Measure of Variation

The mean, median and mode do not by themselves give an adequate description of our data. We need to know how the observations spread out from the average.

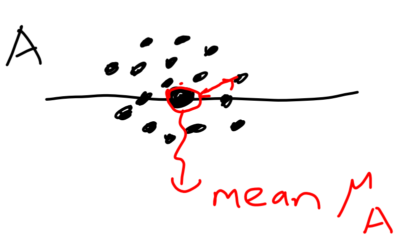
e.g., Consider the following measurements (in L) for two samples of orange juice bottled by companies A and B ⟹

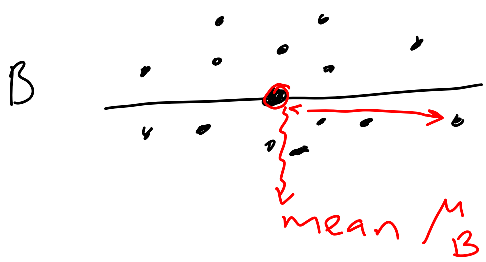
Sample A	0.97	1.00	0.94	1.03	1.06
Sample B	1.06	1.01	0.88	0.91	1.14

$\bar{X}_A = 1, \quad \bar{X}_B = 1$

Both samples have the same mean, 1.0 Liter.

Bottles are more uniform in juice content for A than B.





i.e., the variability or dispersion of the observations from mean is less for sample A than for B.

⊛ Prefer company A

⊛ The range and variance are statistics used for measuring the variability of a set of data.

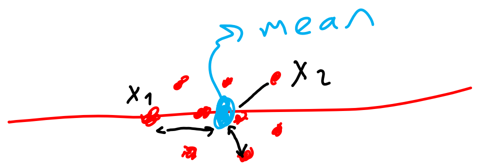
Range = Largest measurement − Smallest measurement.

<u>Example</u>: Find the range for the example of orange juice.

$$R_A = 1.06 - 0.94 = 0.12$$
$$R_B = 1.14 - 0.88 = 0.26$$

} indicating a greater spread in the values for company B.

Variance: It considers the position of each observation relative to the mean of the set (or deviation from the mean = d).



$$\left.\begin{aligned} d &= x_1 - \mu \\ &= x_2 - \mu \\ &\ \vdots \\ &= x_N - \mu \end{aligned}\right\} \text{for population}$$

$$\left.\begin{aligned} d &= x_1 - \bar{x} \\ &= x_2 - \bar{x} \\ &\ \vdots \\ &= x_n - \bar{x} \end{aligned}\right\}$$ for a random sample

<u>Example</u>: Consider the following two sets of data, then, discuss variability of A and B.

Set A	3	4	5	6	8	9	10	12	15
Set B	3	7	7	7	8	8	8	9	15

$R_A = 12,\ R_B = 12;\ \mu_A = 8,\ \mu_B = 8$

Variance or deviation (d)?

Set A	-5	-4	-3	-2	0	1	2	4	7
Set B	-5	-1	-1	-1	0	0	0	1	7

$\circledast$ Most of the deviations of set B are smaller in magnitude than those of set A indicating less variation among the observations of set B.

Population Variance (σ^2): Given the finite population $x_1, x_2, \dots, x_N$, the population variance is

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{\underbrace{N}_{\text{population size}}}$$

Example: Consider the two sets A and B populations ⟹

for set A ⟹ $\sigma_A^2 = \frac{\sum_{i=1}^{N} (x_i - 8)^2}{9} = \frac{(3-8)^2 + (4-8)^2 + \dots + (15-8)^2}{9}$

$\sigma_A^2 = 13.77$

for set B ⟹ $\sigma_B^2 = \frac{\sum_{i=1}^{N} (x_i - 8)^2}{9} = \frac{(3-8)^2 + (7-8)^2 + \dots + (15-8)^2}{9}$

$\sigma_B^2 = 8.66$

Compare σ_A^2 and σ_B^2 ⟹

⊕ Since $\sigma_A^2 > \sigma_B^2$ ⟹ data set for A is more variable than the data of set B.

Population Standard Deviation (σ):

$$\underbrace{\sigma}_{\text{st. dev.}} = \sqrt{\underbrace{\sigma^2}_{\text{variance}}}$$

Example: What is the Standard deviation for the data 7,5, 9, 7, 8, 6 ? (Treat the data as population)

Solution:

$$\mu = \frac{7+5+9+7+8+6}{6} = 7$$

$$\sigma^2 = \frac{\sum_{i=1}^{6}(x_i-7)^2}{6} \Rightarrow \sqrt{\underbrace{\sigma^2}_{\text{variance}}} = \sqrt{\frac{5}{3}} \Rightarrow \underbrace{\sigma}_{\text{st. dev.}} = 1.29$$

Sample Variance (s^2): Given a random sample $x_1, x_2, \ldots, x_n$, the sample variance is

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} \quad \text{OR} \quad s^2 = \frac{n \times \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$$

Example: Coffee prices at 4 randomly selected grocery stores are 12, 15, 17, 20 cents for a jar of 200 g. Find the variance of this random sample of price increase.

