7th December, 2022

DSML : CC Maths

Probability 7 - Distributions - 2.

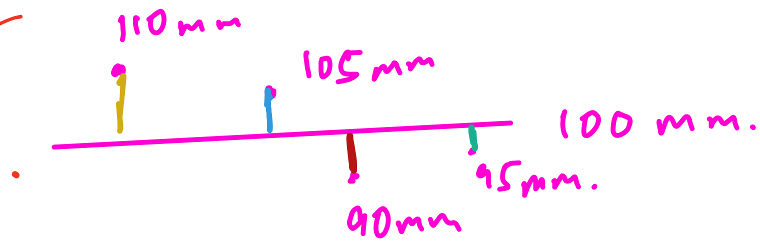
Recap:
(a) Probability theory.
(b) Bayes' theorem.
(c) Combinatorics.
(d) Descriptive statistics.
(e) Binomial Distribution.

Today:
(a) Variance, Standard deviation.
(b) Gaussian distribution.
(c) 68/95/99 rule.
(d) z-score.
(e) z-table
(f) cdf - cumulative distribution function
(g) ppf - percent point function.

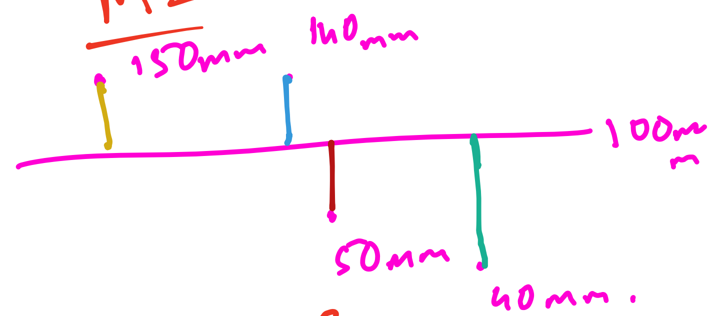
Class starts @ 9:05 p.m.

Variance: Ball manufacturer:

M1



M2



10 mm.
5 mm.
- 10 mm.
- 5 mm.

How much our observations deviate from the mean.

$(50mm)^2$ 2500
$(40mm)^2$ 1600
$(-50\ mm)^2$ 2500
$(-60\ mm.)^2$ 3600

10200

① $|10| + |5| + |-10| + |-5| = 30$

② $(10)^2 + (5)^2 + (-10)^2 + (-5)^2 = 250.$ ✓

$$\frac{10200}{4} \rightarrow 2550 \rightarrow 50.49.$$

$$\frac{250}{4} = 62.5 \rightarrow \text{Variance.}$$

$$\sqrt{\frac{250}{4}} = 7.9 \rightarrow \text{Standard deviation.}$$

① Variance: $\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$.

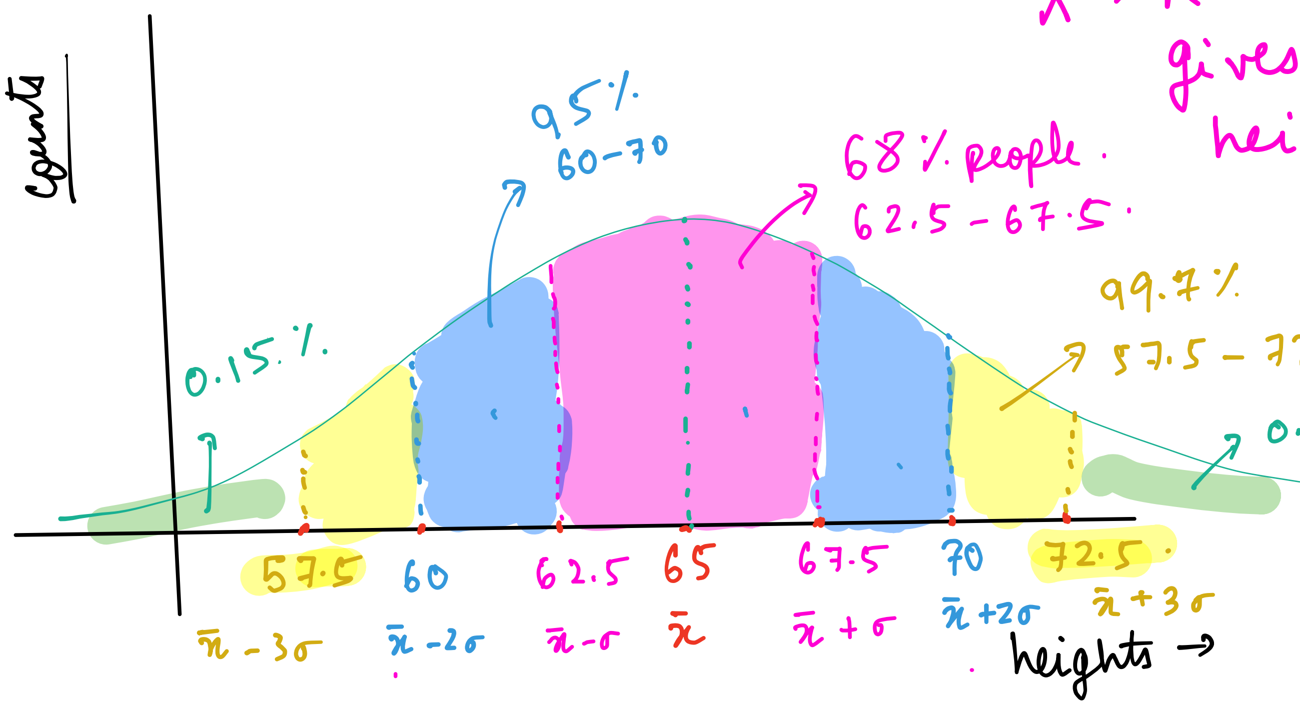
$X = \{x_i\}_{i=1}^{n}$

mean $= \dfrac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$

② Std. : $\sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$.

