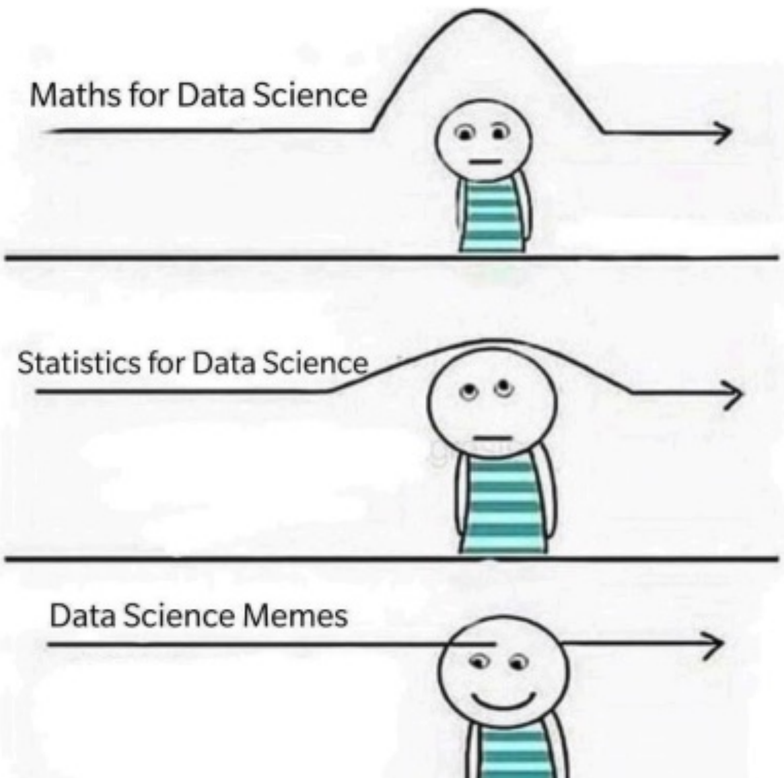
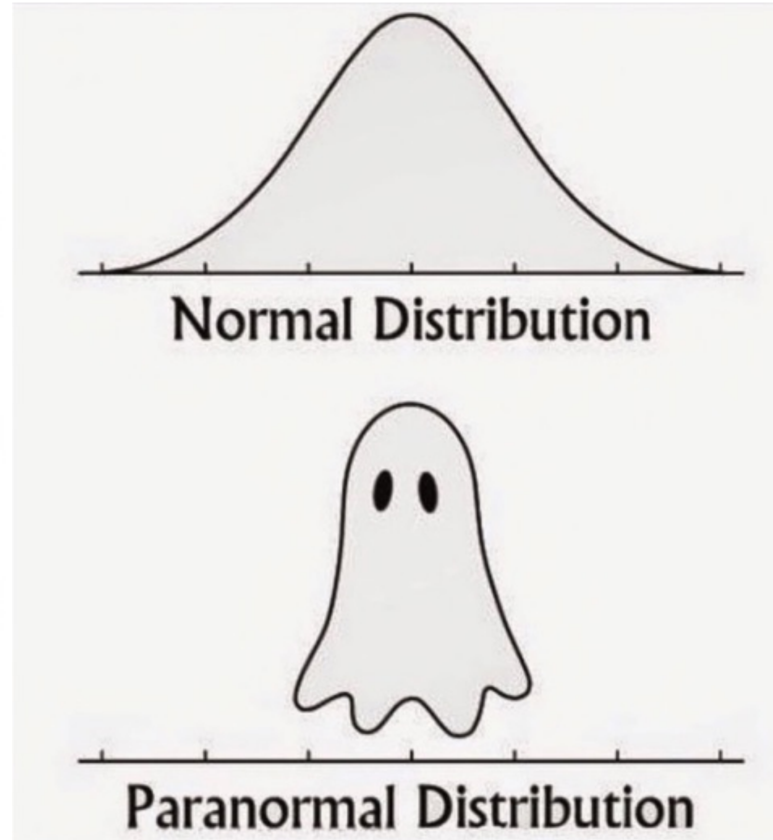

Maths for Data Science
Statistics for Data Science
Data Science Memes


Normal Distribution
Paranormal Distribution

```
df["Height"].describe()
```

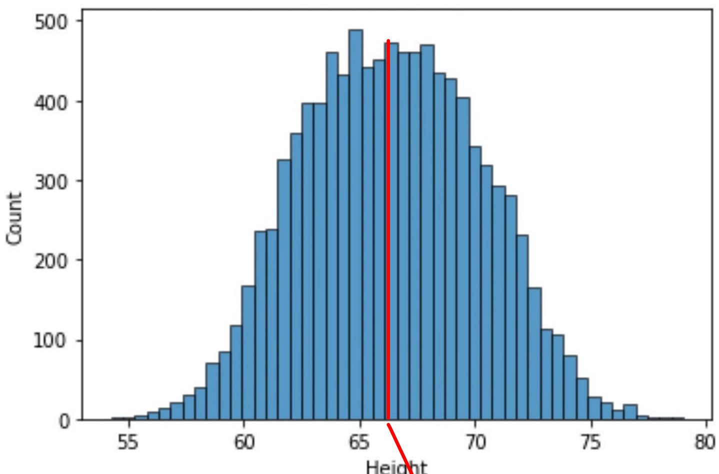
```
count    10000.000000
mean        66.367560
std          3.847528
min         54.263133
25%         63.505620
50%         66.318070
75%         69.174262
max         78.998742
Name: Height, dtype: float64
```

