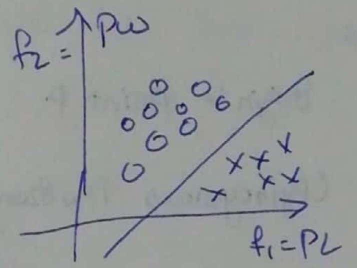
Linear Algebra

Why lear Linear Algebra?

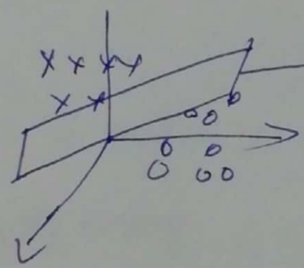
-> Suppose we have 2 dimension

-> To distinguish these 2 dimension we will draw a line to separate two dimension

-> In 1D we have a line to separate



-> In 2D we have a plane to separate



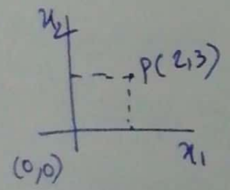
linear surface to separate

-> what about 4D, 5D, 10D, n-D

-> we cannot visualize the data in the higher dimensions space

-> If I have 10 features dimension, It means I am operating at 10 dimensional space.

Point / vector:

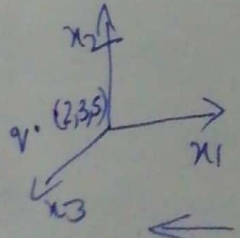


← 2dim

I can represent a point as a vector P = [2, 3]

x_2 Component of vector

x_1 Component vector



← 3dim. point

I can represent my q with a vector of size 3

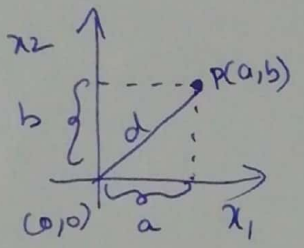
q = [2, 3, 5]

What & how to represent n-dim point

Suppose i have a point x i can represent n dimension point in a n dimensional space

$$x = [2, 3, 4, 1, 5 \dots]$$

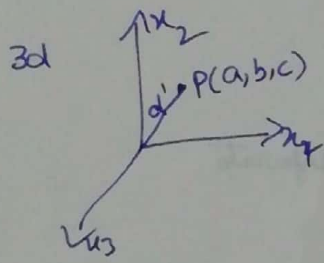
What is the distance of a point from origin?



d = distance b/w origin & point p.

$d = \sqrt{a^2 + b^2}$ (Pythogenous Theorem)

3d



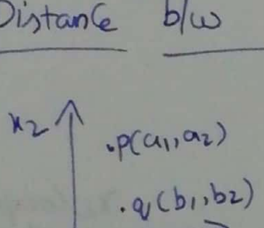
$$d' = \sqrt{a^2 + b^2 + c^2}$$

what about nd ?

$$d = \sqrt{a_1^2 + a_2^2 + \dots a_n^2}$$

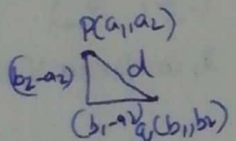
If my point P is $[a_1, a_2, \dots a_n]$

Distance b/w Two points



To find distance b/w Two points P & q

$$d = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$



Similary in 3D

$P(a_1, a_2, a_3)$

$q(b_1, b_2, b_3)$

$$d = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2}$$

Similarly in nD space.

$P(a_1, a_2, \ldots a_n)$

$P(b_1, b_2 \ldots b_n)$

$$\boxed{d = \mathrm{sqrt}\left(\sum_{i=1}^{n} (a_i - b_i)^2\right)}$$

Two Concepts

row vector

I have a vector $A = [a_1, a_2, a_3 \ldots a_n]_{1\times n}$

If I write like this we have 1 rows & n column,

$A_{1\times n}$

Column vector

Suppose I have a vector $B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}_{n\times 1}$

Suppose If I have a matrix A as $A_{m\times n}$

matrix is array of array.

$$A = \begin{array}{c|c} & 1 \quad 2 \quad 3 \ \ldots \ n \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ \vdots \\ m \end{array} & \begin{bmatrix} \quad & & & \\ & & & \\ & & & \\ & & & \\ & & & \end{bmatrix}_{m\times n} \end{array}$$

10.4 Dot product & Angle b/w 2 vectors

$a = [a_1\, a_2 \ldots a_n]$

$b = [b_1, b_2 \ldots b_n]$

$c = a+b = [a_1+b_1, a_2+b_2 \ldots a_n+b_n]$

multiplication

↳ dot product

↳ cross product (we dont use much in machine learning)

Dot product

$a \cdot b = a_1b_1 + a_2b_2 + \cdots + a_nb_n$

If

If we write it in Vector notation

$$a \cdot b = \underset{1 \times n}{[a_1, a_2, a_3 \cdots a_n]} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}_{n \times 1}$$

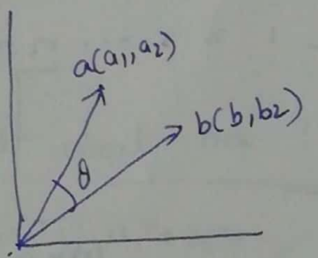
when somebody says it as a vector it is a Column vector by default

$$a \cdot b = a_1 b_1 + a_2 b_2 \cdots + a_n b_n.$$

Suppose if I have a Column vector.

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \qquad a^T = [a_1 \ a_2 \cdots a_n]$$

we can write $a \cdot b$ as $a^T b = \sum_{i=1}^{n} a_i b_i$



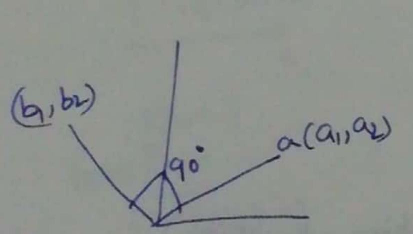
$$a \cdot b = \underset{\text{length of a}}{\|a\|} \ \underset{\text{length of b}}{\|b\|} \cos\theta$$ → cos of angle b/w them

$$a \cdot b = a_1 b_1 + a_2 b_2 = \|a\| \, \|b\| \cos\theta$$

$$\theta = \cos^{-1}\left\{\frac{a_1 b_1 + a_2 b_2}{\|a\| \ \|b\|}\right\}$$

$$\theta = \cos^{-1}\left\{\frac{a_1 b_1 + a_2 b_2}{\sqrt{a_1^2 + a_2^2} + \sqrt{b_1^2 + b_2^2}}\right\}$$

what if I have a two vectors, they are perpendicular to each other.



$$a \cdot b = \|a\| \ \|b\| \cos 90^\circ \qquad \text{i.e., } \cos 90 = 0$$

$$a \cdot b = 0$$

If the dot product b/w two vectors are zero, then those two vectors are perpendicular to each other.

A can be any dimensional vector

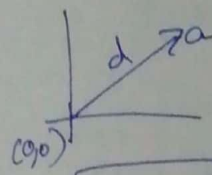
$a = [a_1, a_2, \cdots a_n]$

$b = [b_1, b_2, \cdots b_n]$

$a \cdot b = \|a\| \|b\| \cos\theta$

$\theta = \cos^{-1}\left(\frac{a \cdot b}{\|a\| \|b\|}\right)$

what about $a \cdot a = a_1 a_1 + a_2 a_2 + \cdots a_n a_n$



$= a_1^2 + a_2^2 + a_n^2$

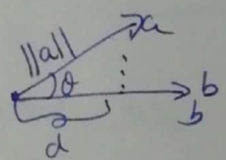
$= \|a\|^2$ ⟸ distance from origin squared.

$d = \sqrt{a_1^2 + a_2^2 + \cdots a_n^2}$

10.5 : Projection & unit vector

Projection

suppose we have a vector a and b, angle b/w them is θ



i.e., let vector a fall on to vector B, & the distance we get on B is the projection of A on B.

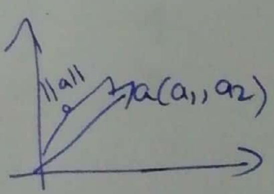
$d = \|a\| \cos\theta \rightarrow (1)$

$a \cdot b = \sum_{i=1}^{n} a_i b_i = \|a\| \|b\| \cos\theta$

$d = \frac{a \cdot b}{\|b\|} = \frac{\|a\| \|b\| \cos\theta}{\|b\|} = \|a\| \cos\theta.$

without known θ, if I know point $a(a_1, b_1)$ & $b(a_2, b_2)$ we can calculate the projection of A and B.

Unit Vector



A unit vector is represented with a "^"

$\hat{a} = \frac{a}{\|a\|}$

$\hat{a}$ is in the same direction as vector a

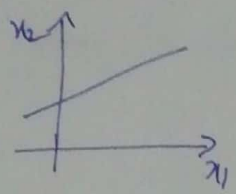
Since we are dividing vector with length of a ($\|a\|$)

The length of unit vector is 1 i.e $\|\hat{a}\| = 1$

10.6 Equation of a line

Line

→ 2D line



$$y = \underbrace{m}_{\text{slope}}x + \underbrace{c}_{\text{intercept}}$$

Other Equation or general form

$ax + by + c = 0$

(or)

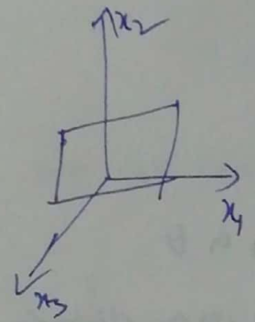
$y = \frac{-c}{b} - \frac{a}{b}x$

$ax_1 + bx_2 + c = 0$

$w_1x_1 + w_2x_2 + w_0 = 0$ → 2D

In 2D line is a linear surface

→ 3D line.