Solution:

$$\bar{X} = \frac{12+15+17+20}{4} = 16 \text{ cents}$$

$$s^2 = \frac{\sum_{i=1}^{4}(x_i-16)^2}{4-1} = \frac{(12-16)^2 + \cdots + (20-16)^2}{3} = \frac{34}{3}$$

St. dev. of sample $\Rightarrow s = \sqrt{\frac{34}{3}} = 3.36$

Z-Score: An observation X from a population with mean μ and st.dev. σ has a Z-score or Z value defined by $Z = \frac{X-\mu}{\sigma}$.

X : any observation.

A z-score measures how many st.deviations an observation is above or below the mean.

Example: If we assume that the student made a grade of 82 in chemistry and a grade of 89 in economics. Can we conclude that she is a beter student in economics than in chemistry ?

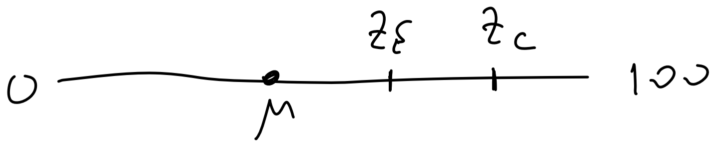
$\mu = 68$, $\sigma = 8$ in chemistry

$\mu = 80$, $\sigma = 6$ in economics

Solution:

$$z_C = \frac{82-68}{8} = 1.75$$

$$z_E = \frac{89-80}{6} = 1.50$$



⊗ grade in chemistry was 1.75 st. dev. above the mean

⊗ " " economics " 1.50 " " " " " .

∴ student's performance in chemistry was higher than her performance in economics.

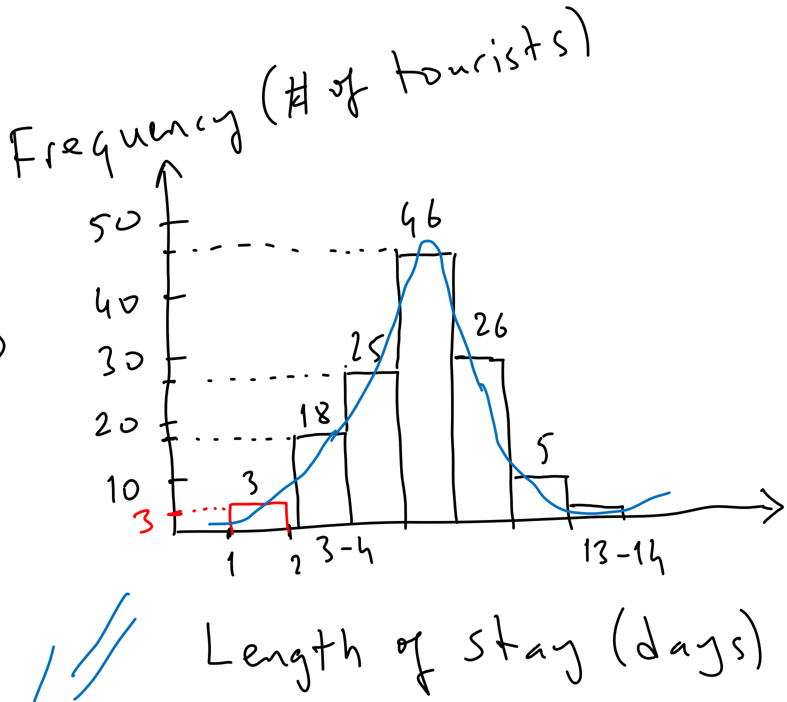
THE NORMAL (Gaussian) DISTRIBUTION (ND)

One of the most important frequency distributions in statistics is the ND.

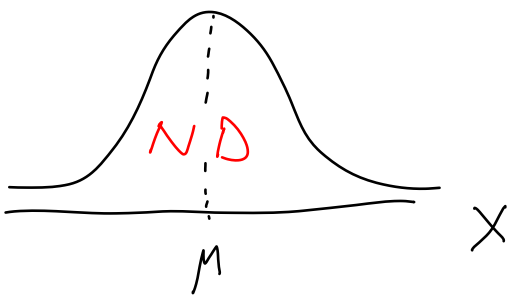
e.g.; The histogram below describes the length of stay of tourists in Antalya.

