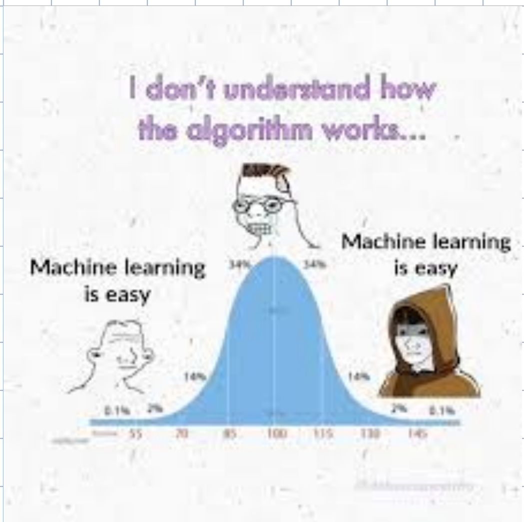
Agenda

Gaussian Distribution



Gaussian Distribution

Normal Distribution
Standard Distribution

Case-study: Remember the heights case study

Q If we plot heights of students collected how will the distribution look like?

Q: Is Height(X) continous or discrete?

Continous

Observations

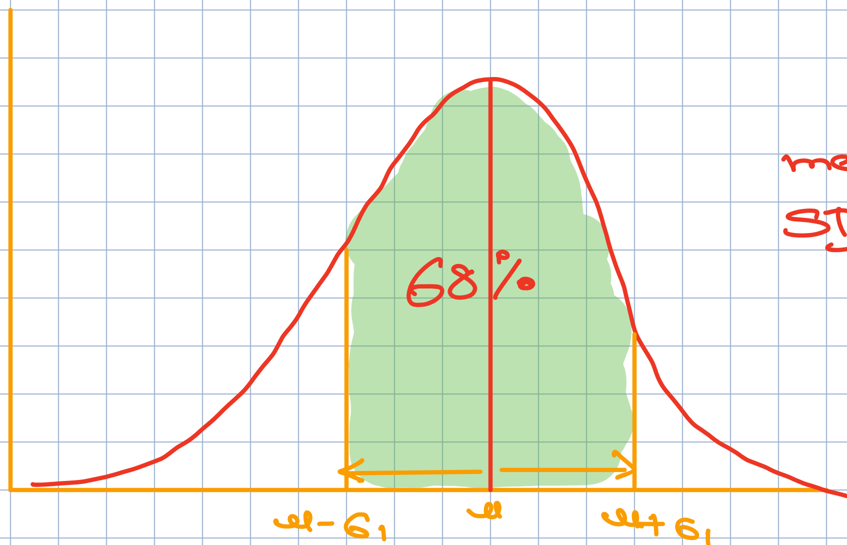
① Curve is Symmetrical around Center

