Probability Distributions

Monday, 8 November 2021 7:52 PM

X = R.V marks =

96, 97, 100

60 - 70 - 75 majority

50, 43, 49

why ?

1000 observation

Summarize data / EDA

follow

K

pdf / cdf

avg / mean

std

Simple Model

function

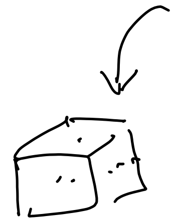
Pdf

cdf

Pmf

Continous r.v

discrete r.v



$\cancel{5/6}$

Coin → head
→ fail

Uniform

↳ equally likely

↳ Discrete Uniform

↳ Continous Uniform

0.561

0.328

0.79

$$X \sim U(a, b)$$

cdf ⇒ $P(X \leq x)$

$$\boxed{P(X \le 20)}$$

Bernaulli Distribution

→ 2 outcomes

- Success (Yes) — Probability → 'p' — '0.9'
- Failure (No) — '1 − p' — '0.1'

Coin → Head, Tail

India → Yes, No

Normal Distribution ← Continuous

heights, weights, - - - - - marks - -

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

$\mu \to$ mean

$\int$

marks
—
—
—

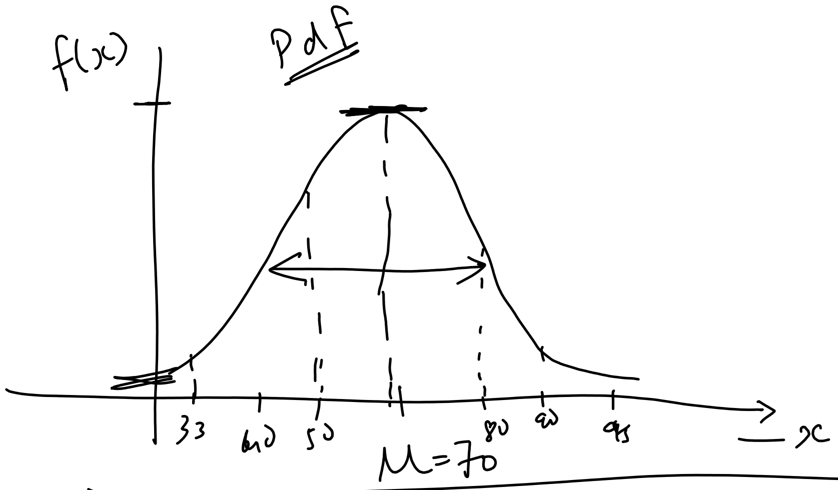
$X \sim N(70, 5)$

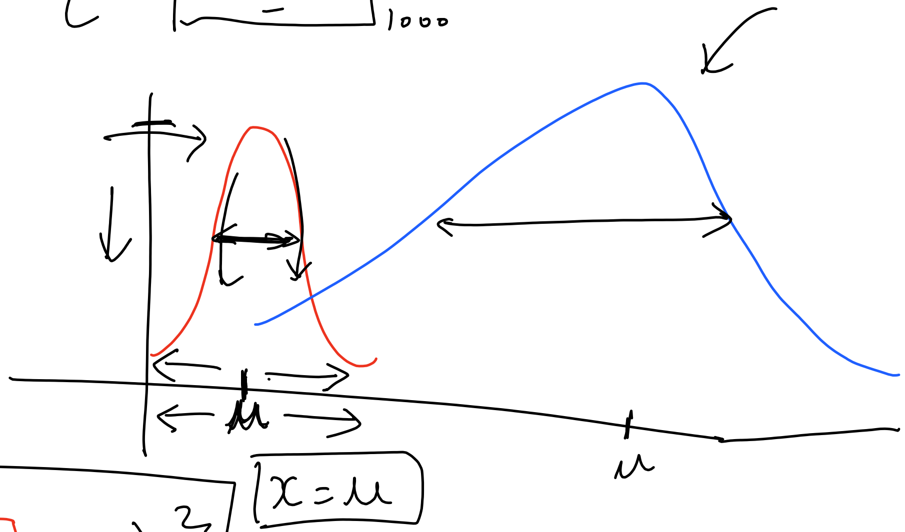
70.5

181.23cm

70 - 75 - 80

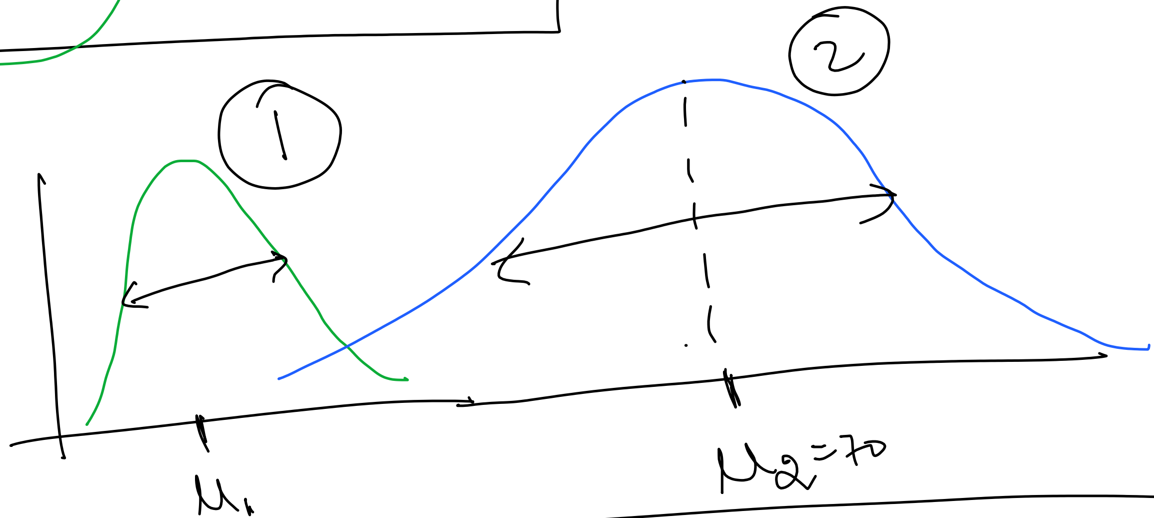
$\sigma \rightarrow$ std.





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

Const



$\mu \rightarrow$

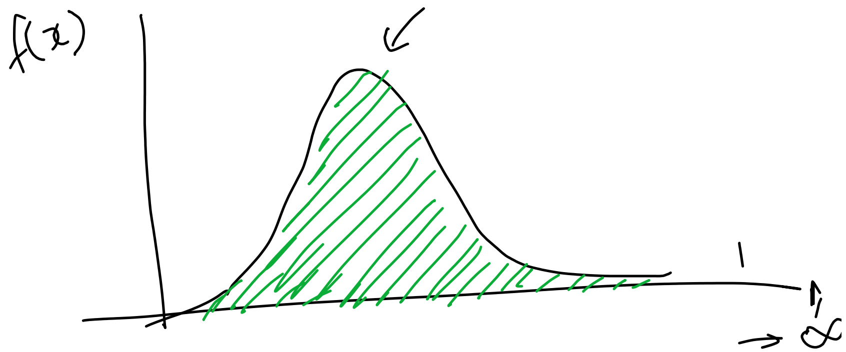
$\sigma \rightarrow$

Pdf $\rightarrow$

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

$\mu = 70$

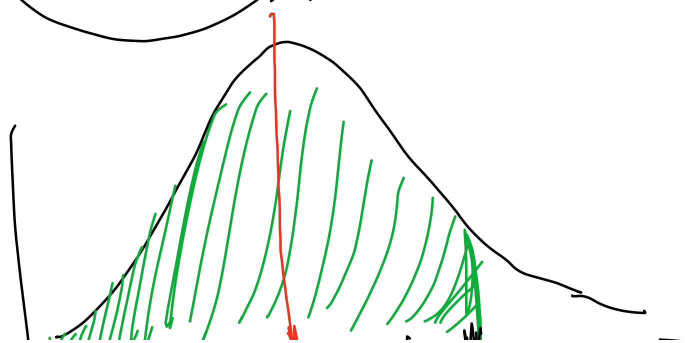
$\sigma = 5$



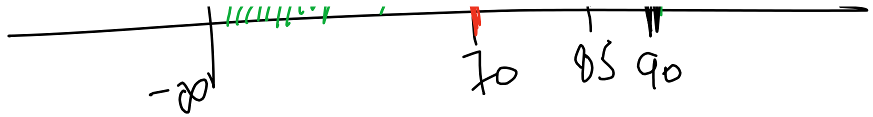
Pdf:

$$\boxed{f(x) \geq 0}$$

$$\int_{-\infty}^{\infty} f(x) \cdot dx = 1 \leftarrow \text{cdf}$$



CDF: $P(X < a)$



$$F(x_0) = \int_{-\infty}^{x_0} f(x)\,dx$$

Integration pdf $\rightarrow$ cdf

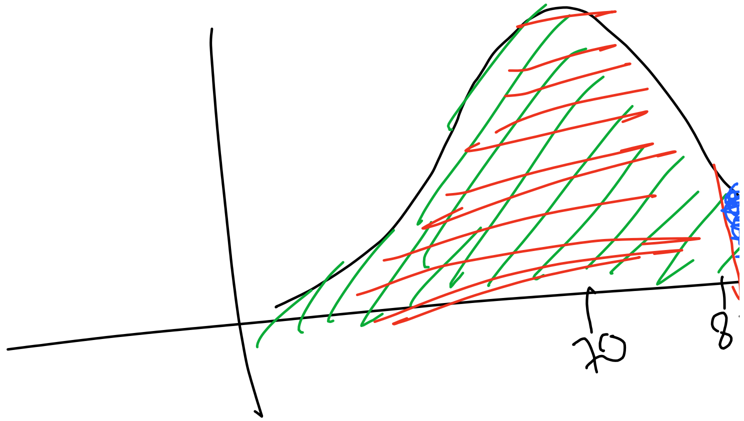
differentiate cdf $\rightarrow$ pdf

$$f(x_0) = \frac{d}{dx} F(x) \Big|_{x=x_0}$$

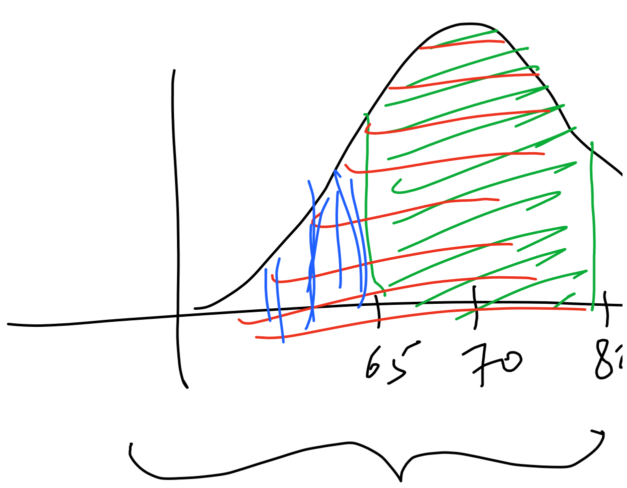
$$P(X > 85) = 1 - P(X \le 85)$$

Red shade

Blue shaded area



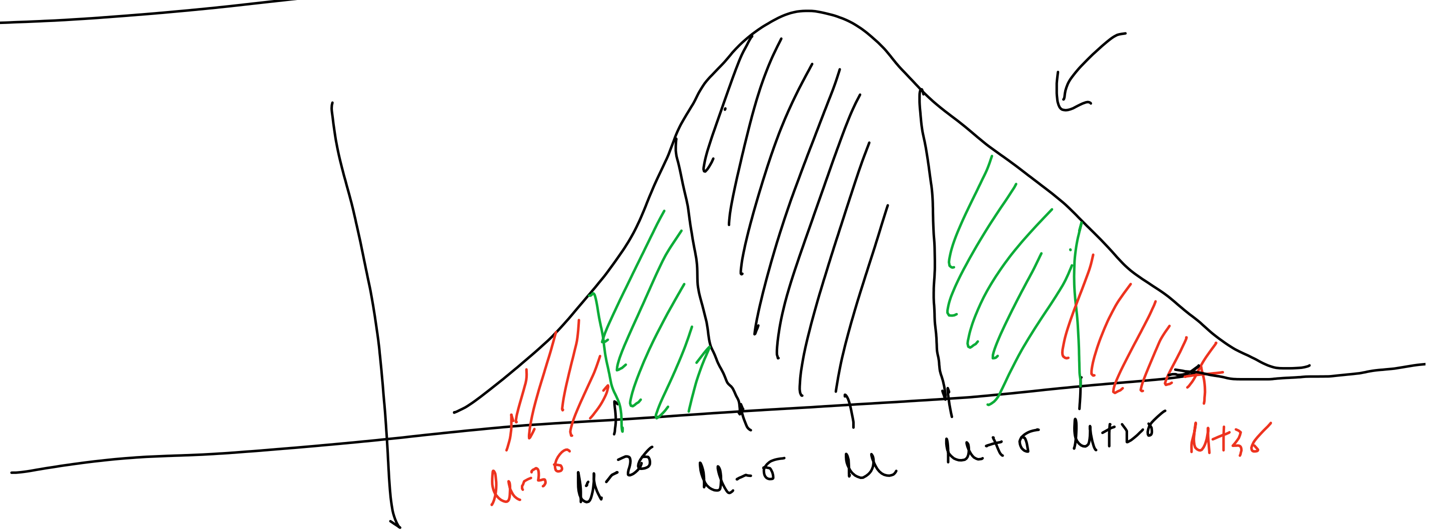
$$P(65 \le X \le 80)$$



$$\boxed{\int_{65}^{80} f(x)\cdot dx}$$

$$\boxed{P(X \le 80) - P(X \le 65)}$$

$1 - p(X < 65)$



$\mu = 70$

$\sigma = 5$

$[60, 80]$