$\mu = 66.36$

game

66.36 → ✓ \$100

63 → ✓ \$3



μ

70 guess

66 actual

$\Rightarrow +4$

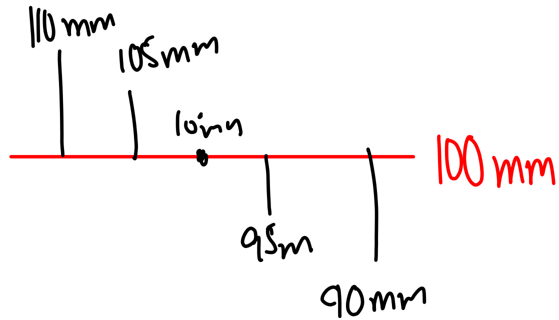
74 → +4

66 → −4

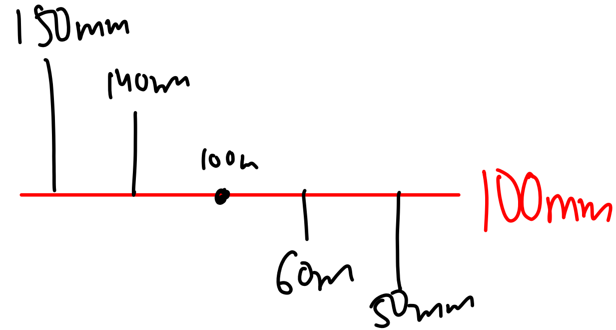
$\}$ +4

Contractor Cricket Balls

M1



M2



How far are these values away on an avg from mean?

$110 - 100 \rightarrow 10$ mm

$105 - 100 \rightarrow 5$ mm

$95 - 100 \rightarrow -5$ mm

$90 - 100 \rightarrow -10$ mm

error

$10 + 5 + (-5) + (-10) = 0$

$\left.\begin{matrix} 10 \\ 5 \\ 5 \\ 10 \end{matrix}\right\}$ error $= \dfrac{30}{4} = \underline{\underline{7.5}}$

Absolute Error

$100 + 25 + 25 + 100 = \dfrac{250}{4} = \underline{\underline{62.5}}$

$$\frac{\text{Sum of Square of errors}}{n}$$

$$\frac{(110-100)^2 + (105-100)^2 + (95-100)^2 + (90-100)^2}{n}$$

$$\text{Variance} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \quad mm^2$$

(σ^2)

$$\Rightarrow \text{Std} = \sqrt{\text{Variance}} = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

(σ)