Bell Curve

② Curve is centred around mean

$\pm\sigma \quad \pm 2\sigma \quad \pm 3\sigma$

$\boxed{68/95/99 \text{ Rule}}$



mean ⇒ μ_1

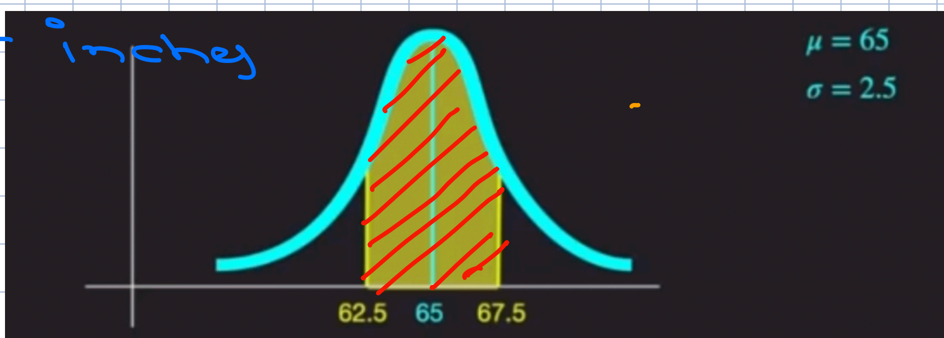
SD ⇒ σ_1

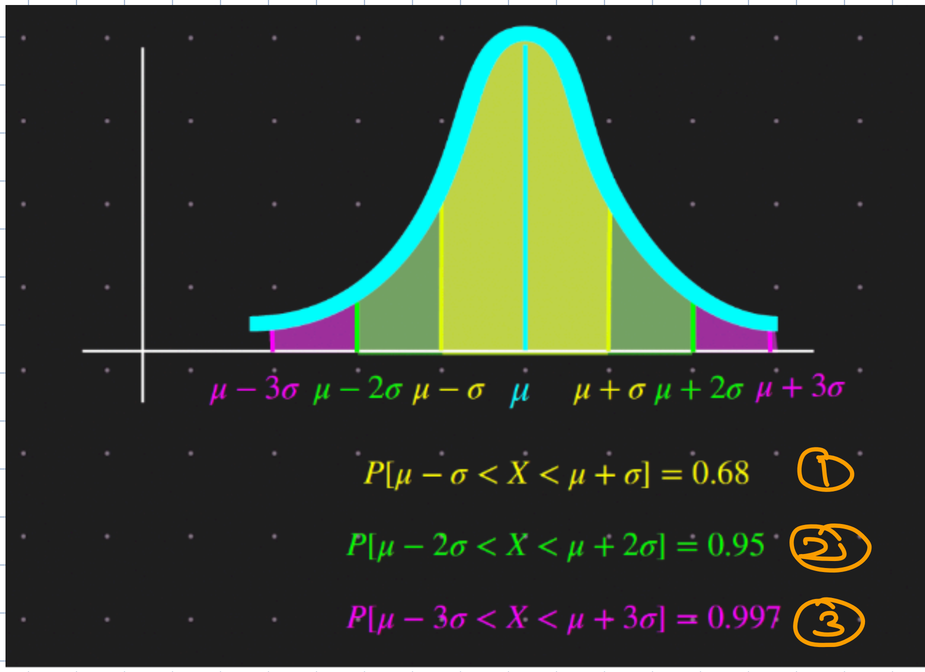
$\mu = 65$ inches

62.5 — 67.5

68% population

$\sigma = 2.5$ inches

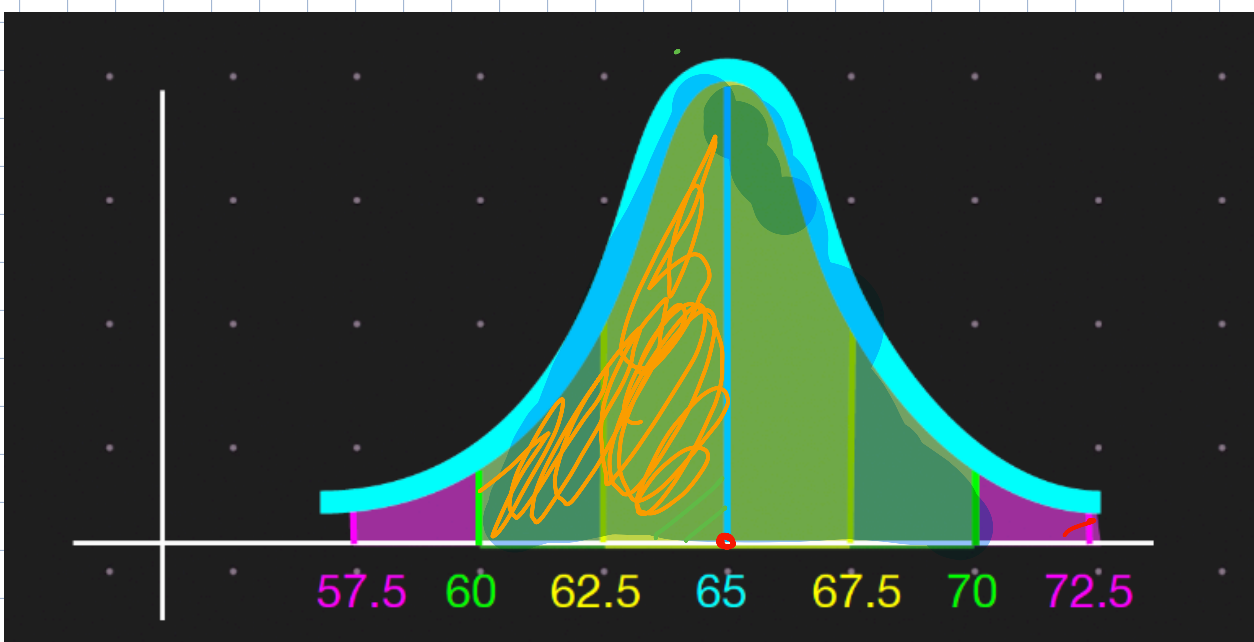




Questions

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?

