

DATA ANALYSIS

Statistics

Juan Antonio Ferriz-Papi



USE OF STATISTICS:

- Help in collecting an appropriate quantitative data
- Organise and present complex data
- Provide a better understanding of nature
- Establish patterns of variability of a phenomenon through quantitative observations

Our purpose:

- Analyse production
- Study probability for anything to happen
- Study variability on the duration and cost of activities

STATISTICAL INFORMATION FOR DATA ANALYSIS:

TENDENCY

DISPERSION

STATISTICAL INFORMATION FOR DATA ANALYSIS:

TENDENCY

DISPERSION

MEAN, MEDIAN and MODE:

They are used for analysing frequencies with numerical data

They are measures of central tendency

There are some differences when doing statistical analysis

Example:

We have the heights of the people in class: 1.75, 1.87, 1.68, 1.62, 1.70, 1.67, 1.75



MEAN

$$\bar{X} = \frac{\text{Sumation of data}}{\text{Number of data}}$$

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$$

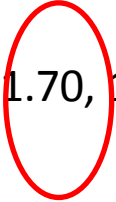
EXAMPLE $\bar{X} = 1.72 \text{ m}$

MEDIAN

This value that divides the set of data into two groups with the same number of elements.

EXAMPLE

1.62, 1.67, 1.68, 1.70, 1.75, 1.75, 1.87



MODE



This value or observation that repeats the most.

It is directly related to the frequency

EXAMPLE

1.62, 1.67, 1.68, 1.70, 1.75, 1.75, 1.87

DISPERSION

Range

Variance

Standard deviation

RANGE

Difference between maximum and minimum

EXAMPLE

1.62, 1.67, 1.68, 1.70, 1.75, 1.75, 1.87

$$\text{Range} = 1.87 - 1.62 = 0.25$$

VARIANCE

It measures how far the data are spread out

$$\text{Variance} = \frac{\text{Sumation of squares}}{N} = \frac{\sum (\bar{X} - X_i)^2}{N}$$

EXAMPLE

1.62, 1.67, 1.68, 1.70, 1.75, 1.75, 1.87

$$V = 0.0199/7 = 0.00284286$$

$X - x_i$	Result	Squared
$1.72 - 1.62$	0.10	0.0100
$1.72 - 1.67$	0.05	0.0025
$1.72 - 1.68$	0.04	0.0016
$1.72 - 1.70$	0.02	0.0004
$1.72 - 1.75$	-0.03	0.0009
$1.72 - 1.75$	-0.03	0.0009
$1.72 - 1.75$	-0.06	0.0036

STANDARD DEVIATION

A more exact way to give dispersion

$$S = \sqrt{V} = \sqrt{\frac{\sum (\bar{X} - X_i)^2}{N}}$$

EXAMPLE

1.62, 1.67, 1.68, 1.70, 1.75, 1.75, 1.87

$$S = 0.053318477$$

THE NORMAL DISTRIBUTION

We want to make C-25 concrete. For that, the mixing plant calculates the proportions needed for this mixture and makes samples with it to study compressive strength. The results are as follows:

MIXING PLANT 1											
Sample											
	29.8	31.8	29.0	24.3	30.3	32.7	27.9	31.2	27.9	28.1	
	27.6	25.9	26.7	26.7	29.8	29.4	26.9	29.1	28.9	29.4	
	32.4	27.6	31.5	28.4	28.6	30.1	27.4	27.4	27.8	30.5	
	26.4	28.2	30.6	26.3	29.5	31.8	31.4	29.2	33.4	34.9	
	27.1	32.6	30.1	28.1	28.2	29.7	30.1	28.1	29.7	33.2	

THE NORMAL DISTRIBUTION

We have very different results, so we need to **establish intervals**

For that:

Step 1. We calculate the range

$$34.9 - 24.3 = 10.6$$

Step 2. We decide about the approximate number of intervals

about 10 intervals is a good number (8-12)

Step 3. We divide the range by the number of intervals and decide

$$10.6/10 = 1.06 \quad \text{Aprox } 1$$

We can establish intervals every 1 N/mm²

THE NORMAL DISTRIBUTION

We organise this data in the intervals and calculate the tally

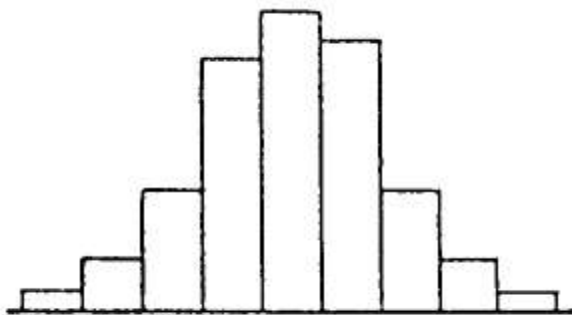
Intervals	Data	Tally
24-25	24.3;	1
25-26	25.9;	1
26-27	26.4; 26.7; 26.7; 26.3; 26.9;	5
27-28	27.6; 27.1; 27.6; 27.9; 27.4; 27.4; 27.9; 27.8;	8
28-29	28.2; 29.0; 28.4; 28.1; 28.6; 28.2; 28.1; 28.9; 28.1;	9
29-30	29.8; 29.8; 29.5; 29.4; 29.7; 29.1; 29.2; 29.7; 29.4;	9
30-31	30.6; 30.1; 30.3; 30.1; 30.1; 30.5;	6
31-32	31.8; 31.5; 31.8; 31.4; 31.2;	5
32-33	32.4; 32.6; 32.7;	3
33-34	33.4; 33.2;	2
34-35	34.9;	1