In 3D line is a plane

$w_1x_1 + w_2x_2 + w_3x_3 + w_0 = 0$ → 3D

→ nD line

In nD line is a hyper plane

~~nD~~ line

$w_1x_1 + w_2x_2 + w_3x_3 + \dots w_nx_n + w_0 = 0$

Is there a more concise way of representing this

Summation notation → $w_0 + \sum_{i=1}^{n} w_ix_i = 0$

Vector notation → $w_0 + \underbrace{[w_1\ w_2\ w_3 \dots w_n]}_{\text{Vector } w_{1\times n}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\text{Vector } x_{n\times 1}} = 0$

$w_0 + w_1x_1 + w_2x_2 + w_3x_3 \dots w_nx_n = 0$

Just a while ago we realized that the eq of a line in any n dimensional space is.

$$w_0 + [w_1, w_2 \cdots w_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = 0$$

If I say a vector w, it is by default is Column vector.

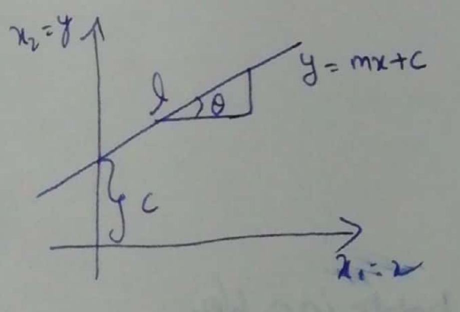
$$w_{n \times 1} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_n \end{bmatrix} \qquad x_{n \times 1} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$w_0 + w^T x = 0$$

Planes are typically written as π

$$\boxed{\pi_n = w_0 + w^T x = 0}$$

What w_0 is ?



Here c is called y intercept

on y axis where does the line 'l' intersect the y-axis

2D: $w_1 x_1 + w_2 x_2 + w_0 = 0$

$$x_2 = -\frac{w_0}{w_2} - \frac{w_1}{w_2} x_1$$

$$\uparrow$$

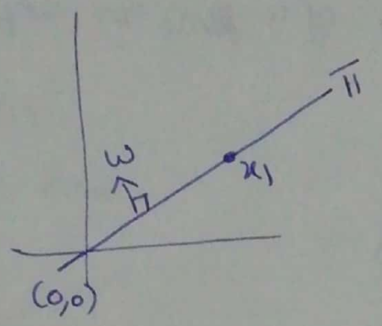
$$y \qquad c + mx$$

If a line is passing through origin

$c = 0$ & $w_0 = 0$

$$w \cdot x = w^T x = \|w\| \|x\| \cos \theta_{w,x} = 0$$

If $w \perp x \Rightarrow \theta_{w,x} = 90^\circ$



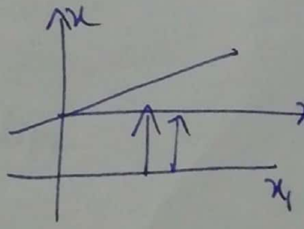
ω is a vector & perpendicular to plane, since ω & x_1 are perpendicular

$\omega \cdot x_i = 0$

If $\omega \perp \pi$ then $\omega \cdot x_i = 0$ for all $x_i \in \pi$

Vector ω is perpendicular to plane.

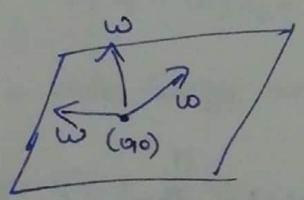
$$\hat{\omega} = \frac{\omega}{\|\omega\|}$$



$\omega^T x + \omega_0 = 0$

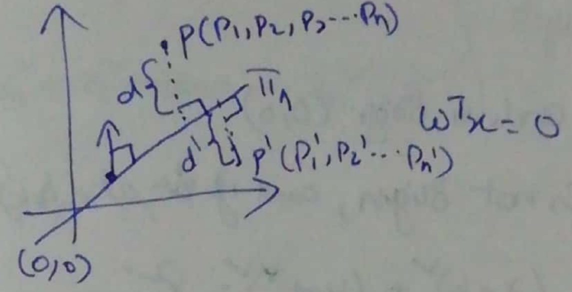
Shifting my axis

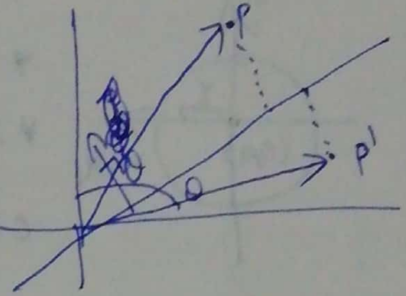
<u>In 3D</u>



It is perpendicular all the points in a plane

10.7 Distance of a point from a plane/hyperplane, Half-spaces.





$w_{n\times1}$, $P_{n\times1}$

$$d = \frac{w^T p}{\|w\|}$$

$\boxed{\text{Exercise: Proof}}$

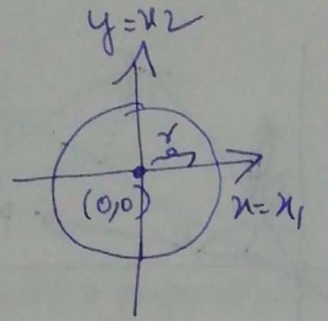
If w is a unit vector, the length of $w = 1$

Since w & P are same side of the plane.

The region above the line and below the line are half spaces.

10.8

Circle



* Center of circle is origin $(0,0)$
* If center is not origin, and if at (h,k)

$c:(h,k) = (x-h)^2 + (y-k)^2 = r^2$

$x^2+y^2=0$ if origin at $(0,0)$

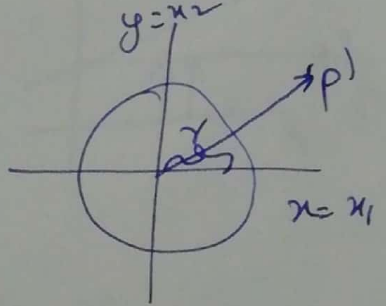
How to find a point is inside or outside the circle.

$P(x_1,x_2) = x_1^2 + x_2^2 < r^2 \Rightarrow P$ lies inside the circle

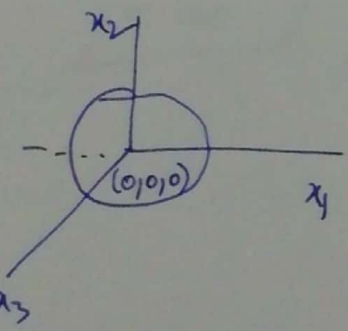
$x_1^2 + x_2^2 > r^2 \Rightarrow P$ lies outside the circle

$x_1^2 + x_2^2 = r^2 \rightarrow P$ lies on the circle.

2D



3D: x_1, x_2, x_3 Sphere



$x_1^2 + x_2^2 + x_3^2 = r^2$

nD: $x_1, x_2, \dots x_n$

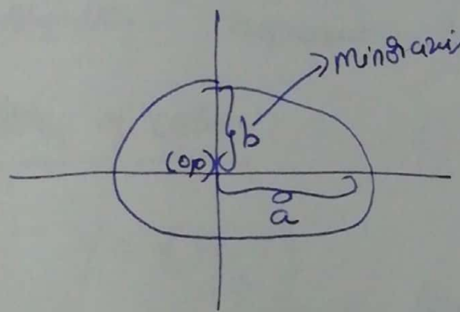
$$\left.\begin{array}{l} x_1^2 + x_2^2 + \cancel{x_3} + \dots x_n^2 = r^2 \end{array}\right\}$$ Equation of a hyper sphere.

$$\sum_{i=1}^{n} x_i^2 = r^2$$

$\sum x_i^2 < r^2$ then given point $P_i = (x_1, x_2 \dots x_n)$ lies inside hyper's sphere

10.9

ellipse: Egg shape. (we dont use ellipse much in ML)



$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

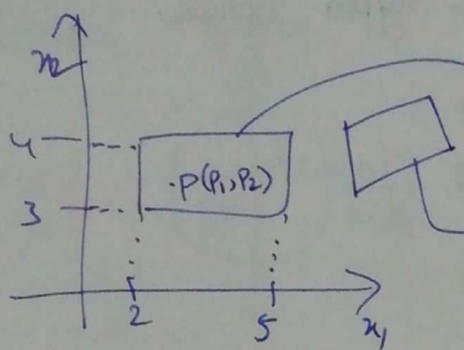
$\frac{x_1^2}{a^2} + \frac{y_1^2}{b^2} < 1$ inside Ellipse

> 1 outside

$= 1$ on

Ellipsoid : 3 dimensional

10.10 Square & Rectangle



Sides of these rectangle is parallel to both axis. ~~this is~~ This is called axis parallel rectangle

Non-axis parallel rectangle

If $P_1 < 5$ & $P_1 > 2$
if $P_2 > 3$ & $P_2 < 4$
} P lies inside this rectangle.

Probability And Statistics

11.1 Introduction to Probability & Statistics :-

-> fundamental Area

-> Histogram, PDF, CDF, mean, Var, std-dev

Random Variable

dice : Six sido -> $\underbrace{\{1, 2, 3, 4, 5, 6\}}$.