0.9735



3σ

95%

2σ 2σ

Answers

$\pm 2\sigma \rightarrow 95\%$

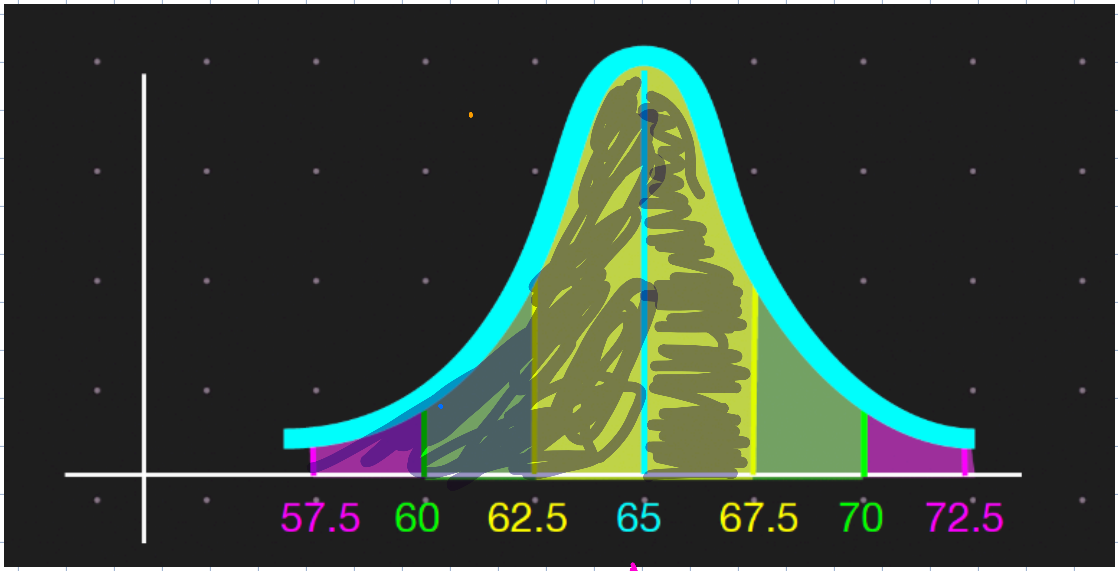
47.5 + 49.85 = 97.35

$\pm 3\sigma \Rightarrow 99\%$

Questions

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?

2 users have participated



50%

16

34%

Answers

50 + 34 → 84 %

Questions

Fraction of people less than 69.1 inches

Z- Score

Q Is the 68/95/99 Rule enough to find P of any Random Value in Gaussian Distribution?

Z Score represent distance of a point from mean in terms of standard Deviation

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

Definition: How far a given value X is from mean in terms of σ (Standard Deviation) as Unit of measurement

Consider the following Question:

```
Suppose The height of people is Gaussian with a mean of 65 inches and a standard deviation of 2.5 inches.
What fraction of people are shorter than 69.1 inches?
```

$$\frac{69.1 - 65}{2.5} = 1.64 \longrightarrow 0.94950$$

94.95 % are shorter than 69.1 inches

https://www.math.arizona.edu/~rsims/ma464/standardnormaltable.pdf

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169

$$\frac{x - 65}{6} \Rightarrow -1.64$$

Z-score to %

Code

from scipy.stats import norm

norm.cdf(z) 1.64

$\xrightarrow{\text{return}}$

.94950

```
u = 65
s = 2.5

x = 69.1

z = round((x - u) / s, 2)

print(z)
```

```
1.64
```

```
from scipy.stats import norm

norm.cdf(z) # 94.94 %of people have height less than 69.1 inches
```