Days	# tourists
1-2 →	3
3-4 →	18
5-6 →	25
7-8 →	46
9-10 →	26
11-12 →	5
13-14 →	2

⇒



⇒



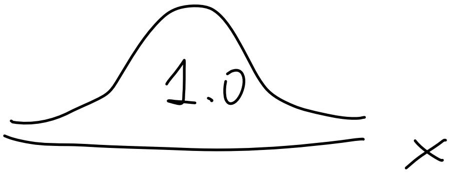
The equation of the normal curve is

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

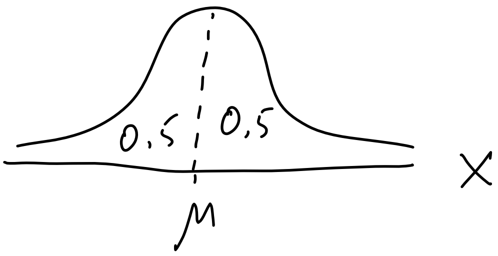
This eqn is not practical to use $\Rightarrow$
Instead, use the TABLES.

Properties of ND Curve:

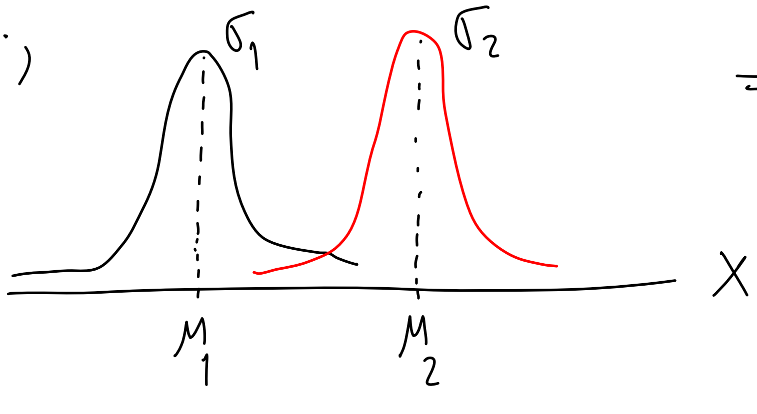
1) The total area under the curve and above the horizontal axis is equal to 1.0.



2) The curve is symmetric about a vertical axis through the mean μ.