$[\mu - \sigma \quad \mu + \sigma] \rightarrow 68\%$

$[\mu-\sigma, \mu+\sigma] \rightarrow 68\%$

$[\mu-2\sigma, \mu+2\sigma] \rightarrow 95\%$

$[\mu-3\sigma, \mu+3\sigma] \rightarrow 99.7\%$

100,000

Cultural event

weight

Height → T Shirt

Data scientist

S →
M →
L →
XL →

2/min

1970's

175cm

180 cm

BMI →

1970's

1000 Employs

500 employs

30cm = 1 feet

Domain Knowledge

1 > 180 cm

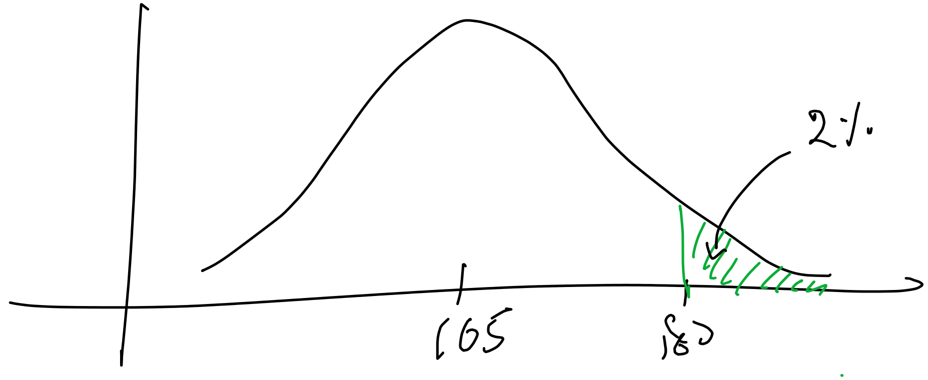
6' feet → XL

L

500

[12 180cm]

② Assuming Heights $\sim \boxed{N(\mu, \sigma)}$



$P(X \geq 180cm)$

$\downarrow$

$\boxed{2\%} \rightarrow \boxed{XL}$

$100000 \times 2\% \Rightarrow \boxed{2000 XL}$

Standard Normal Distrib.

Z-distribution

$$\boxed{Z \sim N(0, 1)}$$

$\mu = 0$

$\sigma = 1$

Normal dist → S.N.D

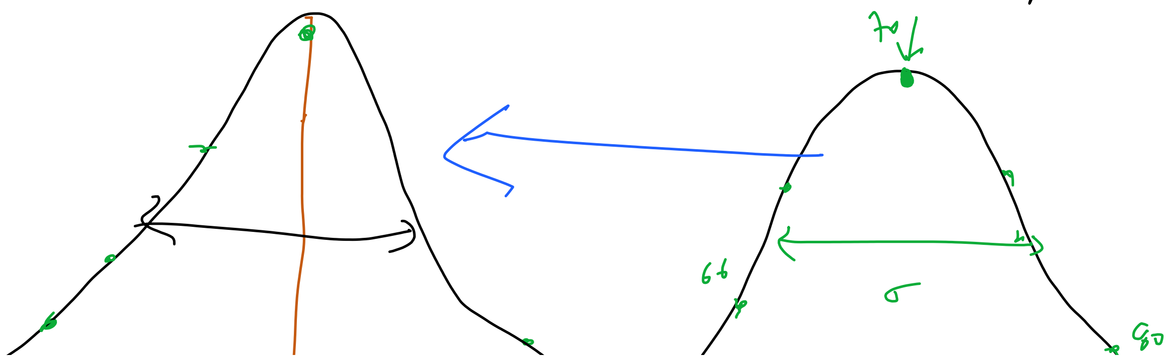
marks

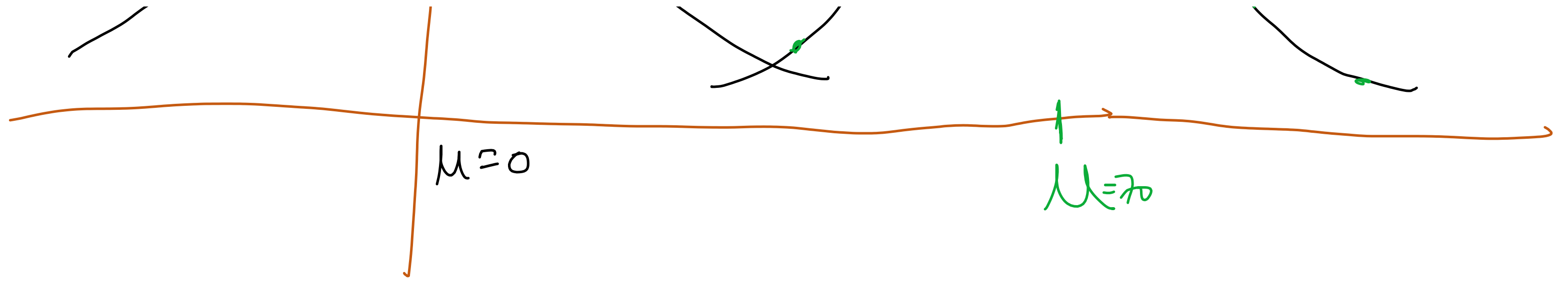
heights

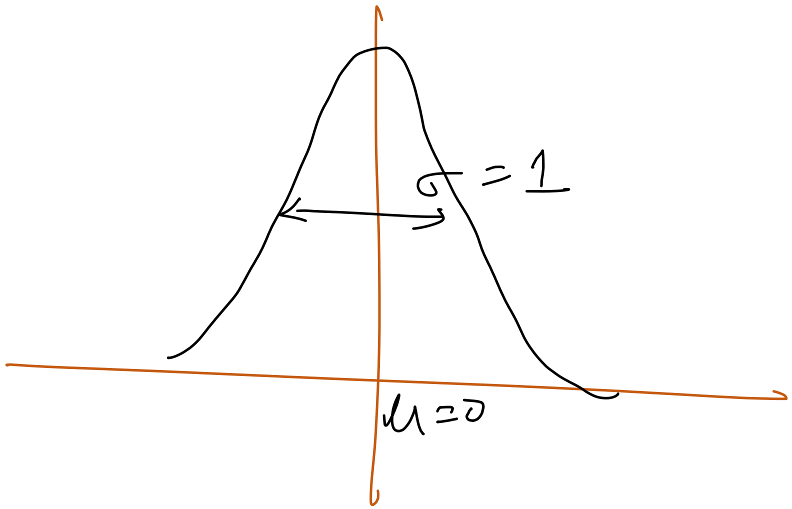
μ, σ

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

z-scores







$$\boxed{Z_i = \frac{x_i - \mu}{\sigma}}$$

$$Z_i \times \sigma + \mu = x_i$$

getting the data Back.

Binomial Distribution → discrete

Bernaulli

1 0

14

10 times

6. ~~5~~ Head

$X \sim \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, \boxed{10}\}$

~~15~~

n-trial (all trials should be independent)
Individual trial should be Bernaulli

no. of trial

$$\boxed{X \sim Binom(n, p)}$$

probab. of success

0.8

Binom (10, 0.5)

8%

open ≈ 0.08

not open ≈ 92%

n = 10000

Binom (10000, 0.08)

~~Pdf~~ PMF

Cdf

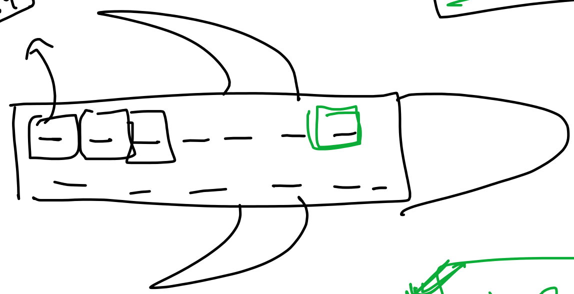
$P(X \geq \boxed{1500}) =$

Overbooking Problem

Vistara Airline

100 Seats

0.95



Maximizing Profit

Overbooking

105 tickets

110, 107

Q: How many extra ticket i can sell?