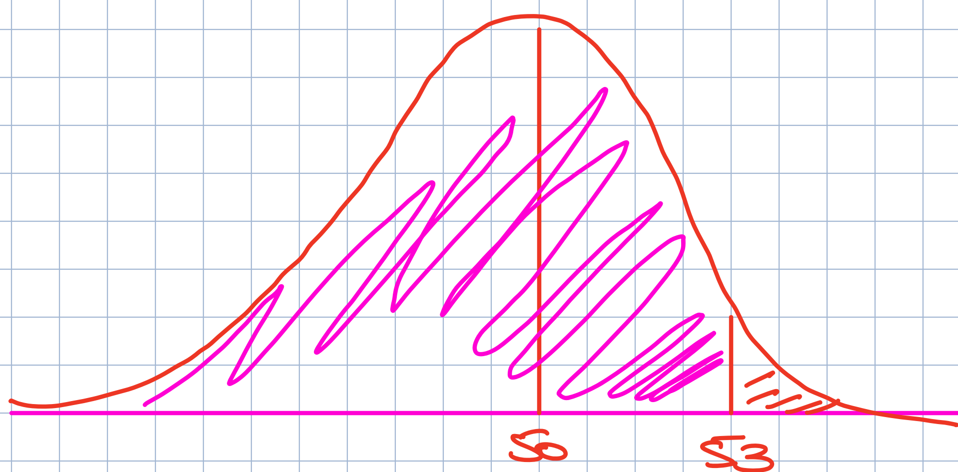
```
0.9494974165258963
```

Balls produced by the manufacturer have a mean diameter of 50 mm and std dev 2 mm. What fraction of balls have a diameter smaller than 53 mm?

13 users have participated

$\mu \rightarrow 50$

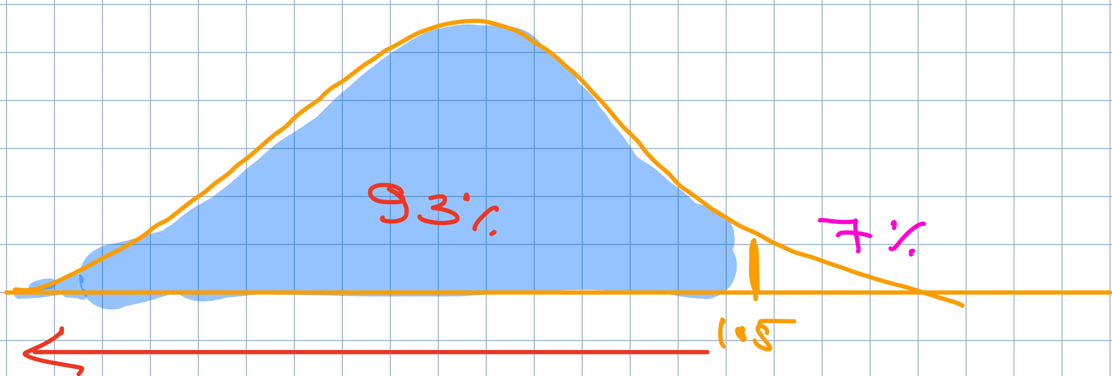
$\sigma \rightarrow 2$



$Z \rightarrow \dfrac{53-50}{2} = 1.5$

Ball $\leq 63 \rightarrow 1.5 \xrightarrow[\text{norm.cdf}]{\text{Z-table}} 93\%$

Balls $> 63 \longrightarrow 100 - 93 \rightarrow 7\%$



PPF

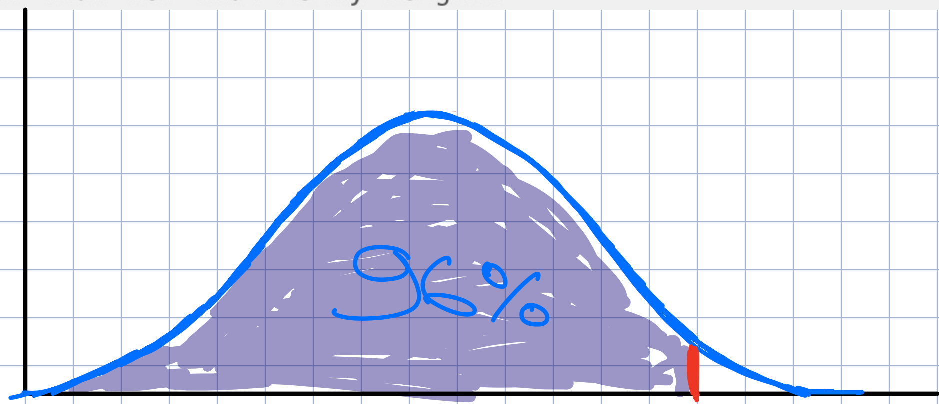
Point Percentile Function

PPF is inverse of Z → CDF

Consider the following Question:

```
The height of people is Gaussian with mean 65 inches and standard deviation 2.5 inches.

One person says:
96% people are shorter than me. What is my height?
```