Gaussian: The height of people is Gaussian distributed. The mean is 65 inches, std is 2.5 inches.

X → R.V. which gives us the height.

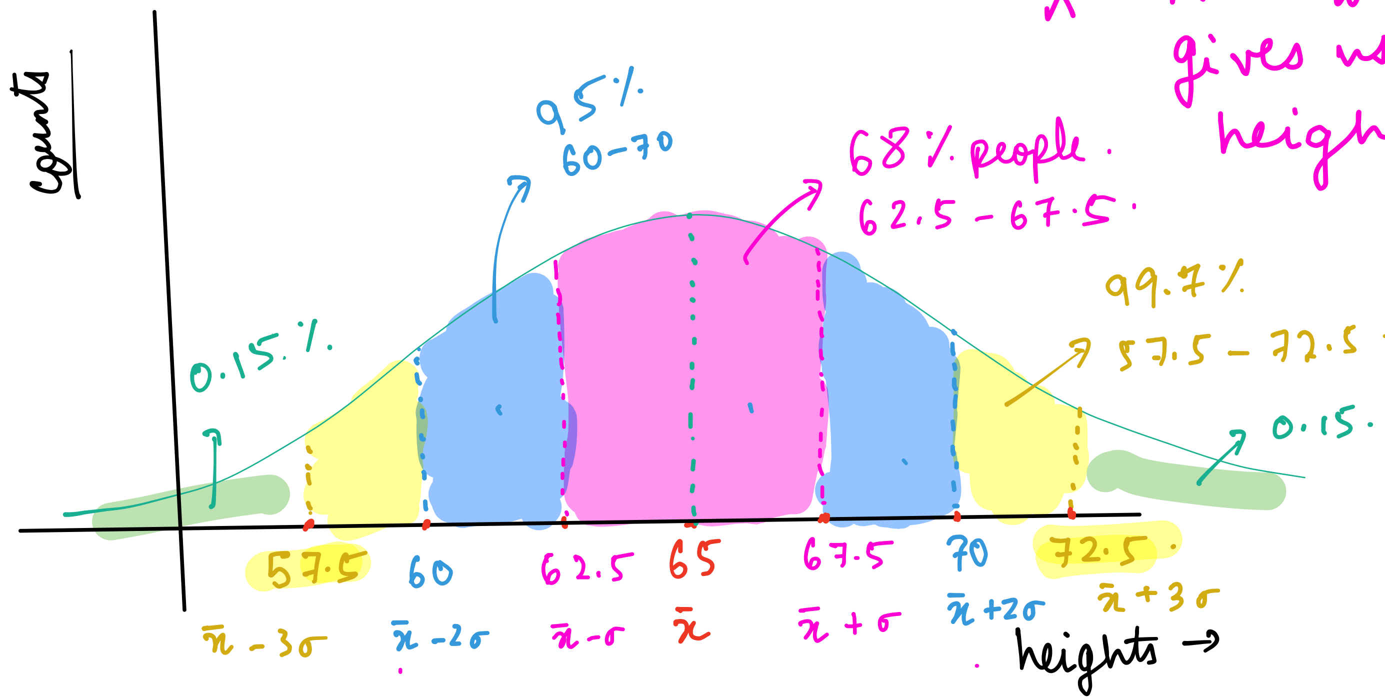


$$P[62.5 \leq X \leq 67.5] = 0.68.$$

$$P[60 \leq X \leq 70] = 0.95.$$

$$P[57.5 \leq X \leq 72.5] = 0.997$$

X → R.V. which gives us the height.



What is the fraction of people between 60 & 72.5?