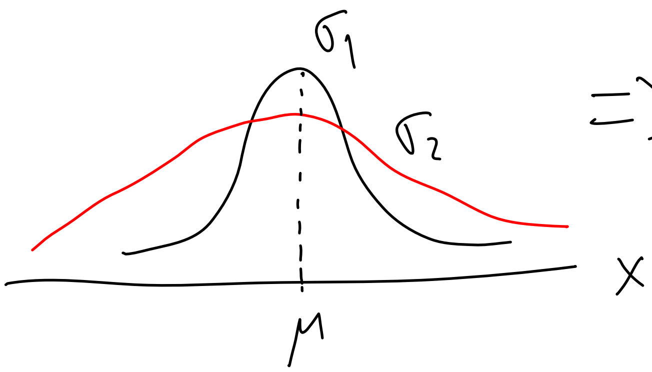


e.g.,



$\Rightarrow \mu_1 \neq \mu_2$

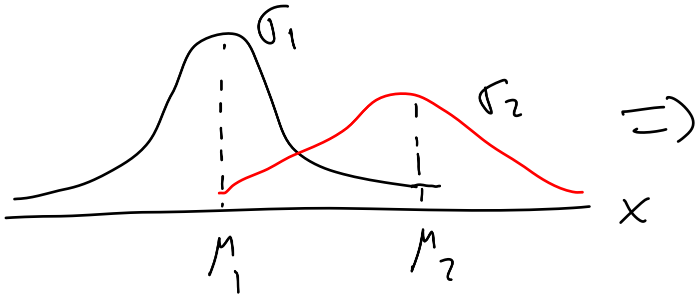
$\sigma_1 \cong \sigma_2$



$\Rightarrow$

$\mu_1 = \mu_2$

$\sigma_1 < \sigma_2$



$\Rightarrow \quad \mu_1 \neq \mu_2$

$\sigma_1 < \sigma_2$

Use the Tables instead of eqn of ND curve. So,

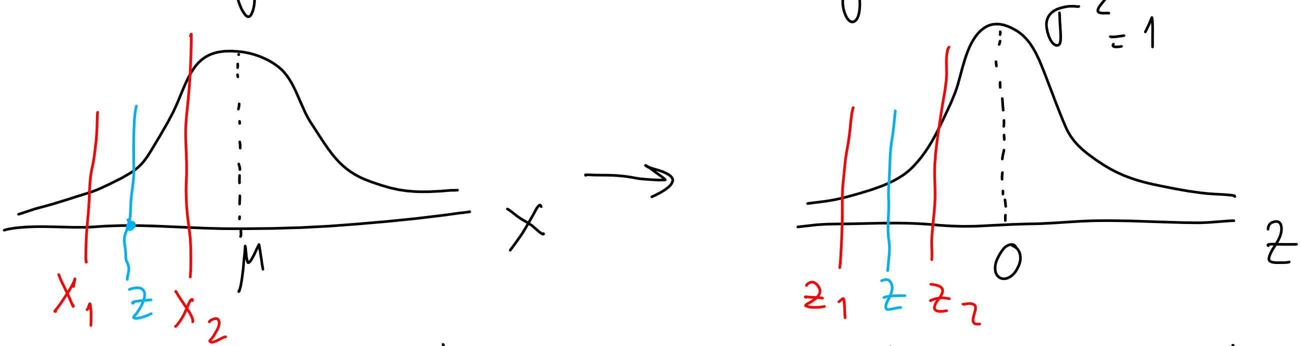
* All the observations of any normal random variables (X) are transformed to a new set of observations of a normal random variable z with mean $\mu = 0$ and $\sigma^2 = 1.0$.

$X \longrightarrow z$, How?

$$z = \frac{X - \mu}{\sigma} \qquad (z_{calculated} = z_{calc})$$

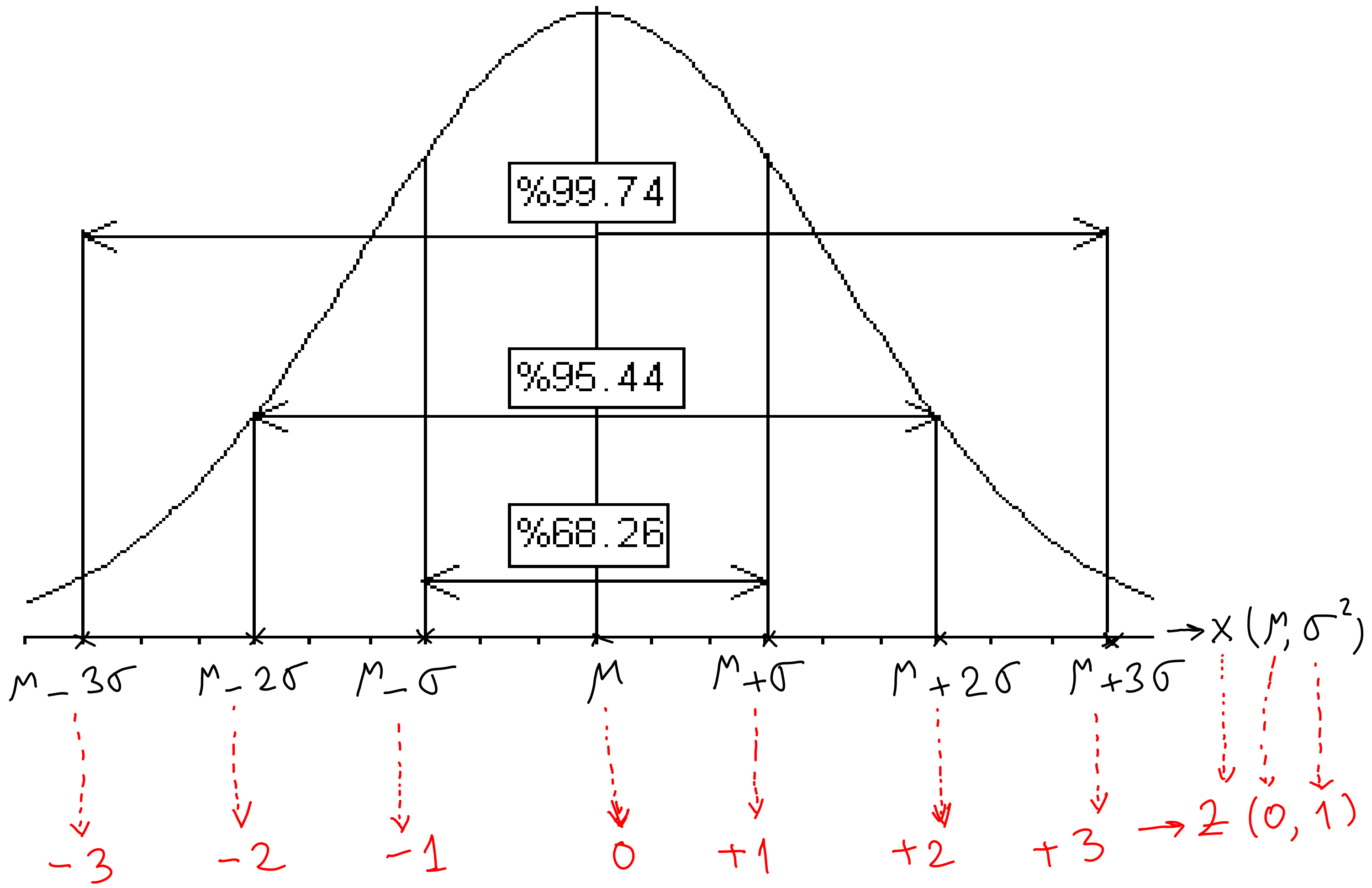
If X is between X_1 and X_2 values, then, variable z will fall between z_1 and z_2.

