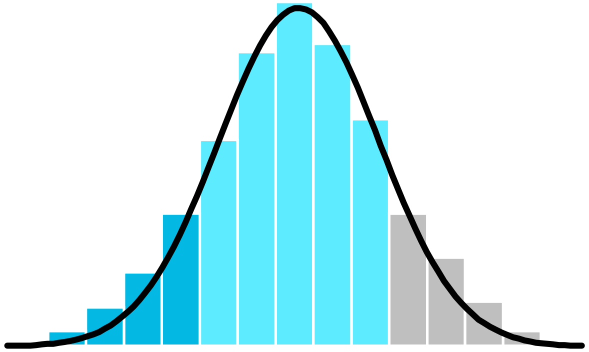
Continuous Improvement Toolkit

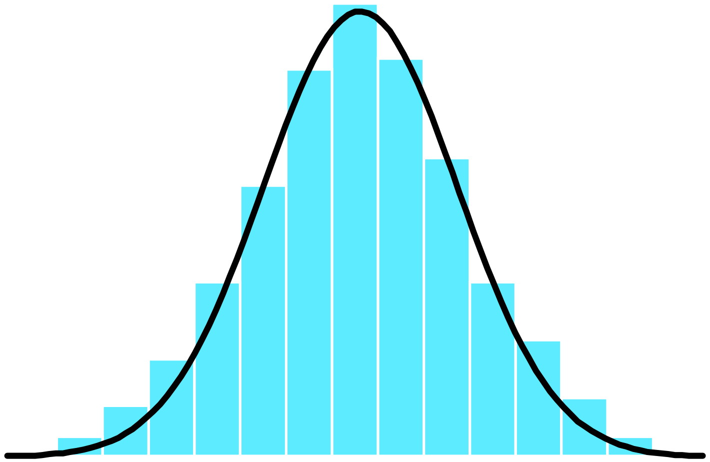
# Normal Distribution



# The Continuous Improvement Map

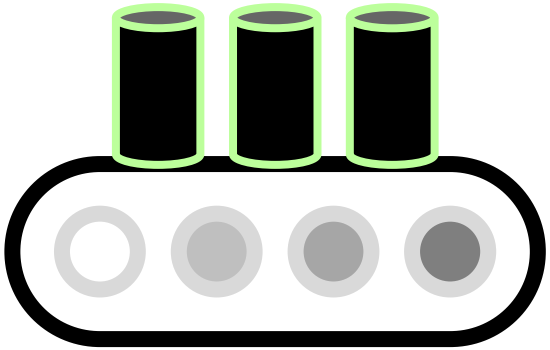
# - Normal Distribution

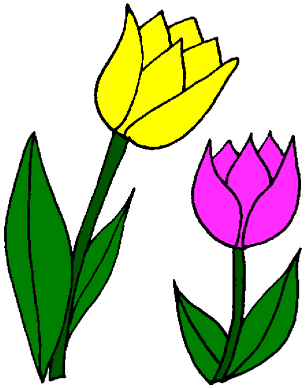
- The commonest and the most useful continuous distribution.
- A symmetrical probability distribution where most results are located in the middle and few are spread on both sides.
- It has the shape of a bell.
- Can entirely be described by its mean and standard deviation.



# - Normal Distribution

- Can be found practically everywhere:
  - In nature.
  - In engineering and industrial processes.
  - In social and human sciences.
- Many everyday data sets follow approximately the normal distribution.

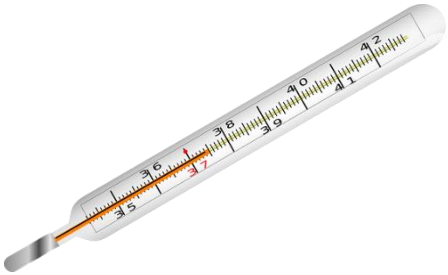


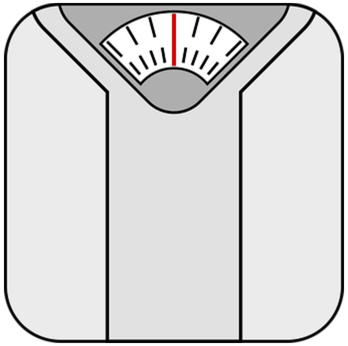


# - Normal Distribution

**Examples:**

- The body temperature for healthy humans.
- The heights and weights of adults.
- The thickness and dimensions of a product.
- IQ and standardized test scores.
- Quality control test results.
- Errors in measurements.



# - Normal Distribution

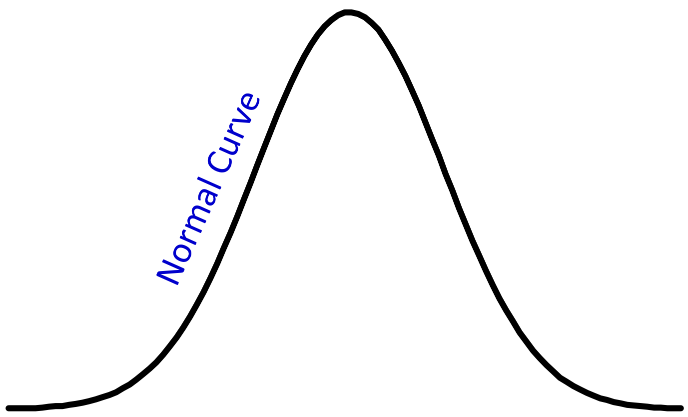
**Why?**

- Used to illustrate the shape and variability of the data.
- Used to estimate future process performance.
- Normality is an important assumption when conducting statistical analysis.
    - Certain SPC charts and many statistical inference tests require the data to be normally distributed.

# - Normal Distribution

**Normal Curve:**

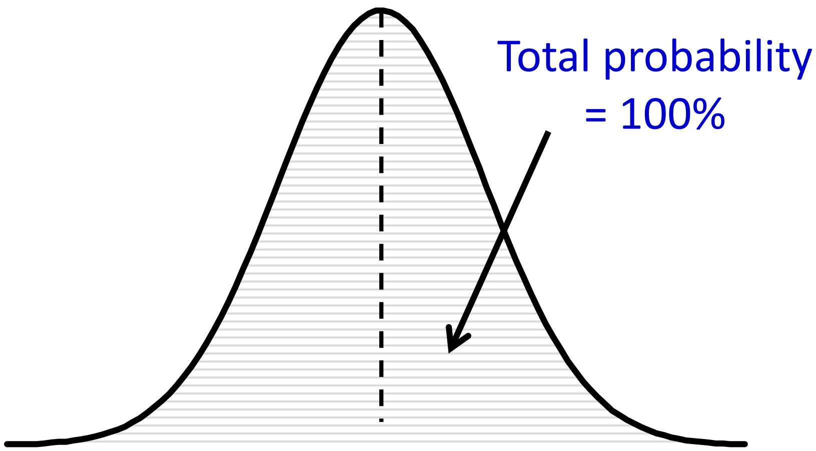
- A graphical representation of the normal distribution.
- It is determined by the mean and the standard deviation.
- It a symmetric unimodal bell-shaped curve.
- Its tails extending infinitely in both directions.
- The wider the curve, the larger the standard deviation and the more variation exists in the process.
- The spread of the curve is equivalent to six times the standard deviation of the process.