M1

$\text{Variance}_{M1} = 62.5\ mm^2$

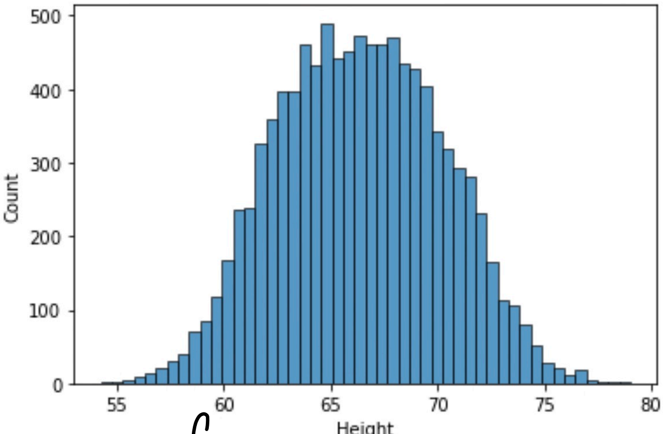
$\text{Std} = 7.905\ mm$

M2

$\text{Variance}_{M2} = 2050\ mm^2$

$\text{Std} = 45.27\ mm$

Distribution

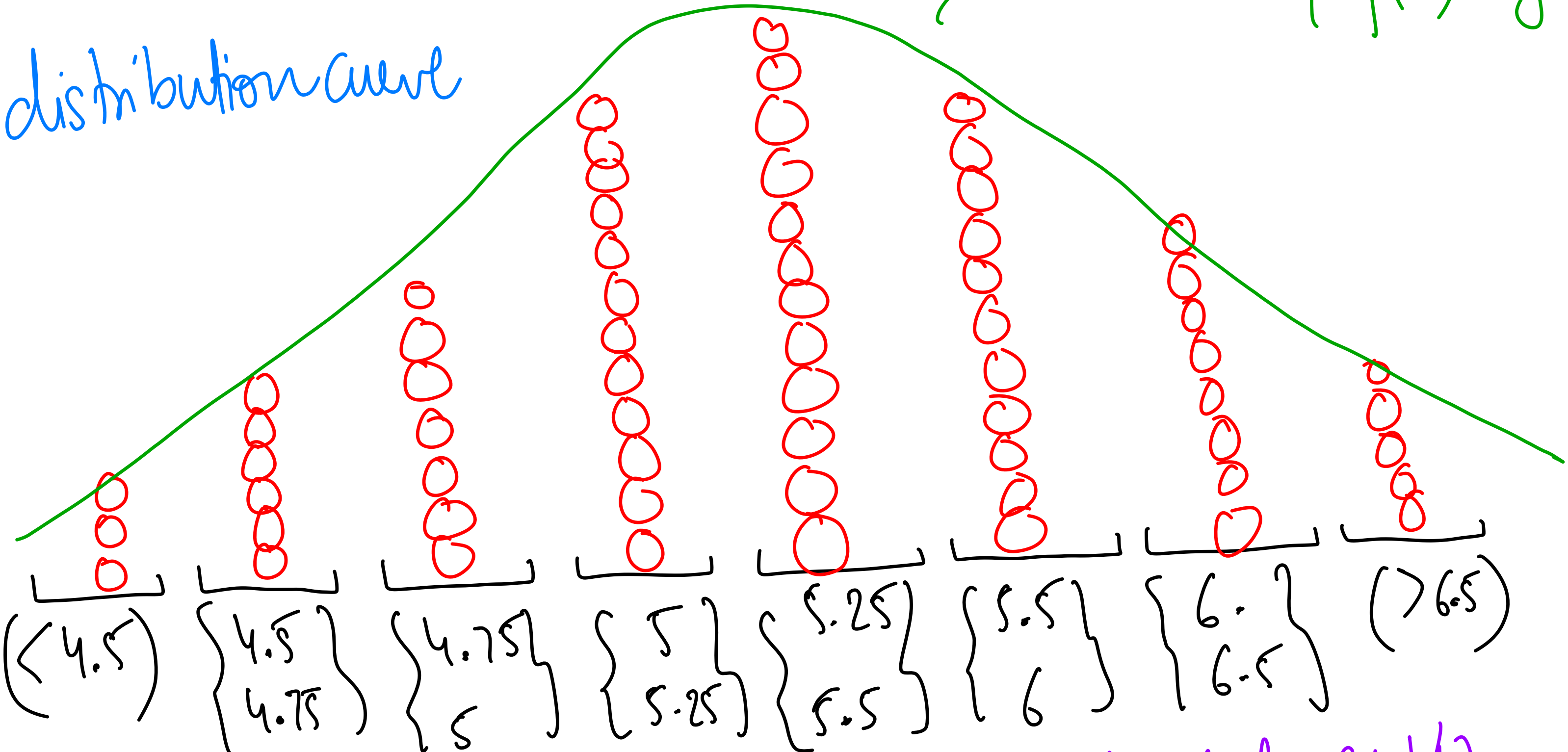


feet

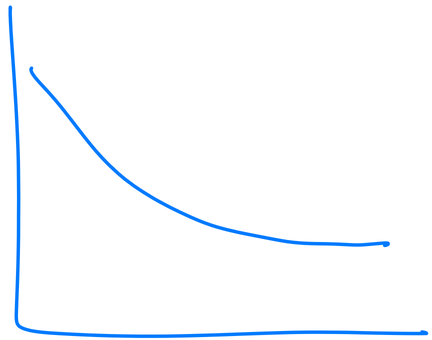
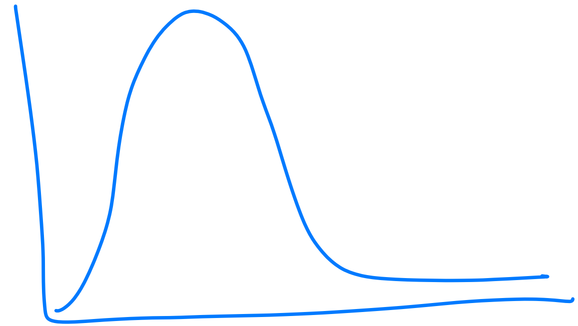
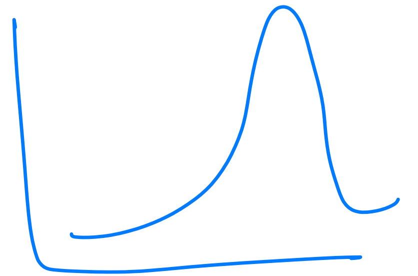