$$z_1 = \frac{X_1 - \mu}{\sigma}, \quad z_2 = \frac{X_2 - \mu}{\sigma}$$



Normal Random Variables $\longrightarrow$ Standard Normal Variables

$$P(x_1 < X < x_2) \cong P(z_1 < Z < z_2)$$



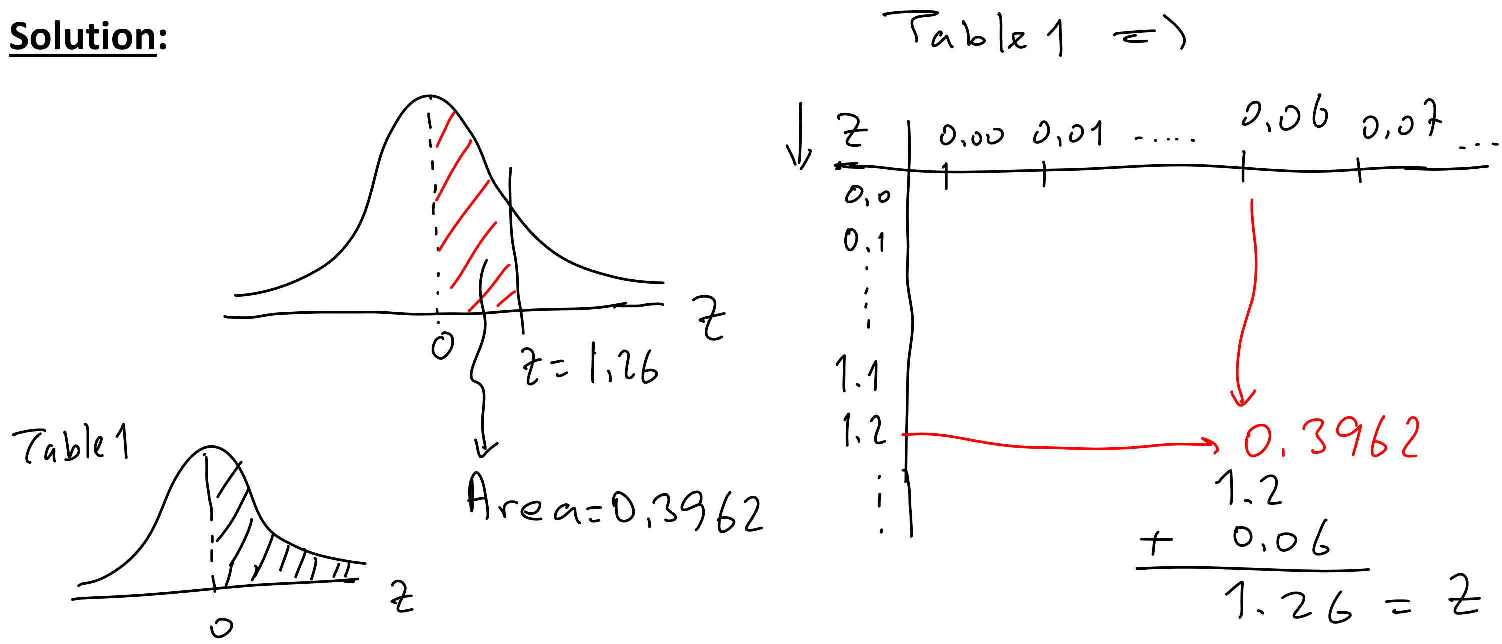
~ 68% of the area under a curve lies within $\mu \mp \sigma$