# - Normal Distribution

## Normal Curve:

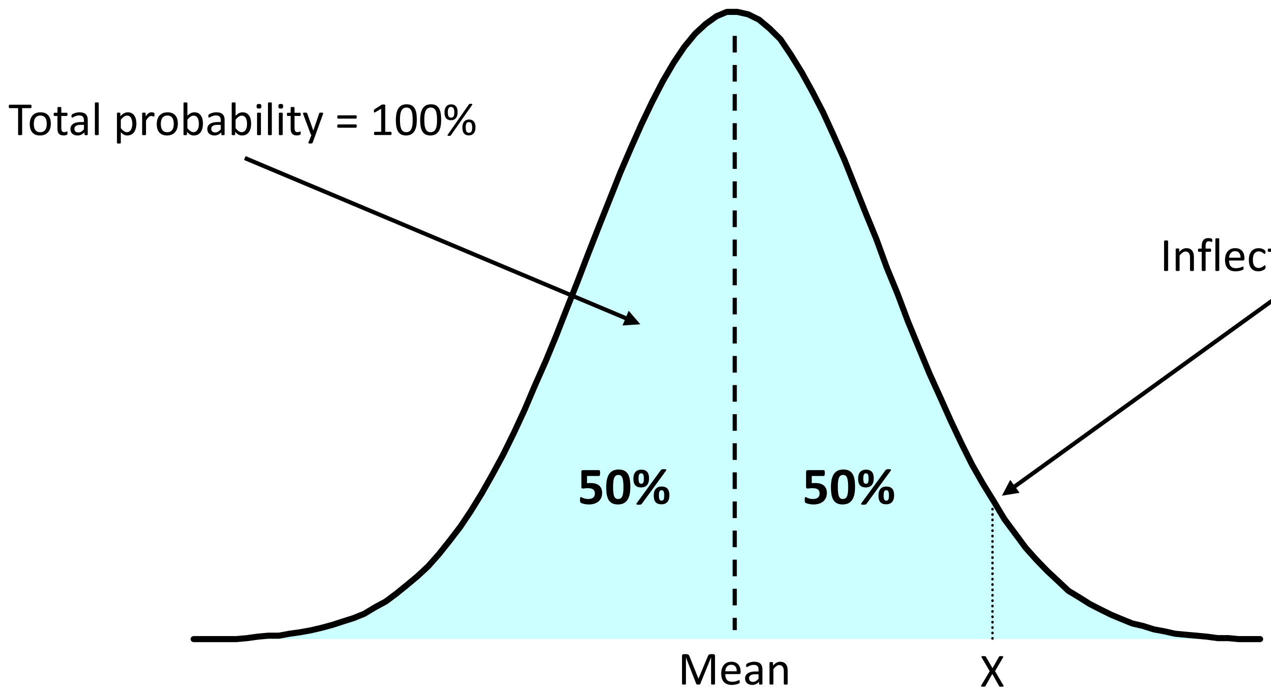
- Helps calculating the probabilities for normally distributed populations.
- The probabilities are represented by the area under the normal curve.
- The total area under the curve is equal to **100%** (or **1.00**).
- This represents the population of the observations.
- We can get a rough estimate of the probability above a value, below a value, or between any two values.



# - Normal Distribution

Normal Curve:

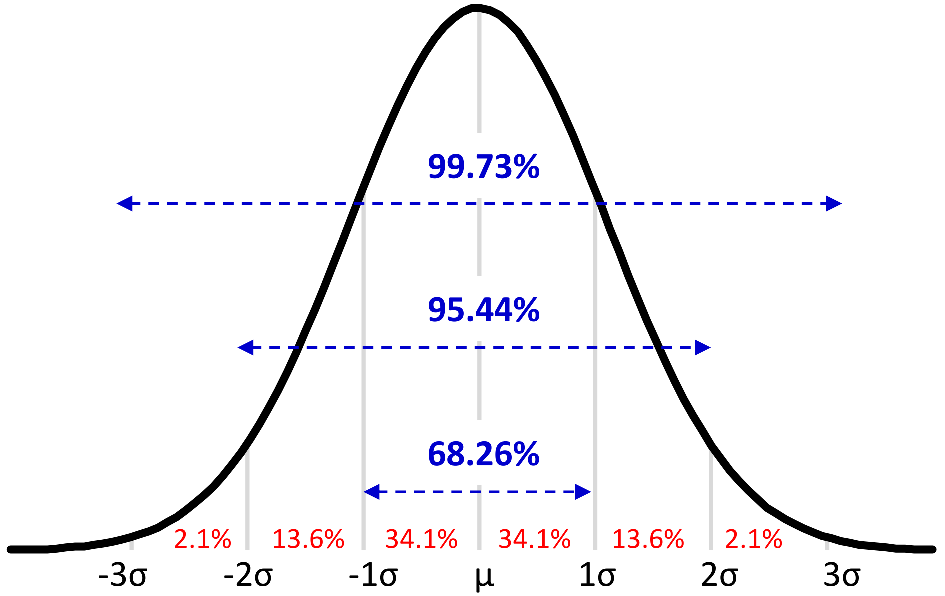
- Since the normal curve is symmetrical, **50 percent** of the data lie on each side of the curve.



# - Normal Distribution

**Empirical Rule**:

- For any normally distributed data:
  - **68%** of the data fall within **1** standard deviation of the mean.
  - **95%** of the data fall within **2** standard deviations of the mean.
  - **99.7%** of the data fall within **3** standard deviations of the mean.

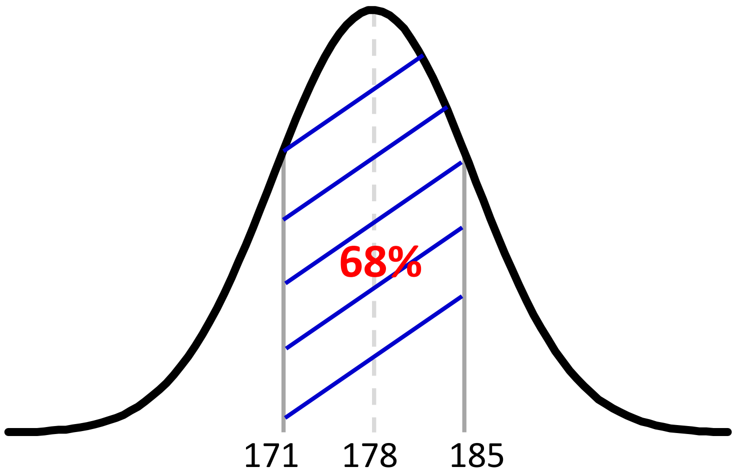


# - Normal Distribution

Empirical Rule:

- Suppose that the heights of a sample men are normally distributed.
- The mean height is **178** cm and a standard deviation is **7** cm.

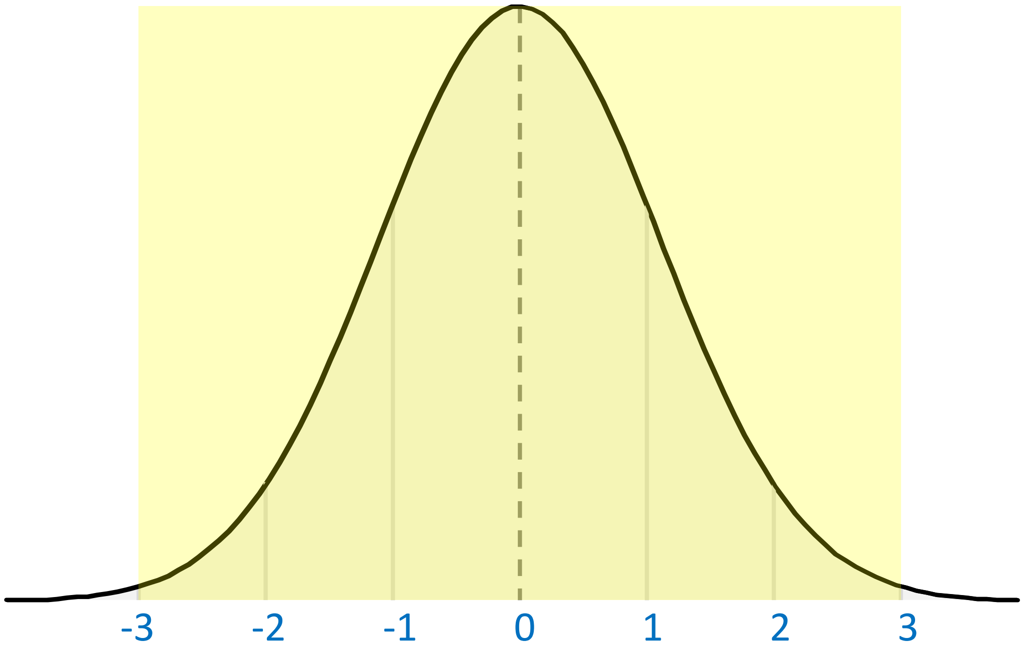
- **We can generalize that:**
  - 68% of population are between 171 cm and 185 cm.
  - This might be a generalization, but it's true if the data is normally distributed.