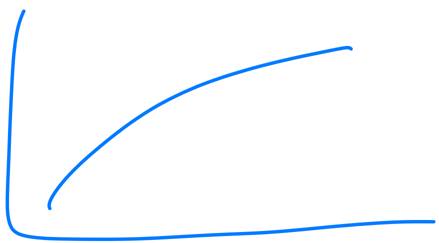
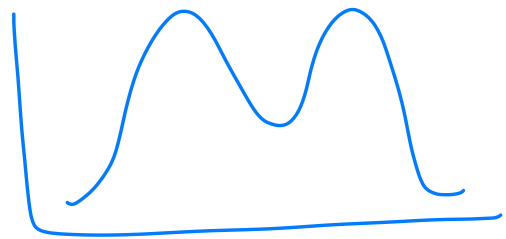
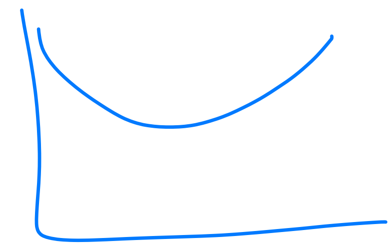
100 people

distribution curve

math eq $f(x) = y$

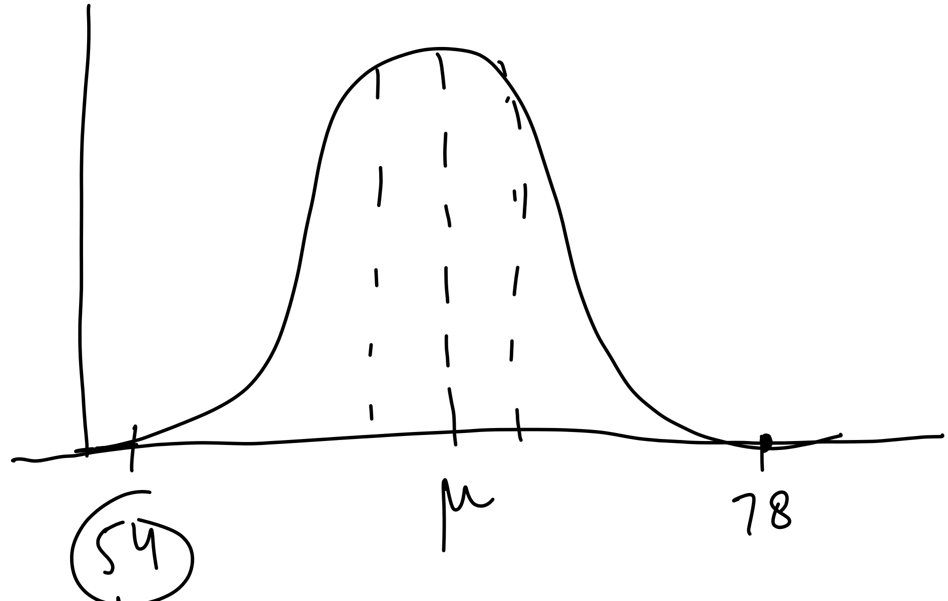


what fracⁿ of people b/w 4.5 - 4.75 ? $\frac{6}{100}$ | prob. of height b/w 4.5 to 4.75 = $\frac{6}{100}$