Percentile ⟶ Z- score

? $Z \rightarrow \frac{x - \mu}{\sigma}$

$$\boxed{x = Z * \sigma + \mu}$$

norm.ppf(percentile) ⟶ Z-score

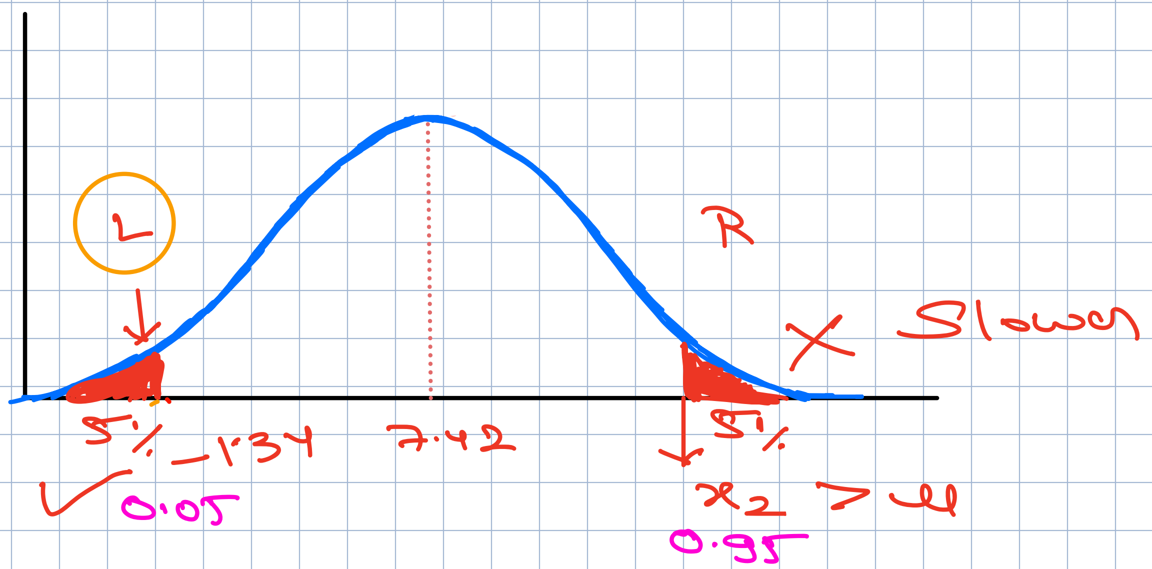
```
z_score = 1.75

x = z_score * 2.5 + 65
x
```

```
69.375
```

Skaters take a mean of 7.42 seconds and std dev of 0.34 seconds for 500 meters.

What should his speed be such that he is faster than 95% of his competitors?



$\mu = 7.42$

σ 0.34

PPF → 0.05

↓

Z-score ⇒ -1.64

↓

Calculate time

6.86

↓

What should his speed be such that he is faster than 95% of his competitors?

4 options

Active Duration (Most preferred: 30 seconds)

Appears for	60 Secs

A	68.6 m/s
B	62.6 m/s
C	72.8 m/s ← Speed
D	83.7 m/s

```python
[9]  z = norm.ppf(0.05)
     z
```