95% " " " " " " " " $\mu \mp 2\sigma$

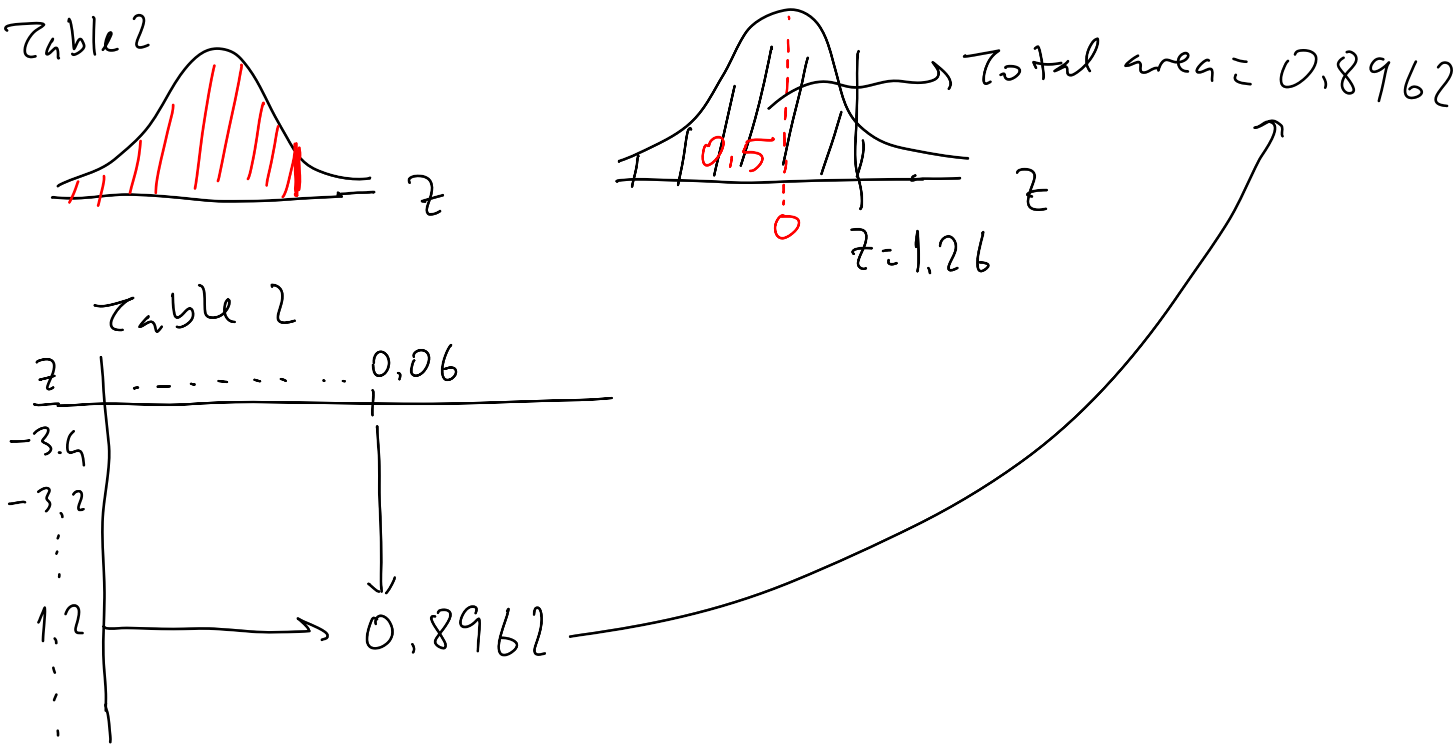
99% " " " " " " " " $\mu \mp 3\sigma$

<u>Example</u>: Find the area under a normal curve between the mean and a point Z = 1.26 standard deviations to the right of the mean.

Solution:



For the same example use Table 2 =)

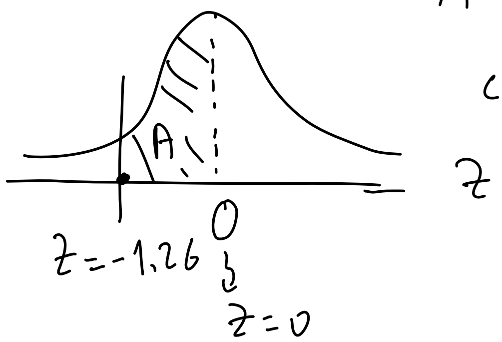


Area between mean (0) and z = 1.26 is equal to 0,8962 − 0,5 = 0,3962.

Example: What is the area between Z = 0 and Z =- 1.26 ?

Solution:

A = 0,3962 since the normal curve is symmetric.