Normal distribution
Gaussian distribution
Bell Curve

~~density~~
~~prob~~
~~fraction~~



pdf → prob density function

Variable

probability Mass function

probability density function

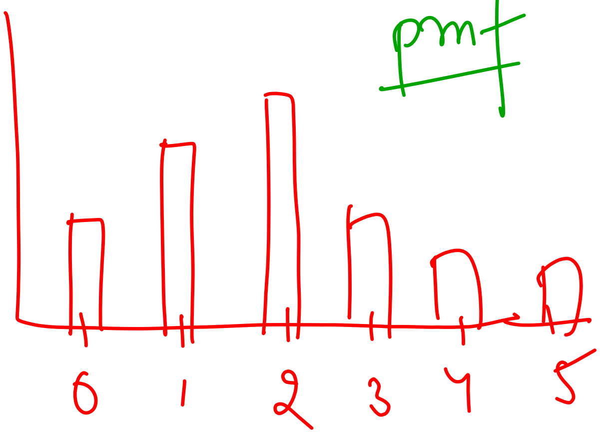
DISCRETE

1, 2, 3, 5, 4

0, 1, 2, 3, 4, 5, 0, 1, 1, 1

pmf

dim

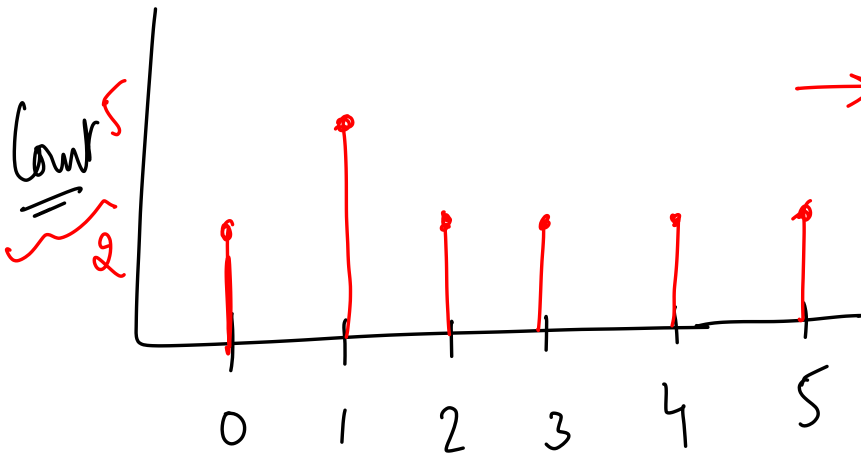


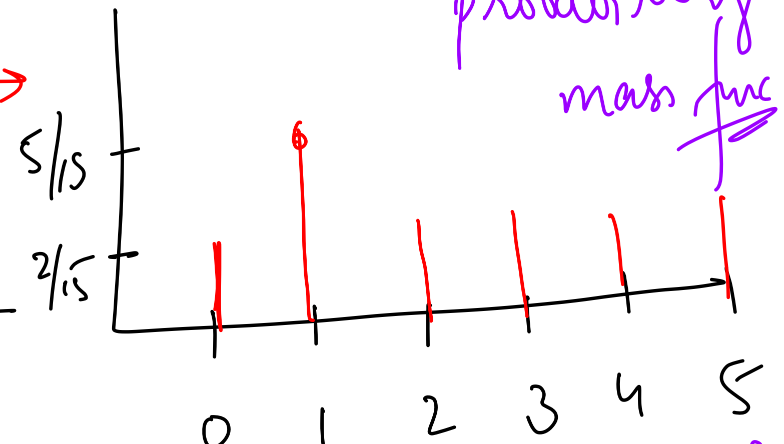
CONTINOUS

64.5, 66.36, 45.9, 54.23

pdf

density prob

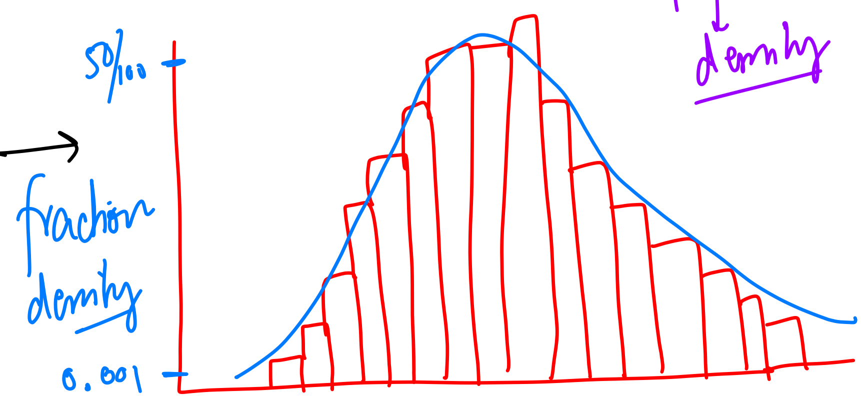


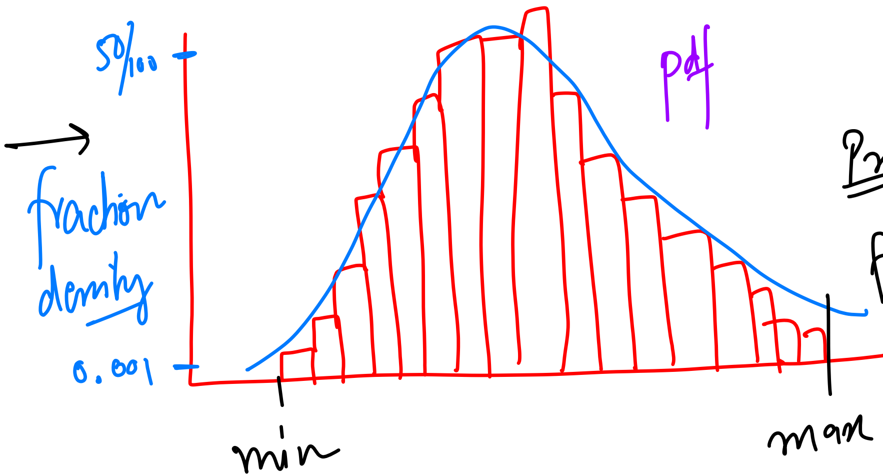


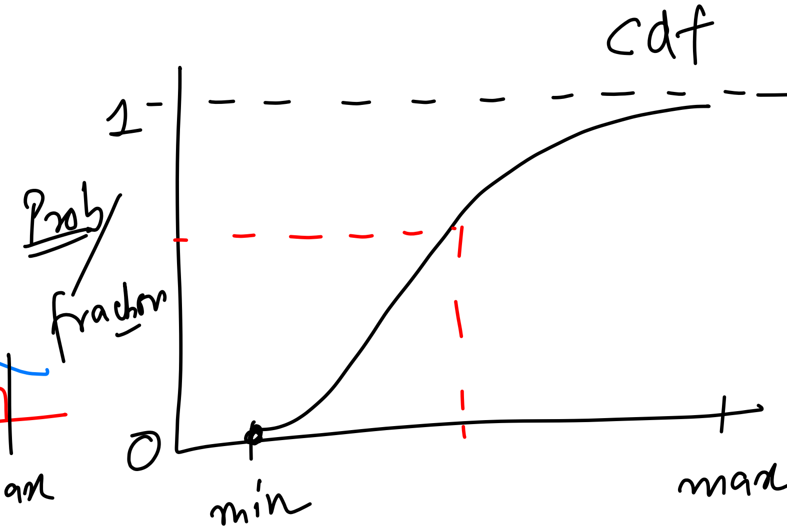
probability mass func.