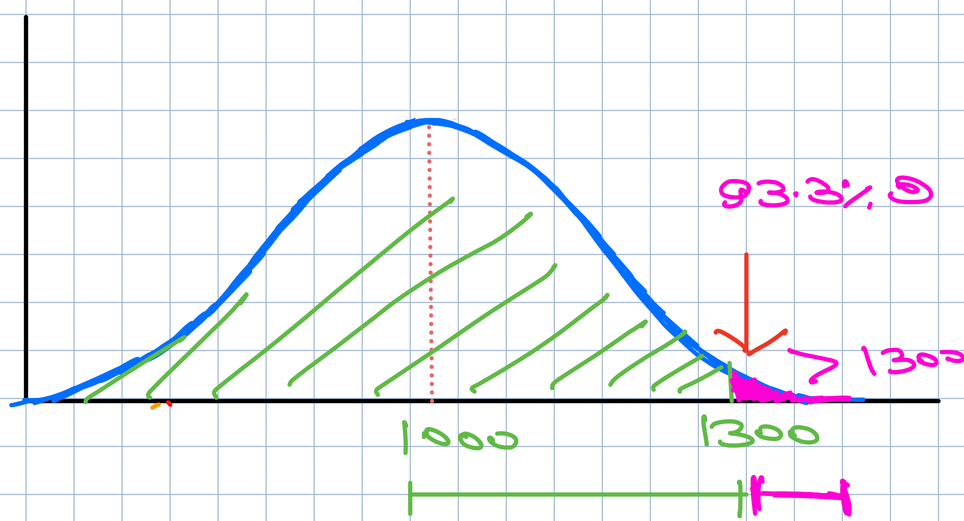
```
np.float64(-1.6448536269514729)
```

```python
[11] mean_time = 7.42
     std = 0.34
     x = z * std + mean_time
     x
```

```
np.float64(6.860749766836499)
```

```python
dist = 500
s = dist/x
s
```

```
np.float64(72.87833210546471)
```



A retail outlet sells around 1000 toothpastes a week, with std dev = 200. If we have 1300 stock units as our inventory, then what is the probability you'd need to replenish stocks within the week?

If sale is less than 1300 Units
we donot Need to Restock

$$\text{Z-score} = \frac{1300-1000}{200} = 1.5 \xrightarrow{\text{cdf}} 93.33\%$$

Probability of Restock $= 1 - 0.933$

$= 0.06$ ✓

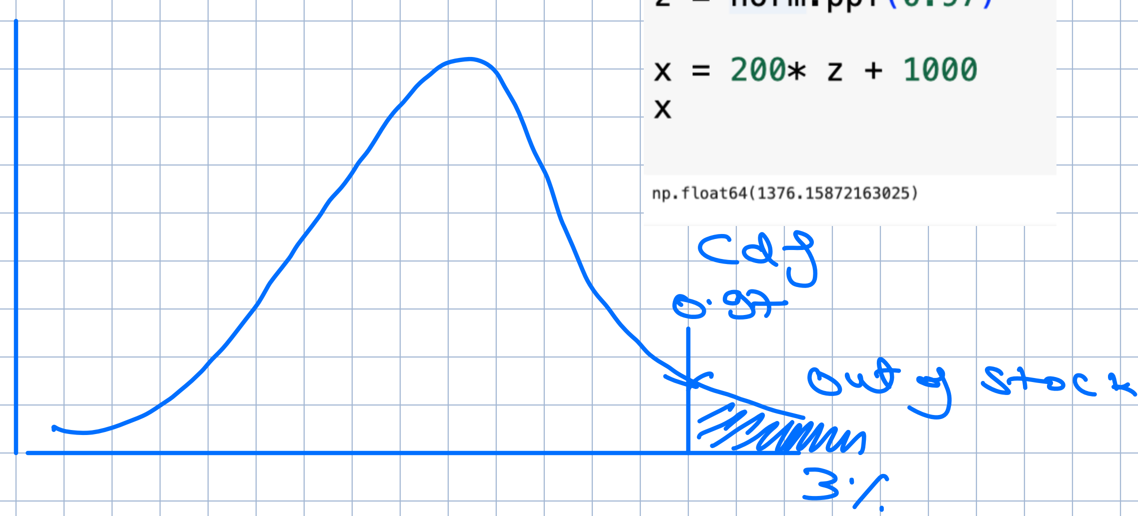
Follow up

How much inventory should you have, such that there is only a 3% chance of running out of stock

```
z = norm.ppf(0.97)

x = 200* z + 1000
x
```

```
np.float64(1376.15872163025)
```