Equally likely can occur.

If i roll a dice

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

Experiment (or) rolling a dic

* random variable are represent in Capital letter

tossing of coin

$Y = \{H, T\}$

Equally likely

rolling a dice Experiment

X = random variable which take values $\{1, 2, 3, 4, 5, 6\}$

$P(X=1) = 1/6 = P(x=2)$

$P(x \text{ is a Even number}) = 3/6 = 1/2 = \overset{1/6}{P(x=2)} + \overset{1/6}{P(x=4)} + \overset{1/6}{P(x=6)}$

$P(x \text{ is odd}) = 1/2$

$P(X = x_1)$ => Random variable X takes a value of x_1

(or)

$P(x_i)$

height of a randomly picked student

$Y = 162.45$ cm.

$Y = 132.62$ cm

* A random variable (X) which can take a value from a finite set of values is called discrete random variable

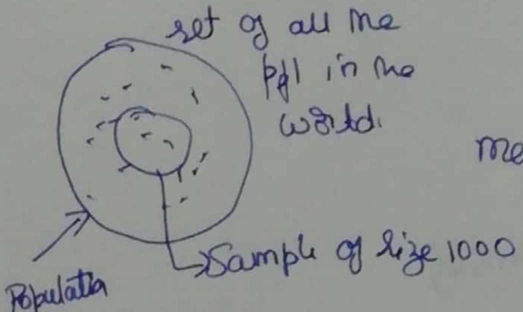
* Random variab Y can take any real value is called continuous random variable.

<u>Outlier:-</u>

Y : height of a student randomly chosen.

$\{122.2, 146.4, 132.5, \ldots \boxed{12.26}, 156.23 \ldots\}$
outliers

Population & Sample



→ Estimate the average/mean height of a human.

mean of ppl ← $\mu = \sum_{i=1}^{7B} h_i \times \frac{1}{7B}$

→ Since we cannot the measure heights in the whole population, we will collect a random sample with size thousand

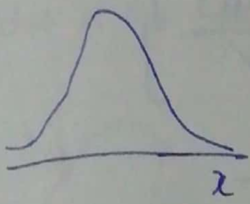
$\bar{h} = \frac{1}{1000} \times \sum_{i=1}^{1000} h_i$ ← heights in my sample.

⇒ we should take care the random sample should contain ppl from every country in equal proportions

For Ex: 1/6th of the world population is in india, so sample should also contain 160 ppl

As sample size increase sample mean will be equal to population mean.

11.3



PDF of a Gaussian distribution of a r.v

X: Continuous r.v (which has a bell shape)

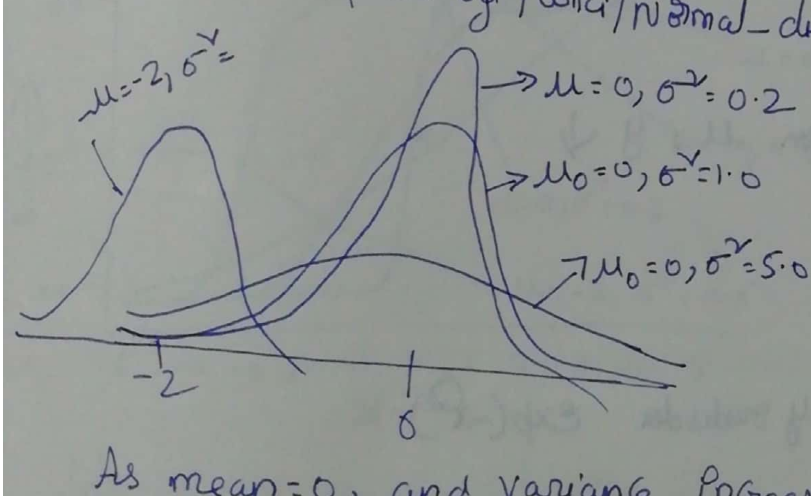
X: has a distribution which is Gaussian distribution.

→ most of the things follow Gaussian distributions.
- → height
- → weight

~~Why distribution~~

→ distributions are simple models

// https://en.wikipedia.org/wiki/Normal_distribution.

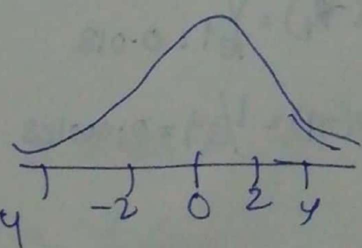


As mean=0, and variance increase the PDF fatter and shorter
↑
measure of spread or scale

Parameters of a Gaussian distribution: μ, σ^2

Normal or Gaussian

$X \sim N(\mu, \sigma^2) \Rightarrow X \sim N(0, 2)$
↑
follows



If we X is normally distributed with μ, σ^2

$X \sim N(\mu, \sigma^2)$

$$P(X = \underbrace{x}_{\text{value}(=160)}) = P(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ \frac{-(x-\mu)^2}{2\sigma^2} \right\}$$

r.v

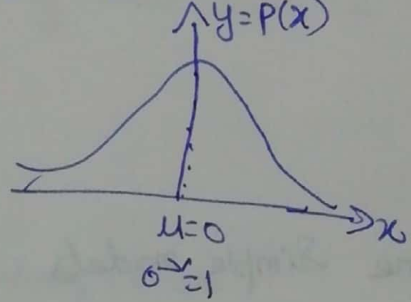
$\text{Exp}(x) = e^x$

Let us assume $\mu_0 = 0, \sigma^2 = 1, \sigma = 1$

$$P(x) = \boxed{\frac{1}{\sqrt{2\pi}(1)}} \exp\left\{ \boxed{-\frac{1}{2}} x^2 \right\} = y$$

Constant

$P(x) = \exp(-x^2)$



→ As x increase $-x^2$ will reduce, Exp of $-x^2$ will also reduce.

Conclusion

1. x ~~means~~ moves away from μ, $y \downarrow$
2. Symmetric
3. ~~Expo reducing Exponentially~~
3. x moves away from μ, y reduces $\text{Exp}(-x^2)$

$y = \exp(-x^2)$

$x = 0, \; y = 1$

$x = 1, \; y = \text{Exp}(-1) = 1/e^1 = 0.3678$

$x = 2, \; y = \text{Exp}(-4) = 1/e^4 = 0.018$

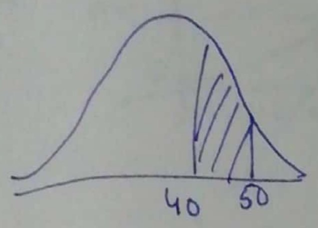
$x = 3, \; y = \text{Exp}(-9) = 1/e^9 = 0.000123$

Ex:- Titanic dataset

Assuming the age is following the normal distribution.

~~pnorm~~

what is the probability of finding a person b/w the age of 40 & 50



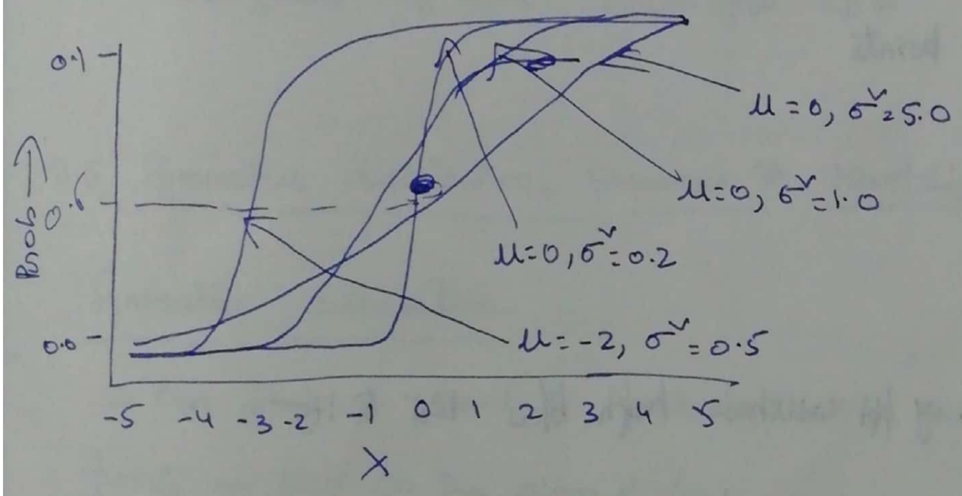
Pnorm(50, mean = mean(titanic age), sd = sd (Age)) −

Pnorm(40, " ")

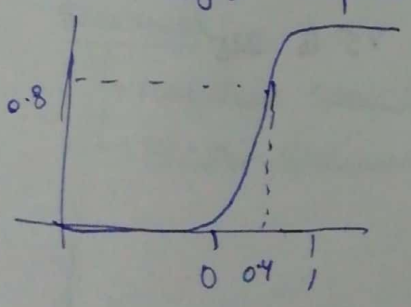
CDF of Gaussian/ normal distribution?