pdf
density

1000







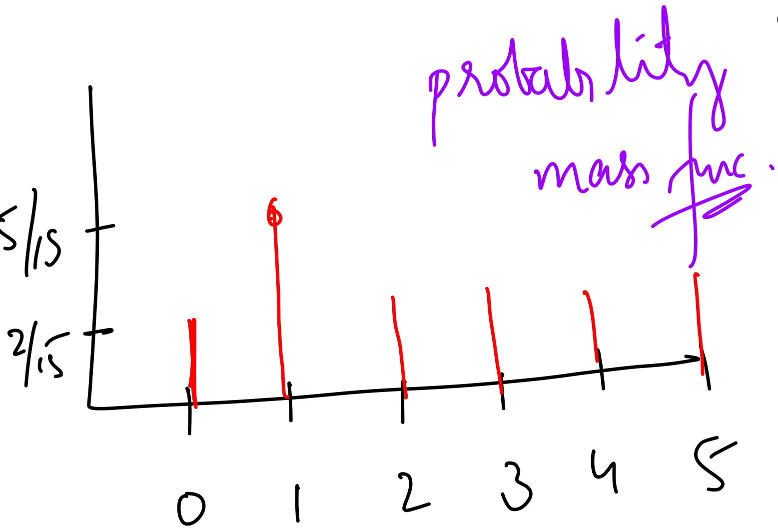
$P[X < x] = \text{Area}$

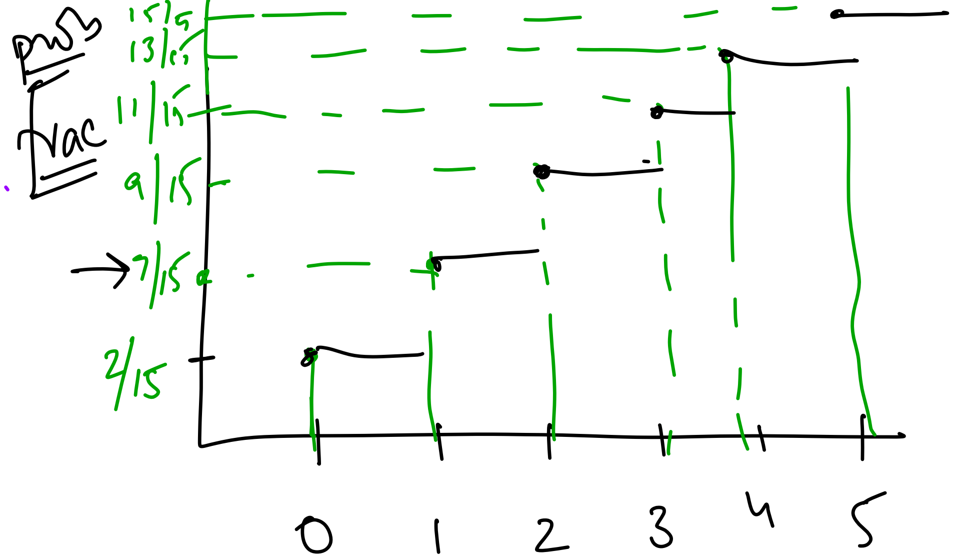
$P[X = 64.56396921\,9] \approx 0$

$P[x_1 < X < x_2]$

$P[X < x] = y \text{ value}$

$P[X \leq x]$





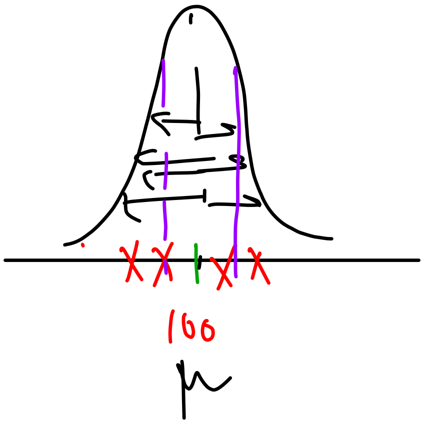
$$P[X=2] = 2/15$$

cdf

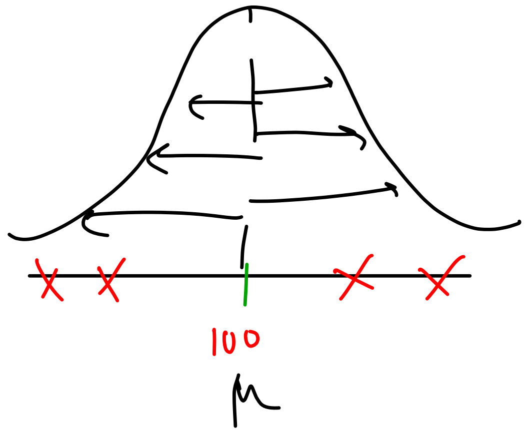
$$P[X \leq 2] = 9/15$$

$$P[X < 2] = 7/15$$

M1



M2



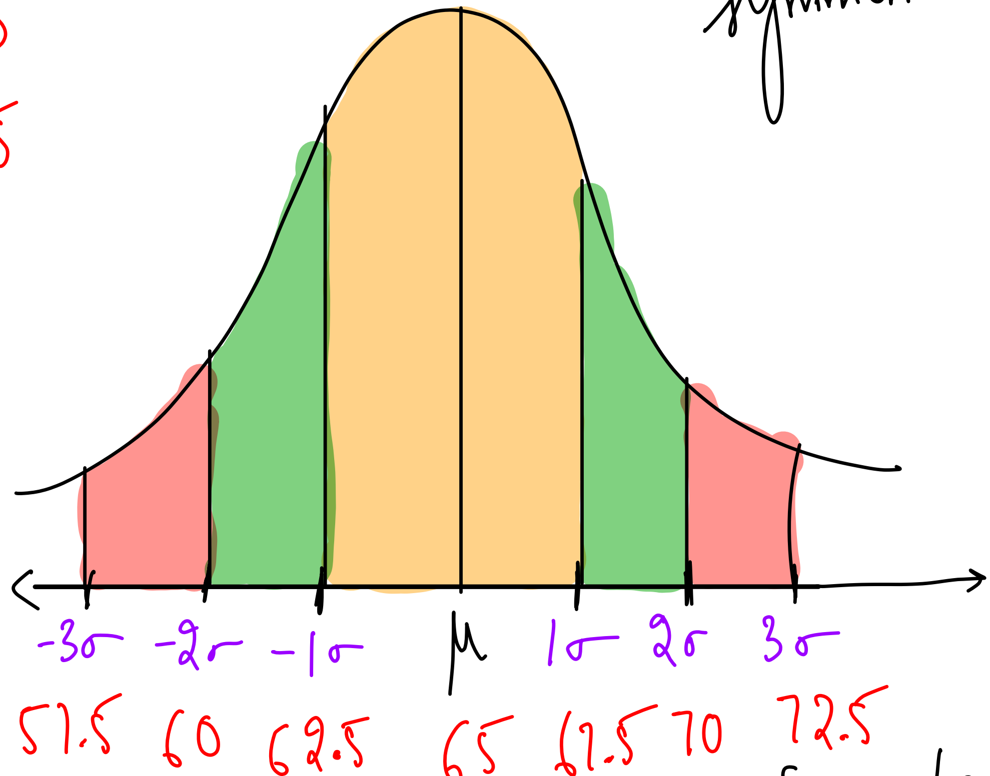
$\sigma(M1) < \sigma(M2)$

- → Normal
- → Exponential
- → Binomial
- → Poisson

$\mu = 65$

$\sigma = 2.5$

symmetrical



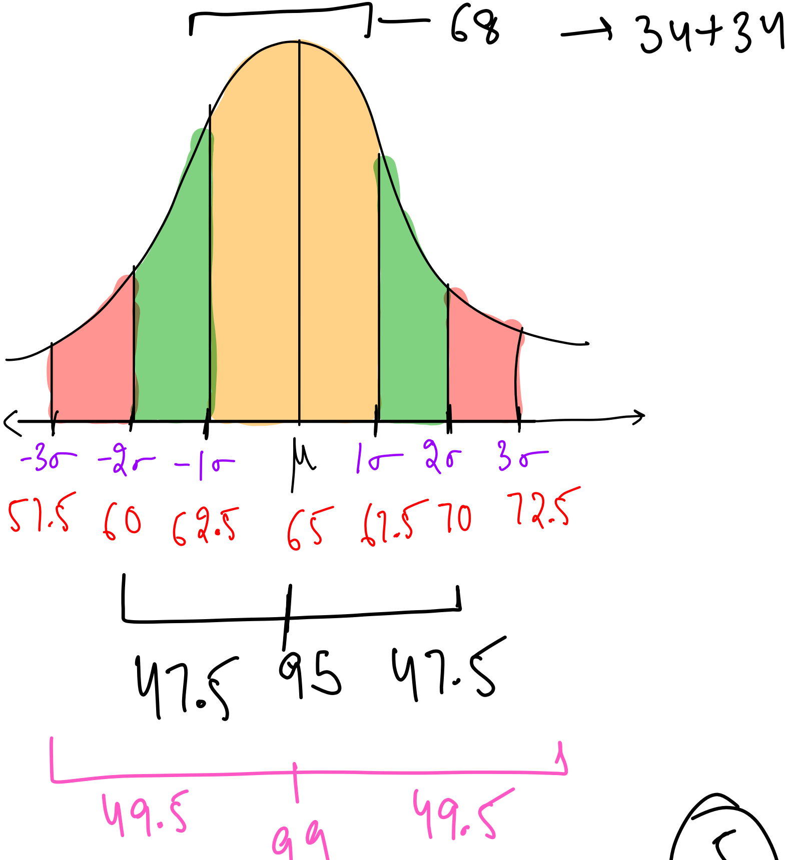
IQ

Shoesize

Appraisal

68/95/99

{68/95/99} → Rule



① $P[62.5 < x < 67.5]$ $(\pm 1\sigma)$