$$P[X \le 65] = 0.5.$$

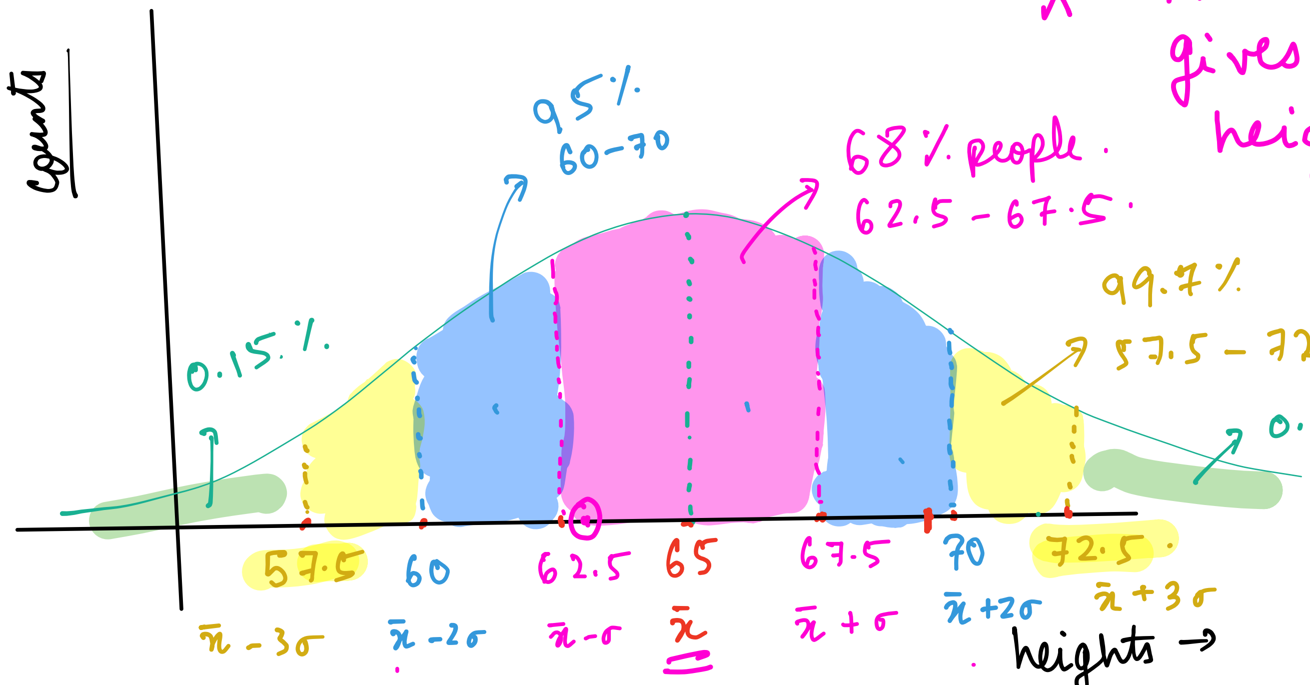
$$P[60 \le X \le 65] = 47.5.$$

$$P[65 \le X \le 72.5] = \frac{99.7}{2} = 49.85.$$

$$47.5 + 49.85 = 97.35\%.$$

std = 2.5.

X → R.V. which gives us the height.



Rajaram → 72

$$\frac{72 - 65}{2.5}$$

* What fraction of people are shorter than Nikhil & Rohit? (69 inches). → 94.52%

Q) How many stds away from the mean is 69 inches?

"Z"