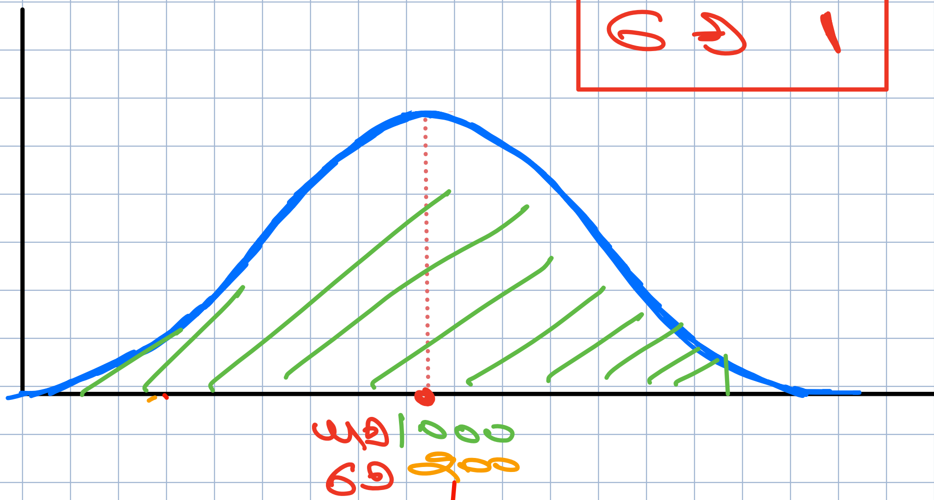


97% won't Need Restocking

Standard Normal Distribution

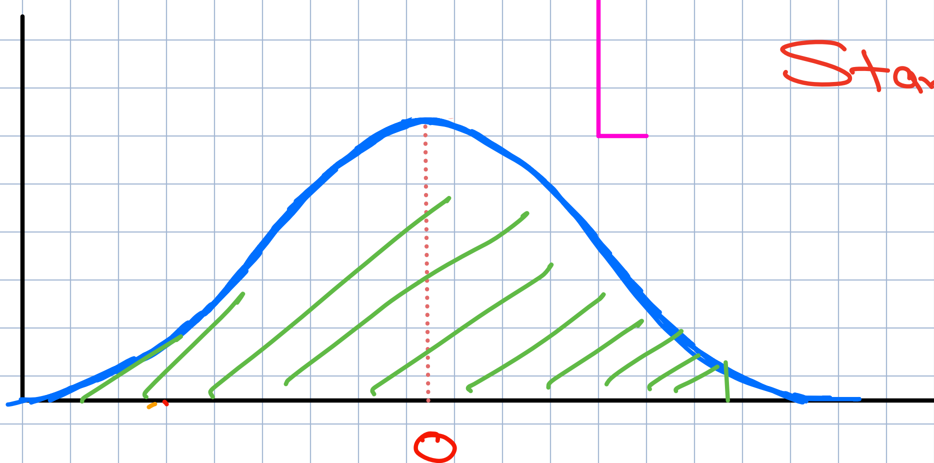
⇒ Special case of Normal Dist

$$\boxed{\begin{array}{l} \mu \Rightarrow 0 \\ \sigma \Rightarrow 1 \end{array}}$$



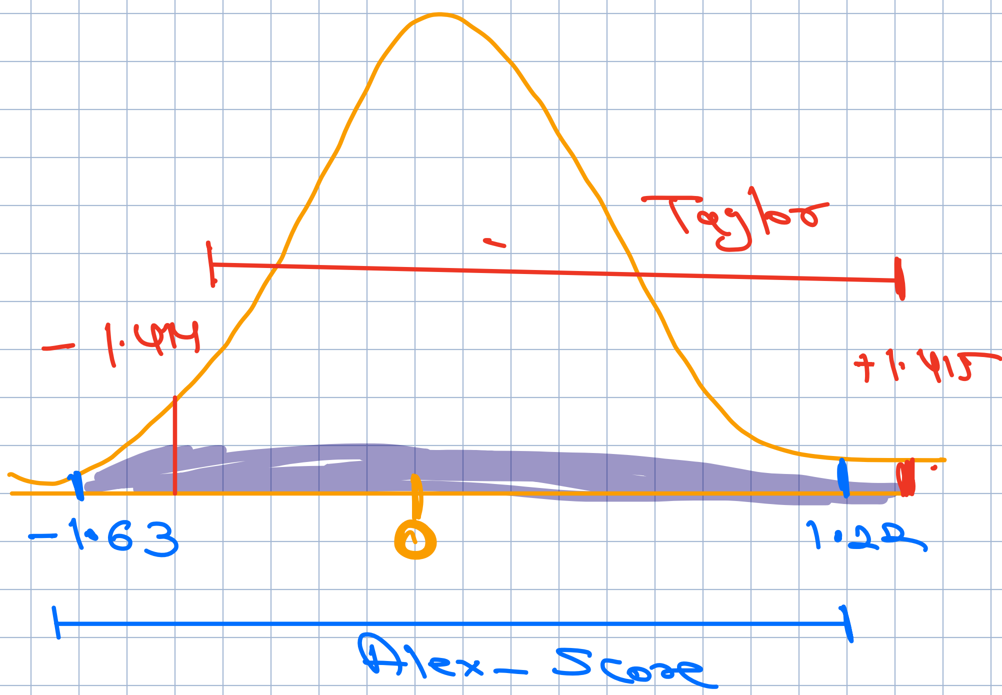
Standardization

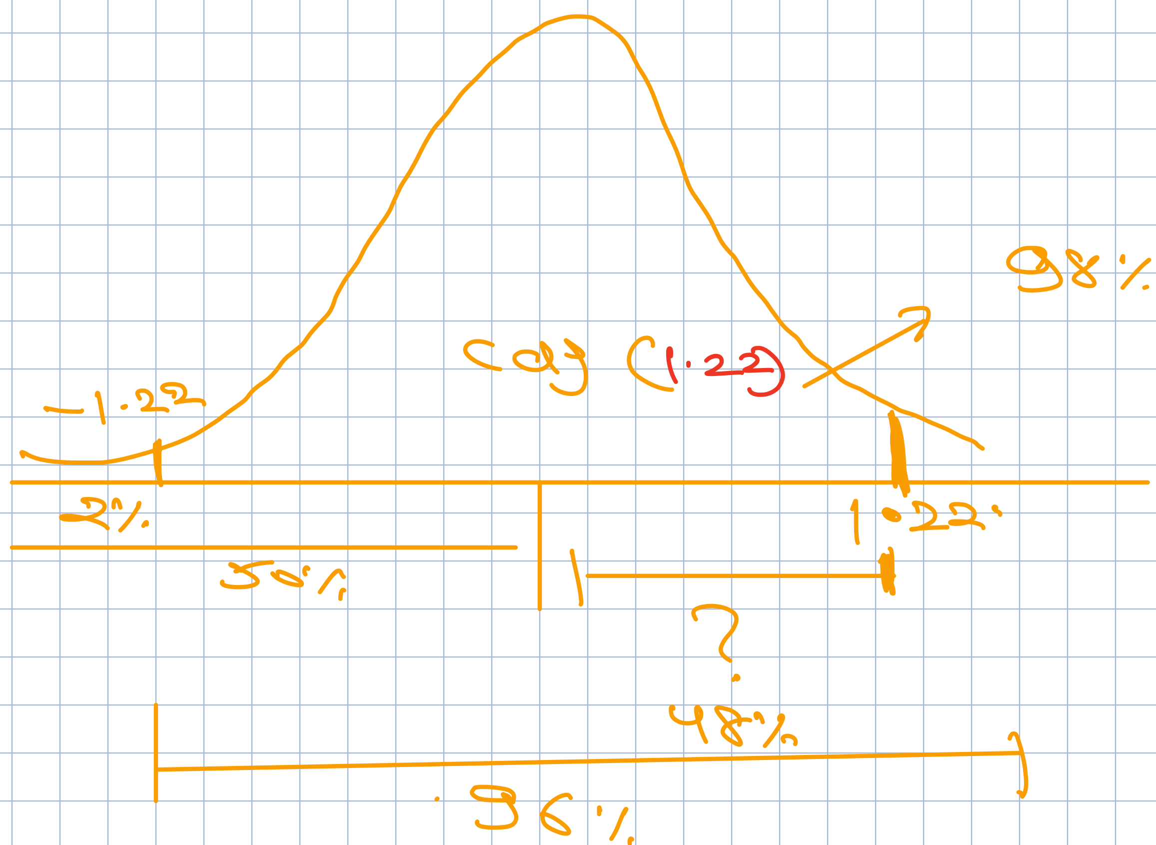
[Process of Converting any Normal Distribution to Standard ND]



How do we perform Standardization

⇒ Convert all Datapoints to the Z-score Values (Z-scaling)


Taylor
- 1.44
+1.415
-1.63
0
1.22
Alex - Score


98%
cdf (1.22)
-1.22
2%
1.22
50%
?
48%
96%