# - Normal Distribution

## Empirical Rule:

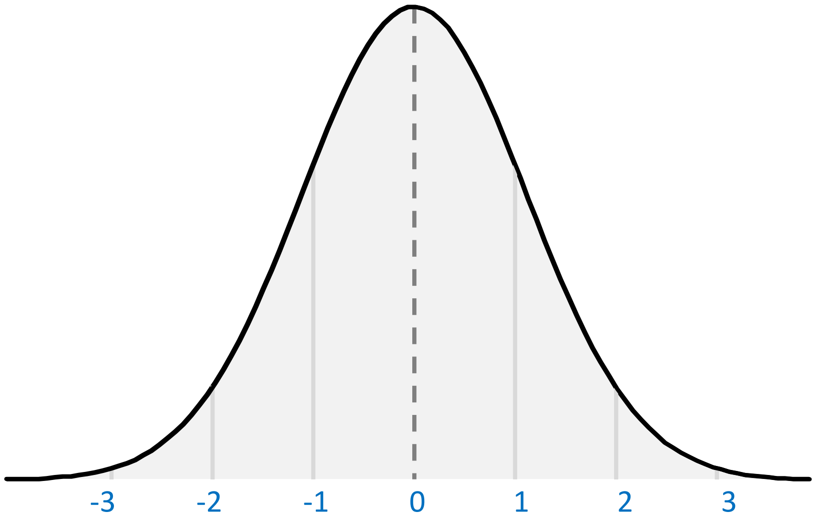
- For a stable normally distributed process, **99.73%** of the values lie within **+/-3** standard deviation of the mean.



# - Normal Distribution

**Standard Normal Distribution**:

- A common practice to convert any normal distribution to the standardized form and then use the standard normal table to find probabilities.
- The **Standard Normal Distribution** (Z distribution) is a way of standardizing the normal distribution.
- It always has a mean of **0** and a standard deviation of **1**.



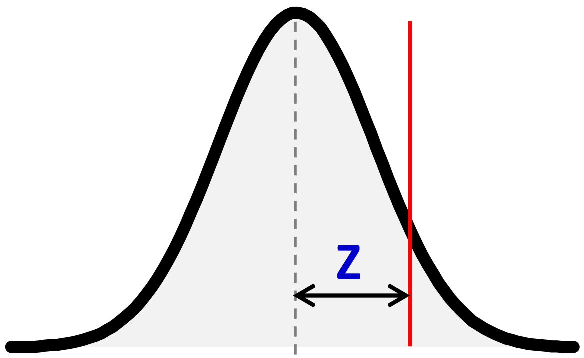
# - Normal Distribution

## Standard Normal Distribution:

- Any normally distributed data can be converted to the standardized form using the formula:

$$Z = (X - \mu) / \sigma$$

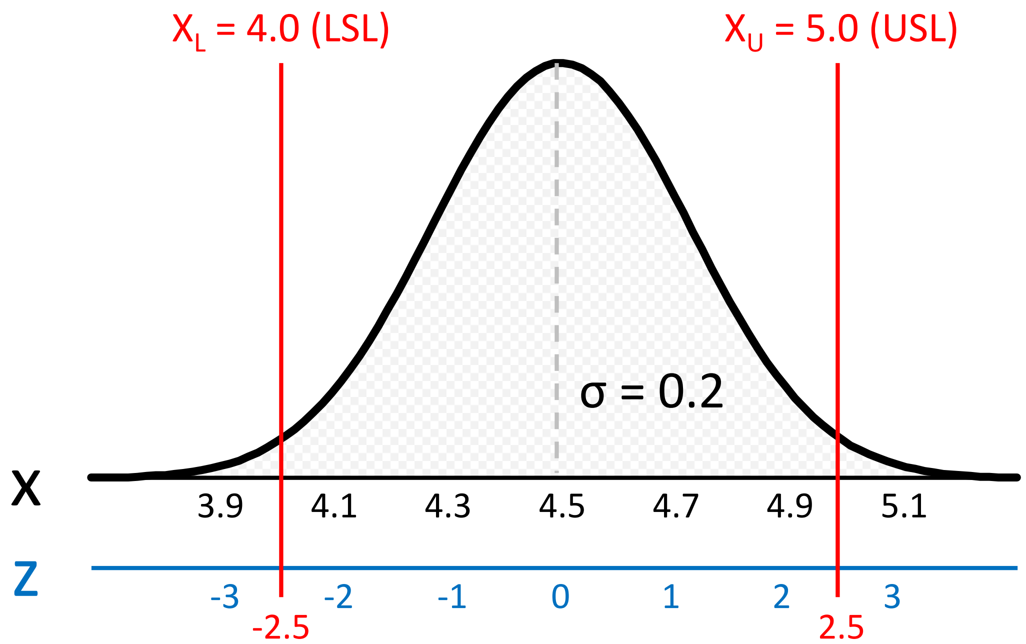
- where:
  - 'X' is the data point in question.
  - 'Z' (or **Z-score**) is a measure of the number of standard deviations of that data point from the mean.



# - Normal Distribution

## Standard Normal Distribution:

- Converting from 'X' to 'Z':



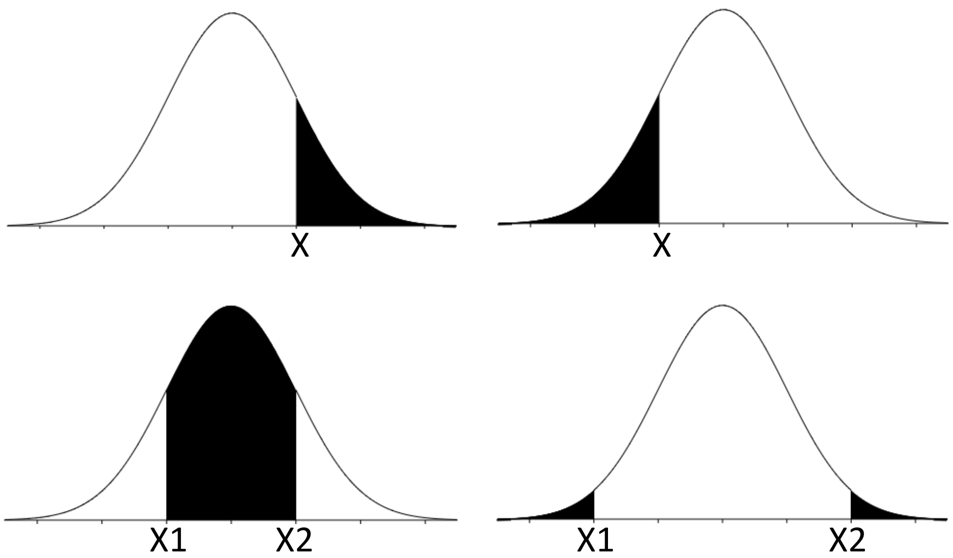

The specification limits at 4.0 and 5.0mm respectively, and the corresponding z values

The X-scale is for the actual values and the Z-scale is for the standardized values

# - Normal Distribution

## Standard Normal Distribution:

- You can then use this information to determine the area under the normal distribution curve that is:
  - To the right of your data point.
  - To the left of the data point.
  - Between two data points.
  - Outside of two data points.



# - Normal Distribution

**Z-Table:**

- Used to find probabilities associated with the standard normal curve.
- You may also use the **Z-table calculator** instead of looking into the Z-table manually.

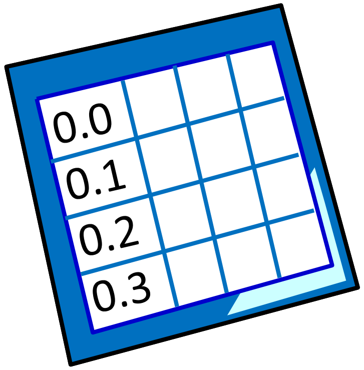
| Z | +0.00 | +0.01 | +0.02 | +0.03 | … |
|---|---|---|---|---|---|
| 0.0 | 0.50000 | 0.50399 | 0.50798 | 0.51197 | |
| 0.1 | 0.53983 | 0.54380 | 0.54776 | 0.55172 | |
| 0.2 | 0.57926 | 0.58317 | 0.58706 | 0.59095 | |
| 0.3 | 0.61791 | 0.62172 | 0.62552 | 0.62930 | |
| … | | | | | |

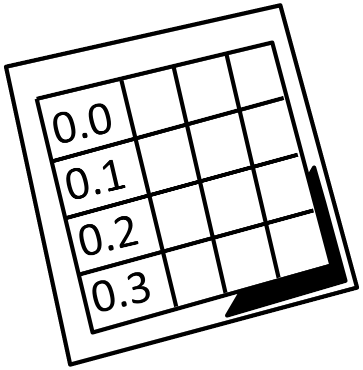
**Cumulative Z-Table**

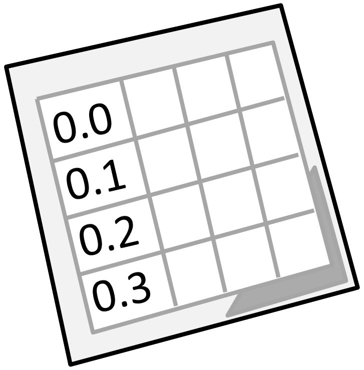
# - Normal Distribution

## Z-Table:

- **There are different forms of the Z-table:**
  - **Cumulative**, gives the proportion of the population that is to the left or below the Z-score.
  - **Complementary cumulative**, gives the proportion of the population that is to the right or above the Z-score.
  - **Cumulative from mean**, gives the proportion of the population to the left of that z-score to the mean only.

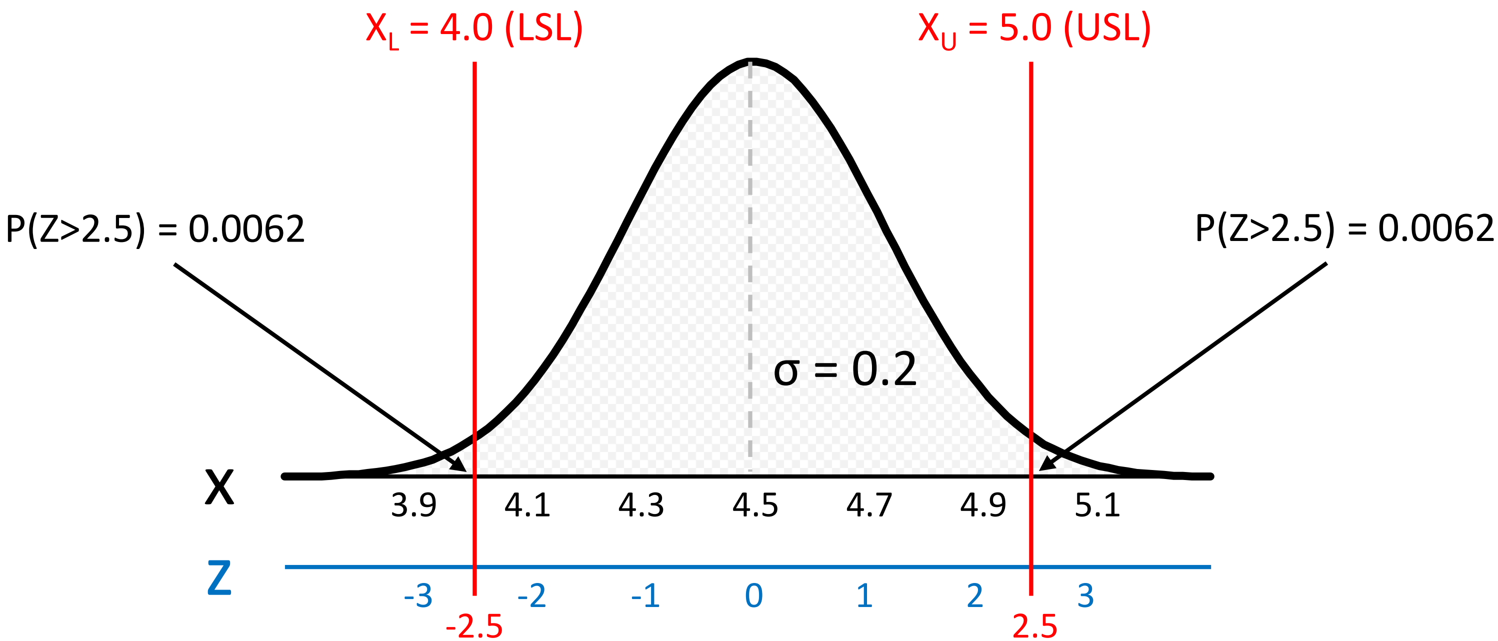


# - Normal Distribution

## Z-Table:



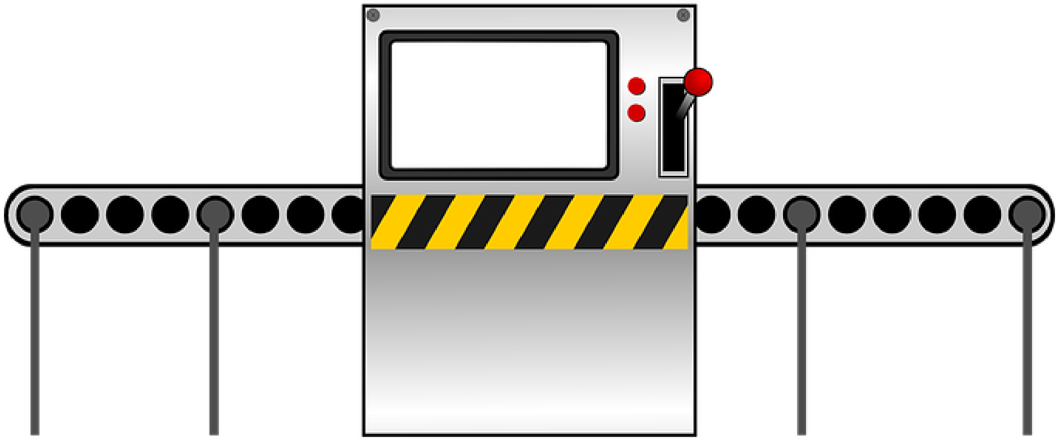

The area between Z = -2.5 & Z = +2.5 cannot be looked up directly in the Z-Table

However, since the area below the total curve is 1, it can be found by subtracting the known areas from 1.

# - Normal Distribution

**Exercise:**

- **Question:** For a process with a mean of **100**, a standard deviation of **10** and an upper specification of **120**, what is the probability that a randomly selected item is defective (or beyond the upper specification limit)?

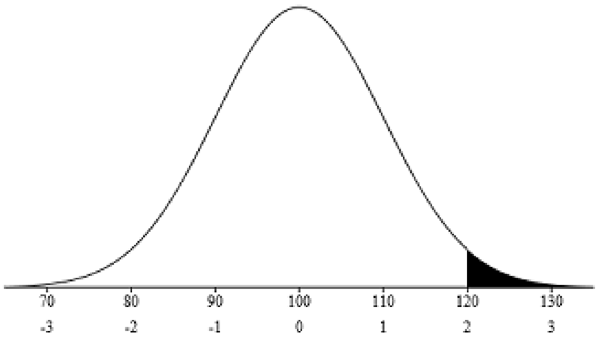


**Time allowed: 10 minutes**

# - Normal Distribution

**Exercise:**

- **Answer:**
  - The Z-score is equal to **= (120 – 100) / 10 = 2**.
  - This means that the upper specification limit is **2** standard deviations above the mean.
  - Now that we have the Z-score, we can use the Z-table to find the probability.
  - From the Z-table (the complementary cumulative table), the area under the curve for a Z-value of **2 = 0.02275** or **2.275%**.
  - This means that there is a chance of **2.275%** for any randomly selected item to be defective.



# - Normality Testing in Minitab

- Many statistical tests require that the distribution is normal.
- Several tools are available to assess the normality of data:
  - Using a **histogram** to visually explore the data.
  - Producing a **normal probability plot**.
  - Carrying out an **Anderson-Darling** normality test.

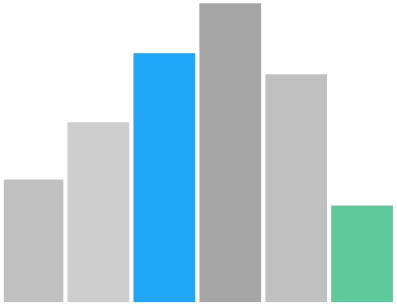
- All these tools are easy to use in **Minitab** statistical software.

Normality Testing

# - Normality Testing in Minitab

**Histograms:**

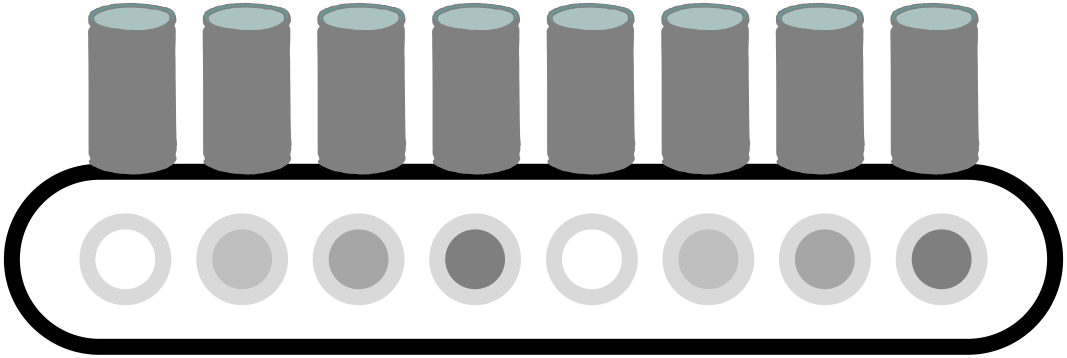
- Efficient graphical methods for describing the distribution of data.
- It is always a good practice to plot your data in a histogram after collecting the data.
- This will give you an insight about the shape of the distribution.
- If the data is symmetrically distributed and most results are located in the middle, we can assume that the data is normally distributed.



# - Normality Testing in Minitab

Histograms:

- Suppose that a line manager is seeking to assess how consistently a production line is producing.
- He is interested in the weight of a food products with a target of **50** grams per item.
- He takes a random sample of **40** products and measures their weights.



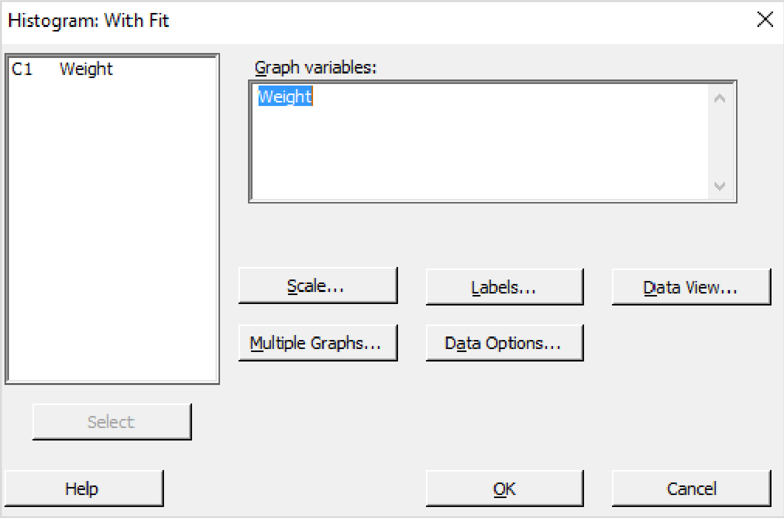
# - Normality Testing in Minitab

## Histograms:

- ❑ Remember to copy the data from the Excel worksheet and paste it into the Minitab worksheet.

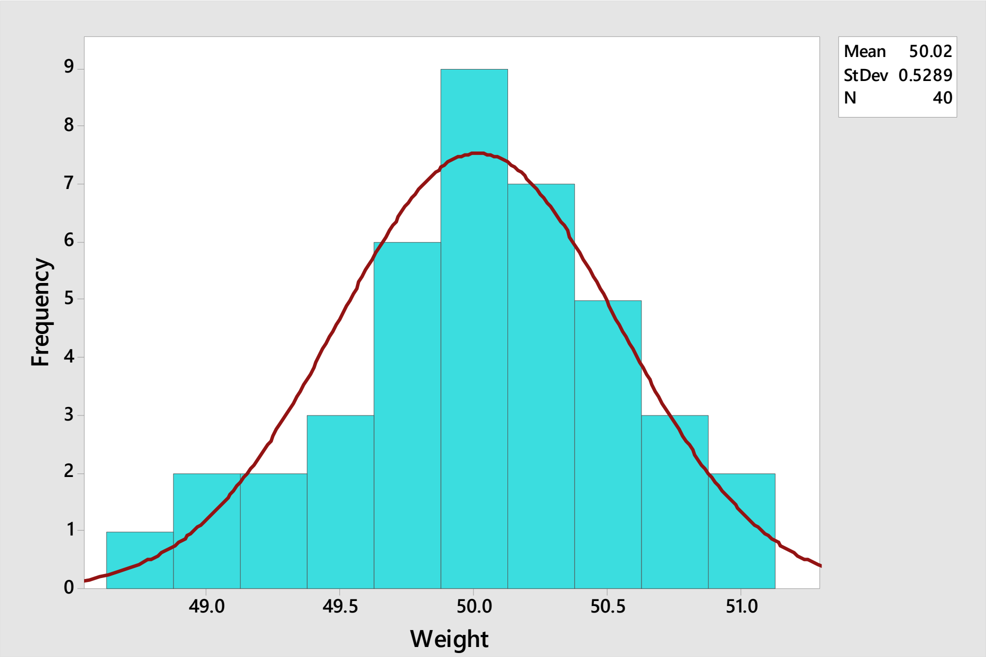
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 50.9 | 49.3 | 49.6 | 51.1 | 49.8 | 49.6 | 49.9 | 49.9 |
| 49.7 | 50.8 | 48.8 | 49.8 | 50.4 | 48.9 | 50.6 | 50.3 |
| 50.4 | 50.7 | 50.2 | 50.0 | 50.8 | 50.3 | 50.4 | 49.9 |
| 49.4 | 50.0 | 50.5 | 50.3 | 50.1 | 49.9 | 49.8 | 50.2 |
| 49.3 | 50.0 | 49.7 | 50.3 | 49.7 | 50.2 | 50.0 | 49.1 |

- ❑ Select **Graph > Histogram > With Fit.**
- ❑ Specify the column of data to analyze
- ❑ Click OK.



# - Normality Testing in Minitab

- The histogram below suggests that the data is normally distribution.

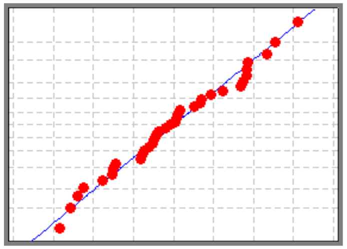


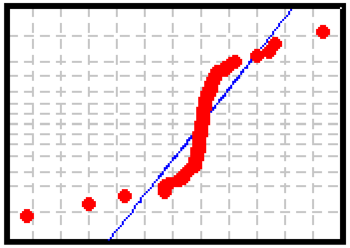

Notice how the data is symmetrically distributed and concentrated in the center of the histogram

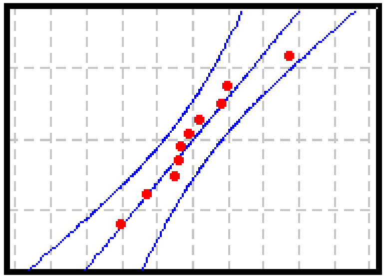
# - Normality Testing in Minitab

**Normal Probability Plots:**

- Used to test the assumption of normality.
- Provides a more decisive approach.
- All points for a normal distribution should approximately form a straight line that falls between **95%** confidence interval limits.



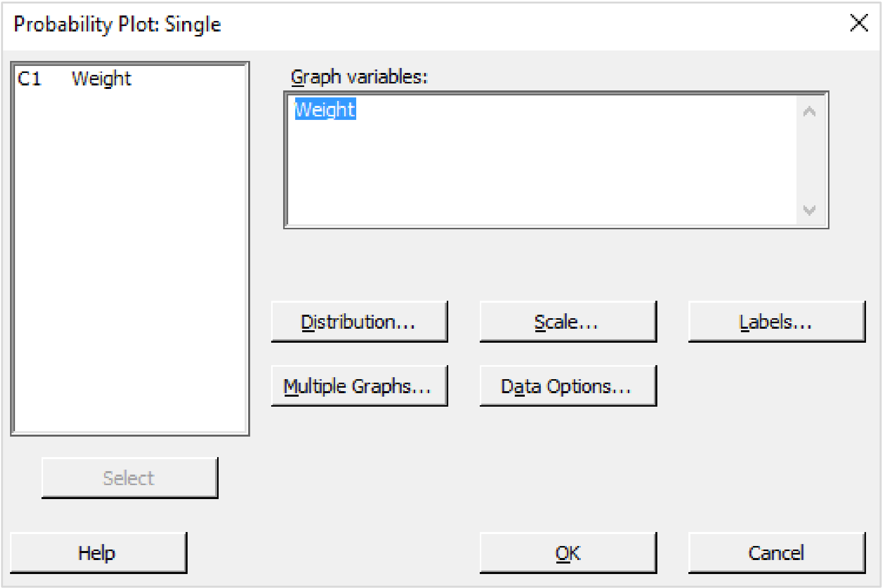




# - Normality Testing in Minitab

## Normal Probability Plots:

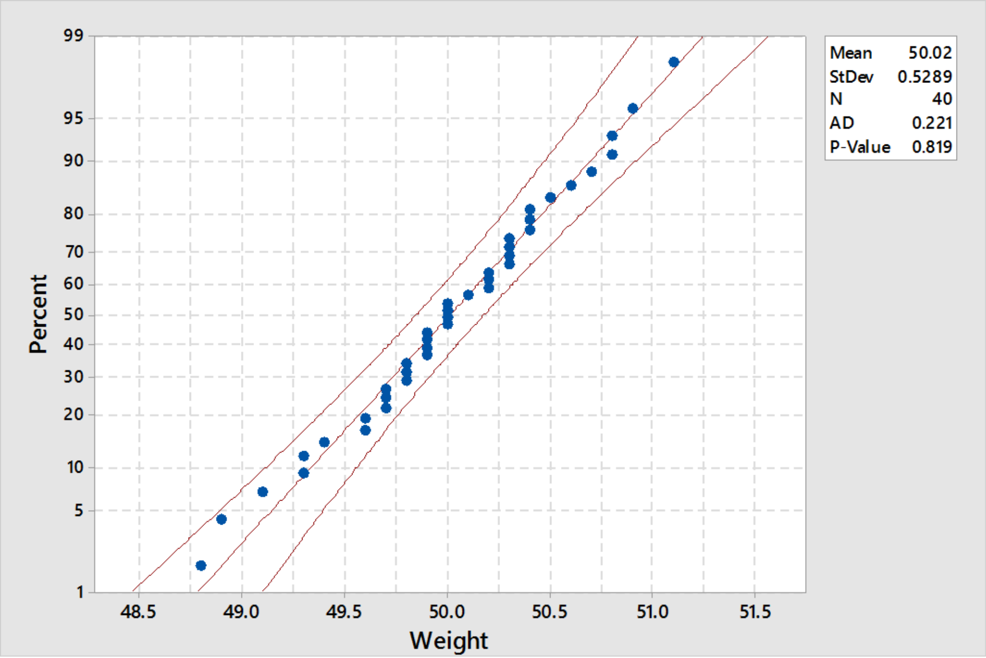
- To create a normal probability plot:
  - Select **Graph > Probability Plot > Single**.
  - Specify the column of data to analyze.
  - Leave the distribution option to be normal.
  - Click OK.



# - Normality Testing in Minitab

## Normal Probability Plots:

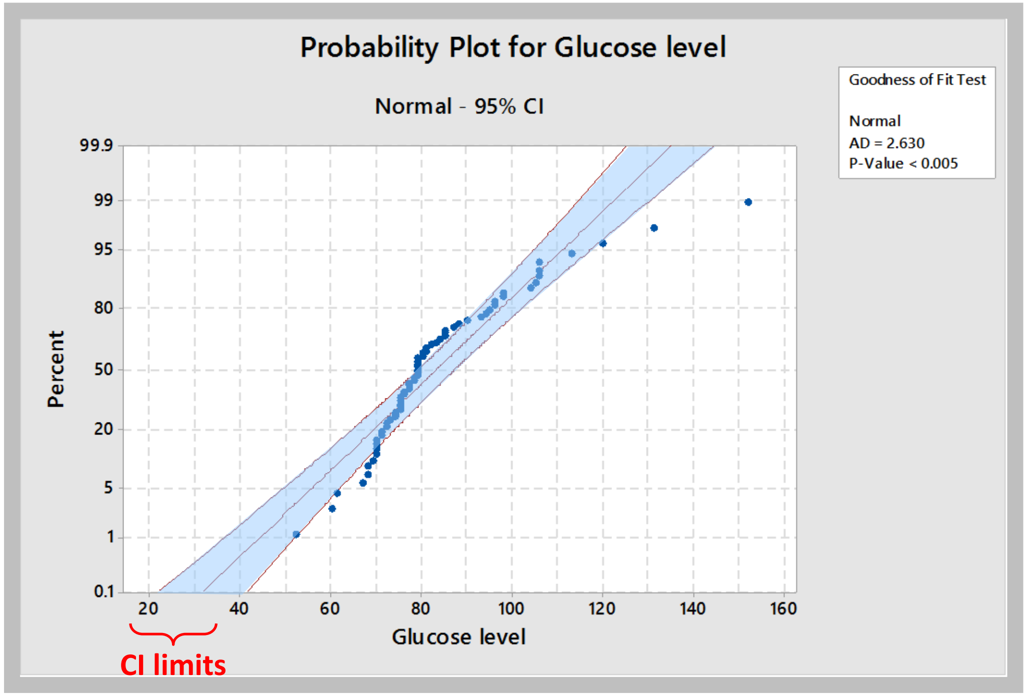
- Here is a screenshot of the example result for our previous example:




The data points approximately follow a straight line that falls mainly between CI limits
It can be concluded that the data is normally distributed

# - Normality Testing in Minitab

- What about the data in the following probability plot?



# - Normality Testing in Minitab

**Anderson-Darling Normality Test:**

- A statistical test that compares the actual distribution with the theoretical distribution.
- Measures how well the data follow the normal distribution (or any particular distribution).
- A lower p-value than the significance level (normally **0.05**) indicates a lack of normality in the data.
- Remember to keep your eyes on the histogram and the normal probability plot in conjunction with the Anderson-Darling test before making any decision.

Goodness of Fit Test

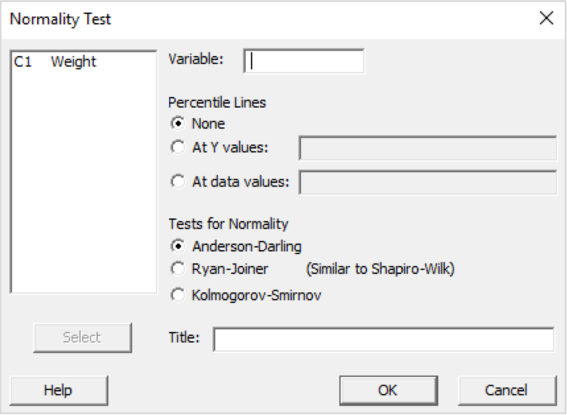
Normal
AD = 2.630
P-Value < 0.005

| | |
|---|---|
| Mean | 50.02 |
| StDev | 0.5289 |
| N | 40 |
| AD | 0.221 |
| P-Value | 0.819 |

# - Normality Testing in Minitab

## Anderson-Darling Normality Test:

- To conduct an Anderson-Darling normality test:
  - Select Stat > Basic Statistics > Normality Test.
  - Specify the column of data to analyze.
  - Specify the test method to be Anderson-Darling.
  - Click OK.



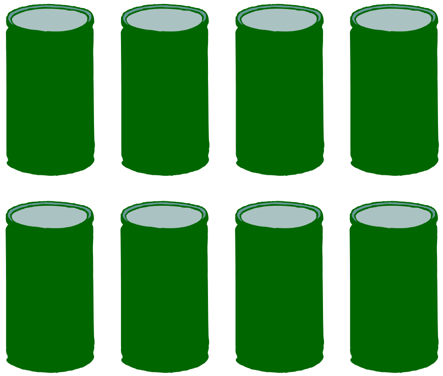

- The p-value in our product weight example was **0.819**.
- This suggests that the data follow the normal distribution.

| | |
|---|---|
| Mean | 50.02 |
| StDev | 0.5289 |
| N | 40 |
| AD | 0.221 |
| P-Value | 0.819 |

# - Normal Distribution

**Further Information:**

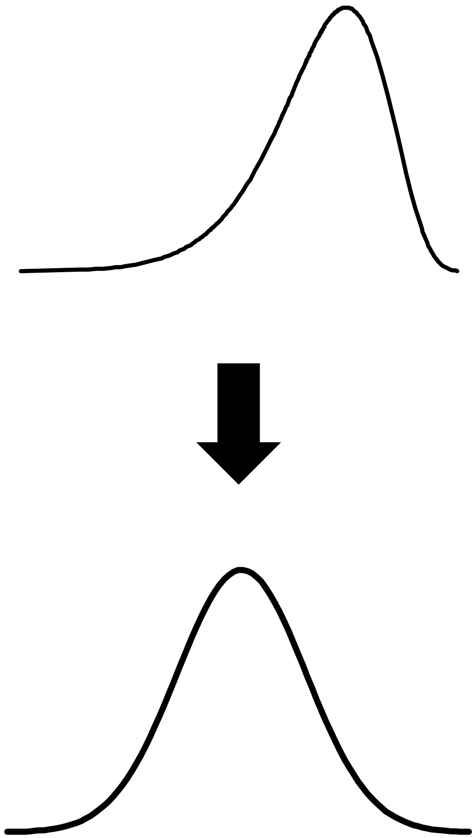
- The normal distribution is also known as the Gaussian Distribution after Carl Gauss who created the mathematical formula of the curve.
- Sometimes the process itself produces an approximately normal distribution.
- Other times the normal distribution can be obtained by performing a mathematical **transformation** on the data or by using means of **subgroups**.



# - Normal Distribution

Further Information:

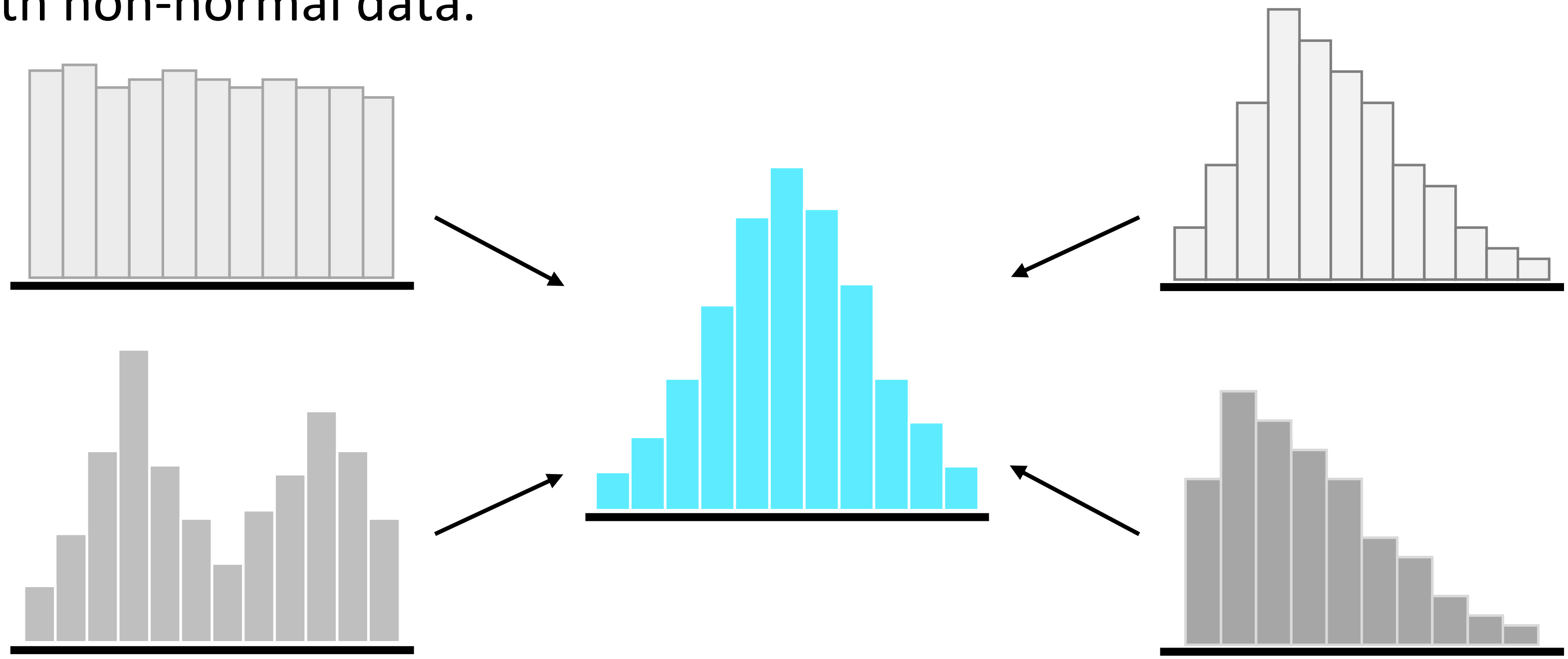
- The **Central Limit Theorem** is a useful statistical concept.
- It states that the distribution of the means of random samples will always approach a normal distribution regardless of the shape or underlying distribution.
- Even if the population is skewed, the subgroup means will always be normal if the sample size is large enough (n >= 30).



# - Normal Distribution

## Further Information:

- This means that statistical methods that work for normal distributions can also be applicable to other distributions.
- For example, certain SPC charts (such as XBar-R charts) can be used with non-normal data.



# - Normal Distribution

Further Information:

- In **Excel**, you may calculate the normal probabilities using the **NORM.DIST** function. Simply write:

=NORM.DIST(x, mean, standard deviation, FALSE)

- where 'x' is the data point in question.

# - Normal Distribution

Further Information:

- In **Excel**, **NORM.INV** is the inverse of the **NORM.DIST** function.
- It calculates the **X** variable given a probability.
- You can generate random numbers that follow the normal distribution by using the **NORM.INV** function. Use the formula:

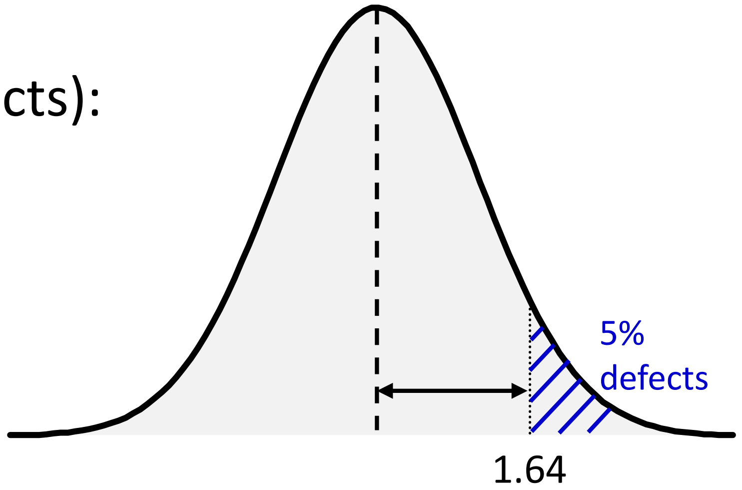
=NORM.INV(RAND(), mean, standard deviation)

# - Normal Distribution

Further Information:

- For **non-continuous data**, we can estimate the Z-value by calculating the area under the curves (DPU) and finding the equivalent Z-value for that area.
- Because discrete data is one-sided; the computed area is the total area.

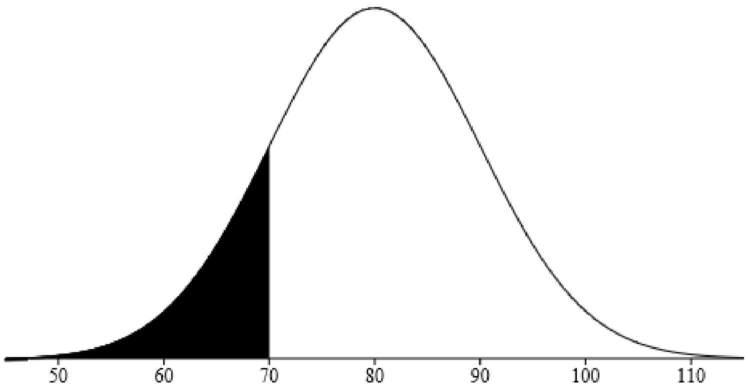
- A process that yields **95%** (or has **5%** defects):
  - The DPU for this process is **0.05**.
  - In the Z-table, an area of **0.05** has a Z-value of **1.64**.



# - Normal Distribution

Further Information - Exercise:

- An engineer collected **30** measurements. He calculated the average to be **80** and the standard deviation to be **10**. The lower specification limit is **70**.
- What is the Z-value?
- What percentage of the product is expected to be out of specification?



# - Normal Distribution

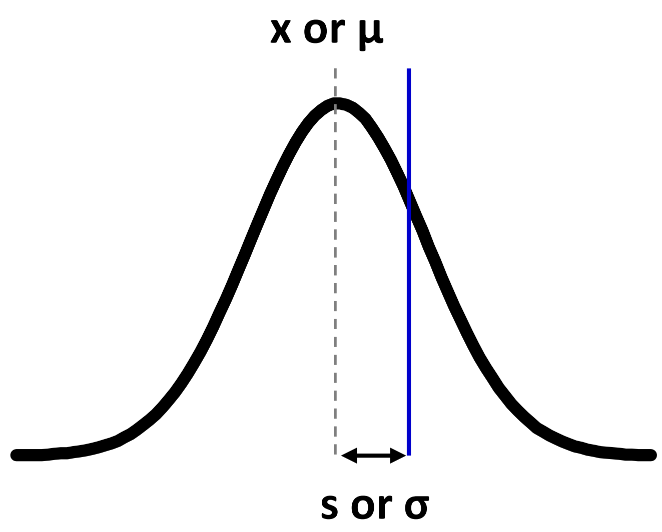
Further Information:

- **More examples using the Z-distribution and the Z-table:**
  - The Z-value corresponding to a proportion defective of **0.05** (in one tail).
  - The proportion of output greater than the process average.
  - The yield of a stable process whose specification limits are set at **+/- 3** standard deviations from the process average.
  - The average weight of a certain product is **4.4** g with a standard deviation of **1.3** g.
  - **Question:** What is the probability that a randomly selected product would weigh at least **3.1** g but not more than **7.0** g.

# - Normal Distribution

Further Information:

- **More examples using the Z-distribution and the Z-table:**
  - The life of a fully-charged cell phone battery is normally distributed with a mean of **14** hours with a standard deviation of **1** hour. What is the probability that a battery lasts at least **13** hours?



# - Normal Distribution

Further Information:

- **More examples using the Z-distribution and the Z-table:**
  - The maximum weight for a certain product should not exceed **52** gm. After sampling **40** products from the line, none of the weights exceeded the upper specification limit.
    - It may look that there will be no overweighed products, however, the normal curve suggests that there might be some over the long term.
    - The normal distribution can be used to make better prediction of the number of failures that will occur in the long term.
    - In our case, the Z-table predicts the area under the curve to be **0.6%** for a Z-value of **2.5**.
    - This is a better prediction than the **0%** assumed earlier.