Shape of CDF

CDF range from 0 to 1



CDF height represent $P(X \leq 0.4)$

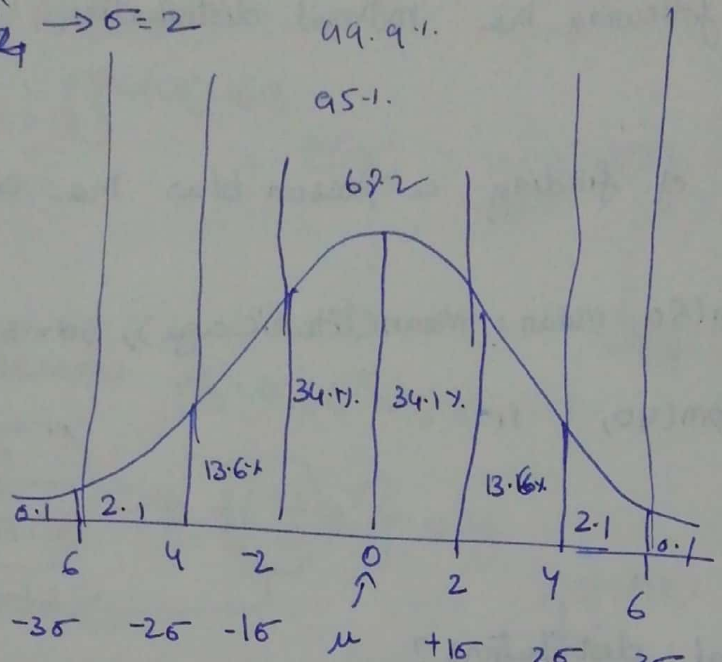


If μ of RV is 0 the central point is 0 which is 0.5

If variance is less, the CDF will be closer to the '0' line

$X \sim N(\mu, \sigma^2)$

0   4 $\rightarrow \sigma = 2$



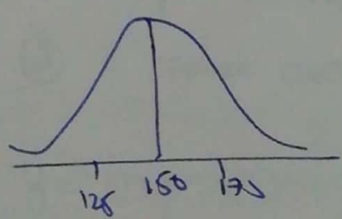
-1σ to 1σ is can find 68% of

-2σ to +2σ there 95% of points

-3σ to +3σ 99.7% of points

Ex:-

$X \sim N(150, \sigma = 25)$ (μ = 150)

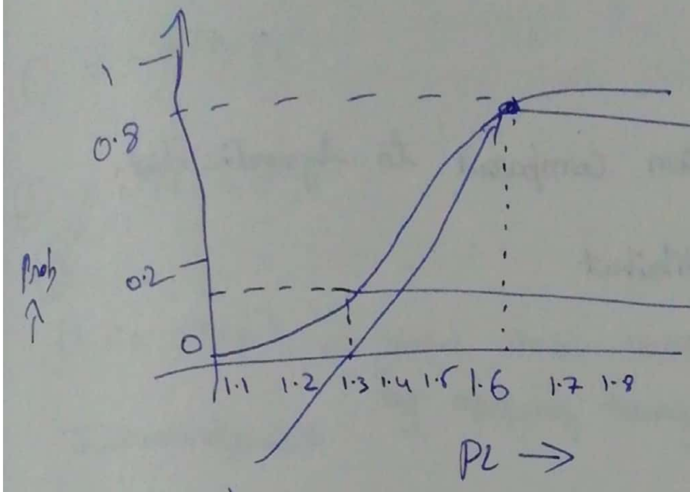


68% of ppl will have heigh b/w 125 to 175

95% of ppl will have height b/w 100 to 200

99.5% of ppl will have height b/w 75 to 225

Cummulative distribution Function.



→ There are 82% of Setosa flower that have petal length < 1.6

→ 15% of the flower that have PL of < 1.3.

How this calculated.

$$\frac{41}{50} = \frac{\text{total \# of flower that have } PL \le 1.6}{\text{total \# of flower}} = 0.82 = 82\%.$$

→ differentiation of CDF will give Pdf

Integration of PDF will give CDF

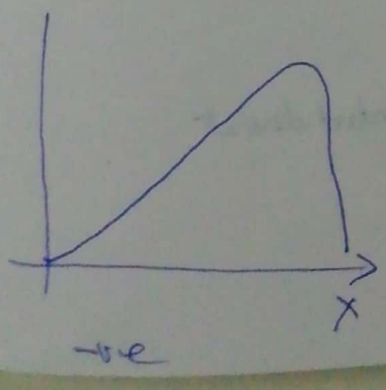
11.5 Symmetric distribution, Skewness & Kurtosis

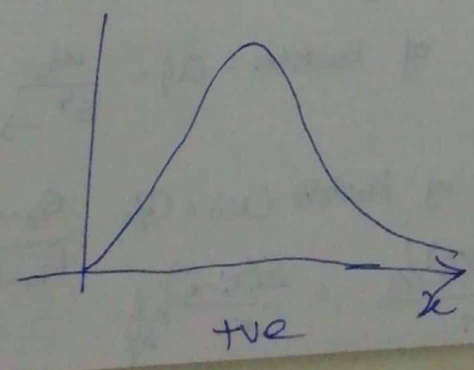
Symmetric distribution

I can find a point x such that one side we have 50% of points & 50% on the other side

Skewness

- negative Skew (Left Skewed distribution)
- Positive Skew. (Right Skewed distributa)



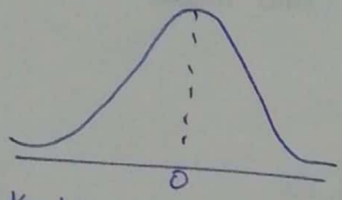


$\mu \rightarrow$ Tells us the Center point

$\sigma^2 \rightarrow$ Tells us the spread.

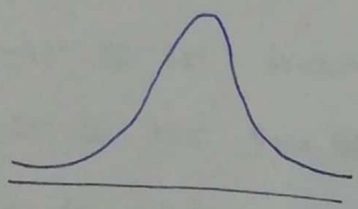
Skewness $\rightarrow$ How dissimilar is distribution compared to symmetric dist.

Kurtosis $\rightarrow$ measure how peaked the distribut.



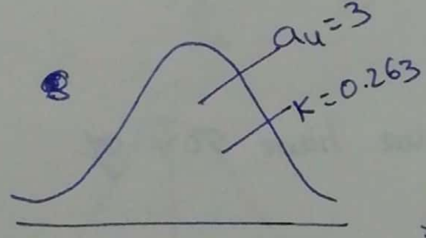
Kurtosis of Gaussian dist is '0'

$X \rightarrow$ kurtosis $(x) = 2$

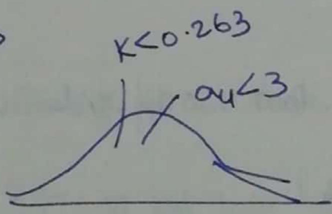


$Y \rightarrow$ kurtosis $= 0.5$

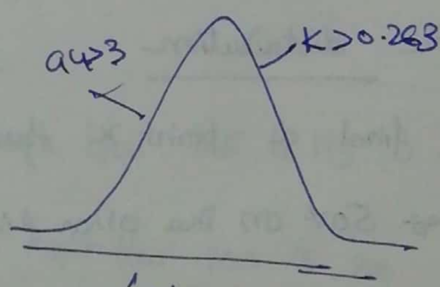
If kurtosis is -ve means, distribution is less peaked when compared to Gaussian dist.



Measo Kurtic



Platy Kurtic



Leptokurtic

$\rightarrow$ Values are bundled in Laptokurtic, and dispersed in measokurtic and Even more dispersed in Platy kurtic.

moment co-efficient of kurtosis $= a_4 = \dfrac{\mu_4}{s^4}$ $\rightarrow$ Standard deviat

Percentail Co-efficent of kurtosis $(K) = \dfrac{Q_3 - Q_1}{P_{90} - P_{10}}$

$$a^4 = \frac{\mu_4}{s_4} = \frac{\Sigma(x_i - \bar{x})^4 / n}{s^4} = \frac{\Sigma(x_i - \bar{x})}{n} \times \frac{1}{s^4}$$

Standard normal Distribution. (Z score)

(Column Standardization)

① $Z \sim N(0, 1)$

② $X \sim N(\mu, \sigma^2)$
PDF

$[x_1, x_2 \cdots x_{50}]$ → given these value, we can standardize these values by applying transformation function.

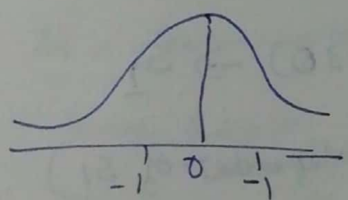
~~To standardize eg~~

Standardization:

$$x_i' = \frac{x_i - \mu}{\sigma} \quad \text{(Z-score)}$$

$x_i' \sim N(0, 1)$
↑
standard normal variate.

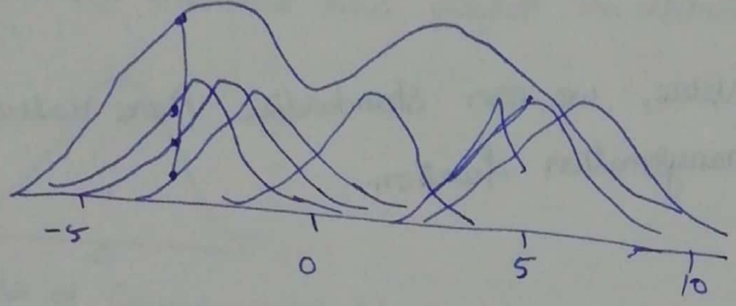
Why we are standardizing?