THE NORMAL DISTRIBUTION

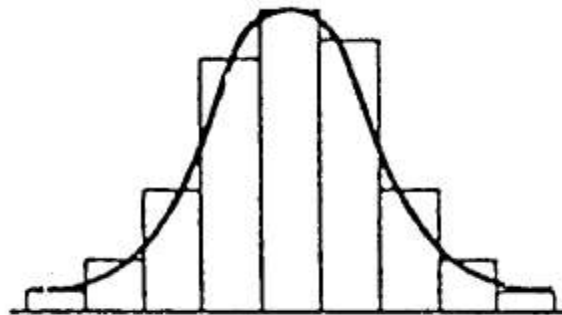
We represent this tally (frequency)

Intervals	Data	Tally
24-25	24.3;	1
25-26	25.9;	1
26-27	26.4; 26.7; 26.7; 26.3; 26.9;	5
27-28	27.6; 27.1; 27.6; 27.9; 27.4; 27.4; 27.9; 27.8;	8
28-29	28.2; 29.0; 28.4; 28.1; 28.6; 28.2; 28.1; 28.9; 28.1;	9
29-30	29.8; 29.8; 29.5; 29.4; 29.7; 29.1; 29.2; 29.7; 29.4;	9
30-31	30.6; 30.1; 30.3; 30.1; 30.1; 30.5;	6
31-32	31.8; 31.5; 31.8; 31.4; 31.2;	5
32-33	32.4; 32.6; 32.7;	3
33-34	33.4; 33.2;	2
34-35	34.9;	1

THE NORMAL DISTRIBUTION



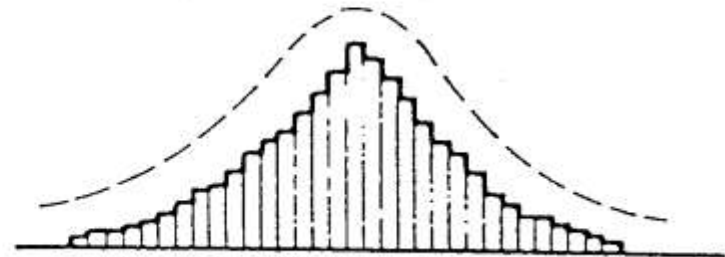
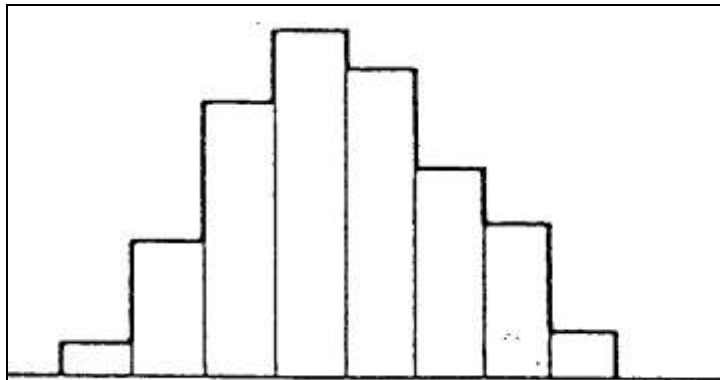
A