$= 0.68$

② $P[60 < x < 70] =$ $(\pm 2\sigma)$

$= 0.95$

③ $P[57.5 < x < 72.5]$ $(\pm 3\sigma)$

$= 0.99$

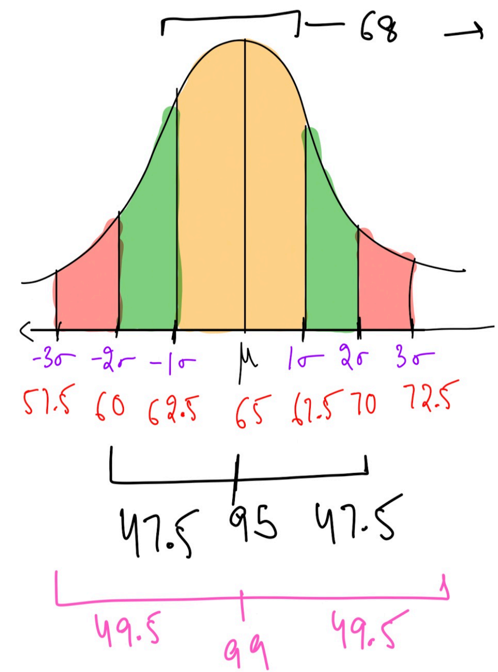
④ $P[60 < x < 67.5] = \underline{0.815}$

⑤ $P[57.5 < x < 70] = 0.97$

The height of people is Gaussian with mean 65 inches and standard deviation 2.5 inches. What is the fraction of people whose height is between 60 and 72.5?

41 users have participated

A	0.68	7%
B	0.895	5%
C	0.9735	78%
D	0.997	10%

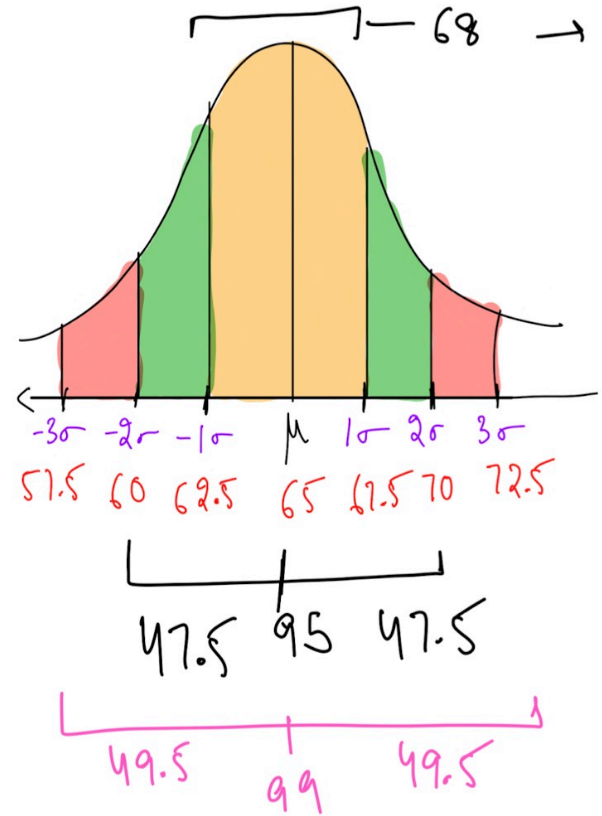


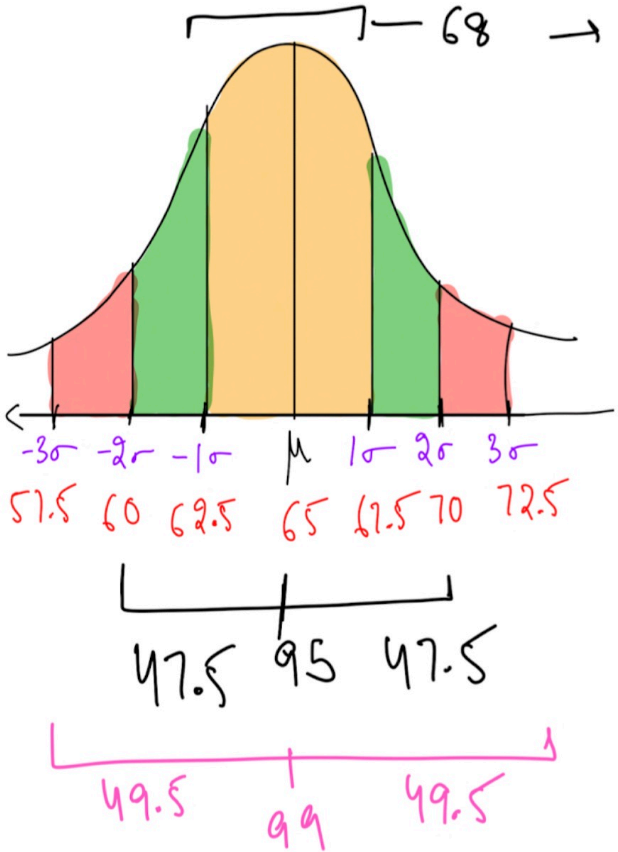
The height of people is Gaussian with mean 65 inches and standard deviation 2.5 inches.
What fraction of people are shorter than 67.5?