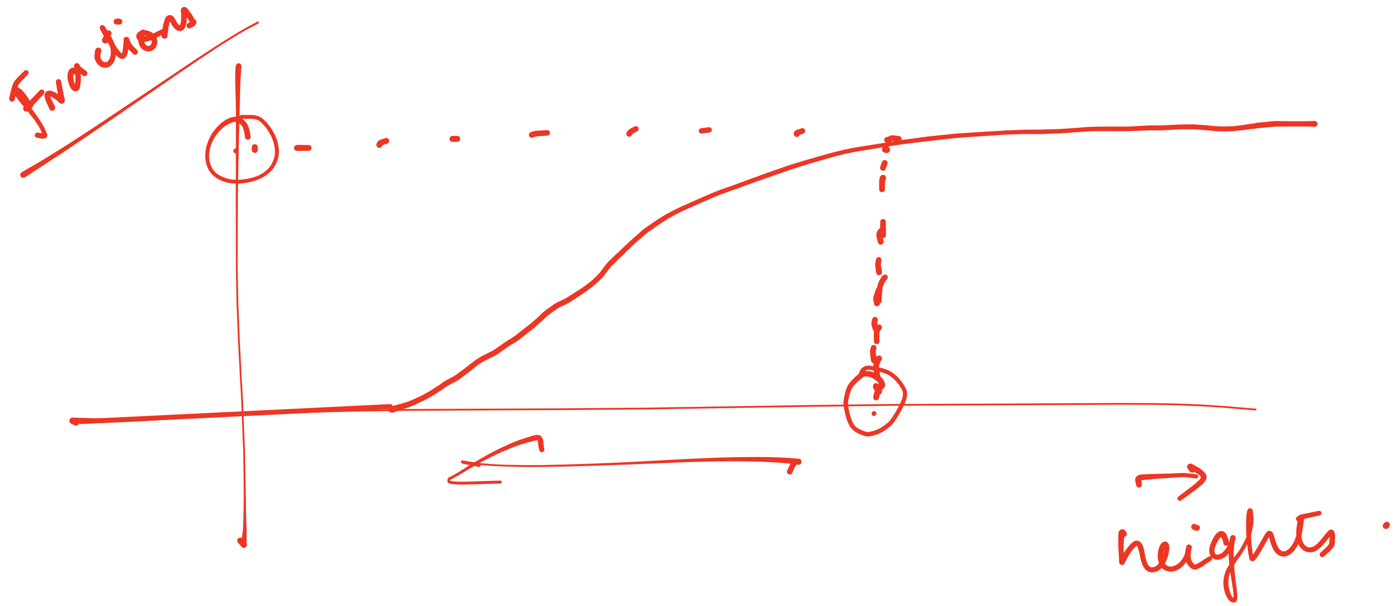
$$69 = 65 + Z \times std.$$

$$\frac{69 - 65}{std} = 7 = \frac{1.6}{1}.$$

Z-score

$$63 = 65 + z\, std.$$

$$\frac{63 - 65}{2.5} = -0.8.$$



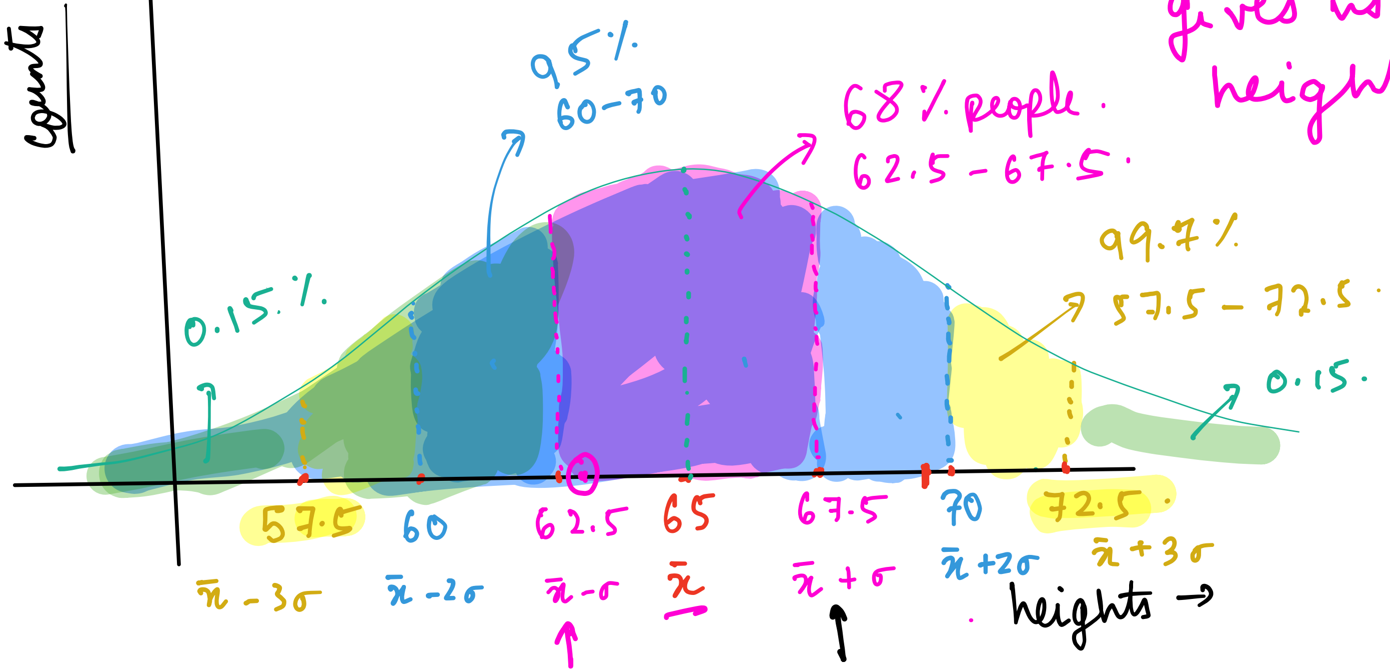
1 → 1 std .

2 → 2 2

2 → 3 → 3

std = 2.5.

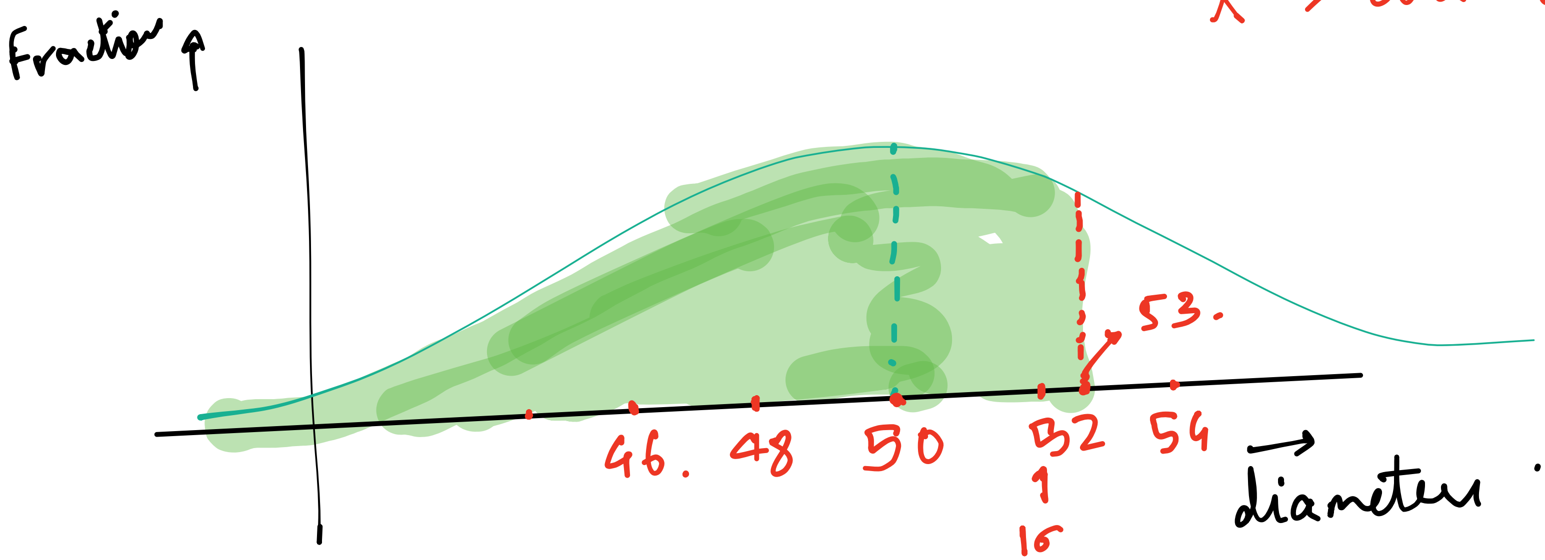
X → R.V. which gives us the height.



norm.cdf(1) - norm.cdf(-1)

Balls produced by M3 have a mean 50mm & std. 2mm, and they are Gaussian distributed.

X → dia. of balls.



What fraction of balls are smaller than 53mm?

$P[X \le 53]$?

Z-score: $\frac{53-50}{2} = \frac{3}{2} = 1.5$.

Z-table → 0.93319.

(Amazon) Skaters take a mean of 7.42 seconds and std = 0.34 s for 500 meters, and this data follows the Gaussian distribution.

What should the speed of a skater be so that she is faster than 95% of the competitors?

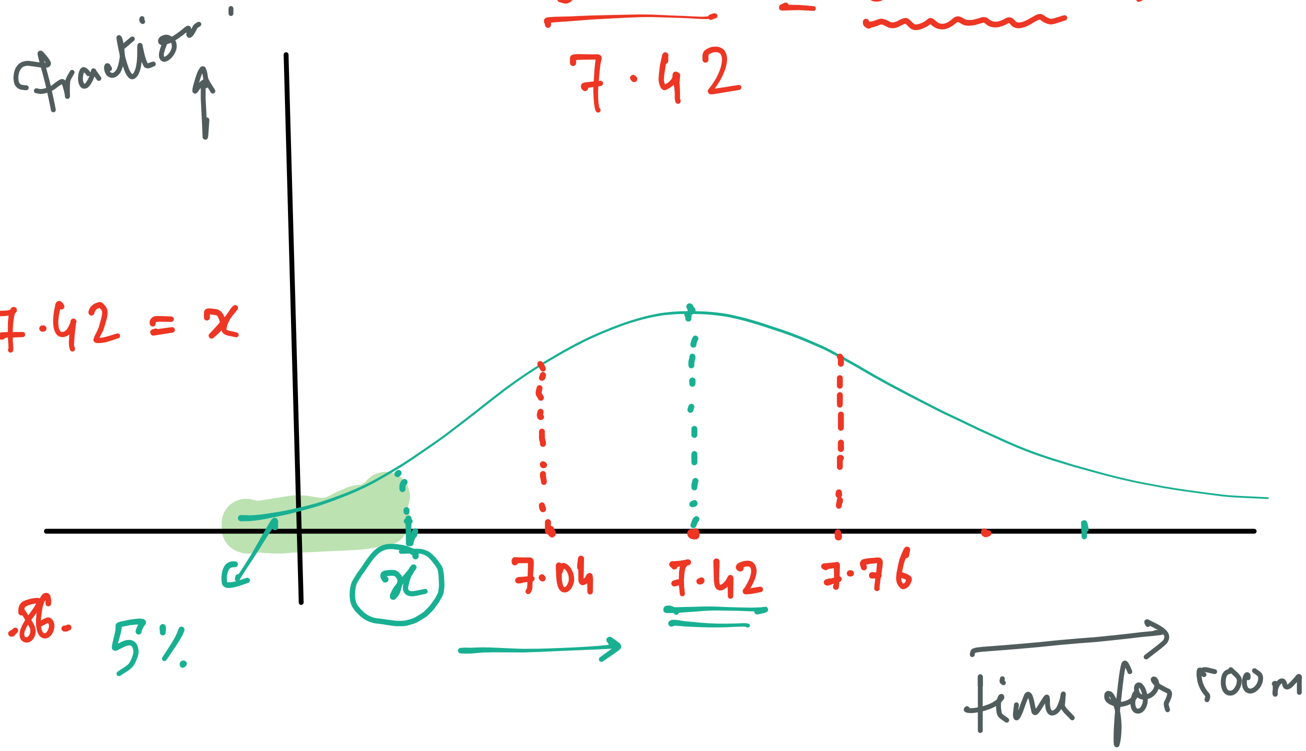
$$\frac{500\,m}{7.42} = 67.39\ m/s.$$

$$Z = \frac{x - \bar{x}}{std.}$$

$$-1.64 \times 0.34 + 7.42 = x$$

$$x = 6.8624.$$

$$\frac{500}{6.8624} = 72.86.$$

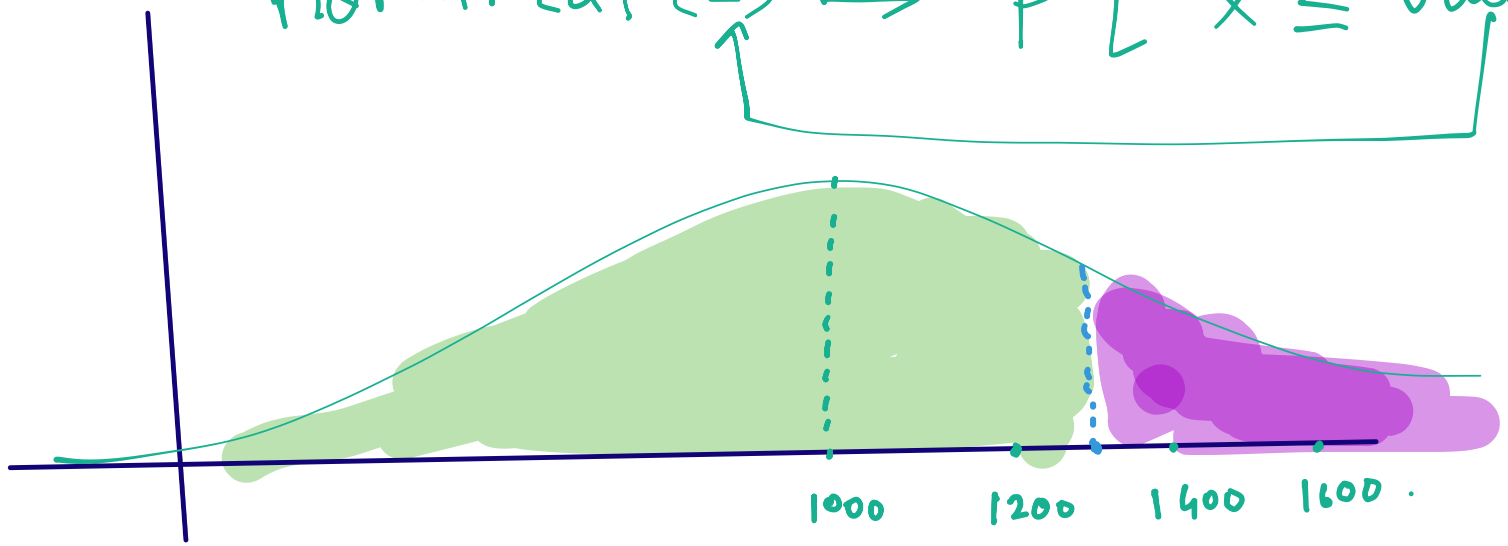


A retail outlet sells around 1000 toothpaste in a week, with std = 200.

If the on-hand inventory is 1300, what is the probability that we have to order more?

$X \rightarrow$ weekly sales.

norm.cdf(z) $\longrightarrow$ $P[X \leq \text{value}]$



$$Z = \frac{1300 - 1000}{200}$$

$$= \frac{300}{200} = 1.5$$

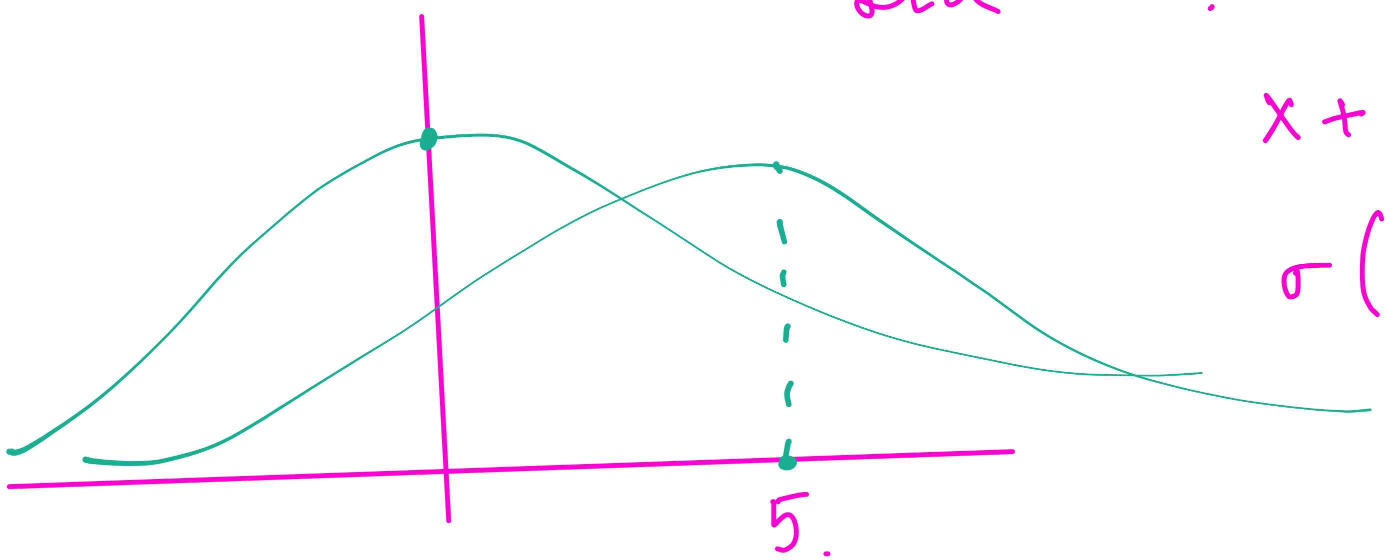
$\rightarrow P[X > 1300]$.

1 − norm.cdf(1.5) $\rightarrow$ 0.0668 ≈ 6.68%

Gaussian → mean : 65
std deviation : 2.5.

Normal → Is nothing but a Gaussian
with mean = 0,
std = 1.

X → Normally distributed.



X + 5

$\sigma(X)$

$$f_{gaussian}(x) = \frac{1}{\sqrt{2\pi s}} e^{-\frac{(x-u)^2}{2s^2}}$$

$u \to$ mean.

$s \to$ std. dev.

$Z \to (0, 1)$

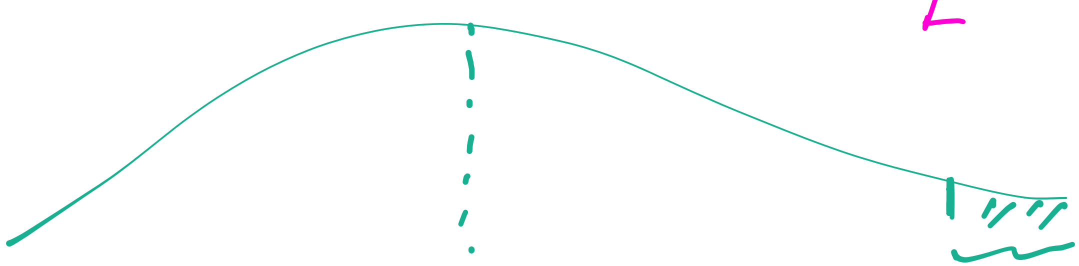
$-$ $+$

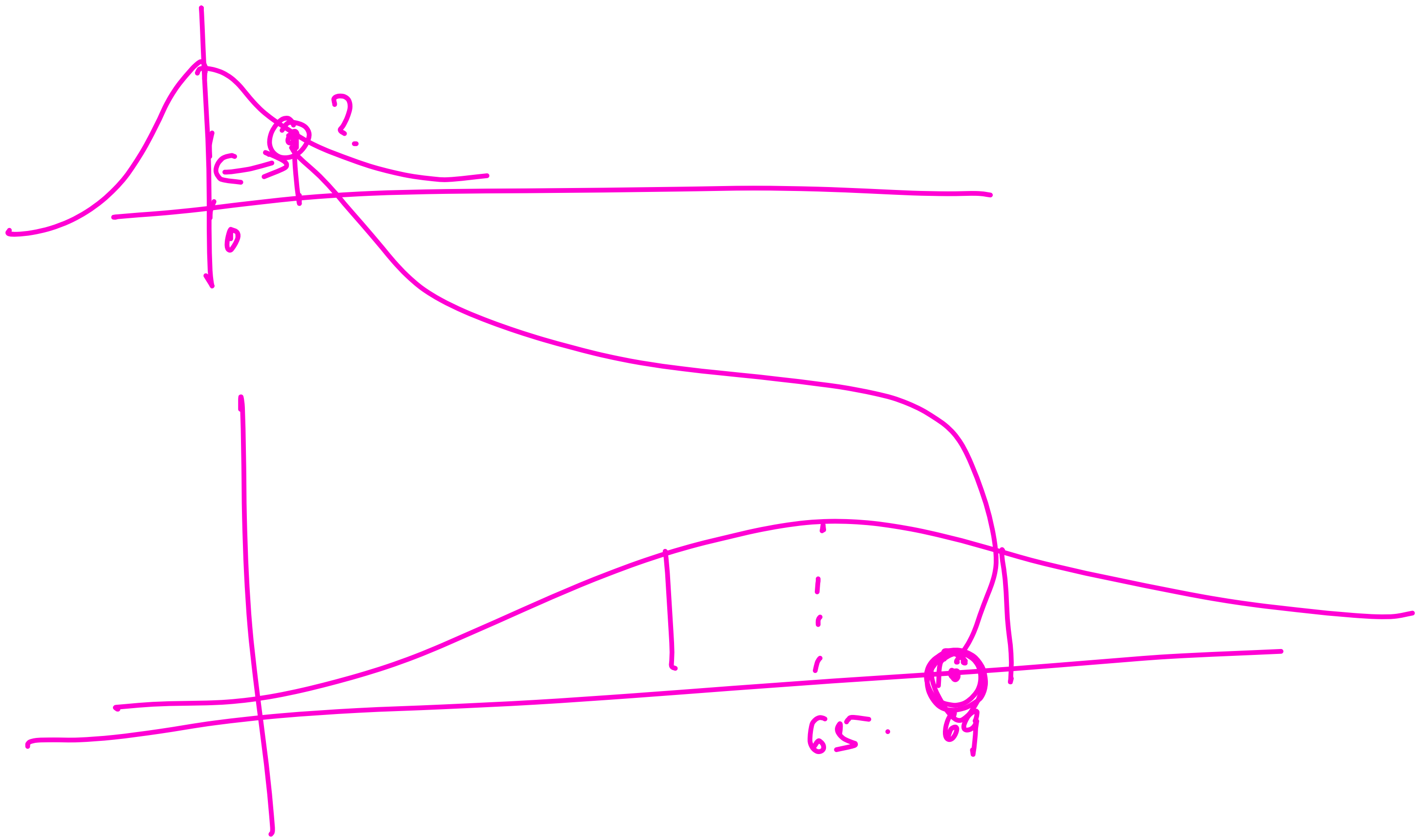
$X \to (65, 2.5)$.

Gaussian $\to$ 69.

$$Z = \frac{x-u}{s}$$

$$Z * std + u = x.$$





$$z = \frac{x - u}{s}.$$

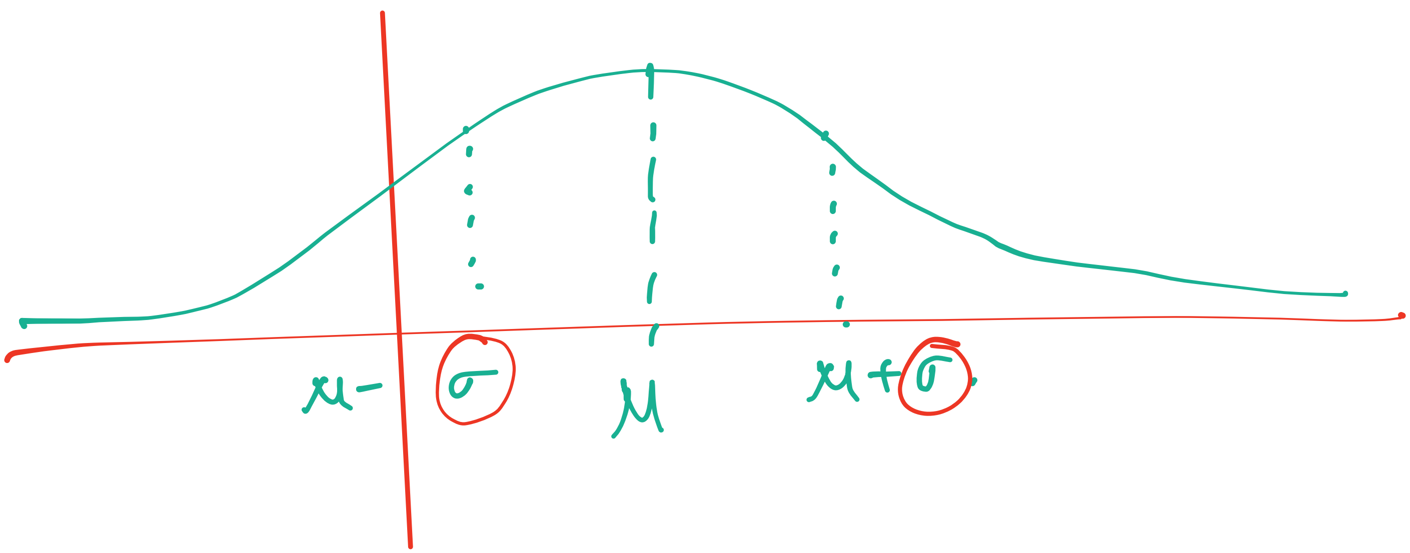
(Amazon) Skaters take a mean of 7.42 seconds and std = 0.34 s for 500 meters, and this data follows the Gaussian distribution.

$X \rightarrow$ Gaussian distributed, time for 500m.

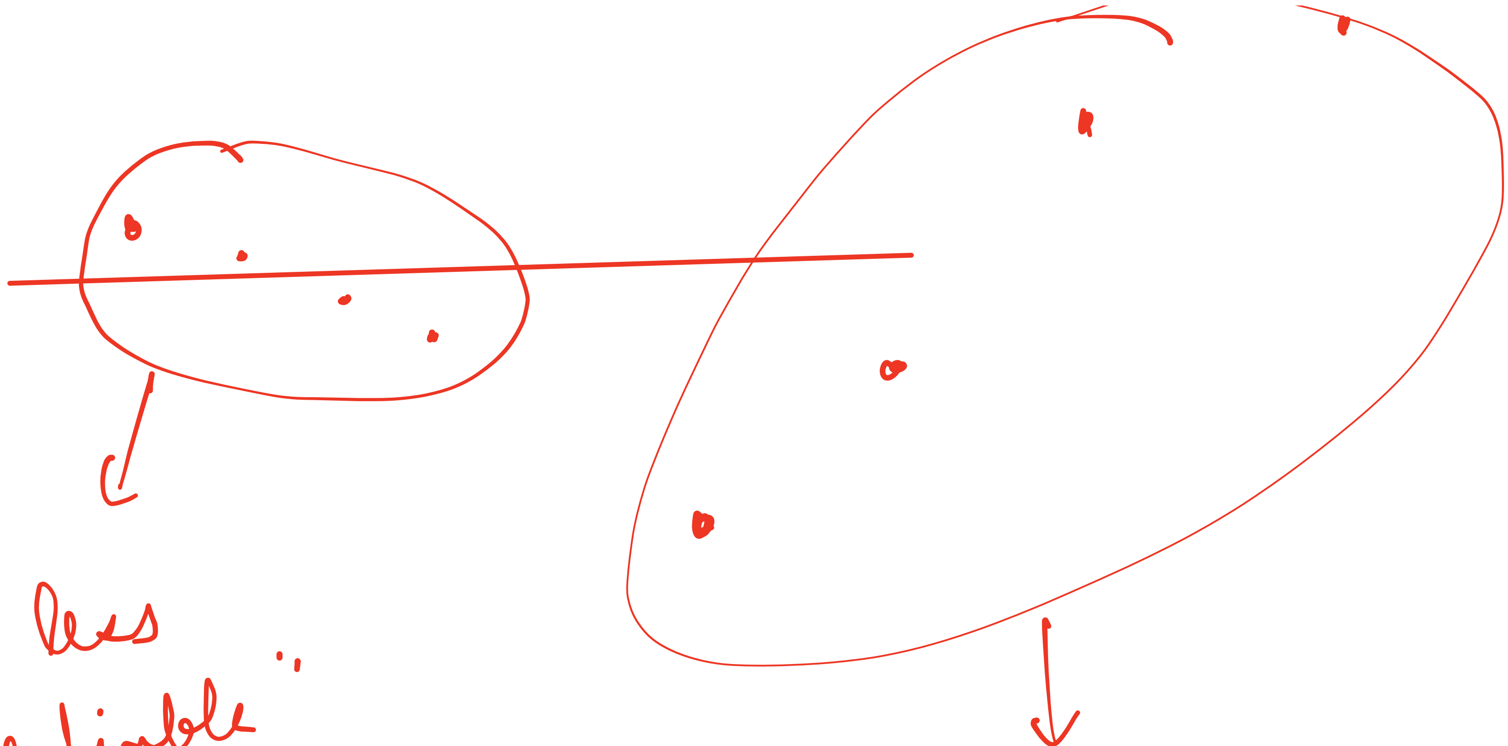
$$S = \frac{500}{X}.$$

Variance

std.



68/95/99.



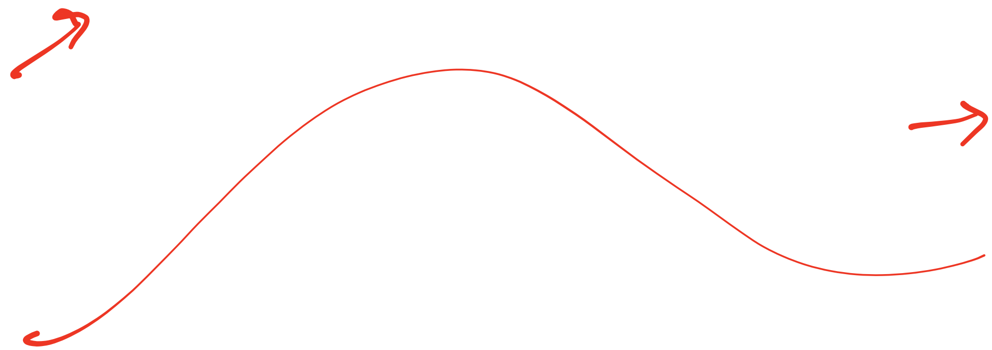
less
"reliable".

more.
"unreliable"

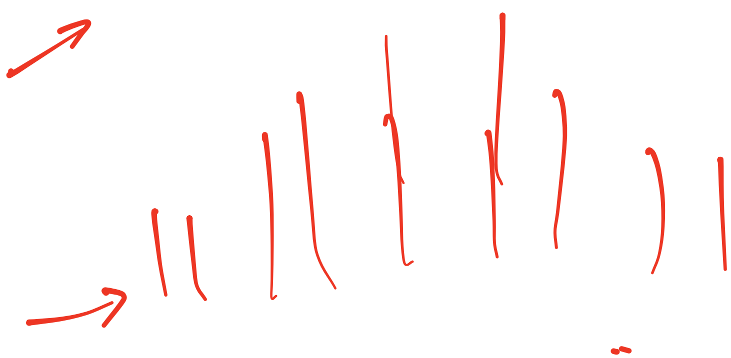
For the variance, we divide by:

(a) n → Biased estimate of the variance.

(b) $(n-1)$ → Unbiased estimate of the variance.



→ True.



→ Estimated distribution.