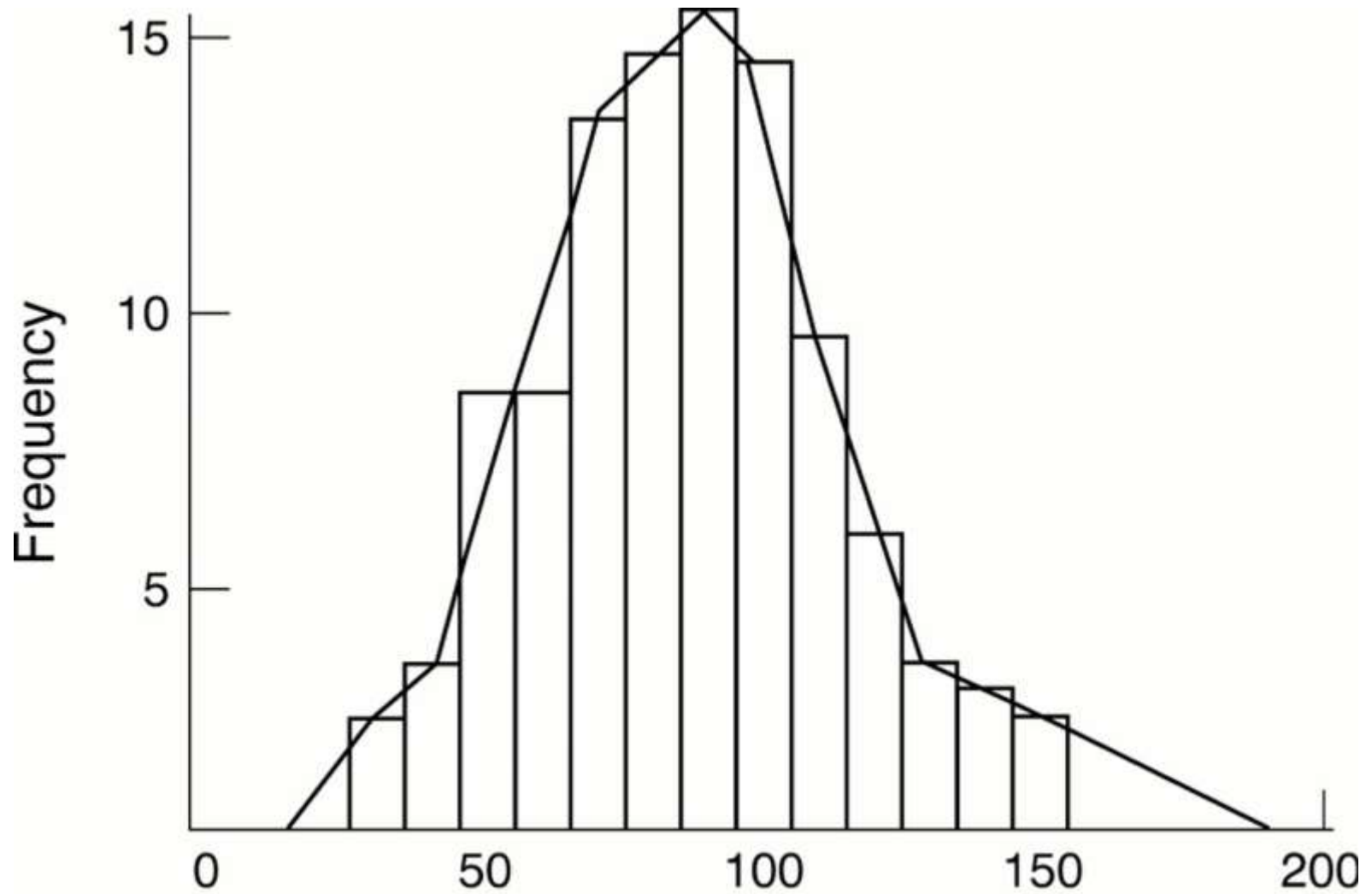


B



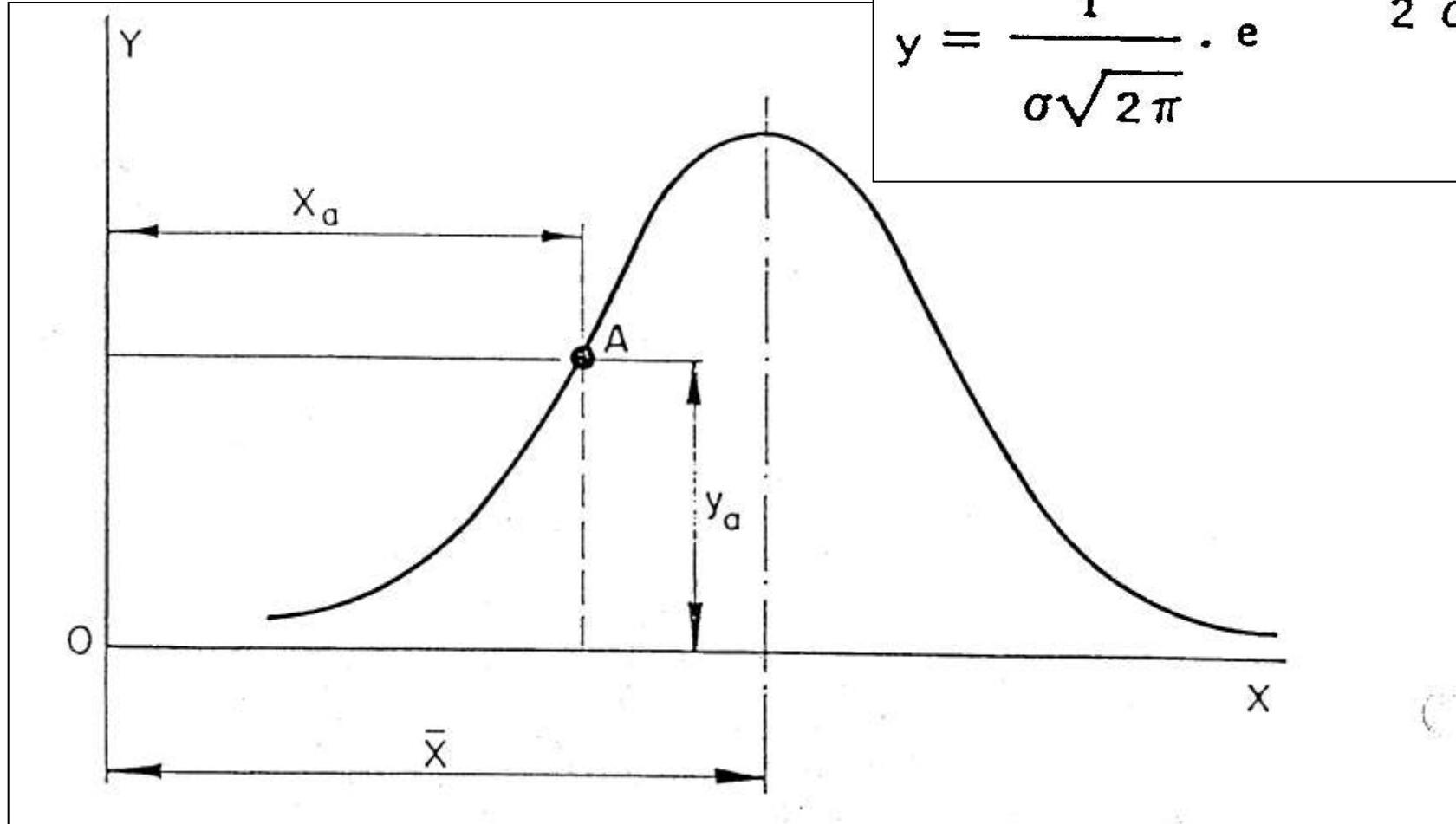
C





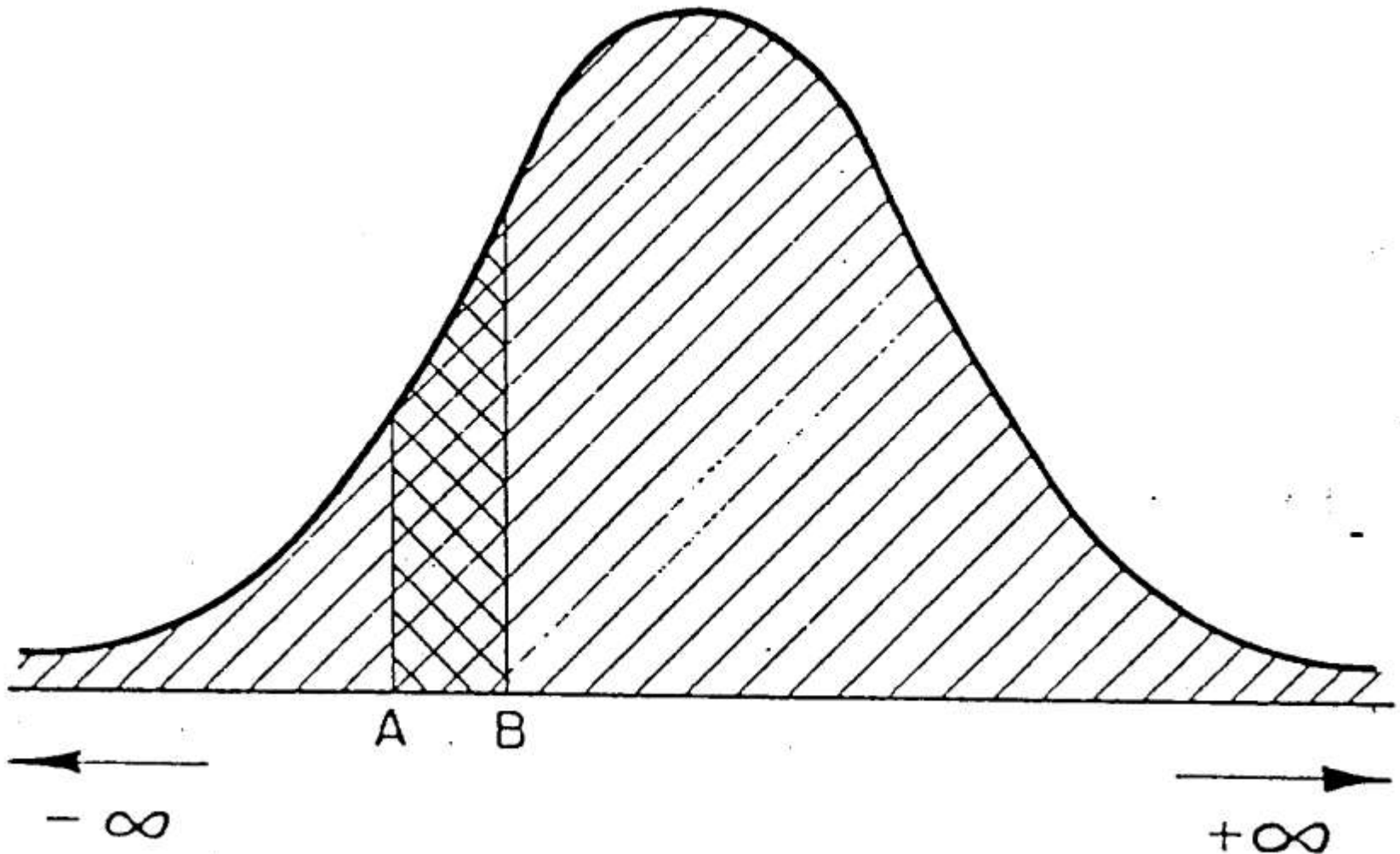
DATA ANALYSIS

$$y = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x - \bar{x})^2}{2\sigma^2}}$$



DATA ANALYSIS

If we calculate the area between AB and this equation, we will know the percentage of product that is inside the interval AB.



¿HOW CAN I CALCULATE THIS AREA?



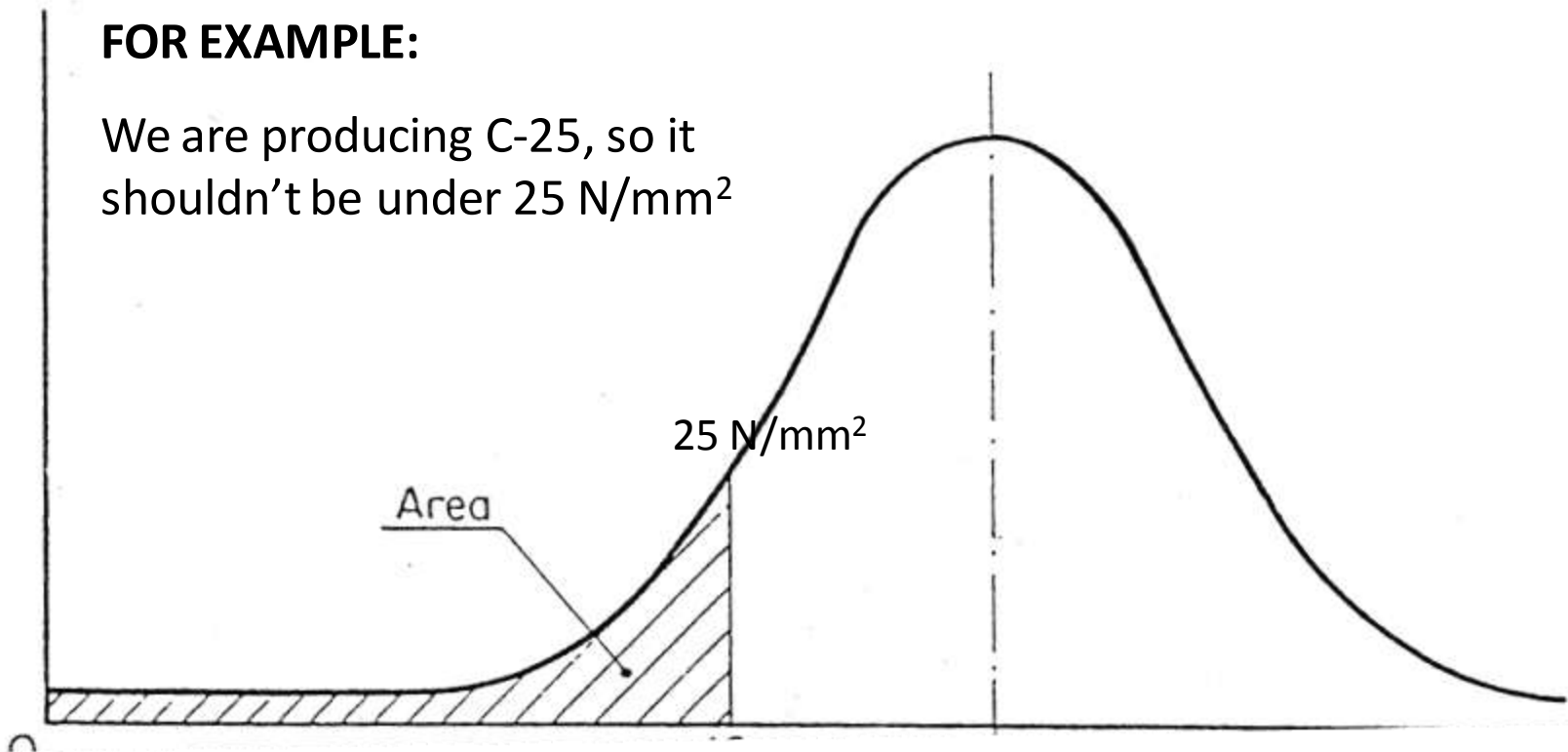
WITH THE FOLLOWING TABLE

TO ENTER THIS TABLE, I MUST CALCULATE:

$$\frac{x - \bar{X}}{\sigma}$$

FOR EXAMPLE:

We are producing C-25, so it shouldn't be under 25 N/mm²



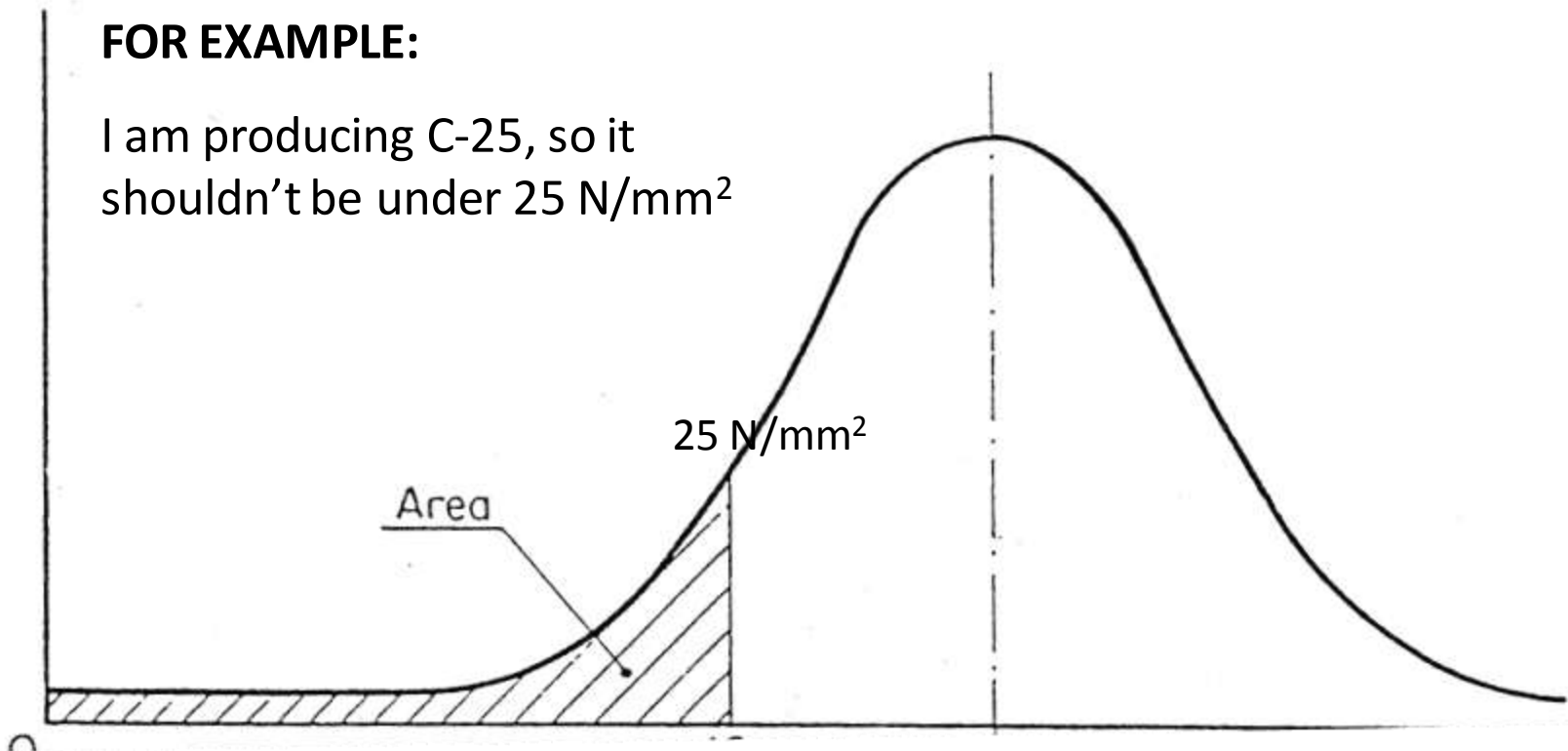
We calculate the mean and the standard deviation

$$\bar{X} = 28.7 \text{ N/mm}^2$$

$$\sigma = 2.16$$

FOR EXAMPLE:

I am producing C-25, so it shouldn't be under 25 N/mm²



$$\frac{x - \bar{x}}{\sigma} = \frac{25 - 28.7}{2.16} = -1.71$$

Table:

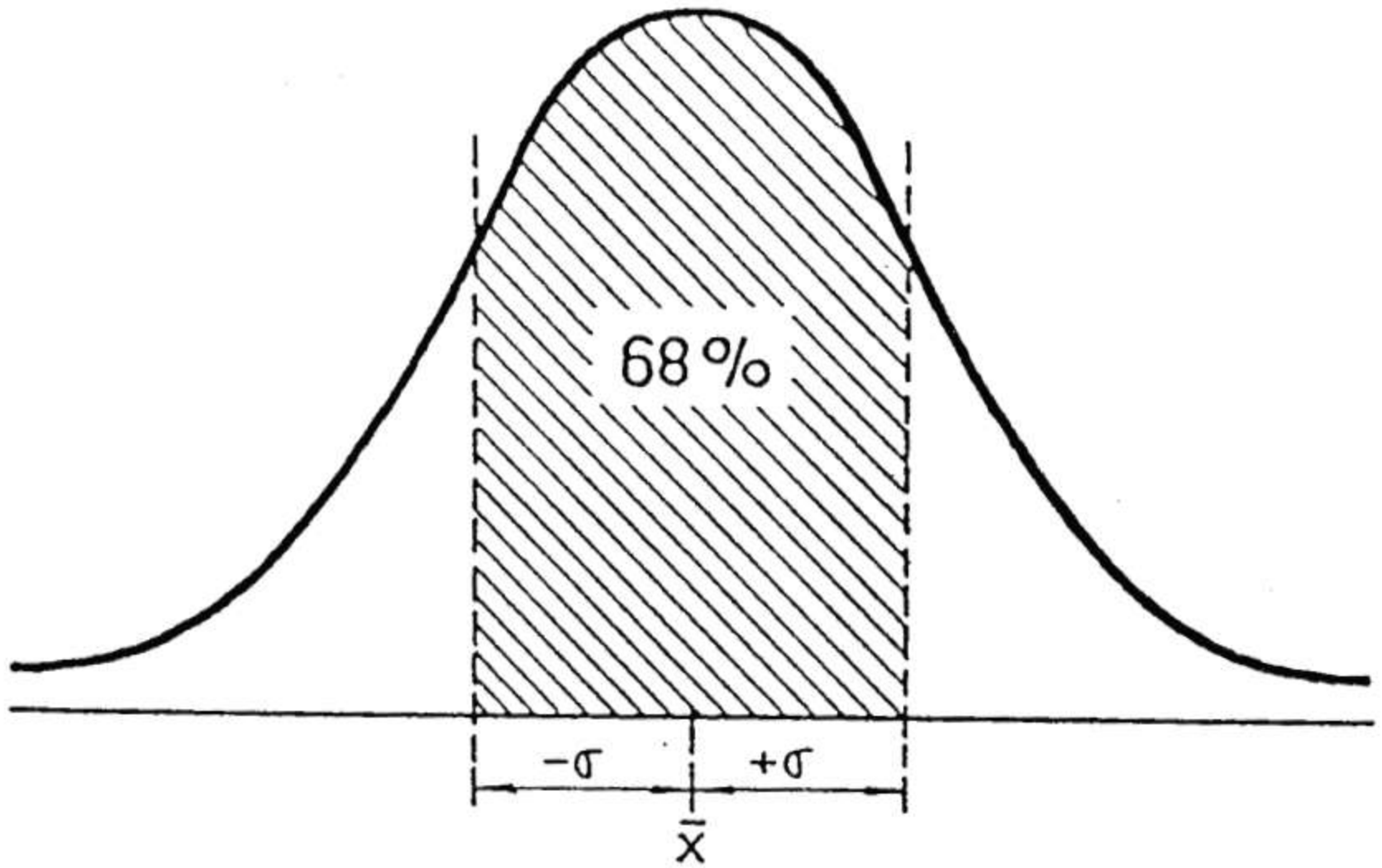
$$-1.80 = 0.0359$$

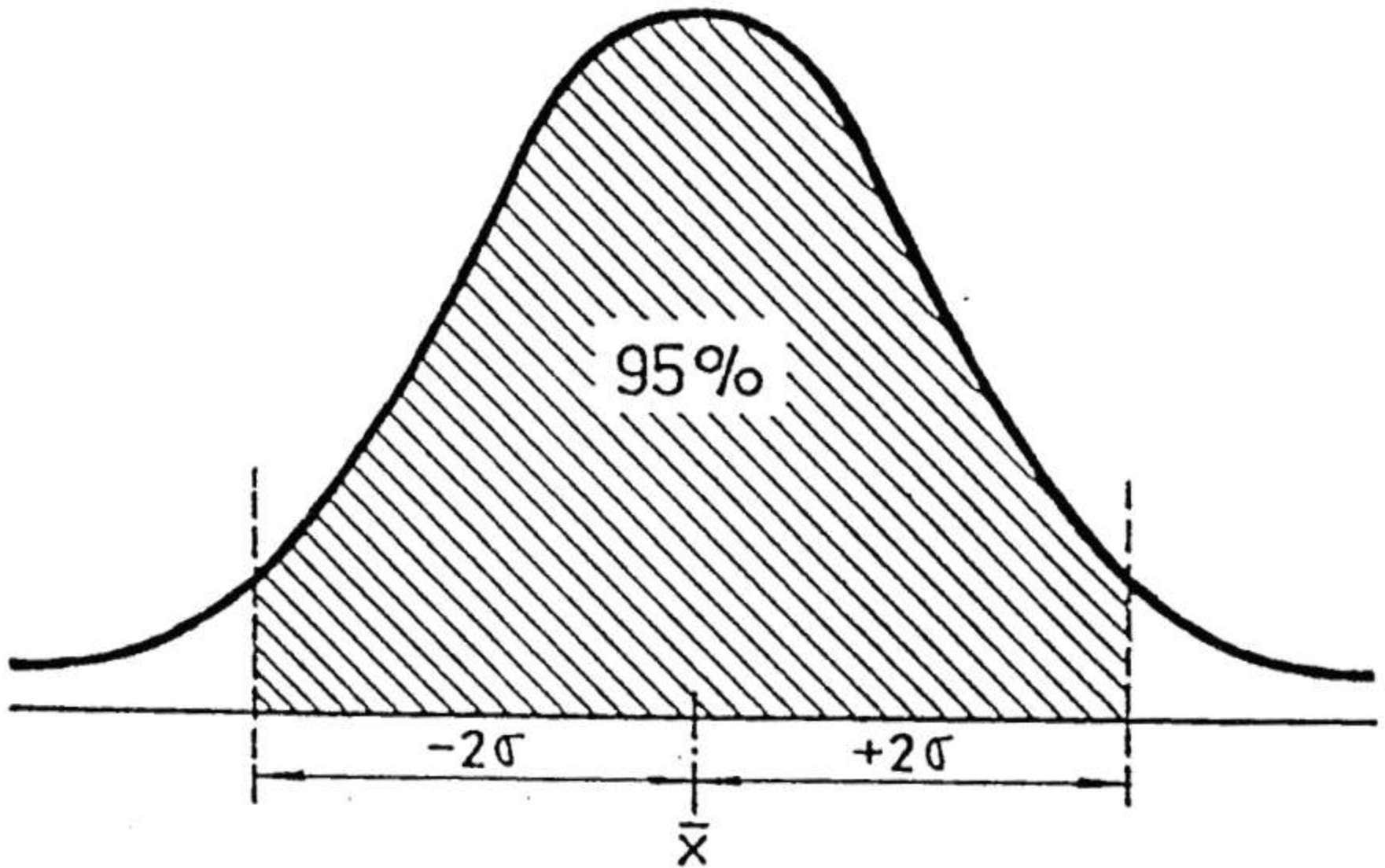
$$-1.71 = 0.0455$$

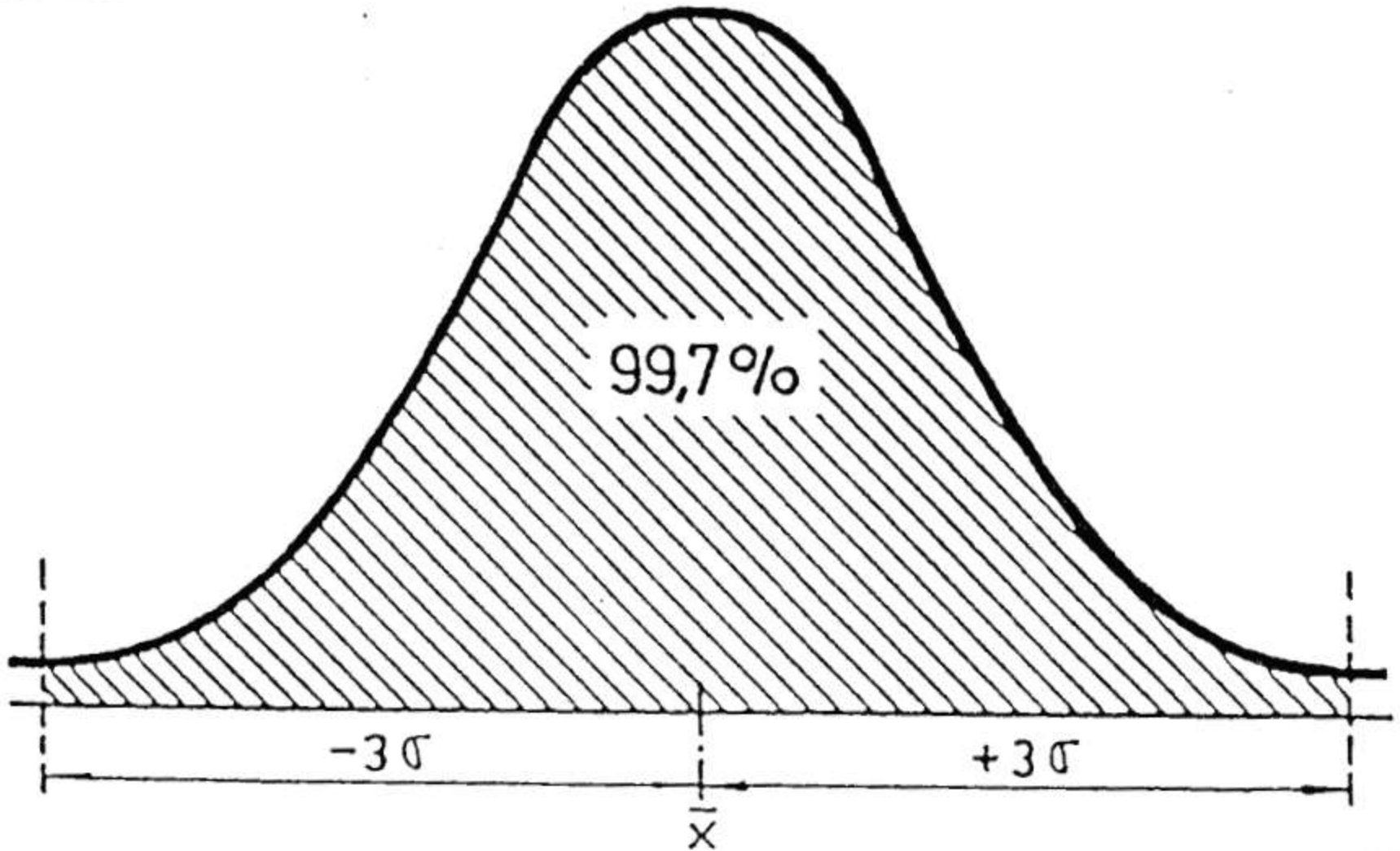
$$-1.70 = 0.0446$$

Area = 4.55%

APPLYING THIS, I CAN GET THE
FOLLOWING RESULTS...







EXERCISE

We want to buy concrete and we have to compare the production quality of two mixing plants.

The product that we want is RC25/30. We ask for the results of the internal control production for every plant and they give us results corresponding to cylindrical-samples compressive strength to 28 days.

With these results, we calculate the mean and the standard deviation, and they are the following:

	Mixing plant 1	Mixing plant 2
Mean	27.5 N/mm ²	28.3 N/mm ²
Standard deviation	1.2	2.6

Calculate:

- 1.The interval where we have 95% of production in each plant
- 2.The percentage out of tolerance in each plant (limit 25 N/mm²)
- 3.Select the mixing plant to buy your concrete and explain why.