40 users have participated

A	0.34	10%
B	0.68	20%
C	0.84	68%
D	0.95	2%

$$P[X < 67.5] \approx 0.84$$





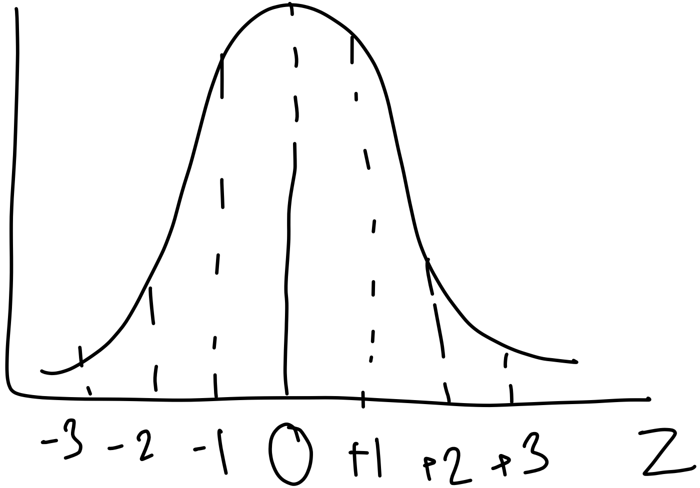
$$57.5 = 65 + (-3)2.5$$
$$60 = 65 + (-2)2.5$$
$$62.5 = 65 + (-1)2.5$$
$$65 = 65 + 0(2.5)$$
$$67.5 = 65 + (+1)(2.5)$$
$$70 = 65 + (+2)2.5$$
$$72.5 = 65 + (+3)2.5$$

\# std dev away from mean → +ve, -ve

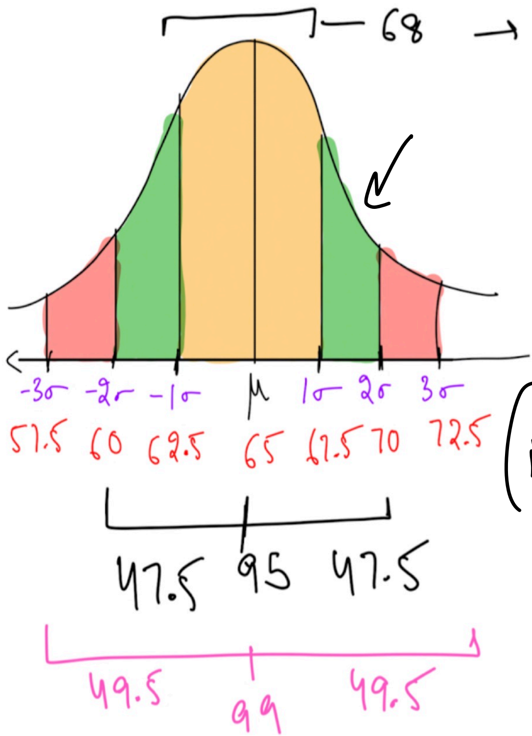
$$x = \mu + (z)\sigma$$

$$\boxed{z = \frac{x-\mu}{\sigma}}$$

Z score

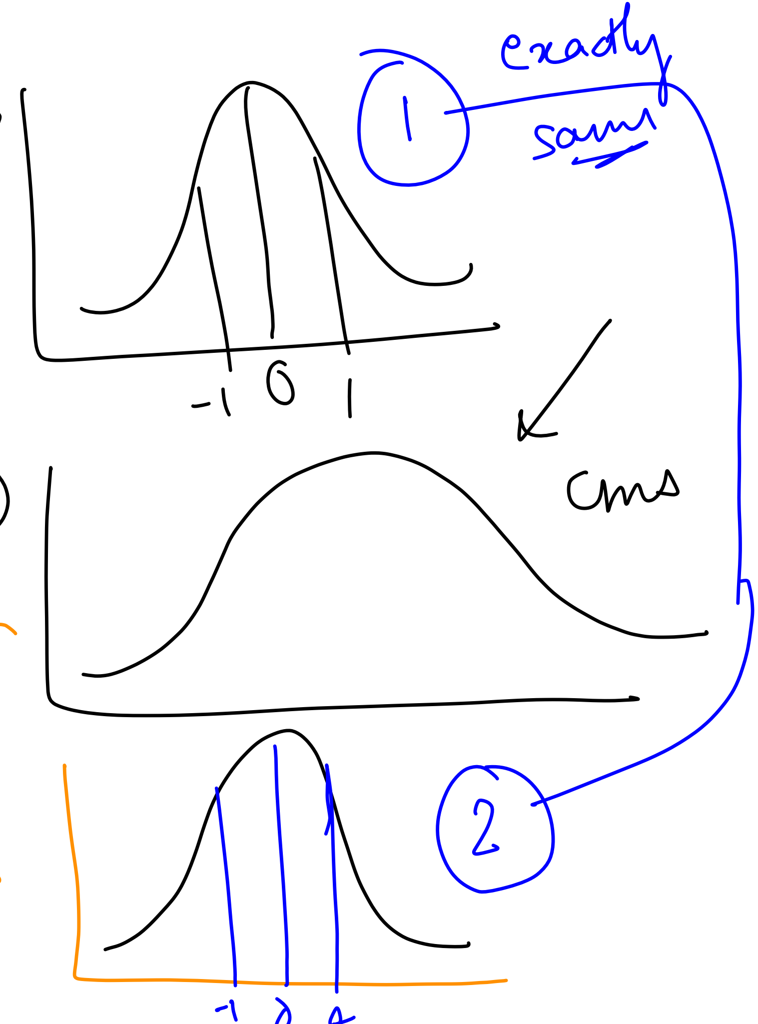


$$Z = \frac{69-65}{2.5} = \frac{4}{2.5} = \underline{1.6}$$



(inches) → (cms)

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



~~Aeshr~~

~~pd. crosstab~~ (margins = normalise)

	KP281	KP481	KP781	
High	200	400	600	1200
Low	700	400	100	1200
	900	800	700	2400

$P[High / KP481] = 400/800$

① Marginal Prob

$P[KP481] = 800/2400$

② Joint Prob

$P[KP481 \cap High] = 400/2400$

③ Conditional Prob

$P[KP481 / High] = 400/1200$

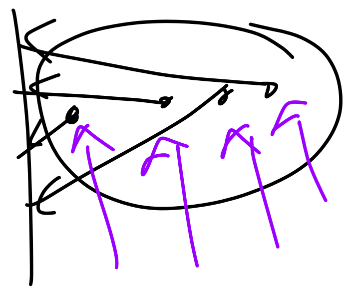
* Outliers

~~Clipping~~

28

O

M



DBSCAN

0, 2, 3, 5, 6, 7, 209

0, 2, 3, 5, 6, 7, 7