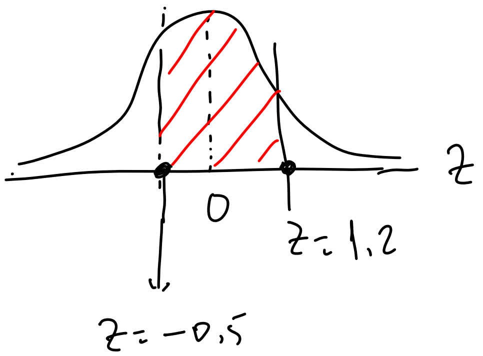
Example: Given a normal distribution with μ = 50 and σ = 10. Find the probability (P) that X assumes a value between 45 and 62.

Solution:

$x_1 = 45$, $X_2 = 62$ => Convert into z.

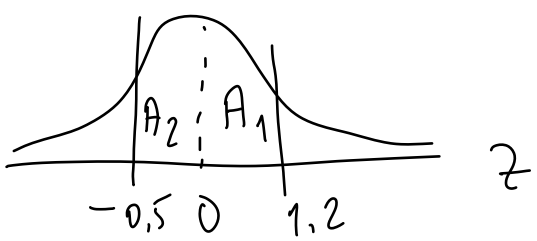
$$z_1 = \frac{45-50}{10} = -0,5 \quad , \quad z_2 = \frac{62-50}{10} = +1.2$$

$$P(45 < X < 62) \cong P(-0,5 < z < 1.2)$$
$$= P(z < 1.2) - P(z < -0.5)$$
$$= 0.8849 - 0.3085 \quad \text{(Table 2)}$$
$$= 0.5764$$



OR

Use Table 1 =>



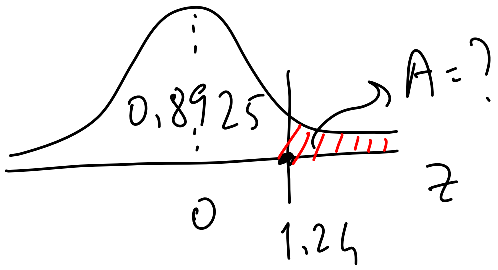
Total are $P = A_1 + A_2$
$$= 0,3849 + 0,1915$$
$$= 0.5764$$

Example: Given μ = 300 and σ = 50. Find the probability that X assumes a value greater than 362.

Solution:

$$P(X > 362), \quad X = 362$$

$$z = \frac{362 - 300}{50} = 1{,}24$$



$$P(X > 362) \cong P(z > 1{,}24)$$
$$= 1 - P(z < 1{,}24)$$
$$= 1 - 0{,}8925 \quad (\text{Table 2})$$
$$= 0{,}1075 = \text{Area} = \text{Probability}$$

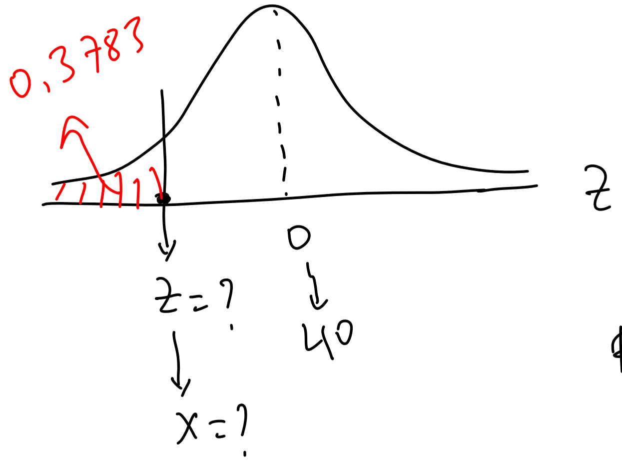
Example: Given a normal distribution with μ = 40 and σ = 6. Find the value of X that has

a) 37.83 % of the area below it

b) 5 % of the area above it

Solution:

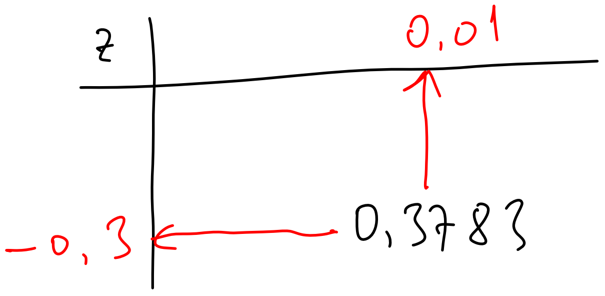
Find z, then, calculate X.



a) $z = \dfrac{X - \mu}{\sigma} \Rightarrow$

$$X = \sigma \times z + \mu$$

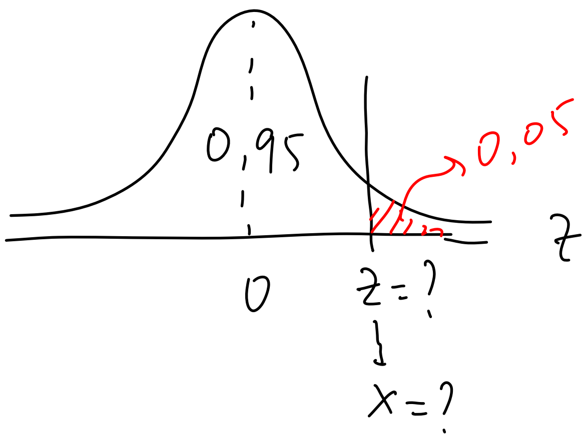
Find the z value having an area of 0,3783.