Standard normal distribution is a method, that i can transform any RV X with any mean & sd into a standard form which $\mu = 0, \sigma = 1$

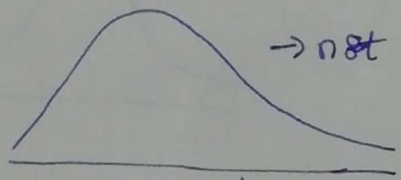
(11.7) KERNEL DENSITY ESTIMATION

→ From histogram we can calculate PDF by using KDE



(11.8) Sampling Distribution and Central limit theorem

X:



→ not necessarily Gaussian.

Population distribution

→ Income

↳ Picking the random Sample of size n (n=30) → S_1

↳ Again sample (n=30) → S_2 (Independent of S_1)

⋮

S_m

mean of $S_1 = \overline{x_1}$ → Sample mean.

$S_2 = \overline{x_2}$

$S_m = \overline{x_m}$

$\overline{x_1}, \overline{x_2} \dots \overline{x_m}$; m-sample means

$\overline{x_i}$ will has dist

↑ dist of sample means

dist of $\overline{x_i}$ = Sampling distribution of Sample means

Central Limit Theorem:

If my original populatation has finite μ & σ^2

X = finite μ & σ^2

Sample of size $n \leftarrow S_1, S_2 \cdots S_m$.

Sample means $\leftarrow \underbrace{\bar{x}_1, \bar{x}_2 \cdots \bar{x}_m}_{\bar{x}_i}$

CLT:

$\bar{x}_i$ is distributed with a gaussian ~~trit~~ distributeta.

$$\bar{x}_i \rightarrow N\left(\mu, \frac{\sigma^2}{n}\right) \text{ as } n \rightarrow \infty$$

Using CLT - Any distribution can be converted to normal distribution with mean $= \mu$ & sd $= \frac{\sigma^2}{n}$

11.9

Q-Q Plot

$X: x_1, x_2, x_3 \cdots x_{500}$

(Q) Is X Gaussian distributed?

using QQplt, statistical testing (KS Test, Anderson darling test)

How to plot Q-Q plot?

Step ① Sort your x_i's & compute percentiles

$x_1, x_2 \cdots x_{500}$

$\downarrow$

small sort (asc) largest

$x_1', x_2' \cdots x_{500}$

$x_1' \leq x_2' \cdots$

Compute Percentile

1st percentile $= x_5' \rightarrow x^{(1)}$ → first percentile value of x_i's

2nd " $= x_{10}' \rightarrow x^{(2)}$ → second "

⋮ → third "

500 " $= x_{500}' \rightarrow x^{(500)}$

what is percentile?

Index → 1 2 - - - 100

$x_s = \{ \quad \quad \}$

$n = 100$

Median(x) = ~~sorted~~ mean (x_1, x_2)

50^{th} percentile value of $x = x_s[50]$ = median.

10^{th} percentile " $x = x_s[10]$

10th percentage of value less the value

90% of value greater than value

25^{th} percentile is the first quantile

50^{th} percentile is the mean or second quantile

75^{th} " third quantile

100^{th} " fourth "

Ex

np. percentile (DS[colum], np. arrange (0, 100, 25))

returns a list

[1, 1.4, 1.5, 1.575]

Ex

e-commerce (Amazon)

delivery time

$[1, 1.5, 2 \cdots\cdots]_{10k}$

95^{th} percentile = 4 days

99^{th} percentile = 5.6 days

99% of the people are receiving the order with in 5.6 days step ② $Y \sim N(0,1)$ i.e., random variable which is normal distributed

$y_1, y_2, y_3 \dots y_{1000} \rightarrow$ 1000 obs from $N(0,1)$

↓ sort (asc)

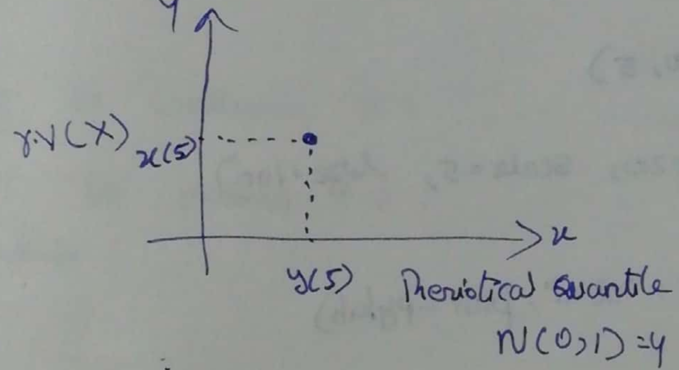
$y_1', y_2', y_3' \dots y_{1000}'$

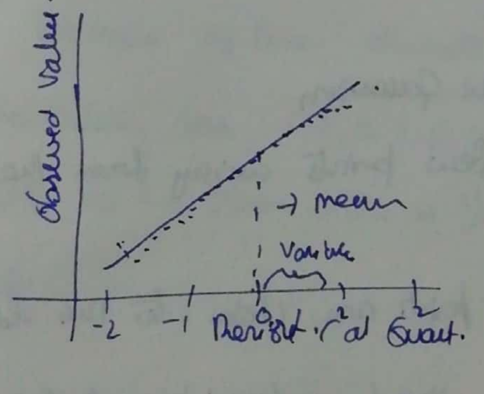
↓ Percentile

$y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)} \dots y^{(1000)}$

step ③ Plot Q-Q plot using $x^{(1)}, x^{(2)} \dots x^{(1000)}$ & $y^{(1)}, y^{(2)} \dots y^{(1000)}$

we will use this percentiles to plot Q-Q plot





If $(y^{(i)}, x^{(i)})$ $i=1 \rightarrow 100$ lie on a straight line then X & Y have similar distribution.

We can calculate mean & variance from the plot

Q-Q Plot

```
import numpy as np
import pylab
import scipy.stats as stats
```

N(0,1)

mean   std-dev   sample size

```
std_normal = np.random.normal(loc=0, scale=1, size=1000)
```

↳ random observation with gaussian distribution

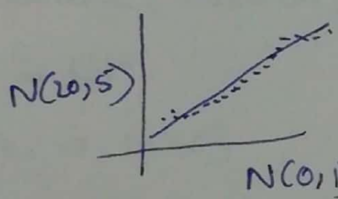
0 to 100th percentile of std-normal

```
for i in range(0, 101):
    print(i, np.percentile(std_normal, i))
```

generate 100 sample from N(20, 5)

```
measurement = np.random.normal(loc=20, scale=5, size=100)

stats = probplot(measurements, dist="norm", plot=pylab)
pylab.show()
```



x & y are Gaussian

There are few points away from the line.

As # of sample points increases more and more points are closer to the line

<u>Limitation</u> :- # of observation are small its hard to interpret QQ plot

<u>Consider the uniform distribution</u>

```
measurement = np.random.uniform(low=1, high=1, size=10,000)
```

Q-Q plot

(Q) is r.v X is normally distributed $X \sim N(\mu, \sigma^2)$

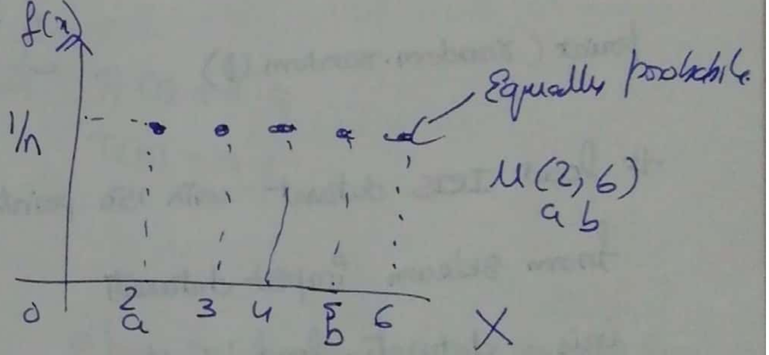
(Q2) X, Y are two r.v
using Q-Q plot we can define does X & Y have the same distribution.

11.10 Discrete & Continuous uniform distribution

Uniform Distribution. types
↳ Discrete
↳ Continous.

PDF for Continuous r.v

PMF for discrete r.v
↑ mass function



$n = 5$, when $n = b - a + 1$

A simple example of the discrete uniform distribution is throwing a dice. The possible values are 1, 2, 3, 4, 5, 6 and each time the die is thrown the probability of a given score is 1/6.

~~Ex two di~~

notation. $U\{a, b\}$ or unif (a, b)

Parameter: $a \in \{\ldots -2, -1, 0, 1, 2, \ldots\}$

$b \in \{\ldots -2, -1, 0, 1, 2, \ldots\}$, $b \geq a$

$n = b - a + 1$

Eg: Consider $a = 1$, $b = 6$

$n = 6$ # of outcomes

pmf $= 1/n$

CDF: $\lfloor k \rfloor - a + 1 / n$

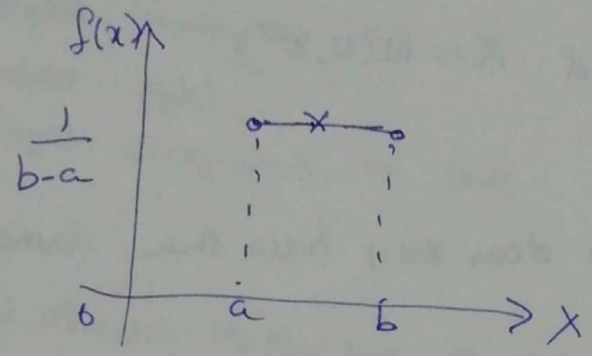
mean $= \frac{a+b}{2}$, median $= \frac{a+b}{2}$

Mode = NA

Variance $= \frac{(b-a+1)^2 - 1}{12}$

Skewness = 0

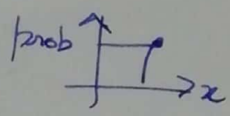
Uniform Continious distribution.



11.11 How to randomly Sample points (Uniform Distribution)

```
import random
print ( random.random ())
```

} It gives us a value b/w 0 & 1




```
# load IRIS dataset with 150 points
from sklearn import datasets
iris = datasets. load_iris ()
d= iris.data
d. shape.
```

```
# Sample 30 points randomly from the 150 point dataset
n=150
m=30
p= m/n #0.2
Sample_data [ ]
for i in range (0,n):
    if random. random <= p :
        Sample_data. append (d [i, :])
```

$\mathcal{U}(0,1)$

dataset

$D = x_1, x_2, x_3 \ldots x_{150}$

↓ Sample

$D' = x_1', x_2', \ldots x_3'$

len (sampled_data)

↳ need not be same every run; it will be around 30

11.12 Bernouli & Binomial distribution

Bernouli

Coin Toss

↳ chance of getting head is $1/2$

tais $= 1/2$

Bernouli distribution has only two out comes

X is Bernouli distributed $(P = 0.5)$

suppose we have a biased Cooints $H(1) = 0.9$

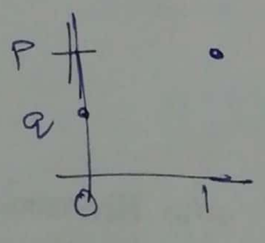
$T(0) = 0.1$

Bernouli distributon uses pmf

$$\text{Pmf} \begin{cases} q = (1-P) & \text{for } k=0 \\ p & \text{for } k=1 \end{cases}$$

P has to less than 1 & greater than 0

Suppor $k \in \{0, 1\}$



$P + q = 1$

mean $= P$

Variaa $= pq$

Binomial distribution (not used Extensively)

$X \sim$ Bernouli $(p=0.5)$

n times $(n=10)$

Q I want to Create a random Variable y

y = no. of times i get a when I toss my fare coins n times

$\underbrace{Y \in \{0, 1, 2, \ldots 10\}}_{\text{heads}}$

Y is distrubute coin binomial distribution

$Y \sim \text{Binomial}(n, p)$

↳ Prob of gettin a head (@ 1)

Parameter = n, P

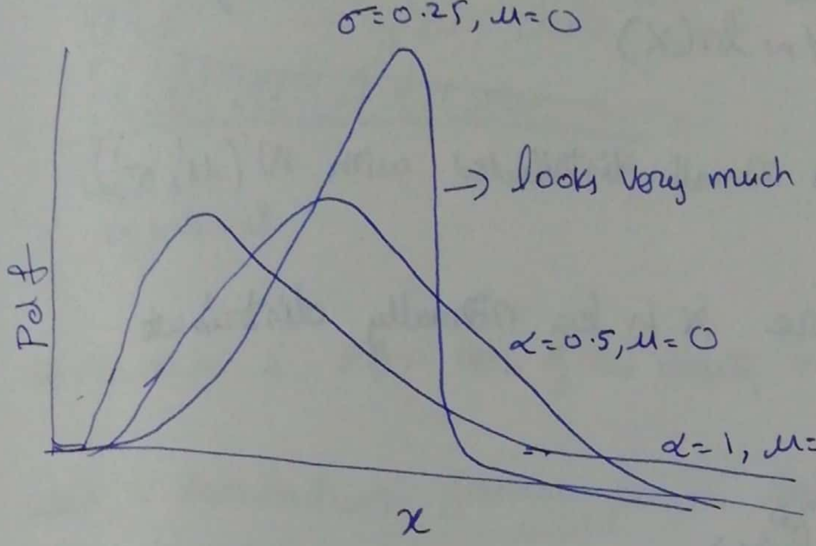
pmf $= \binom{n}{k} p^{k} (1-p)^{n-k}$

11.13

Log normal Distribution

→ Random X considered to be log normal if log(x) is normally distributed

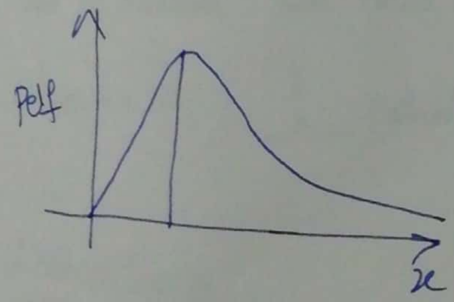
natural logarithm often referred as ln.



→ looks very much like Gaussian distribution.

As σ ↑, the ~~fatt~~ PDF tends to have a fatter tail (Skewed)

→ The length of comments posted in internet discussion forums (Reddit, Quora) follows a log normal distribution



→ lot of comments are short, but a very small set of comments are large.

→ The user's dwell time on the online article (jokes, news, Etc.) followed a log normal distribution.

→ Human behaviours

→ Economics (97% to 99%) income are log normal distribution

Imagine we have given a r.v "X" which is log normally distributed

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

~~(Q) How do you~~ As we came to know X is lognormal, we can convert to gaussian distribution by applying log

$Y \sim \log_e(X) \Rightarrow Y \sim \ln(X)$

Y is normall distributed with $N(\mu', \sigma')$

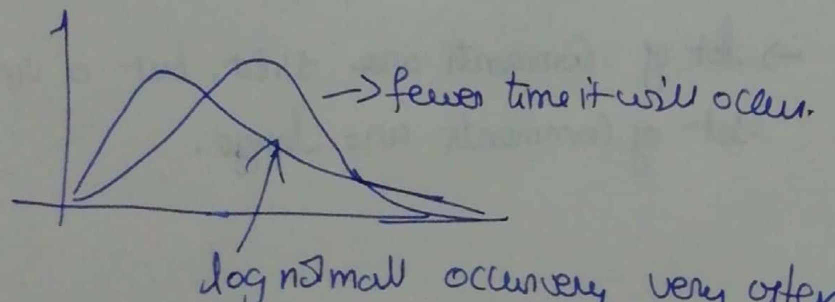
(Q) How do you make sure X is log normally distributed

$$\begin{array}{ccccc} x_1 & x_2 & x_3 & \cdots & x_n \\ \downarrow & \downarrow & & & \\ \ln(x_1) & \ln(x_2) & \ln(x_3) & \cdots & \ln(x_n) \\ \wr & \wr & \wr & & \wr \\ y_1 & y_2 & y_3 & & y_n. \end{array}$$

y_i's are gaussian then X is log normal

-> apply QQ plot y_i's are gaussian

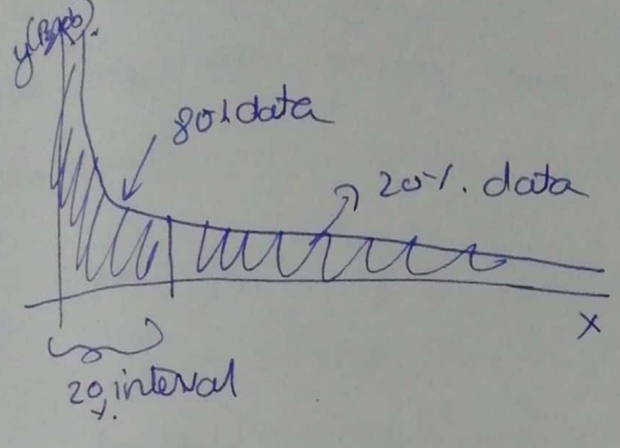
-> when we plot the pdf of any variable.



11.4

Power Law Distribution

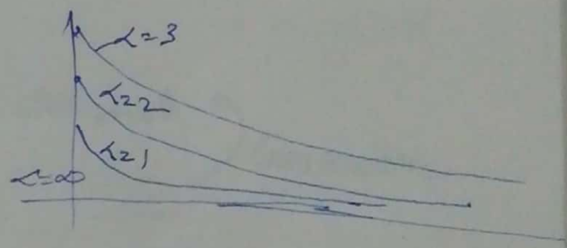
Imagine I have two variables x & y.



Relation b/w x and y, if it doesn't end very fast & long tail.

80-20 rule : 80% of the density in first 20% of intervals

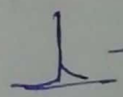
when a distribution follows power law then it is called ~~pareto~~ Pareto distribution



Parameter $x_m > 0$ (scale) $\rightarrow \mu$

$\alpha > 0$ shape (real)

As α reduces we will have fatter & fatter tail.

when $\alpha = \infty$  → delta function.

Applications

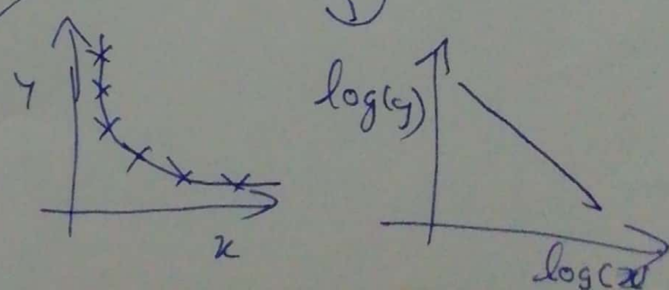
→ The sizes of human settlements (cities, many ~~xx~~ hamlets/villages)

→ file size distribution of internet traffic

→ Hard disk drive errors.

How to check

→ Using log-log plot



Box Cox Transform (Power Transform)

Given a random Variable X which is log normally distributed.

we ~~convert~~ can convert into random variable Y (Gaussian) by taking the ln operator

$x_1, x_2 \cdots x_n$ $\qquad$ $(\ln(x_1), \ln(x_2) \cdots \ln(x_3) \cdots)$

$X \xrightarrow{\ln} Y \quad y_1 \quad y_2 \quad y_3$

$\Downarrow$ log-normal $\qquad$ $\Downarrow$ Gaussian

Given a r.v "X" which is pareto of power distribu, is there any way to normal dist

$X = x_1, x_2 \cdots x_n$ ← x follow pareto distributi. Conversion.

$Y = y_1, y_2 \cdots y_n$ ← Gaussian

① $\text{box-cox}(X) \Rightarrow$ ~~alpha~~ lambda (λ)

$x_1, x_2 \cdots x_n$

② $$y_i = \begin{cases} \dfrac{x_i^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(x_i) & \text{if } \lambda = 0 \end{cases} \quad \forall i = 1 \to n$$

Gaussian dist

scipy.stats.boxcox (x, lambda=None, alpha=1

```
from scipy import stats
import matplotlib.pyplot as plt
```

→ one of the power law dist

```
x = stats.loggamma.rvs(5, size=50) +5

prob= stats.probplot(x, dist= stats.norm, plot=ax1)

plt.show()
```

```
xt, _ = stats.boxcox(x)
prob = stats.probplot(xt, dist = stats.norm, plot = ax2)
ax2.set_title('probplot after Box-cox transformation')
plt.show()
```

11.16 Covariance

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

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

$Cov(x,y) = +ve \quad x\uparrow, y\uparrow$

" $= -ve \quad x\uparrow, y\downarrow$

11.17 ~~Spearman Rank~~ Pearson Correlation Co-efficient

$$\rho_{x,y} = \frac{cov(x,y)}{\sigma x \; \sigma y} \qquad \sigma x = \sqrt{Var(x)}$$

$cov(x,y) = x\uparrow, y\uparrow = +ve$ (How much positive)

11.18 Spearman Rank Correlation Co-efficient

$\rho_{x,y}$ = linear relationship.

→ Monotonically Increasing function $x\uparrow, y\uparrow$

$\gamma = \rho_{r_x, r_y}$

$\rho = 1 \leftarrow$ linear $x\uparrow \; y\uparrow$ $\gamma = 1$

$\rho = -1 \leftarrow$ linear. $(\gamma = -1)$

	X	Y	r_x	r_y
s_1	160	52	4	3
s_2	150	66	2	4
s_3	170	68	5	5
s_4	140	46	1	1
s_5	158	51	3	2

11.20. Confidence Interval Introduction

X : heights of ppl
↑ don't know the distribution

$\{x_1, x_2, \ldots, x_{10}\}$

Estimate the population mean of $X = \mu$

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

population mean — sample mean — As n increases $\bar{x}$ will move closer to μ

* $\mu = \bar{x}$ is called point Estimate

$\{x_1, x_2, \ldots, x_{10}\} = \{180, 162, 158, 172, 168, 150, 171, 183, 165, 176\}$

Point Estimate of $\mu = \frac{1}{10} \sum_{i=1}^{10} x_i = 168.5$ cm

If I can say

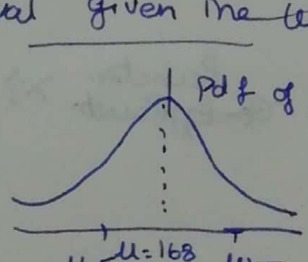
$\mu \in [162.1, 174.9]$ with 95% of probability

population mean lies b/w — Interval — Confidence

11.21 Computing Confidence Interval given the underlying distribution.

$X \sim N(\mu, \sigma^2)$ — σ = 5cm, μ = 168cm

heights of ppl — Gaussian



(Q) C.I of heights

95% of my data/observation lies b/w $(\mu - 2\sigma, \mu + 2\sigma)$ with 95% probability

158 ↓ 173 ↓

↑ Confidence (C)

11.22 **Confidence Interval for mean (μ) of a normal random Variable.**

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

Some distribution

Sample

$\{x_1, x_2 \dots x_{10}\}$ sample of size $n = 10$

$\{180, 162, 158, 172, 168, 150, 171, 183, 165, 176\}$

(Q) what is the 95% Confidence Interval of μ

Case 1 $\sigma = 5$ cm {we know pop-std dev}

Central limit theorem: $\bar{x}$ = sample mean $= \frac{1}{n}\sum_{i=1}^{10} x_i$

$\bar{x} = N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Sample mean

pop mean

pop std-dev / $\sqrt{n}$

$\mu \in \left[\bar{x} - \frac{2\sigma}{\sqrt{n}}, \bar{x} + \frac{2\sigma}{\sqrt{n}}\right]$ with 95% Confidence Interval

$\bar{x} = 168.5$ cm.

$\mu \in [168.34, 171.66]$ with 95% Confidence Interval.

Case 2: If we donot know σ

Sample of size n

t-distribution → Student's t-dist. → CI of mean of r.v when σ is unknown.

follows

$\bar{x} \sim t(n-1)$ — degrees of freedom.

Sample mean t-dist

(Q) How to I Estimate C.I for σ of a r.v

11.23 Confidence Interval using bootstrapping

→ Computing C.I for median, var, std-dev, 90th percentile.

$X \sim F$ Task ÷ Estimate 95% C.I for median of X

Sample of Size 10: $\{x_1, x_2 \dots x_n\}$ $n = 10$

Generating new samples from the sample

$S_1 : \underbrace{x_1^{(1)}\ x_2^{(1)}\ x_3^{(1)} \dots x_m^{(1)}}$ such that $m \le n$ (5, 10)

Sampling with repetition.

$S_2 = x_1^{(2)}, x_2^{(2)}\ x_3^{(2)} \dots x_m^{(2)} \to m_2$ = median of Sample 2

⋮

$S_k = x_1^{(k)}, x_2^{(k)} \dots x_m^{(k)} \to m_k \to$ Median of Sample S_k

Using discrete uniform distribution.

These are called boot strap sample.

$m_1\ m_2 \dots m_{1000} \Leftarrow$ 1000 medians gen using bootstrap samples

$m_1' \le m_2' \le m_3' \le \dots \le m_{1000}'$ (increasing order)

25 values $\le$ $\underbrace{m_{25}' \qquad m_{975}'}_{950}$ $\ge$ 25 values

95% C.I of median of X is $[m_{25}', m_{975}']$

* This is a non-parametric technique

↑ doesn't make any assumption about the dist of data

Empirical bootstrap based Confidence Interval

```
import numpy
from pandas import read_csv
from sklearn.utils import resample
from sklearn.metrics import accuracy_score
from matplotlib import pyplot

X = numpy.array([180, 162, 158, 172, 168, 150, 171, 183, 165, 176])
```
↳ Sample S

```
# Configure bootstrap
n_iterations = 1000          = K
n_size = int(len(x))         = m

# run bootstrap
median = list()
for i in range(n_iterations):
    # prepare train & test sets
    s = resample(X, n_samples = n_size);
    m = numpy.median(s)
    medians.append(m)
```

→ Example
first iteration might contain
[171, 172, 183, 158, 171, 168, 180
172, 168, 172]

```
# Confidence intervals
alpha = 0.95
p = ((1.0 - alpha)/2.0) * 100
lower = numpy.percentile(median, p)
```
→ 2.5%

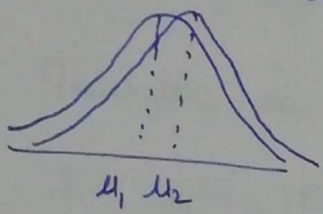
```
p = (alpha + ((1.0 - alpha)/2.0)) * 100
upper = np.percentile(median, p)
```
→ 95%

11.24 Hypothesis testing methodology, null-hypothesis, p-value

Is have two classes (class room 1 & 2)

(Q) Is there a difference in heights of students in cl1 & cl2

	cl1	cl2
1	160	162
2	152	156
:	:	
50	148	182



① Choosing a test statistic

$\mu_2 - \mu_1$

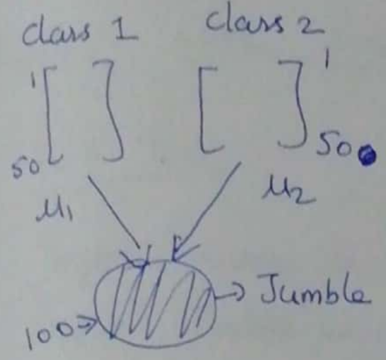
mean height of cl2 stu

mean heih of st2

11.25

Resampling and Permutation Test

How to Compute P-value ?



$\mu_1 - \mu_2 = \Delta$ (delta)

After Jumbling "randomly" sample 50 points

X $[\ ]_{50}^{1}$ Y $[\ ]_{50}^{1}$ $\mu_2 - \mu_1 = \delta_1$

μ_1 μ_2

X $[\ ]_{50}^{1}$ Y $[\ ]_{50}^{1}$ $\mu_2 - \mu_1 \cdots \delta_2$

μ_1 μ_2

⋮

Repeat for 10k times.

Sort the 10k deltas $(\delta_1\ \delta_2 \cdots \delta_{10000})$

$\delta_1\ \delta_2\ \delta_3 \cdots \Delta \overset{x\%}{\frown} \delta_{10k}$

→ Sorted (increasing)   ⟵ 5% ⟶ δ P-value = 0.05

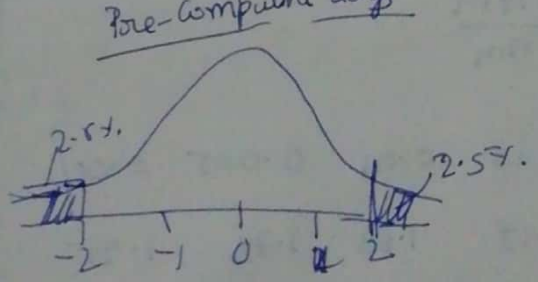
If this ~~delta~~ is greater than 95% of the points

In old days   Ex

given δ_i's

$\delta \sim N(0,1)$

$\Delta = 2$



$P(\delta \geq 2) = 0.025 \Rightarrow$ Then we can say p value = 0.025

11.26 K-S Test for Similarity of two distributions

Kolmogorov-Smirnov test

X_1 : $[x_1', x_2' \dots x_n']$ (n observations)

X_2 : $[x_1' \ x_2' \dots x_m']$ (m observations)

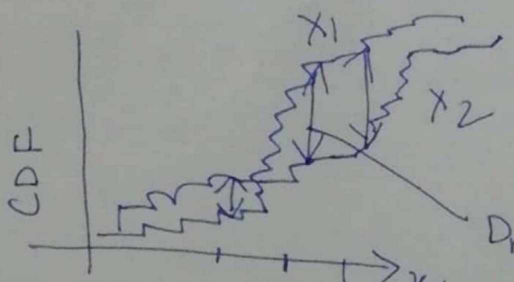
(Q) Is the distribution X_1 same as X_2

Petal length

$\frac{PL - \mu_{PL}}{\sigma_{PL}} \sim N(0,1)$

H_0 : X_1 & X_2 have same distributi

H_A : X_1 & X_2 have different distributi.



If two CDF's overlap the difference is zero

$D_{n,m}$ (large difference) b/w CDF's.

If n is large & m is very large, X_1 & X_2 will over lap on Each other then $D_{n,m} = 0$, H_0 is True

$$\boxed{D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|,}$$

The null hypothesis is rejected at level α if

$$D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{nm}}$$

α	0.10	0.05	0.025	0.01	0.005	0.001
$c(\alpha)$	1.22	1.36	1.48	1.63	1.73	1.95

$n = 1000$

$m = 5000$

$D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{nm}}$

$D_{n,m} > 1.36\sqrt{\frac{1000+5000}{1000 \times 5000}}$

If $D_{n,m} > 0.047$

If $D_{n,m} > 0.047$ then reject H_0 at 0.05 significance level.

what is $D_{n,m}$ if $n=50, m=30$

$D_{n,m} > 0.31$

11.27 Code Snippet K-S Test

K-S Test

```
import numpy as np
import seaborn as sns
from scipy import stats
import matplotlib.pyplot as plt

# generate a gaussian rv X
X = stats.norm.rvs(size=1000);          # X ~ N(0,1)
sns.set_style('white grid)
sns.kdeplot(np.array(x), bw=0.5)
plt.show()
```

kernal dist plot

```
stats.kstest(x, 'norm')
```

→ o/p KstestResult(statistic = 0.0213083972, pValue = 0.7542103145)

H_0 since p-value is very high. we can say R.V is $N(0,1)$

```
y = np.random.uniform(0, 1, 10000);          Y ~ U(0,10)
sns.kdeplot(np.array(y), bw=0.1)
plt.show()

stats.kstest(y, 'norm')
```

o/p : (p.value = 0.0)

H_a since p-value is very less.

we can say R.V is $U(0,1)$

11.28 Hypothesis testing intuition with Coin toss Example

Ex:- 1

Coin, determine if the coin is biased towards heads or not

biased towards heads $P(H) > 0.5$

not biased $P(H) = 0.5$

Experiment

To flip a coin 5 times and I count the # of heads

count # heads = X ← test_statistic

Perform Experiment flip, flip, flip, flip, flip

H H H H H ⇒ X = 5 ← observation

$P(\underbrace{X=5}_{obs} \mid \underbrace{\text{Coin is not biased towards H}}_{assumption}) = P(obs \mid H_0)$

null hypothesis (H_0)

H_0: Coin is not biased towards heads

$P(X=5 \mid H_0) = \frac{1}{2^5} = \frac{1}{32} = 0.03 \approx 3\%$

5 head in 5 flips — coin is not biased towards head

P-value

HHHHH
HHHHT
HHHTH
⋮
32 Combinations

$P(H) = 1/2 = 0.5$

There is a 3% chance of getting 5 heads in 5 flips if the coin is not biased towards heads

if $P(obs \mid H_0) < 5\%$

Then H_0 may be incorrect

⇓

assumption, H_0 is not true or we reject null hypothesis

H_A: accept the idea that coin is biased why only 5 times, we can do it 10, 5, 50 times then the probabilities will change.

This is often called as sample size.

Expt 1 Let's flip the coin 3 time Count # heads

Count # heads = X

f, f, f

↓ ↓ ↓

H H H

X = 3 ⇐ observation

P(obs | assumption) = $\frac{1}{2^3} = \frac{1}{8}$ = 12.5% > 5%

↑ Coin is not biased

Ans ÷ Accept null hypothesis & reject my H_A

Important things

1. design of the Expt
2. defining H_0

   P(obs | H_0) should be easy

P-value = 3% = reject H_0

P(obs | H_0) ✓

P(H_0 is true) ✗

3. Design of X

11.29 Hypothesis testing means differences Example.

Task ÷ you have two cities C_1 & C_2

determine if the population means of heights of people in these two cities is same or not

C_1 1mln ppl → μ_1 C_2 2million ppl → μ_2 | Determine μ_1 & μ_2 are same or not

It is very expensive to collect the heights of 3million ppl..

Exp ÷

$$C_1 \begin{bmatrix} h_1 \\ \vdots \\ h_{50} \end{bmatrix} \quad C_2 \begin{bmatrix} h_1 \\ \vdots \\ h_{50} \end{bmatrix}$$

Sample heights of 50 random ppl

Sample mean of C_1 is $\mu_1 = \frac{h_1 + h_2 \cdots h_{50}}{2}$ → 162cm (observed ppl)

Sample mean of C_2 is $\mu_2 = \frac{h_1' + h_2' \cdots h_{50}'}{2}$ → 167cm → observed ppl

Test statistic $= |\mu_1 - \mu_2| = 162 - 167 = 5cm$

$P(X = 5cm \mid H_0)$

Prob. of observing a difference of 5cm in sample mean height of of sample size 50 b/w C_1 & C_2. If there is no population difference in mean heights

case 1 $P(x=5 \mid H_0) = 0.2 = 20\%$.

There is a 20% chance of observing a difference of 5cm in sample mean heights of C_1 & C_2 (with sample of 50?) If there is no population mean diff

$\boxed{\Rightarrow \text{accept our } H_0}$

Case 2: $P(x=5 \mid H_0) = 0.03 = 3\%$.

$P(\text{obs} \mid \text{assumption}) = 3\% \rightarrow$ small

$\rightarrow$ assumption must be incorrect

$\rightarrow$ reject $H_0 \Rightarrow$ accept H_1

11.30: Resampling & Permutation test for mean difference example

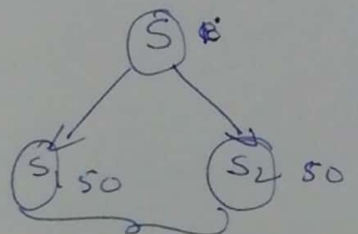
$P(x=5 \mid H_0) \rightarrow$ How to compute it.

H_0: no diff in population mean.

$x = |\overset{162}{\mu_1} - \overset{167}{\mu_2}|$

(1) club two classes and create a new set S city

(2) From S we will create S_1 & S_2



randomly select.

Resampling

Calculat $\mu_2 - \mu_1 = 3\text{ cm} \rightarrow \delta_1$

repeat (2) $\mu_2 - \mu_1 = -2\text{ cms} \rightarrow \delta_2$

(3) $\mu_2 - \mu_1 = 1\text{ cm} \rightarrow \delta_3$

⋮

(K) $\mu_2 - \mu_1 = 6\text{ cm} \rightarrow \delta_K$ $\boxed{\text{Let's } K = 1000}$

(3) sort δ_i's in increasi

$$\delta_1' \le \delta_2' < \delta_3' \le \delta_4' \le \dots \delta_K'$$

Case 1: Obs-difference $= x = 5\text{cm}$ (167-162)

~~P(diff ≥ 5cm | H0)~~

$$\underbrace{\delta_1' \le \delta_2' \le \delta_3' \cdots \delta_{800}}_{80\text{ \% of Simulation differences } \le 5cm} \le 5cm < \underbrace{\delta_{801} \cdots \delta_{1000}}_{20\text{ \% of sim differences } > 5cm}$$

$P(\text{obs-diff} \mid \text{assumption}) = 20\%$ ← P.value.

$\boxed{\text{Accepting } H_0}$

Case2 $\underbrace{\delta_1' \le \delta_2' \cdots\cdots\cdots}_{97\%} 5cm \le \underbrace{\delta_{1000}'}_{3\%}$

$P(\text{obs-diff} \ge 5cm \mid H_0) = 3\%$.

$\boxed{\text{Rejecting } H_0}$ Assumption must be incorrect