %

Bernoulli distribution



→ Outcome → 2

Eg: Tossing a coin → H, T

$x = 1 \Rightarrow$ success

$x = 0 \Rightarrow$ fail (↳ T)

$P(x=x) = p^x (1-p)^{1-x}$

This is called [pmf] ↓ outcome discrete

gd in Pdf outcome → cts

$P(\text{success}) = p$, $P(\text{fail}) = 1-p = q$

$P(x=0) = p^0 (1-p)^{1-0} = (1-p) = q$

$P(x=1) = p^1 (1-p)^{1-1} = p$

$$PMF = \begin{cases} q = 1-p & \text{if } x=0 \\ p & \text{if } x=1 \end{cases}$$

$p^x (1-p)^{1-x} \Rightarrow$ pmf

Mean

$$E(x) = \sum_{i=1}^{x} x_i P(x)$$

$x = 0$ or 1

$P(x=0) = 0.4 = 1-p$

$P(x=1) = 0.6$ ↳ p

$= 0 \times (0.4) + 1 \times (0.6)$

$= 0.6 \rightarrow p$

⇒ Variance: $P(1-P) = pq$

$\sigma = SD = \sqrt{pq}$

Median Mode.

Eg: Colleges in Karnataka have 85% placement rate. A new college was opened and it was found that a sample of 150 students had a placement rate of 88% with a std 4%. Does this clg has a diff placement rate? $\alpha = 0.05$

2 tail test.

① Null hypothesis (H_0): It has diff placement rate

② Alt Hyp (H_1): It has some placement rate

③ Significance value = 0.05

④ Decision Rule: $Z = \frac{\bar{x} - M}{\frac{\sigma}{\sqrt{n}}} = \frac{88-85}{\frac{4}{\sqrt{150}}} = \frac{3}{4} \times \sqrt{150} = 9.18$

→ 9.18 falls outside of interval (-1.96, 1.96)

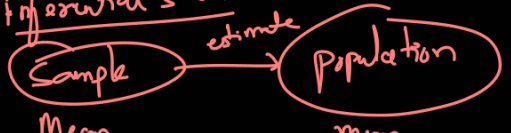
→ We reject the null hypothesis

* So, the clg don't have diff placement rate.

Point Estimate:

The value of any statistic that estimates the value of a parameter.

Inferential stats



$\bar{x}$ → μ

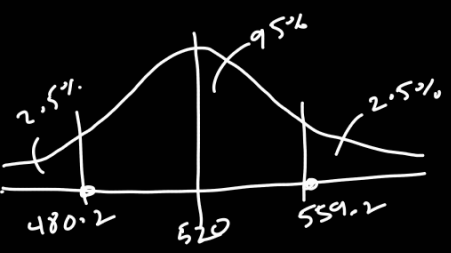
$\bar{x} = 2.9$ $M = 3$

→ Confidence Intervals

Point Estimate ± Margin of error

Q) On the quant test of CAT Exam, the std. deviation is known to be 100. A sample of 25 test has a mean of 520 score. Construct a 95% CI about the mean?

→ $\sigma = 100$, $\bar{x} = 520$, $n = 25$, $\alpha = 0.05$



→ CI = point estimate ± margin of error.

→ population std given

→ $\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

std error.

→ this formula usually use when $n \geq 30$

upper bound $= \bar{x} + z_{\frac{0.05}{2}} \frac{100}{\sqrt{25}}$

lower bound $= \bar{x} - z_{0.025} \frac{100}{\sqrt{25}}$

$z_{@0.975} = 1.96$

UB = $520 + 1.96(20) = 559.2$

LB = $520 - 1.96(20) = 480.2$

Q> On the quant test of CAT exam, a sample of 25 test takers has a mean of 520 with a std deviation of 80. Construct 95% CI about the mean?

$n = 25$, $\bar{x} = 520$ $s = 80$

Here population std is not given → t-test

$CI = \bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$

→ degree of freedom = $n - 1 = 25 - 1 = 24$

$t_{0.025} = 2.064$

UB = $520 + 2.064 \left(\frac{80}{5}\right)$

$= 553.024$

LB = $\bar{x} - 2.064 \left(\frac{80}{5}\right)$

$= 520 - 2.064 \left(\frac{80}{5}\right)$

$= 486.97$

[486.97 ⟷ 553.024]

Chi Square Test

① Chi square Test claims abt population proportions.

→ It is a parametric test that is performed on categorical, [nominal or ordinal] data.

Q> In the 2000 Indian Census, the age of the individual in a small town were found to be the following:

Less than 18	18 - 35	>35
20%	30%	50%

In 2010, ages of $n = 500$ individuals were sampled. Below are the results:

<18	18-35	>35
121	288	91

Using $\alpha = 0.05$, would you conclude the population distribution of ages has changed in the last 10 years?

→ Here proportion of population is given — chi²

Ans>

<18	18-35	>35
90%	30%	50%

population 2000

<18	18 - 35	>35	n: 500
121	288	91	Observed
$\frac{20\% \times 500}{100}$	$\frac{30\% \times 500}{100}$	$\frac{50\% \times 500}{100}$	Expected

①
H_0 = The data meets the distribution 2000 data

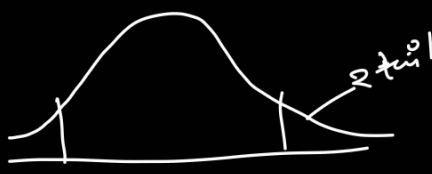
H_1 = The data doesn't meet . . . . .

② $\alpha = 0.05$ [95% CI]

③ Degree of freedom = $n - 1 = 3 - 1 = 2$ (3 categories)

→ df = 2, $\alpha = 0.05$, chi² table

→ ④ Decision Boundary



→ χ^2 is greater than 5.99, reject H_0

⑤ Calculate Test statistics

$$\chi^2 = \sum \frac{(f_o - f_c)^2}{f_c}$$

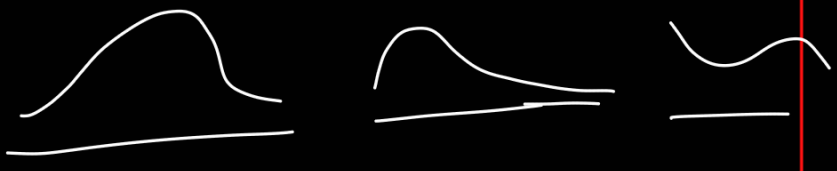
$$= \frac{(121-100)^2}{100} + \frac{(288-150)^2}{150} + \frac{(91-200)^2}{200}$$

$$= 232.494$$

$\chi^2 = 232.494 \leq 5.99$ { reject null Hyp }

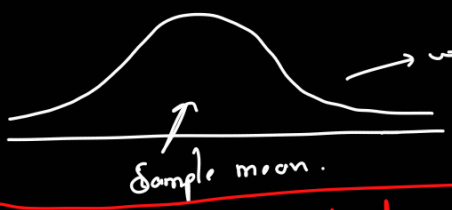
Central Limit theorem

→ Any kind of distribution



→ $n \geq 30$ — $n \geq 30$ — sample size.

$S_1 \rightarrow \bar{x}_1$
$S_2 \rightarrow \bar{x}_2$ } sample mean
$S_3 \rightarrow \bar{x}_3$
⋮
$S_m \rightarrow \bar{x}_m$ — more bigger 'm' more CLT holds good



→ with $\bar{x}_1, \bar{x}_2 \ldots \bar{x}_m$.

Poisson distribution :

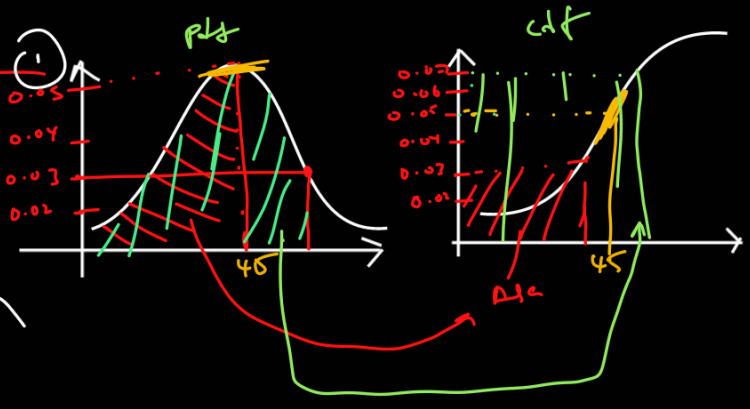
Imp Stats Qn

→ Pdf Vs cdf Vs Pmf

→ Pdf → Probability density function

→ cdf → Cummulative distribution fn

→ Pmf → Probability mass fn.



Pdf ⇒ is the derivative of a cdf

② Pmf :- Discrete Random Variable.

Eg: Tossing a coin
Rolling a dice