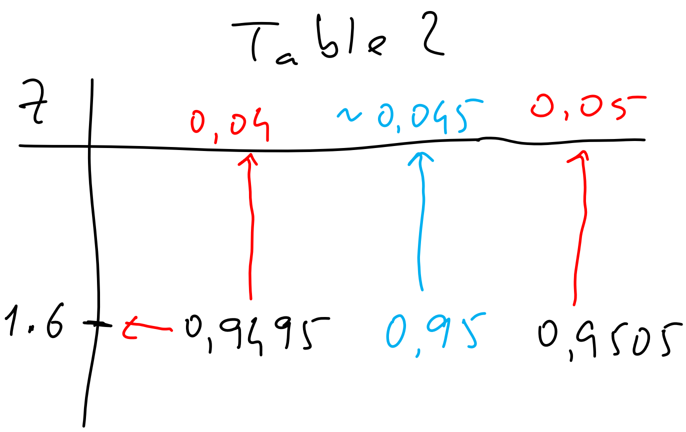


$$\Rightarrow \begin{array}{r} 0,3 \\ + \ 0,01 \\ \hline -0.31 \end{array} = z$$

$$X = 6 \times (-0,31) + 40 = 38.14$$

b)





$$1.6 + 0.045 = 1.645 \cong z$$

$$X = 6 \times (1.645) + 40 = 49.87$$

Example: **An electric firm manufactures light bulbs that have a length of life that is normally distributed with mean equal to 800 hr and a st.dev. of 40 hrs. Find the probability (P) that a bulb burns between 778 and 834 hrs.**

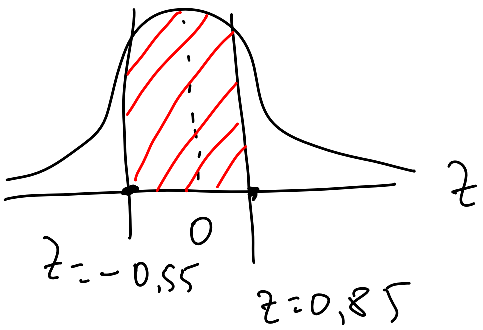
Solution:

$$X_1 = 778, \quad X_2 = 834$$

$$z_1 = \frac{778 - 800}{40} = -0,55, \quad z_2 = \frac{834 - 800}{40} = 0,85$$

$$P(778 < X < 834) \cong P(-0,55 < z < 0,85)$$

$$= P(z < 0,85) - P(z < -0,55)$$



$$= 0{,}8023 - 0{,}2912 \quad (\text{Table 2})$$

$$= 0{,}5111 \text{ or } 51.11\%$$

<u>Example</u>: A certain type of storage battery lasts on the average 3 years with a st.dev. of 0.5 years. Find the P that a given battery will last less than 2.3 years.

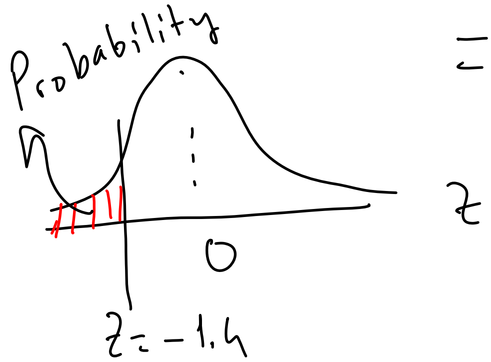
<u>Solution</u>:

$$\mu = 3,\ \sigma = 0{,}5\ ,\ X = 2{,}3 \Rightarrow$$

$$z = \frac{2{,}3 - 3}{0{,}5} = -1.4$$

$$P(X < 2{,}3) \cong P(z < -1{,}4)$$

$$= 0{,}0808 = \text{Area} = \text{Probability}$$



<u>Homework</u>: On an examination the average grade was 74 and σ = 7. If 12 % of the class are given A's, and the grades are curved to follow a normal distribution, what is the lowest possible A and the highest possible B ?

<u>Solution</u>: