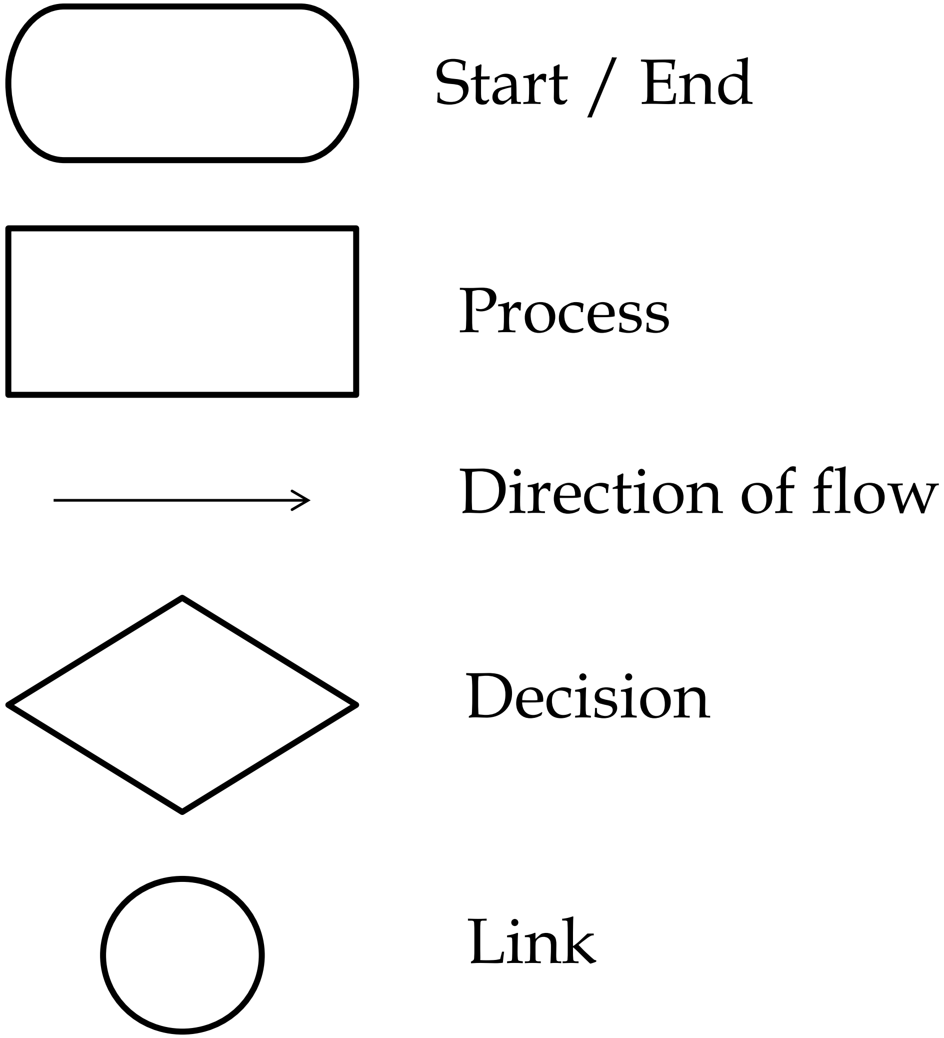
# 3A
# Process Analysis and Documentation

01 Process maps

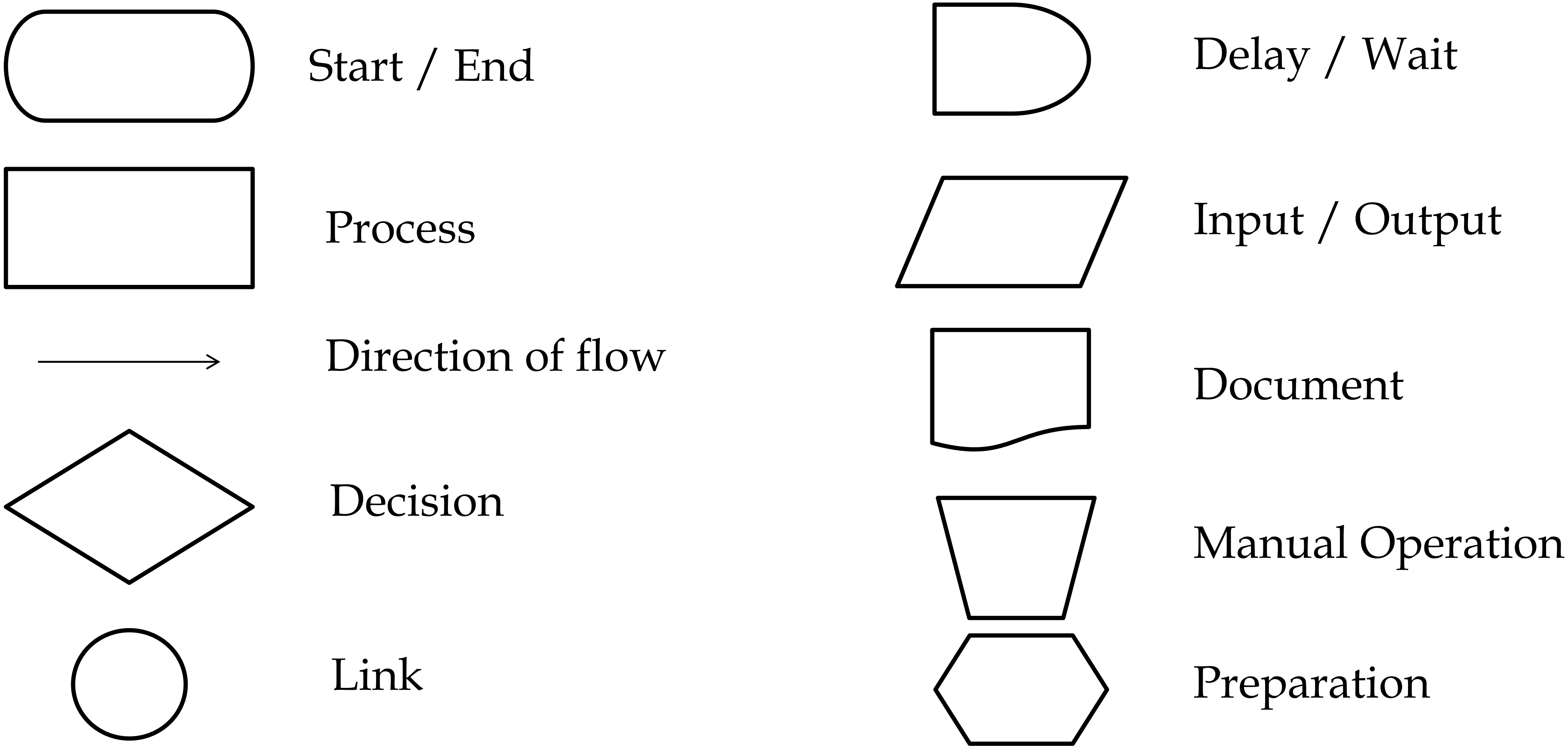
02 Documentation

# *Process Maps*

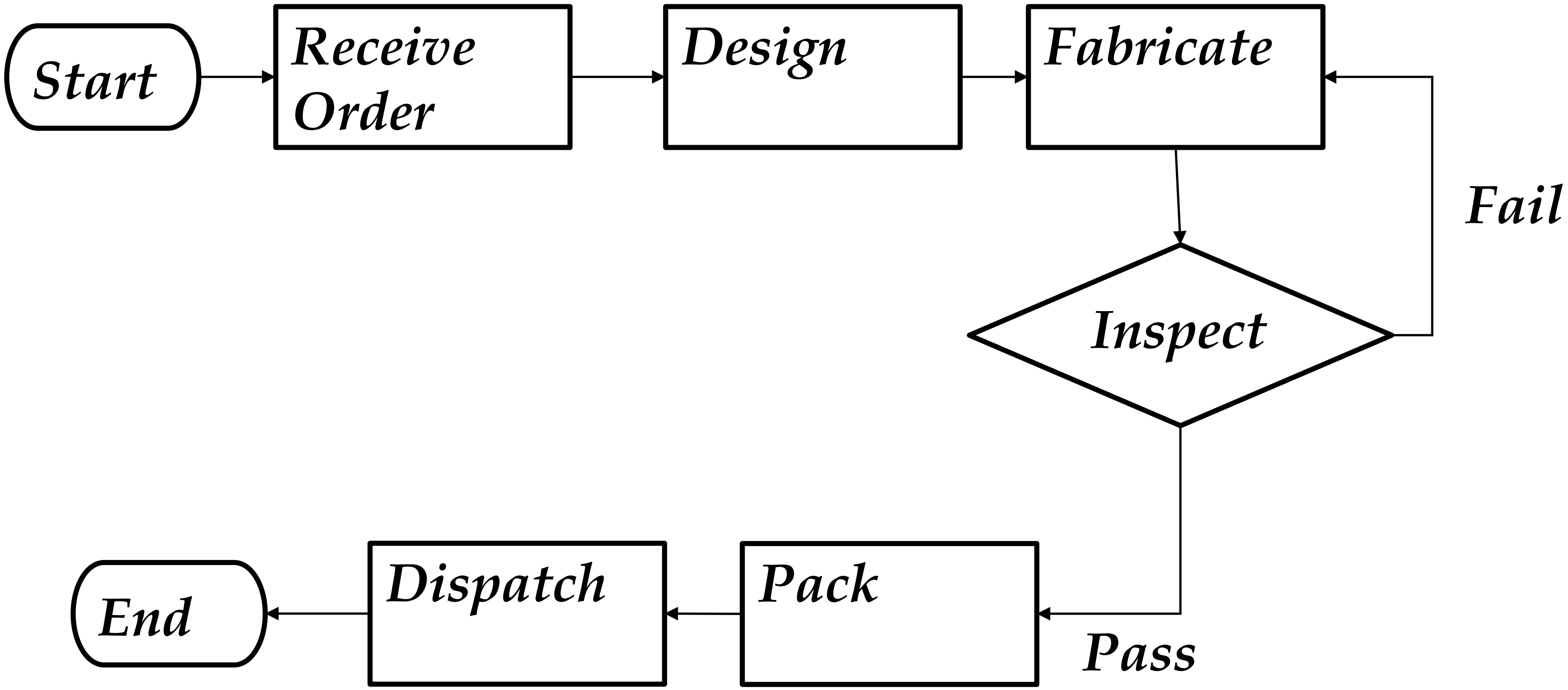
- Flow chart and process map are used interchangeably
- Process mapping is the process of creating a diagram; the diagram itself is called a flow chart.



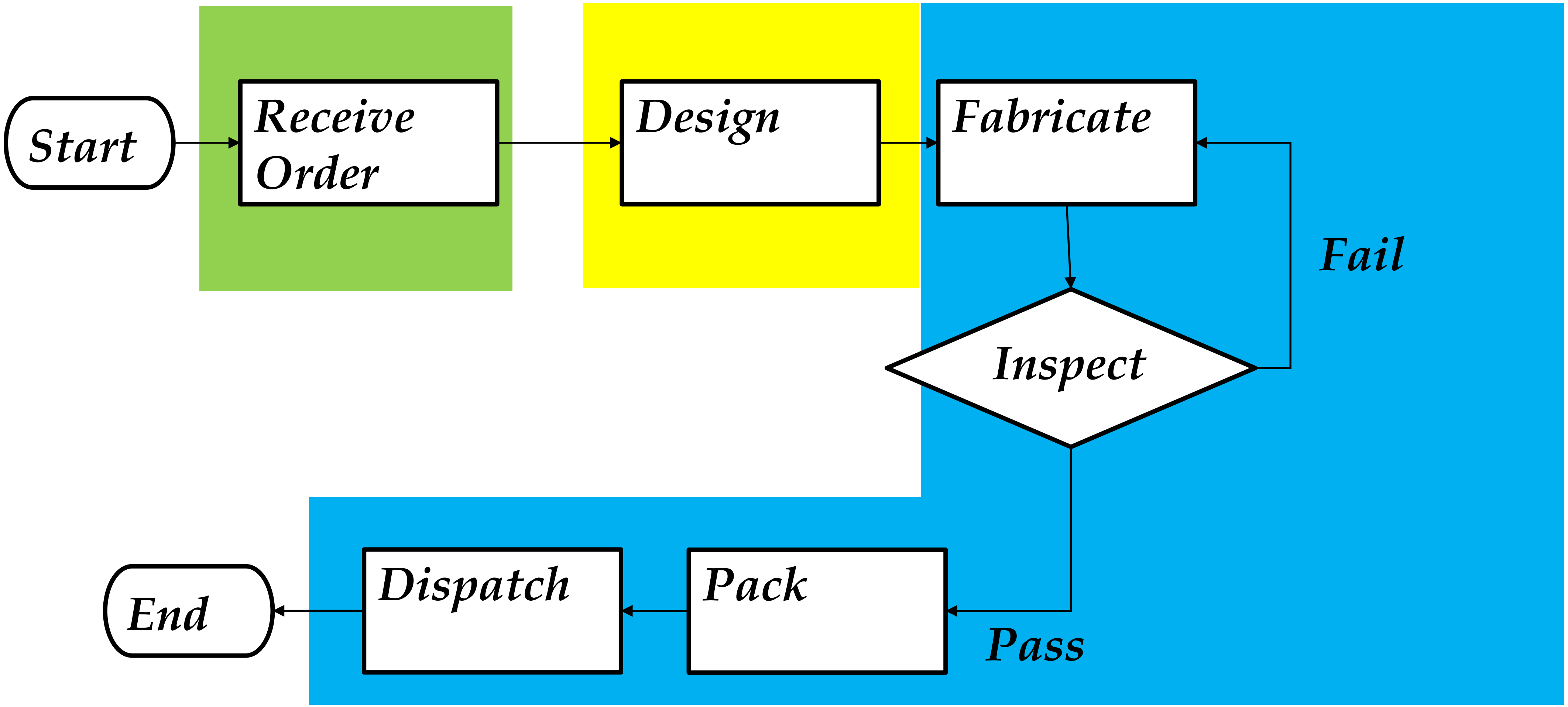

# Process Map - Elements



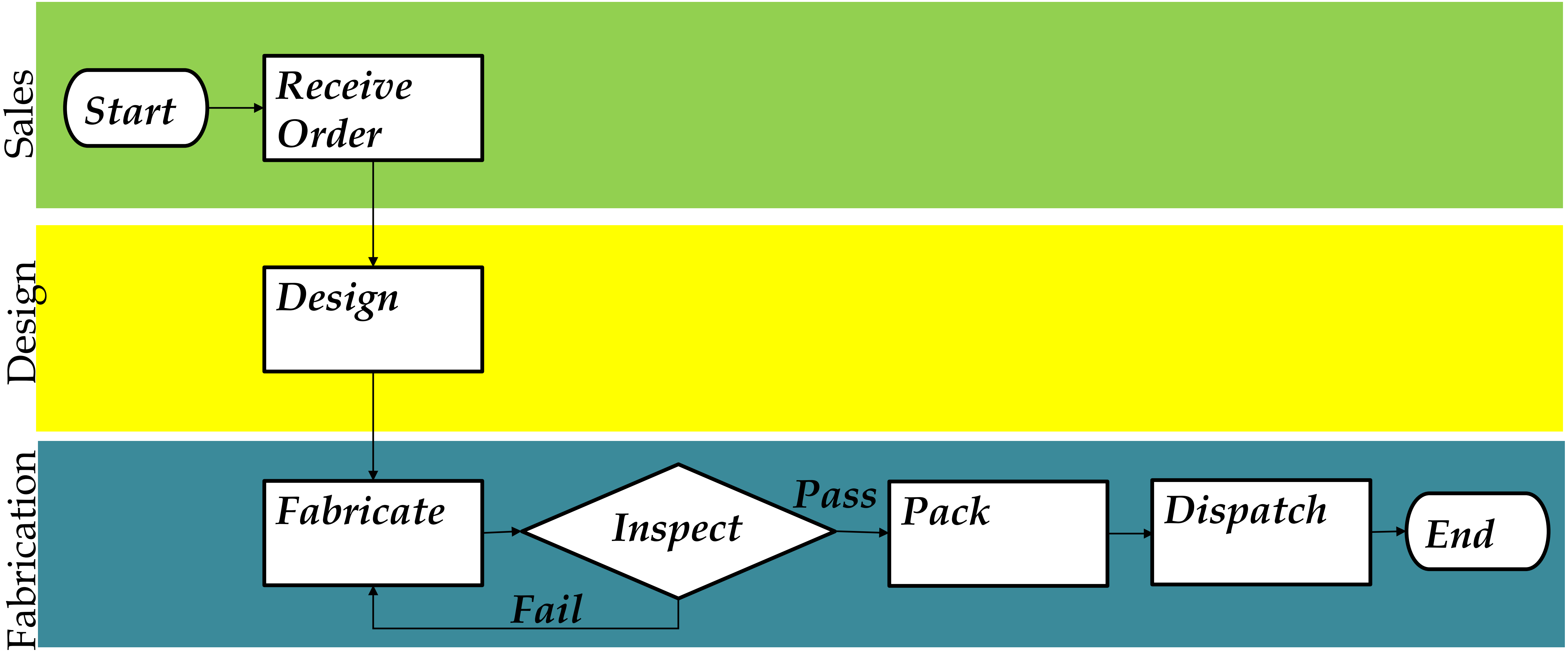

# Process Map - Elements




# Process Map – Hand drawn Example




# Process Map – Hand drawn Example

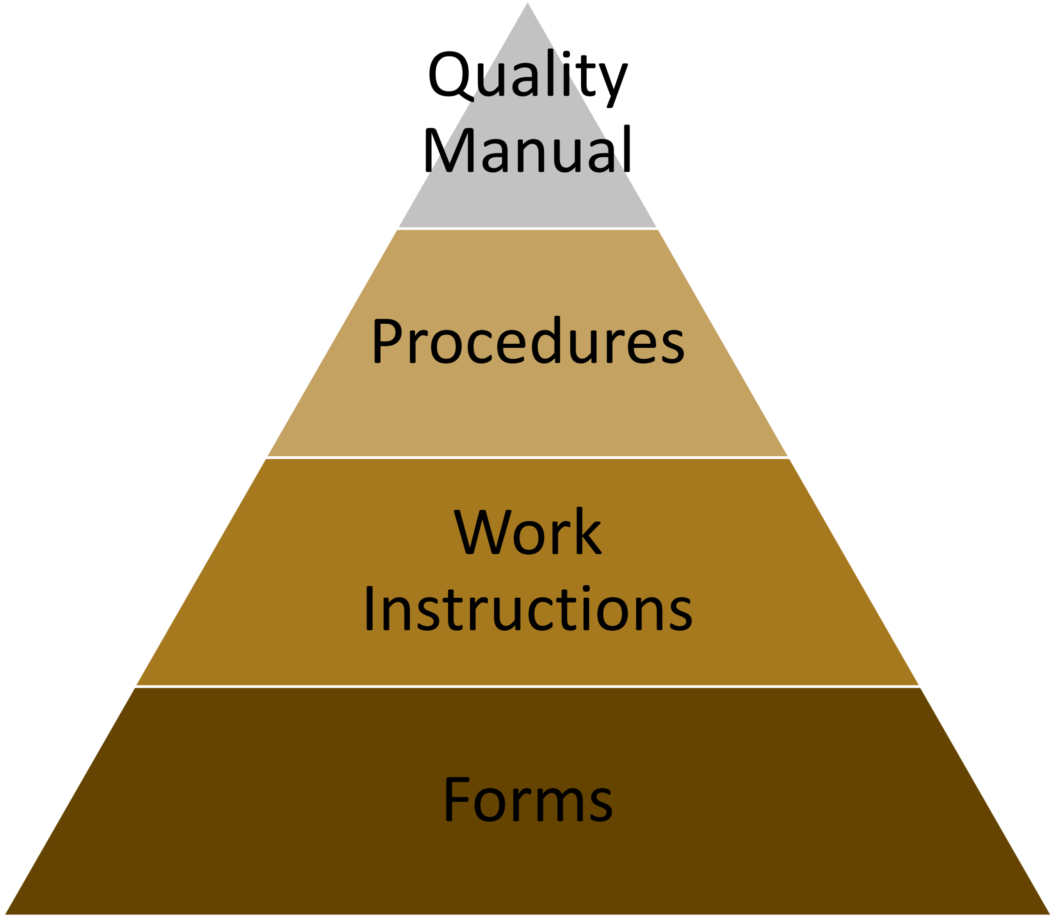



# Process Map – Swim Lane

| Suppliers | Inputs | Process | Outputs | Customers |
|---|---|---|---|---|
| | | | | |

# SIPOC

# *Documentation*




- Procedures specify what is done, when, where and why
- Work Instructions specify the details such as how and who.

# 3B Probability And Statistics

01 Basic probability concepts

02 Central limit theorem

# 3B-1 Basic Probability Concepts

01 Independent events

02 Mutually exclusive events

03 Multiplication rule

04 Permutations & combinations

- Classic Model

$$\frac{\text{Number of outcomes in which the event occurs}}{\text{Total Number of possible outcomes of an experiment}}$$

- Relative Frequency of Occurrence

$$\frac{\text{Number of times an event occurred}}{\text{Total number of opportunities for an event to occur}}$$

# *Probability*

- Experiment/Trial: Some thing done with an expectation of result.

- Event or Outcome: Result of experiment

- Sample Space: A sample space of an experiment is the set of all possible results of that random experiment.

$$\{1, 2, 3, 4, 5, 6\}$$

# *Probability*

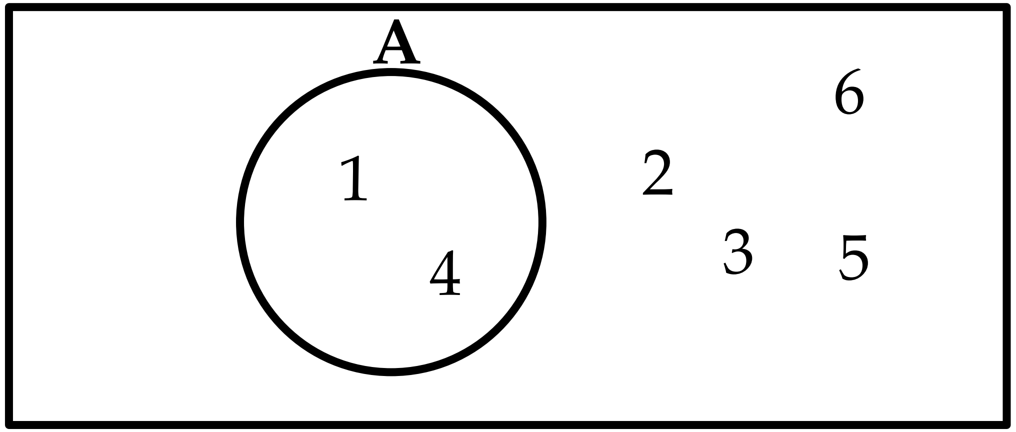
- Sample space: In roll of two dices

{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6),
(2,1), (2,2), (2,3), (2,4), (2,5), (2,6),
(3,1), (3,2), (3,3), (3,4), (3,5), (3,6),
(4,1), (4,2), (4,3), (4,4), (4,5), (4,6),
(5,1), (5,2), (5,3), (5,4), (5,5), (5,6),
(6,1), (6,2), (6,3), (6,4), (6,5), (6,6),}

*Probability*

- Venn diagram:

  Event A: Probability of getting 1 or 4 in the roll of a dice.

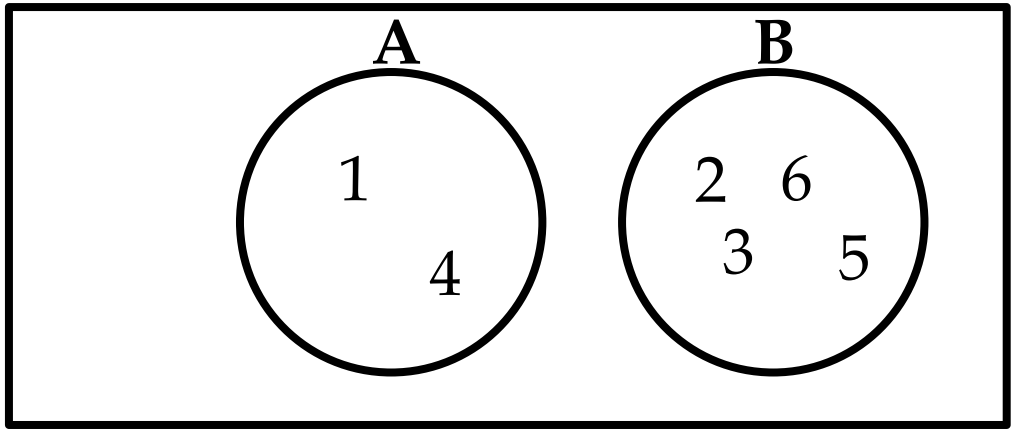



**Probability**

- Venn diagram:

  Event A: Probability of getting 1 or 4 in the roll of a dice.

  Event B: Probability of getting 2, 3, 5 or 6



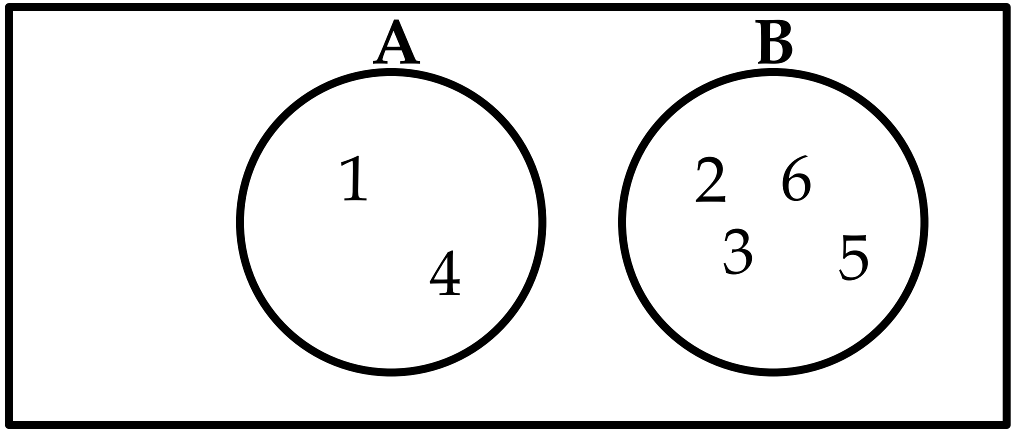

**Probability**

- Venn diagram:

  Event A: Probability of getting 1 or 4 in the roll of a dice.

  Event B: Probability of getting 2, 3, 5 or 6

**Mutually Exclusive Events: When two events cannot occur at the same time**

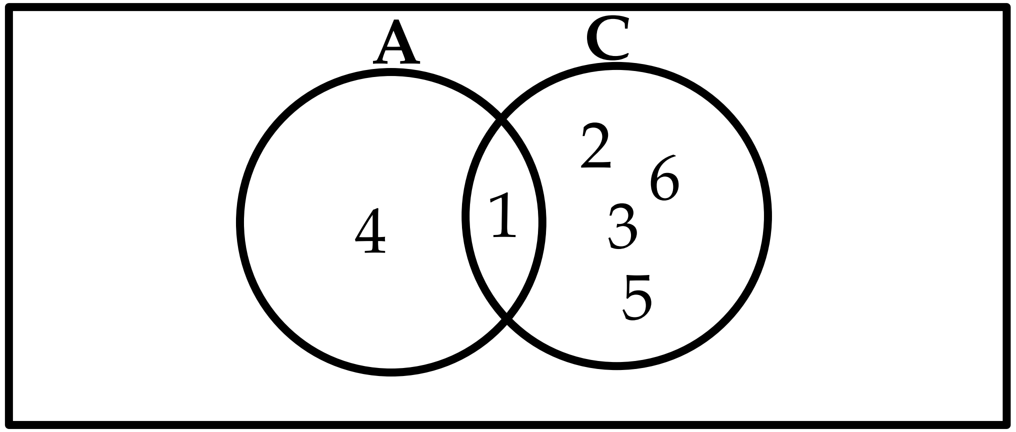



***Probability***

❖ Venn diagram:

Event A: Probability of getting 1 or 4 in the roll of a dice.

Event C: Probability of getting 1, 2, 3, 5 or 6




**Probability**

❖ Union:

Probability that events A <u>or</u> B occur

P(A ∪ B)

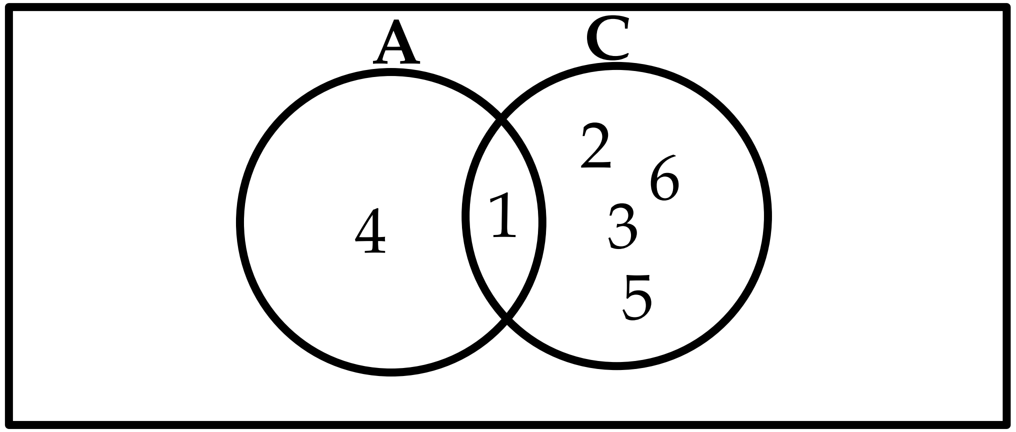
{1, 2, 3, 4, 5, 6}

❖ Intersection:

Probability that events A <u>and</u> B occur

P(A ∩ B)

{1}




***Probability***

- Mutually Exclusive Events: When two events cannot occur at the same time

- Independent Events: The occurrence of Event A does not change the probability of Event B

- Complementary Events: The probability that Event A will <u>NOT</u> occur is denoted by P(A').

# *Probability*

- Rule of Multiplication:

The probability that Events A **and** B both occur =

Probability that Event A occurs

x

Probability that Event B occurs, given that A has occurred

$$\mathbf{P(A \cap B) = P(A)\, P(B|A)}$$

*Probability*

- Rule of Multiplication:

The probability that Events A **and** B both occur =

Probability that Event A occurs

x

Probability that Event B occurs, given that A has occurred

$$\mathbf{P(A \cap B) = P(A)\, P(B|A)}$$

- In two rolls of dice what is the probability of getting 6 in both? (independent event)

***Probability***

- Rule of Multiplication:

The probability that Events A **and** B both occur =

Probability that Event A occurs

x

Probability that Event B occurs, given that A has occurred

$$P(A \cap B) = P(A)\, P(B|A)$$

- There are 10 candies in the plate (5 Green, 2 Yellow, 2 Orange and 1 Red). If I pick 2 random ones, what is the probability of getting both Yellow?

*Probability*

## ❖ Rule of Addition

The probability that Event A **or** Event B occurs

=

Probability that Event A occurs

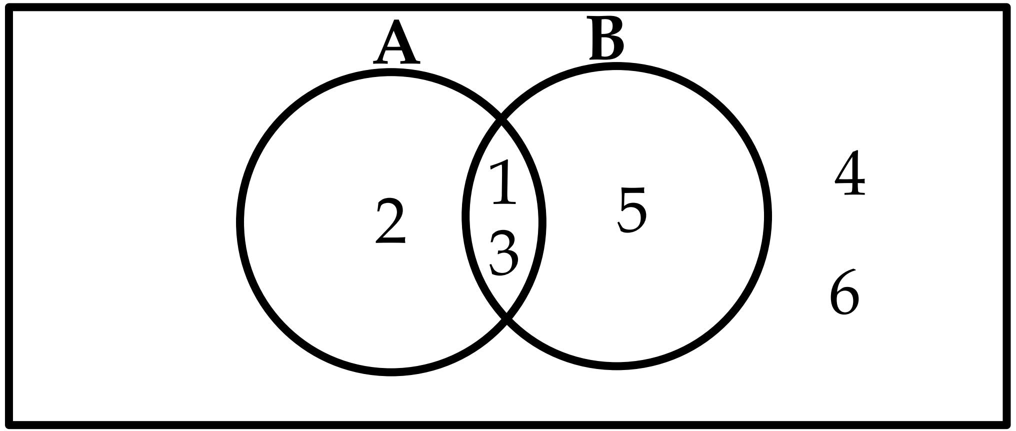
+

Probability that Event B occurs

-

Probability that both Events A and B occur

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$




*Probability*

- **Factorial** of a non-negative integer n, denoted by n!, is the product of all positive integers less than or equal to n

$$5! = 5 \times 4 \times 3 \times 2 \times 1$$

$$1! = 1$$

$$0! = 1$$

# *Factorial*

- **Permutation:** A set of objects in which position (or order) is important.

- **Combination:** A set of objects in which position (or order) is NOT important.

*Permutations*
*Combinations*

**Permutation:** With repetition

- In case of the lock, total permutations are
- 10 x 10 x 10 x 10 = 10,000

$$n^r$$

Excel Formula =PERMUTATIONA(10,4)

***Permutations***

***With repetition***

**Permutation:** Without repetition

- How many ways we can select 3 players out of 5. The first selected becomes the captain, second the vice-captain and third the treasurer.

$$n_{P_r} = \frac{n!}{(n-r)!}$$

$$\mathbf{5}_{P_\mathbf{3}} = \frac{\mathbf{5}!}{(\mathbf{5}-\mathbf{3})!} = 5 \times 4 \times 3 = 60$$

Excel Formula = PERMUT(5,3)

***Permutations***

***Without repetition***

- **Combination:** Without repetition:

- How many ways we can select 3 players out of 5

$$n_{C_r} = \frac{n!}{(n-r)!\,\mathrm{r}!}$$

$$\mathbf{5}_{C_3} = \frac{5!}{(5-3)!\,3!}$$

$$\mathbf{5}_{C_3} = \frac{5 \text{ x } 4 \text{ x } 3}{3 \text{ x } 2 \text{ x } 1} = 10$$

Excel Formula = COMBIN(5,3)

***Combinations***

***Without repetition***

- **Combination:** With repetition:

  - e.g. In the store there are 5 varieties of juice bottles. You want to buy 3 bottles. How many possible combinations you can buy?

~~Without repetition:~~ $\cancel{{}^{n}C_{r} = \frac{n!}{(n-r)!\,r!}}$

$$\frac{(r+n-1)!}{r!\,(n-1)!}$$

$$\frac{(5+3-1)!}{3!(5-1)!} = \frac{7!}{3!4!} = \frac{7 \times 6 \times 5}{3 \times 2 \times 1} = 35$$

Excel Formula = COMBINA(5,3)

# *Combinations*

## *With repetition*

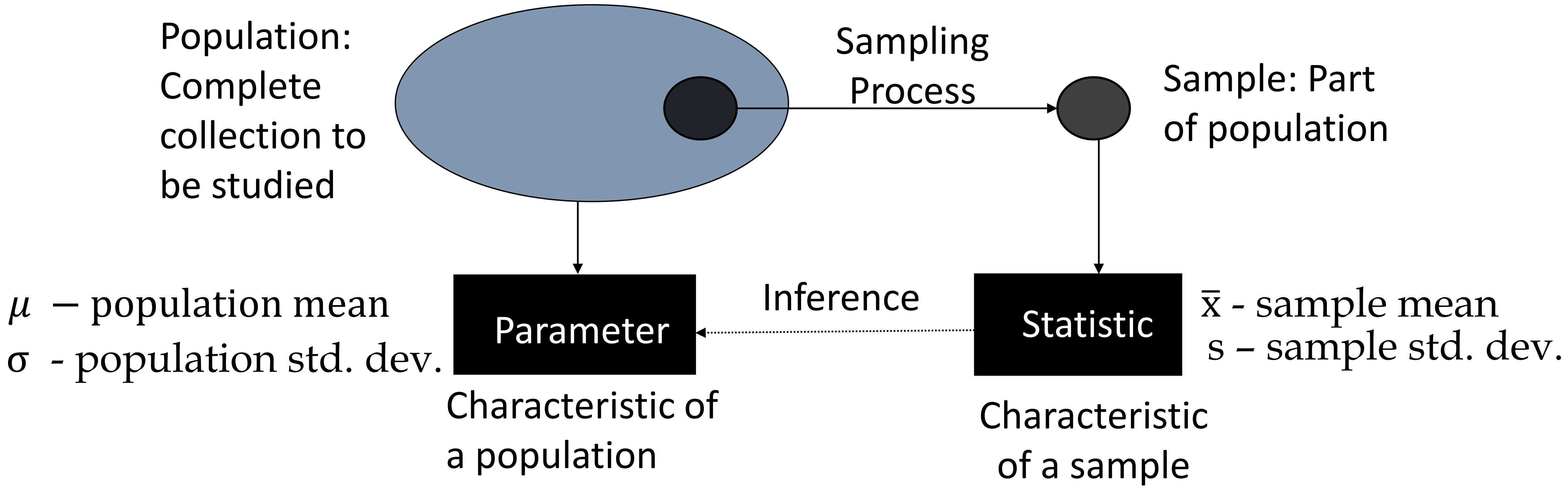
| | With repetition | Without repetition |
| --- | --- | --- |
| ❖ **Permutation:** Order is important. | $n^r$ | $n_{P_r} = \dfrac{n!}{(n-r)!}$ |
| MS Excel Formula >> | PERMUTATION**A**(n,r) | PERMUT(n,r) |
| ❖ **Combination:** Order is NOT important. | $\dfrac{(r+n-1)!}{r!\,(n-1)!}$ | $n_{C_r} = \dfrac{n!}{(n-r)!\,r!}$ |
| MS Excel Formula >> | COMBINA**A**(n,r) | COMBIN(n,r) |

# PERMUTATIONS / COMBINATIONS SUMMARY

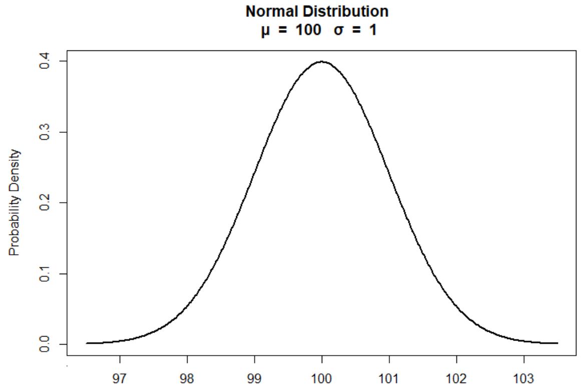
# 3B Probability And Statistics

01 Basic probability concepts

02 Central limit theorem


Population:
Complete
collection to
be studied
Sampling
Process
Sample: Part
of population
$\mu$ – population mean
σ - population std. dev.
Parameter
Inference
Statistic
x̄ - sample mean
s – sample std. dev.
Characteristic of
a population
Characteristic
of a sample

# SAMPLING




102 101
99 100

Mean = 100.5

$$CI = \bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$CI = 100.5 \pm 1$$

Population: Complete collection to be studied

Sampling Process

Sample: Part of population

Inference

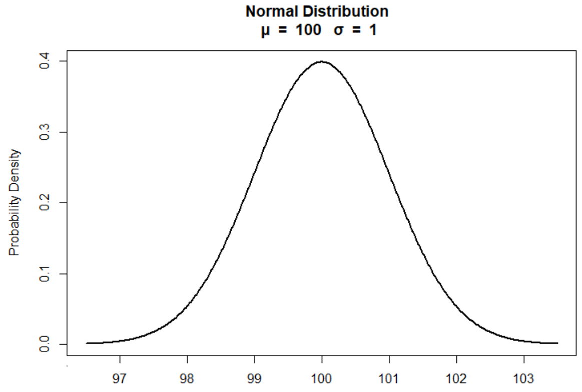
**Parameter**

Characteristic of a population

**Statistic**

Characteristic of a sample

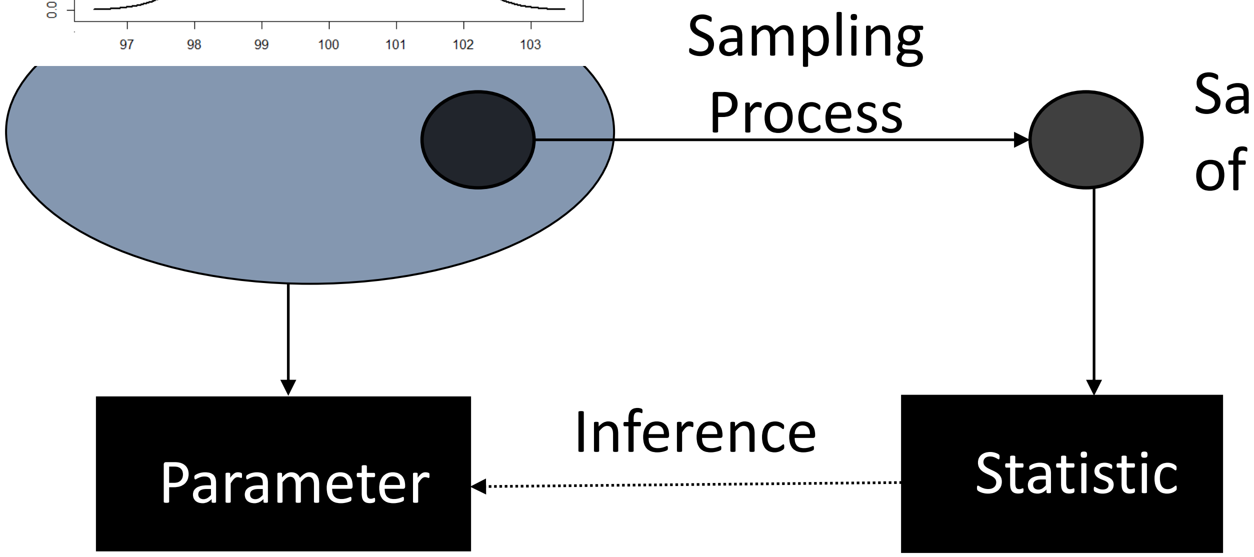
# SAMPLING – CONFIDENCE INTERVAL




102 101
99 100

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Population: Complete collection to be studied




Sample: Part of population

Characteristic of a population

Characteristic of a sample

# SAMPLING – HYPOTHESIS TEST

| Xbar Control Limits | Range Control Limits |
|---|---|
| Control limits $\bar{x} \pm A_2 \bar{R}$ | Upper control limit $D_4 \bar{R}$<br>Lower control limit $D_3 \bar{R}$ |

| Sample Size = m | $A_2$ | $A_3$ | $d_2$ | $D_3$ | $D_4$ | $B_3$ | $B_4$ |
|---|---|---|---|---|---|---|---|
| 2 | 1.880 | 2.659 | 1.128 | 0 | 3.267 | 0 | 3.267 |
| 3 | 1.023 | 1.954 | 1.693 | 0 | 2.574 | 0 | 2.568 |
| 4 | 0.729 | 1.628 | 2.059 | 0 | 2.282 | 0 | 2.266 |
| 5 | 0.577 | 1.427 | 2.326 | 0 | 2.114 | 0 | 2.089 |

# CONTROL CHARTS

- For almost all populations, the sampling distribution of the mean can be approximated closely by a normal distribution, provided the sample size is sufficiently large.

# *Central Limit Theorem*

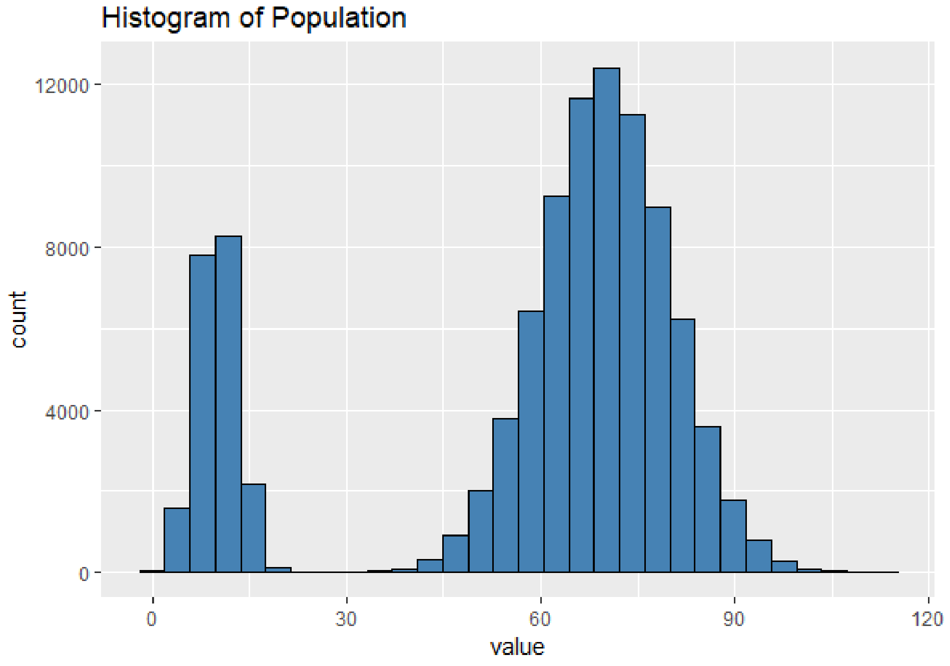
- If a variable has a mean of μ and the variance $\sigma^2$, as the sample size n increases, the sample mean approaches a normal distribution with mean $\mu_{\bar{x}}$ and variance $\boldsymbol{\sigma_{\bar{x}}^2}$

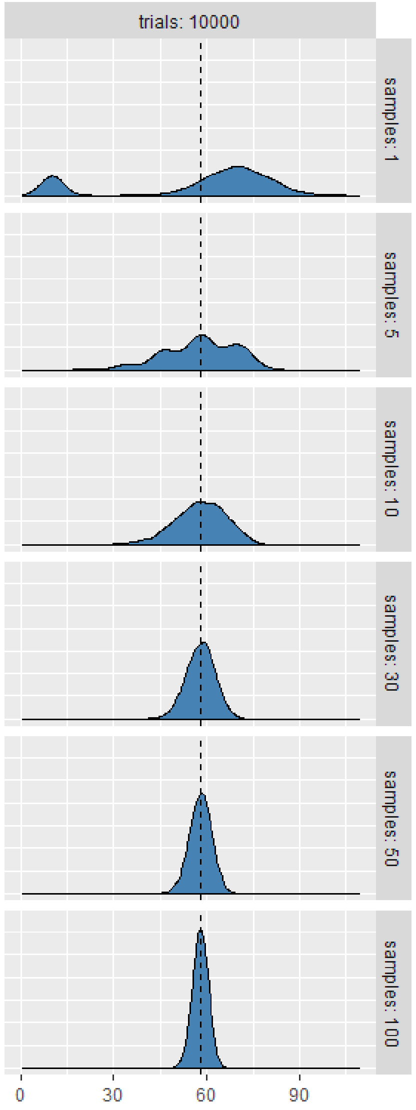
$$\mu_{\bar{x}} = \mu$$

$$\boldsymbol{\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}}$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

# *Central Limit Theorem*




# CENTRAL LIMIT THEOREM

- Standard deviation of the sampling distribution of the sample mean
  - Called “standard error of the mean”

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

# Central Limit Theorem

# 3C Statistical Distributions

01 Normal distribution

02 Binominal distribution

03 Poisson distribution

04 Chi square distribution

05 Student's t distribution

06 F distribution

QG

CONTINUOUS

Normal distribution

Chi square distribution

Student's t distribution

F distribution

DISCRETE

Binominal distribution

Poisson distribution

# CONTINUOUS vs DISCRETE DATA

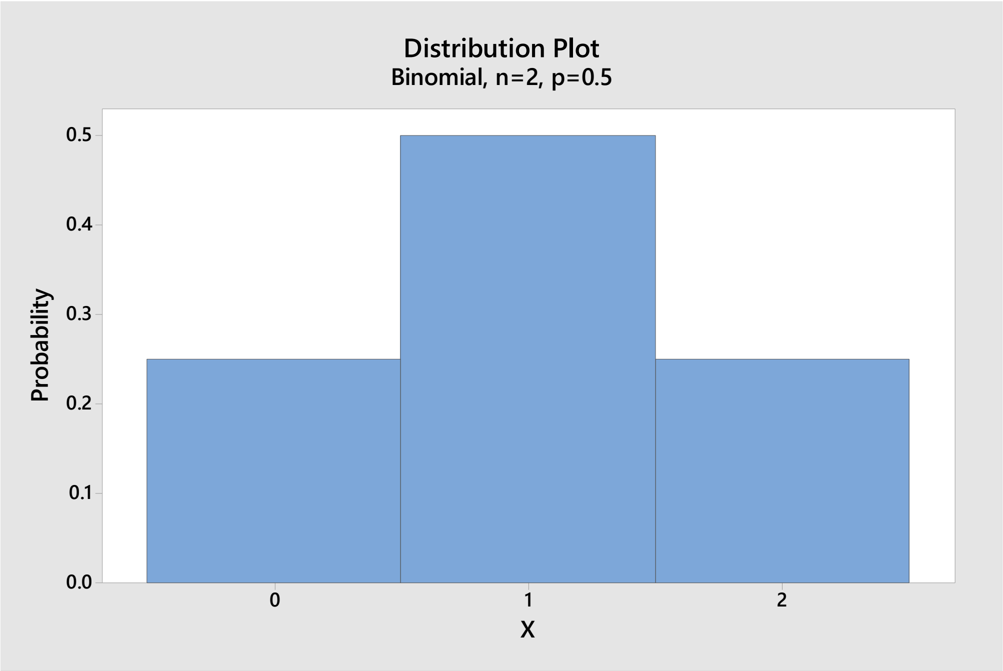
❖ If you flip a coin 2 times, what is the probability of getting 1 head?

| | |
|---|---|
| H | H |
| H | T |
| T | H |
| T | T |

# Binomial Distribution

- If you flip a coin 2 times, what is the probability of getting 1 head?

| | |
|---|---|
| H | H |
| H | T |
| T | H |
| T | T |




***Binomial Distribution***

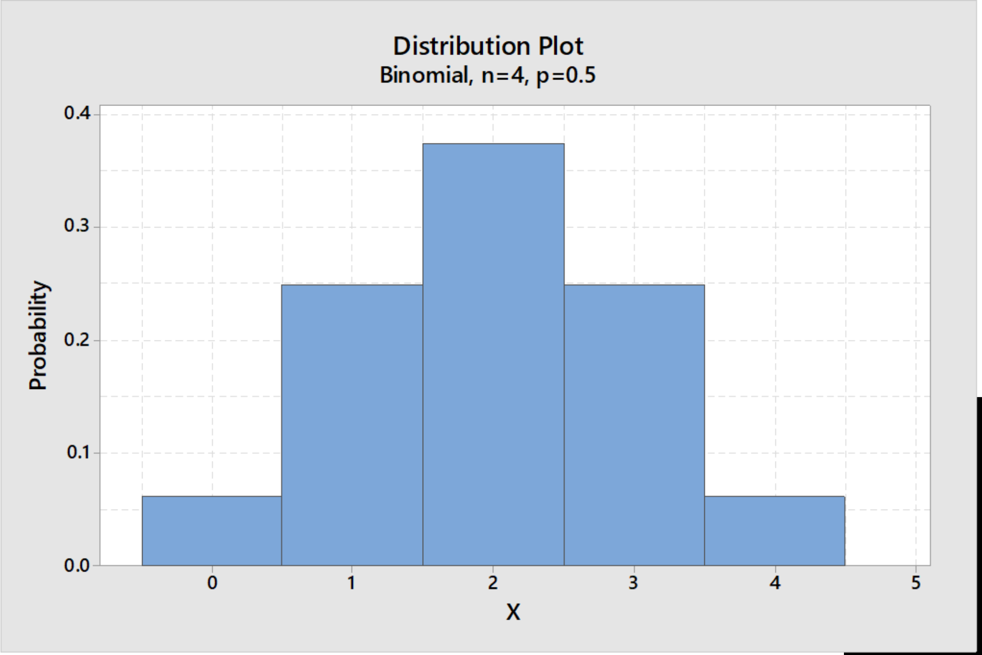
- If you flip a coin 4 times, what is the probability of getting 1 head?

| | | | |
|---|---|---|---|
| H | H | H | H |
| H | H | H | T |
| H | H | T | T |
| H | T | T | T |
| T | H | H | H |
| T | H | H | T |
| T | H | T | T |
| T | T | T | T |
| .. | .. | .. | .. |

# *Binomial Distribution*

- If you flip a coin 4 times, what is the probability of getting 1 head?

| | | | |
|---|---|---|---|
| H | H | H | H |
| H | H | H | T |
| H | H | T | T |
| H | T | T | T |
| T | H | H | H |
| T | H | H | T |
| T | H | T | T |
| T | T | T | T |
| .. | .. | .. | .. |




# *Binomial Distribution*

- **A binomial experiment** has the following properties:
  - The experiment consists of *n* repeated trials.
  - Each trial can result in just two possible outcomes. We call one of these outcomes a success and the other, a failure.
  - The probability of success, denoted by *p*, is the same on every trial.
  - The trials are independent; that is, the outcome on one trial does not affect the outcome on other trials.

## *Binomial Distribution*

$$\mathbf{P(x) = n_{C_x} \cdot p^x \cdot (1 - p)^{n-x}}$$

$$\mathbf{P(x) = \frac{n!}{x!\,(n - x)!} . p^x \cdot (1 - p)^{n-x}}$$

- **x**: The number of successes that result from the binomial experiment.
- **n**: The number of trials in the binomial experiment.
- **p**: The probability of success on an individual trial.
- **P(x)** : Binomial probability - the probability that an *n*-trial binomial experiment results in <u>exactly</u> *x* successes, when the probability of success on an individual trial is *p*.

# *Binomial Distribution*

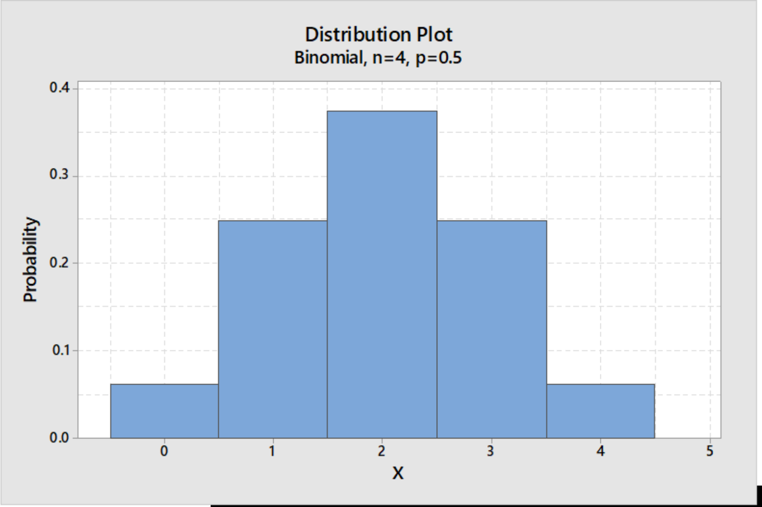
- If you flip a coin 4 times, what is the probability of getting 1 head?

$$\mathbf{P(x)} = \frac{\mathbf{n!}}{\mathbf{x!\,(n-x)!}}\,.\,\mathbf{p^{x} \cdot (1-p)^{n-x}}$$

$$\mathbf{P(1)} = \frac{\mathbf{4!}}{\mathbf{1!\,(4-1)!}}\,.\,\mathbf{0.5^{1} \cdot (1-0.5)^{4-1}}$$

$$\mathbf{P(1) = (4).(0.5)^{1} \cdot (0.5)^{3}}$$

$$\mathbf{P(1) = (4).(0.5)^{4} = 0.25}$$




BINOM.DIST(1,4,0.5,FALSE) = 0.25

BINOM.DIST(2,4,0.5,FALSE) = 0.375

***Binomial Distribution***

- The mean of the distribution ($\mu_x$) is

  $\boldsymbol{n \,.\, p}$

- The variance ($\sigma^2_x$) is

  $\boldsymbol{n \,.\, p \,.\, (\, 1 - p \,)}$

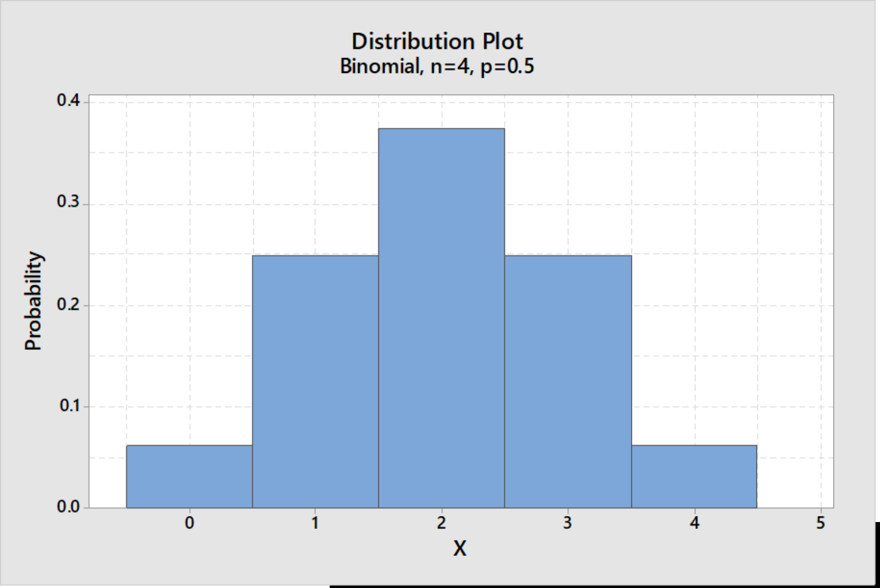
- The standard deviation ($\sigma_x$) is

  $$\sqrt{\mathbf{n \cdot p \cdot (1 - p)}}$$

# *Binomial Distribution*

- The mean of the distribution ($\mu_x$) is

$$\boldsymbol{n} \cdot \boldsymbol{p} = 4 \times 0.5 = 2$$




- The variance ($\sigma^2_x$) is

$$\boldsymbol{n} \cdot \boldsymbol{p} \cdot (1 - \boldsymbol{p})$$

$$4 \times 0.5 \times 0.5 = 1$$

- The standard deviation ($\sigma_x$) is

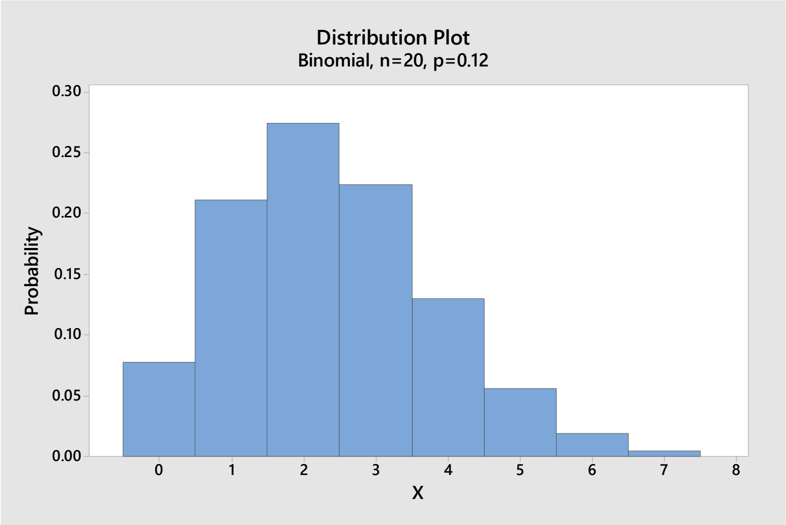
$$\sqrt{\mathbf{n} \cdot \mathbf{p} \cdot (\mathbf{1} - \mathbf{p})}$$

$$\sqrt{4 \times 0.5 \times 0.5} = 1$$

# *Binomial Distribution*

- A manufacturer has 12% defects rate in production. The buyer decides to test 20 random pieces and will accept the supplier if there are 2 or less defectives. What is the probability of getting accepted?

***=BINOM.DIST(2,20,0.12,TRUE) = 0.56***




***Binomial Distribution***

- p = 0.12, n = 20, x = 0, 1, 2

$$P(x) = \frac{n!}{x!\,(n-x)!}\,.\,p^{x}\cdot(1-p)^{n-x}$$

$$P(0) = \frac{20!}{0!\,(20-0)!}\,.\,0.12^{0}\cdot(1-0.12)^{20-0}$$

$$P(1) = \frac{20!}{1!\,(20-1)!}\,.\,0.12^{1}\cdot(1-0.12)^{20-1}$$

$$P(2) = \frac{20!}{2!\,(20-2)!}\,.\,0.12^{2}\cdot(1-0.12)^{20-2}$$

**P (0,1, 2) = 0.077563 + 0.211535 + 0.274034 = 0.563132**

***Binomial Distribution***

- $p = 0.12,\ n = 20,\ x = 0, 1, 2$
- The mean of the distribution ($\mu_x$) is

  $\boldsymbol{n\ .\ p} =$ 20 x 0.12 = 2.4

- The variance ($\sigma^2_x$) is

  $\boldsymbol{n\ .\ p\ .\ (\ 1 - p\ )}$

  20 x 0.12 x 0.88 = 2.112

- The standard deviation ($\sigma_x$) is

  $\sqrt{\mathbf{n} \cdot \mathbf{p} \cdot (\mathbf{1} - \mathbf{p})}$ = 1.453

***Binomial Distribution***

# 3C Statistical Distributions

01 Normal distribution

02 Binominal distribution

03 Poisson distribution

04 Chi square distribution

05 Student's t distribution

06 F distribution

## CONTINUOUS

Normal distribution

Chi square distribution

Student's t distribution

F distribution

## DISCRETE

Binominal distribution

Poisson distribution

# CONTINUOUS vs DISCRETE DATA

**Binomial vs Poisson Distribution**

Similarities:

- Both are for discrete distribution
- Both measure the number of successes

Differences:

- In Poisson distribution the possibilities of success are infinite.

Binomial distribution is one in which the probability of repeated number of trials are studied.

Poisson Distribution gives the count of independent events occur randomly with a given period of time.

*Poisson Distribution*

A **Poisson experiment** has the following properties:

- ❖ The experiment results in outcomes that can be classified as successes or failures.
- ❖ The average number of successes ($\mu$) that occurs in a specified region is known.
- ❖ Outcomes are random. Occurrence of one outcome does not influence the chance of another outcome of interest.
- ❖ The outcomes of interest are rare relative to the possible outcomes.
  - ❖ Example: Road accidents, queue at the counter

# *Poisson Distribution*

$$P(x, \mu) = e^{-\mu} \cdot \frac{\mu^x}{x!}$$

- *e*: A constant equal to approximately 2.71828. (Actually, *e* is the base of the natural logarithm system)
- μ: The mean number of successes that occur in a specified region.
- *x*: The actual number of successes that occur in a specified region.
- P(*x*; μ): The **Poisson probability** that exactly *x* successes occur in a Poisson experiment, when the mean number of successes is μ.

# *Poisson Distribution*

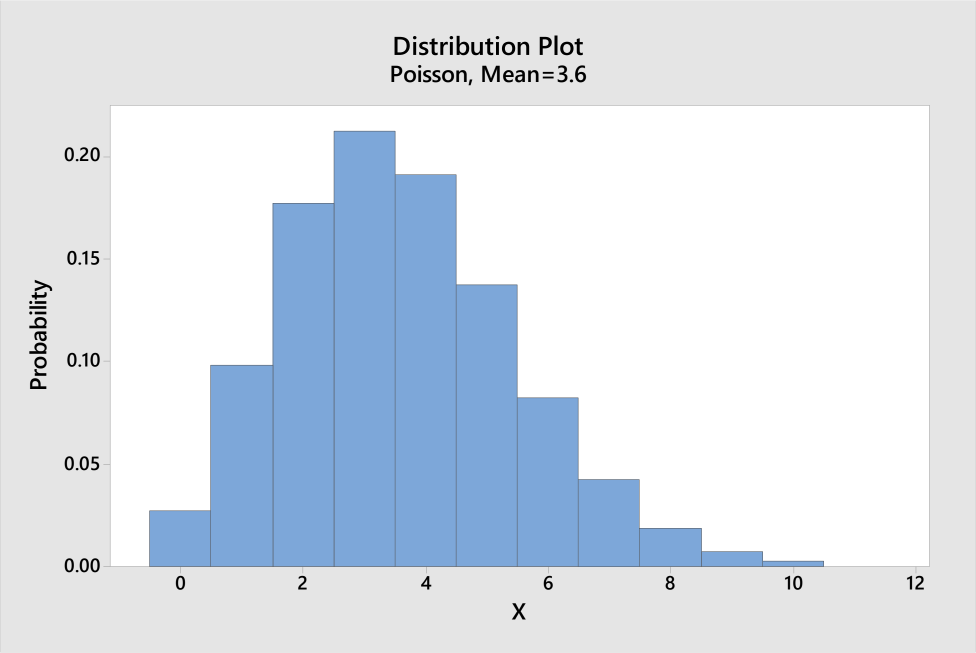
$$P(x, \mu) = e^{-\mu} \cdot \frac{\mu^x}{x!}$$

- On a booking counter on the average 3.6 people come every 10 minute on weekends. What is the probability of getting 7 people in 10 minutes?
- $\mu = 3.6$, $x=7$
- $P(x; \mu) = (e^{-\mu}) (\mu^x) / x! = (e^{-3.6}) (3.6^7) / 7!$
- $=0.02732 \times 7836.41 / 5040 = 0.0424$

=POISSON.DIST(7, 3.6, FALSE) = 0.042484

# Poisson Distribution

- On a booking counter on the average 3.6 people come every 10 minute on weekends. What is the probability of getting 7 people in 10 minutes?
- P(*7*; 3.6) = 0.0424




# *Poisson Distribution*

- The Poisson distribution has the following properties:
- The mean of the distribution is equal to $\mu$ .
- The variance is also equal to $\mu$ .

***Poisson Distribution***

# 3C Statistical Distributions

01 Normal distribution

02 Binominal distribution

03 Poisson distribution

04 Chi square distribution

05 Student's t distribution

06 F distribution

## CONTINUOUS

**Normal distribution**

**Chi square distribution**

**Student's t distribution**

**F distribution**

## DISCRETE

**Binominal distribution**

**Poisson distribution**

# CONTINUOUS vs DISCRETE DATA

- Symmetrically distributed
- Long Tails / Bell Shaped
- Mean/ Mode and Median are same

# *Normal Distribution*

- Two factors define the shape of the curve:
  - Mean
  - Standard Deviation

*Normal Distribution*

- About 68% of the area under the curve falls within **1 standard deviation** of the mean.

- About 95% of the area under the curve falls within **2 standard deviations** of the mean.

- About 99.7% of the area under the curve falls within **3 standard deviations** of the mean.

## *Normal Distribution*

- The total area under the normal curve = 1.
- The probability of any particular value is 0.
- The probability that X is greater than or less than a value = area under the normal curve in that direction

# *Normal Distribution*

$$P(x) = \frac{\mathbf{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

- where x is a normal random variable,
- μ = mean,
- σ = standard deviation,
- π is approximately 3.14159,
- e is approximately 2.71828.

# *Normal Distribution*

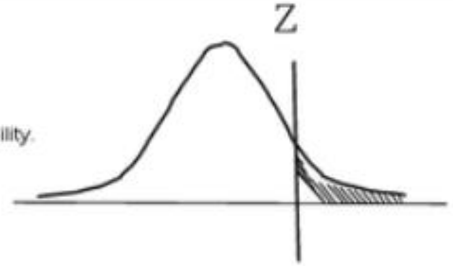
## Z Value / Standard Score

- How many standard deviations an element is from the mean.
- $z = (x - \mu) / \sigma$

- z is the z-score,
- x is the value of the element,
- μ is the population mean,
- σ is the standard deviation.

*Normal Distribution*

The value in the table is the right tail probability.




Hundredth place for Z-value →

| Z-Value | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.50000 | 0.49601 | 0.49202 | 0.48803 | 0.48405 | 0.48006 | 0.47608 | 0.47210 | 0.46812 | 0.46414 |
| 0.1 | 0.46017 | 0.45620 | 0.45224 | 0.44828 | 0.44433 | 0.44038 | 0.43644 | 0.43251 | 0.42858 | 0.42465 |
| 0.2 | 0.42074 | 0.41683 | 0.41294 | 0.40905 | 0.40517 | 0.40129 | 0.39743 | 0.39358 | 0.38974 | 0.38591 |
| 0.3 | 0.38209 | 0.37828 | 0.37448 | 0.37070 | 0.36693 | 0.36317 | 0.35942 | 0.35569 | 0.35197 | 0.34827 |
| 0.4 | 0.34458 | 0.34090 | 0.33724 | 0.33360 | 0.32997 | 0.32636 | 0.32276 | 0.31918 | 0.31561 | 0.31207 |
| 0.5 | 0.30854 | 0.30503 | 0.30153 | 0.29806 | 0.29460 | 0.29116 | 0.28774 | 0.28434 | 0.28096 | 0.27760 |
| 0.6 | 0.27425 | 0.27093 | 0.26763 | 0.26435 | 0.26109 | 0.25785 | 0.25463 | 0.25143 | 0.24825 | 0.24510 |
| 0.7 | 0.24196 | 0.23885 | 0.23576 | 0.23270 | 0.22965 | 0.22663 | 0.22363 | 0.22065 | 0.21770 | 0.21476 |
| 0.8 | 0.21186 | 0.20897 | 0.20611 | 0.20327 | 0.20045 | 0.19766 | 0.19489 | 0.19215 | 0.18943 | 0.18673 |
| 0.9 | 0.18406 | 0.18141 | 0.17879 | 0.17619 | 0.17361 | 0.17106 | 0.16853 | 0.16602 | 0.16354 | 0.16109 |
| 1.0 | 0.15866 | 0.15625 | 0.15386 | 0.15151 | 0.14917 | 0.14686 | 0.14457 | 0.14231 | 0.14007 | 0.13786 |
| 1.1 | 0.13567 | 0.13350 | 0.13136 | 0.12924 | 0.12714 | 0.12507 | 0.12302 | 0.12100 | 0.11900 | 0.11702 |
| 1.2 | 0.11507 | 0.11314 | 0.11123 | 0.10935 | 0.10749 | 0.10565 | 0.10383 | 0.10204 | 0.10027 | 0.09853 |
| 1.3 | 0.09680 | 0.09510 | 0.09342 | 0.09176 | 0.09012 | 0.08851 | 0.08691 | 0.08534 | 0.08379 | 0.08226 |
| 1.4 | 0.08076 | 0.07927 | 0.07780 | 0.07636 | 0.07493 | 0.07353 | 0.07215 | 0.07078 | 0.06944 | 0.06811 |
| 1.5 | 0.06681 | 0.06552 | 0.06426 | 0.06301 | 0.06178 | 0.06057 | 0.05938 | 0.05821 | 0.05705 | 0.05592 |
| 1.6 | 0.05480 | 0.05370 | 0.05262 | 0.05155 | 0.05050 | 0.04947 | 0.04846 | 0.04746 | 0.04648 | 0.04551 |

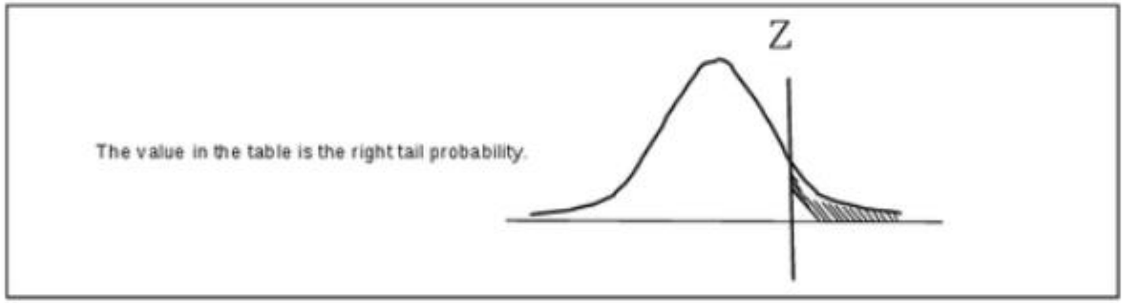
# *Normal Distribution*

**Z Value / Standard Score**

- Perfume bottles are filled with the average volume of 150 cc and the standard deviation of 2 cc.
- What percent of bottles will have the volume more than 153 cc?

$\mu = 150$ cc
$\sigma = 2$ cc

$z = (x - \mu) / \sigma = (153 - 150)/2 = 1.5$




Hundredth place for Z-value →

| Z-Value | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.50000 | 0.49601 | 0.49202 | 0.48803 | 0.48405 | 0.48006 | 0.47608 | 0.47210 | 0.46812 | 0.46414 |
| 0.1 | 0.46017 | 0.45620 | 0.45224 | 0.44828 | 0.44433 | 0.44038 | 0.43644 | 0.43251 | 0.42858 | 0.42465 |
| 0.2 | 0.42074 | 0.41683 | 0.41294 | 0.40905 | 0.40517 | 0.40129 | 0.39743 | 0.39358 | 0.38974 | 0.38591 |
| 0.3 | 0.38209 | 0.37828 | 0.37448 | 0.37070 | 0.36693 | 0.36317 | 0.35942 | 0.35569 | 0.35197 | 0.34827 |
| 0.4 | 0.34458 | 0.34090 | 0.33724 | 0.33360 | 0.32997 | 0.32636 | 0.32276 | 0.31918 | 0.31561 | 0.31207 |
| 0.5 | 0.30854 | 0.30503 | 0.30153 | 0.29806 | 0.29460 | 0.29116 | 0.28774 | 0.28434 | 0.28096 | 0.27760 |
| 0.6 | 0.27425 | 0.27093 | 0.26763 | 0.26435 | 0.26109 | 0.25785 | 0.25463 | 0.25143 | 0.24825 | 0.24510 |
| 0.7 | 0.24196 | 0.23885 | 0.23576 | 0.23270 | 0.22965 | 0.22663 | 0.22363 | 0.22065 | 0.21770 | 0.21476 |
| 0.8 | 0.21186 | 0.20897 | 0.20611 | 0.20327 | 0.20045 | 0.19766 | 0.19489 | 0.19215 | 0.18943 | 0.18673 |
| 0.9 | 0.18406 | 0.18141 | 0.17879 | 0.17619 | 0.17361 | 0.17106 | 0.16853 | 0.16602 | 0.16354 | 0.16109 |
| 1.0 | 0.15866 | 0.15625 | 0.15386 | 0.15151 | 0.14917 | 0.14686 | 0.14457 | 0.14231 | 0.14007 | 0.13786 |
| 1.1 | 0.13567 | 0.13350 | 0.13136 | 0.12924 | 0.12714 | 0.12507 | 0.12302 | 0.12100 | 0.11900 | 0.11702 |
| 1.2 | 0.11507 | 0.11314 | 0.11123 | 0.10935 | 0.10749 | 0.10565 | 0.10383 | 0.10204 | 0.10027 | 0.09853 |
| 1.3 | 0.09680 | 0.09510 | 0.09342 | 0.09176 | 0.09012 | 0.08851 | 0.08691 | 0.08534 | 0.08379 | 0.08226 |
| 1.4 | 0.08076 | 0.07927 | 0.07780 | 0.07636 | 0.07493 | 0.07353 | 0.07215 | 0.07078 | 0.06944 | 0.06811 |
| 1.5 | 0.06681 | 0.06552 | 0.06426 | 0.06301 | 0.06178 | 0.06057 | 0.05938 | 0.05821 | 0.05705 | 0.05592 |
| 1.6 | 0.05480 | 0.05370 | 0.05262 | 0.05155 | 0.05050 | 0.04947 | 0.04846 | 0.04746 | 0.04648 | 0.04551 |

***Normal Distribution***

## Z Value / Standard Score

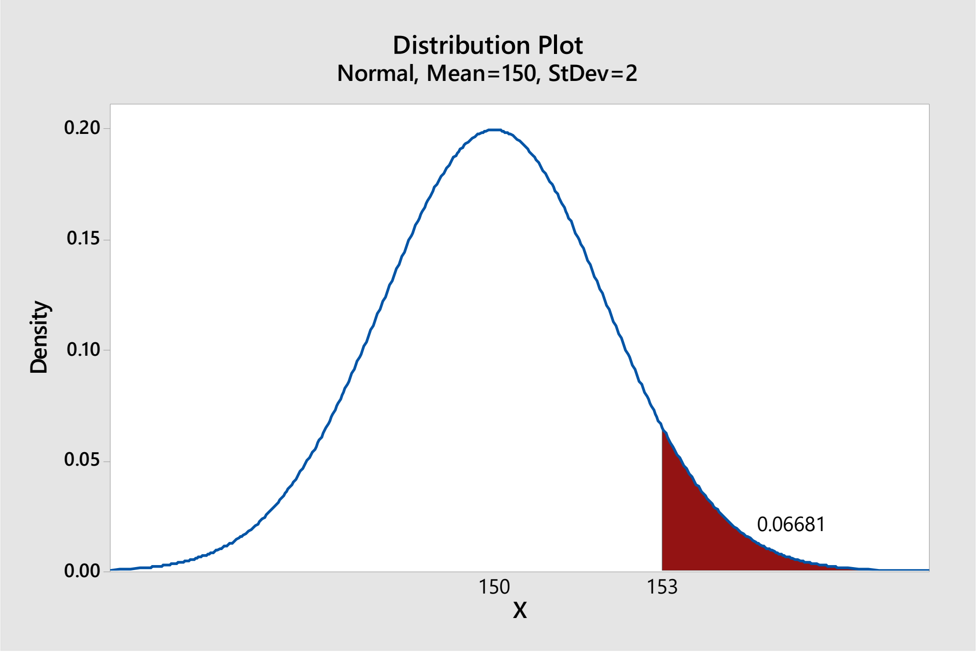
- Perfume bottles are filled with the average volume of 150 cc and the standard deviation of 2 cc.
- What percent of bottles will have the volume more than 153 cc?

$\mu = 150$ cc
$\sigma = 2$ cc

$z = (x - \mu) / \sigma = (153 – 150)/2 = 1.5$

$$P(x) = 0.06681 \text{ or } 6.681\%$$

$=NORM.DIST(153,150,2,TRUE) = 0.933193$




**Normal Distribution**

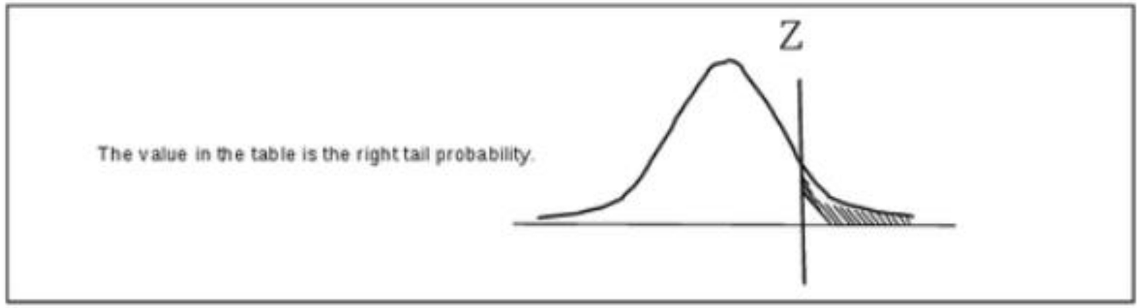
## Z Value / Standard Score

- ❖ Perfume bottles are filled with the average volume of 150 cc and the standard deviation of 2 cc.
- ❖ What percent of bottles will have the volume between 148 and 152 cc?

$\mu = 150$ cc
$\sigma = 2$ cc

$z1 = (x - \mu) / \sigma = (148 – 150)/2 = -1$
$z2 = (x - \mu) / \sigma = (152 – 150)/2 = 1$

$$P(x) = 1 - 0.15866 - 0.15866 = 0.68268$$




Hundredth place for Z-value →

| Z-Value | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.50000 | 0.49601 | 0.49202 | 0.48803 | 0.48405 | 0.48006 | 0.47608 | 0.47210 | 0.46812 | 0.46414 |
| 0.1 | 0.46017 | 0.45620 | 0.45224 | 0.44828 | 0.44433 | 0.44038 | 0.43644 | 0.43251 | 0.42858 | 0.42465 |
| 0.2 | 0.42074 | 0.41683 | 0.41294 | 0.40905 | 0.40517 | 0.40129 | 0.39743 | 0.39358 | 0.38974 | 0.38591 |
| 0.3 | 0.38209 | 0.37828 | 0.37448 | 0.37070 | 0.36693 | 0.36317 | 0.35942 | 0.35569 | 0.35197 | 0.34827 |
| 0.4 | 0.34458 | 0.34090 | 0.33724 | 0.33360 | 0.32997 | 0.32636 | 0.32276 | 0.31918 | 0.31561 | 0.31207 |
| 0.5 | 0.30854 | 0.30503 | 0.30153 | 0.29806 | 0.29460 | 0.29116 | 0.28774 | 0.28434 | 0.28096 | 0.27760 |
| 0.6 | 0.27425 | 0.27093 | 0.26763 | 0.26435 | 0.26109 | 0.25785 | 0.25463 | 0.25143 | 0.24825 | 0.24510 |
| 0.7 | 0.24196 | 0.23885 | 0.23576 | 0.23270 | 0.22965 | 0.22663 | 0.22363 | 0.22065 | 0.21770 | 0.21476 |
| 0.8 | 0.21186 | 0.20897 | 0.20611 | 0.20327 | 0.20045 | 0.19766 | 0.19489 | 0.19215 | 0.18943 | 0.18673 |
| 0.9 | 0.18406 | 0.18141 | 0.17879 | 0.17619 | 0.17361 | 0.17106 | 0.16853 | 0.16602 | 0.16354 | 0.16109 |
| 1.0 | 0.15866 | 0.15625 | 0.15386 | 0.15151 | 0.14917 | 0.14686 | 0.14457 | 0.14231 | 0.14007 | 0.13786 |
| 1.1 | 0.13567 | 0.13350 | 0.13136 | 0.12924 | 0.12714 | 0.12507 | 0.12302 | 0.12100 | 0.11900 | 0.11702 |
| 1.2 | 0.11507 | 0.11314 | 0.11123 | 0.10935 | 0.10749 | 0.10565 | 0.10383 | 0.10204 | 0.10027 | 0.09853 |
| 1.3 | 0.09680 | 0.09510 | 0.09342 | 0.09176 | 0.09012 | 0.08851 | 0.08691 | 0.08534 | 0.08379 | 0.08226 |
| 1.4 | 0.08076 | 0.07927 | 0.07780 | 0.07636 | 0.07493 | 0.07353 | 0.07215 | 0.07078 | 0.06944 | 0.06811 |
| 1.5 | 0.06681 | 0.06552 | 0.06426 | 0.06301 | 0.06178 | 0.06057 | 0.05938 | 0.05821 | 0.05705 | 0.05592 |
| 1.6 | 0.05480 | 0.05370 | 0.05262 | 0.05155 | 0.05050 | 0.04947 | 0.04846 | 0.04746 | 0.04648 | 0.04551 |

# *Normal Distribution*

## Z Value / Standard Score

- Perfume bottles are filled with the average volume of 150 cc and the standard deviation of 2 cc.
- Four bottles are selected from it. What is the probability that the average of 4 bottles will be more than 153 cc?

$\mu = 150$ cc

$\sigma_{\bar{x}} = 2 / \sqrt{4}$

$z = (\bar{x} - \mu) / \sigma_{\bar{x}} = (153 - 150)/1 = 3$

$P(x) = 0.00135 \text{ or } 0.135\%$

*Normal Distribution*

# 3C Statistical Distributions

01 Normal distribution

02 Binominal distribution

03 Poisson distribution

04 Chi square distribution

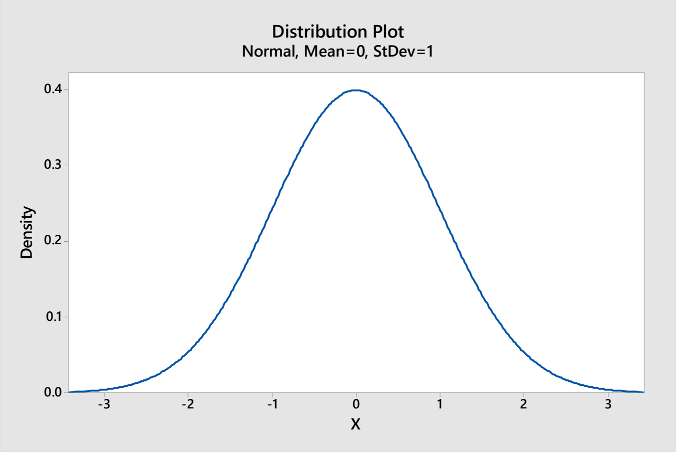
05 Student's t distribution

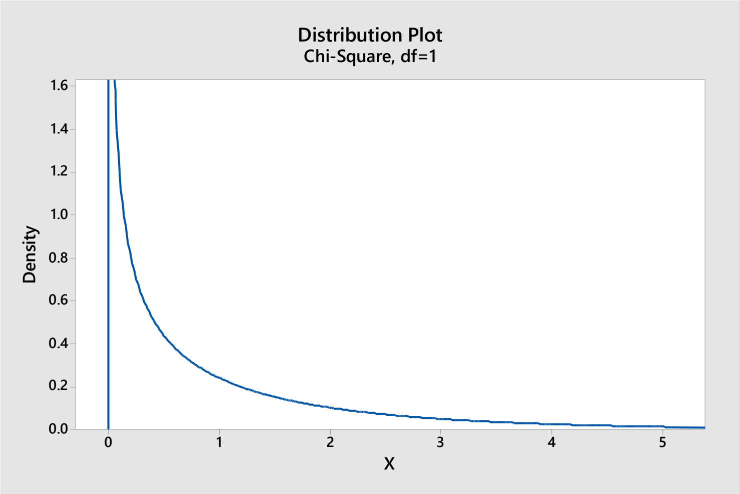
06 F distribution

| CONTINUOUS | DISCRETE |
|---|---|
| Normal distribution | Binominal distribution |
| Chi square distribution | Poisson distribution |
| Student’s t distribution | |
| F distribution | |

# CONTINUOUS vs DISCRETE DATA

❖ If a random variable Z has the Standard Normal Distribution, then $Z^2$ has a $\chi^2$ distribution with 1 degree of freedom.



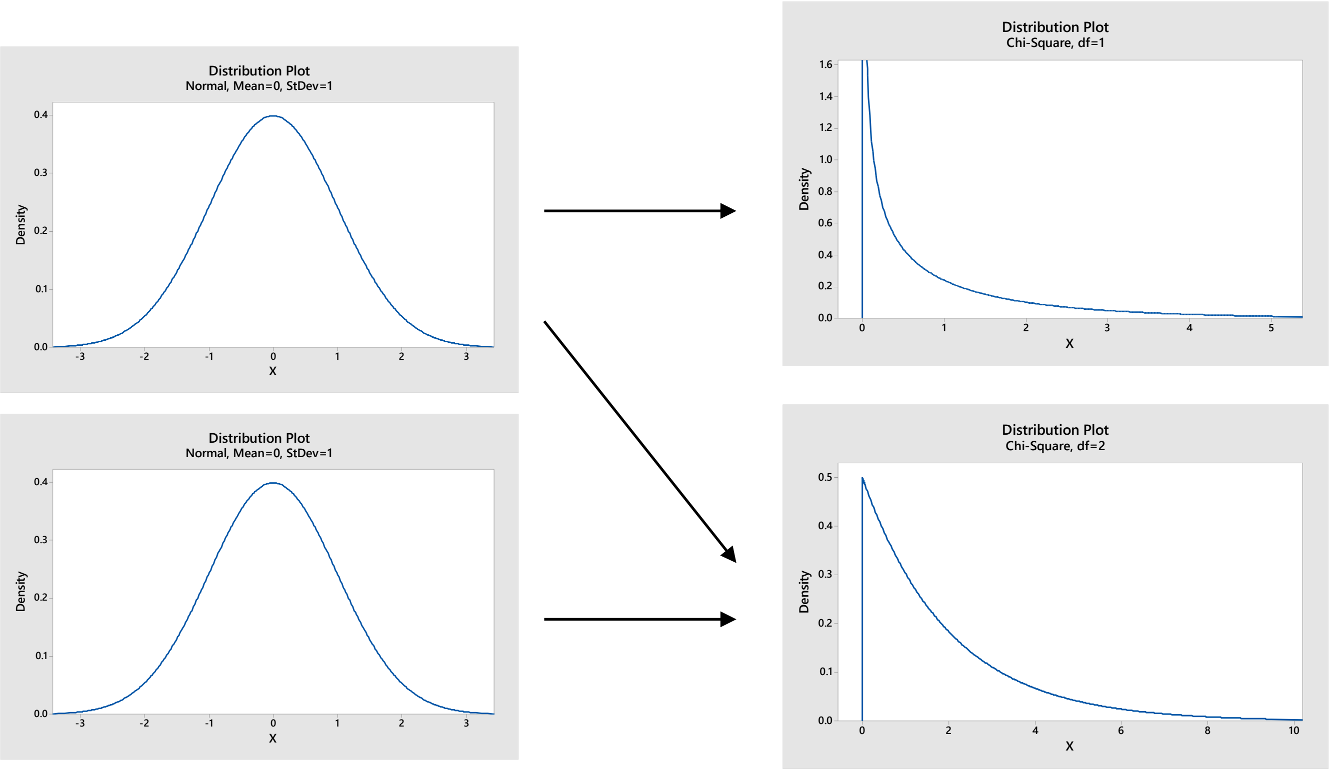

***$\chi^2$***

***Chi Square Distribution***

- If a random variables $Z_1, Z_2, Z_3 \ldots Z_k$ are the Standard Normal Distributions, then

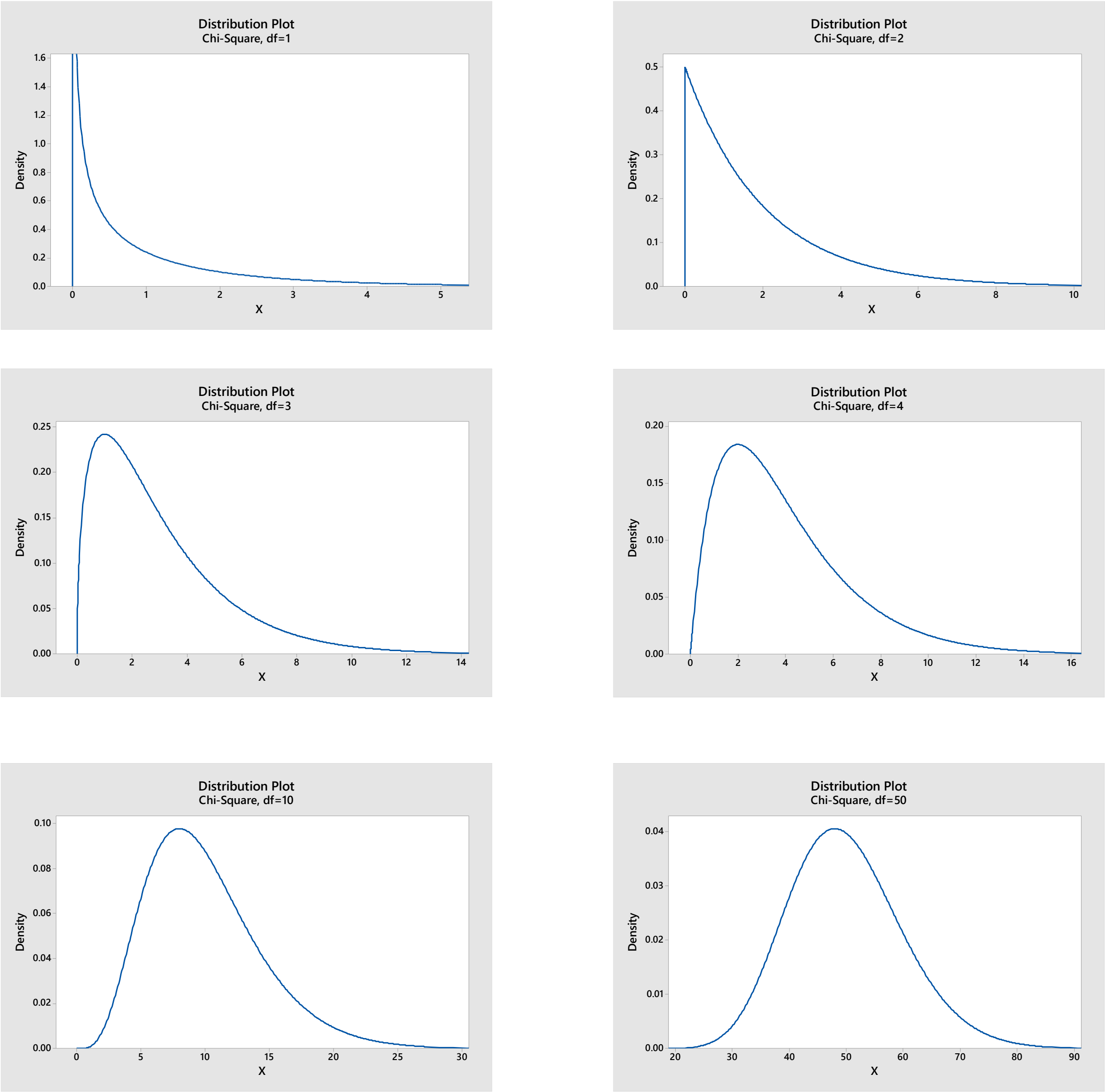
$$Z_1^2 + Z_2^2 + Z_3^2 + \ldots + Z_k^2$$

- has a $\chi^2$ distribution with k degree of freedom.

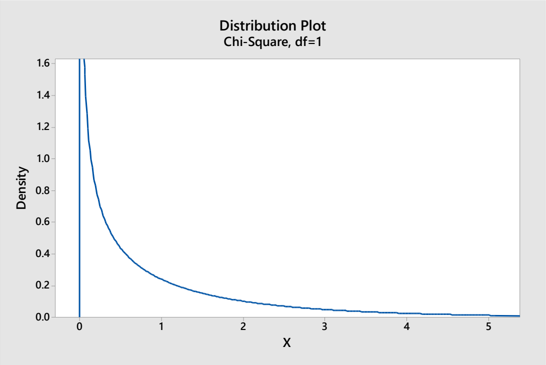


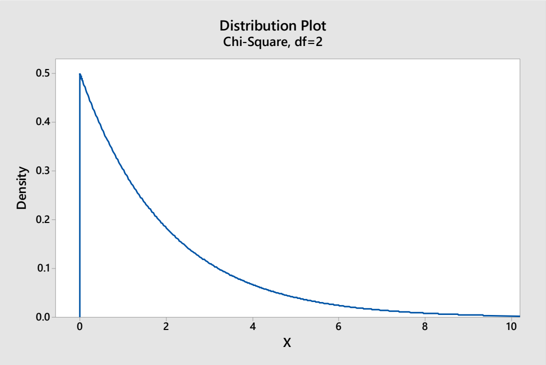

# $\chi^2$ *Chi Square Distribution*

# $\chi^2$ Chi Square Distribution

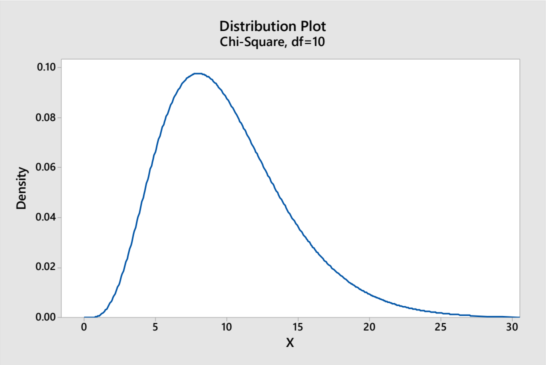


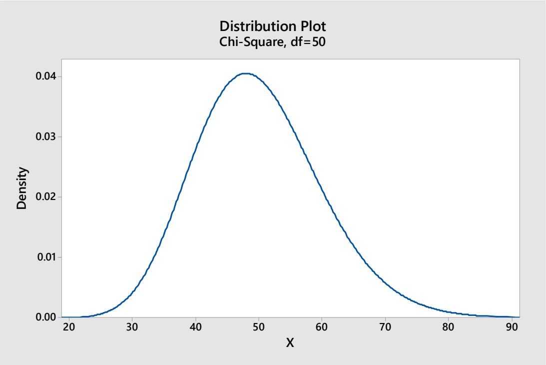

Min zero, max infinity




- Mean = df
- Variance = 2. df
- Mode = df – 2 (for min 2 df)



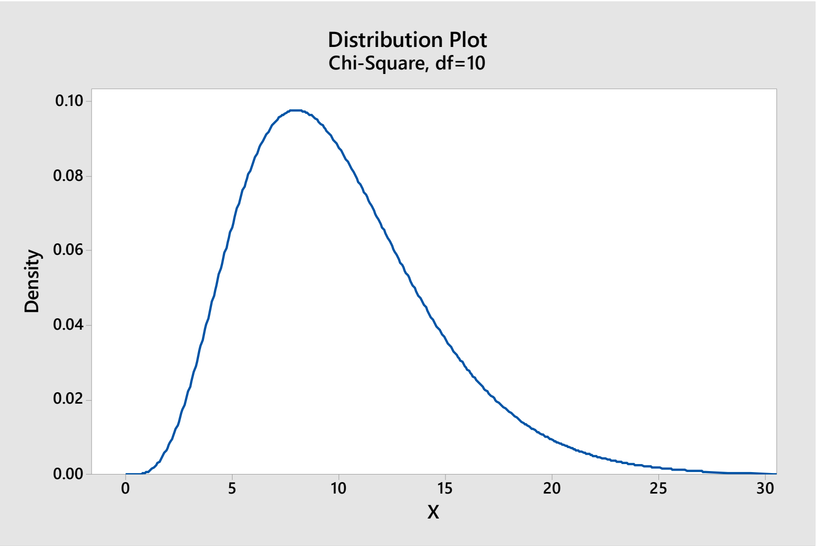




Min zero, max infinity

# $\chi^2$ *Chi Square Distribution*

$$f(x;\,k) = \begin{cases} \dfrac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{k}{2}}\,\Gamma\left(\frac{k}{2}\right)}, & x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Gamma Function
k = degrees of freedom




**Percentage Points of the Chi-Square Distribution**

| Degrees of Freedom | Probability of a larger value of $x^2$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0.99 | 0.95 | 0.90 | 0.75 | 0.50 | 0.25 | 0.10 | 0.05 | 0.01 |
| 1 | 0.000 | 0.004 | 0.016 | 0.102 | 0.455 | 1.32 | 2.71 | 3.84 | 6.63 |
| 2 | 0.020 | 0.103 | 0.211 | 0.575 | 1.386 | 2.77 | 4.61 | 5.99 | 9.21 |
| 3 | 0.115 | 0.352 | 0.584 | 1.212 | 2.366 | 4.11 | 6.25 | 7.81 | 11.34 |
| 4 | 0.297 | 0.711 | 1.064 | 1.923 | 3.357 | 5.39 | 7.78 | 9.49 | 13.28 |
| 5 | 0.554 | 1.145 | 1.610 | 2.675 | 4.351 | 6.63 | 9.24 | 11.07 | 15.09 |
| 6 | 0.872 | 1.635 | 2.204 | 3.455 | 5.348 | 7.84 | 10.64 | 12.59 | 16.81 |
| 7 | 1.239 | 2.167 | 2.833 | 4.255 | 6.346 | 9.04 | 12.02 | 14.07 | 18.48 |
| 8 | 1.647 | 2.733 | 3.490 | 5.071 | 7.344 | 10.22 | 13.36 | 15.51 | 20.09 |
| 9 | 2.088 | 3.325 | 4.168 | 5.899 | 8.343 | 11.39 | 14.68 | 16.92 | 21.67 |
| 10 | 2.558 | 3.940 | 4.865 | 6.737 | 9.342 | 12.55 | 15.99 | 18.31 | 23.21 |
| 11 | 3.053 | 4.575 | 5.578 | 7.584 | 10.341 | 13.70 | 17.28 | 19.68 | 24.72 |
| 12 | 3.571 | 5.226 | 6.304 | 8.438 | 11.340 | 14.85 | 18.55 | 21.03 | 26.22 |
| 13 | 4.107 | 5.892 | 7.042 | 9.299 | 12.340 | 15.98 | 19.81 | 22.36 | 27.69 |
| 14 | 4.660 | 6.571 | 7.790 | 10.165 | 13.339 | 17.12 | 21.06 | 23.68 | 29.14 |
| 15 | 5.229 | 7.261 | 8.547 | 11.037 | 14.339 | 18.25 | 22.31 | 25.00 | 30.58 |
| 16 | 5.812 | 7.962 | 9.312 | 11.912 | 15.338 | 19.37 | 23.54 | 26.30 | 32.00 |
| 17 | 6.408 | 8.672 | 10.085 | 12.792 | 16.338 | 20.49 | 24.77 | 27.59 | 33.41 |
| 18 | 7.015 | 9.390 | 10.865 | 13.675 | 17.338 | 21.60 | 25.99 | 28.87 | 34.80 |
| 19 | 7.633 | 10.117 | 11.651 | 14.562 | 18.338 | 22.72 | 27.20 | 30.14 | 36.19 |
| 20 | 8.260 | 10.851 | 12.443 | 15.452 | 19.337 | 23.83 | 28.41 | 31.41 | 37.57 |
| 22 | 9.542 | 12.338 | 14.041 | 17.240 | 21.337 | 26.04 | 30.81 | 33.92 | 40.29 |
| 24 | 10.856 | 13.848 | 15.659 | 19.037 | 23.337 | 28.24 | 33.20 | 36.42 | 42.98 |
| 26 | 12.198 | 15.379 | 17.292 | 20.843 | 25.336 | 30.43 | 35.56 | 38.89 | 45.64 |
| 28 | 13.565 | 16.928 | 18.939 | 22.657 | 27.336 | 32.62 | 37.92 | 41.34 | 48.28 |
| 30 | 14.953 | 18.493 | 20.599 | 24.478 | 29.336 | 34.80 | 40.26 | 43.77 | 50.89 |
| 40 | 22.164 | 26.509 | 29.051 | 33.660 | 39.335 | 45.62 | 51.80 | 55.76 | 63.69 |
| 50 | 27.707 | 34.764 | 37.689 | 42.942 | 49.335 | 56.33 | 63.17 | 67.50 | 76.15 |
| 60 | 37.485 | 43.188 | 46.459 | 52.294 | 59.335 | 66.98 | 74.40 | 79.08 | 88.38 |

# $\chi^2$ *Chi Square Distribution*

# 3C Statistical Distributions

01 Normal distribution

02 Binominal distribution

03 Poisson distribution

04 Chi square distribution

05 Student’s t distribution

06 F distribution

| CONTINUOUS | DISCRETE |
|---|---|
| Normal distribution | Binominal distribution |
| Chi square distribution | Poisson distribution |
| Student's t distribution | |
| F distribution | |

# CONTINUOUS vs DISCRETE DATA

- Student's t distribution is formed by combining the Standard Normal Distribution and the Chi Square Distribution.
  - Suppose Z has Standard Normal distribution
  - U has a Chi square distribution with $\nu$ degrees of freedom as
  - and Z and U are independent

$$T = \frac{\boldsymbol{Z}}{\sqrt{\frac{U}{\nu}}}$$

Has t distribution with $\nu$ degrees of freedom

# *Student's t Distribution*

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

$\nu$ is the number of degrees of freedom
$\Gamma$ is the gamma function

# *Student's t Distribution*

$z = (\bar{x} - \mu) / \sigma_{\bar{x}}$

$$z = \frac{(\bar{x} - \mu)}{\sigma/\sqrt{n}}$$

- We used the above equation in the Normal distribution example.

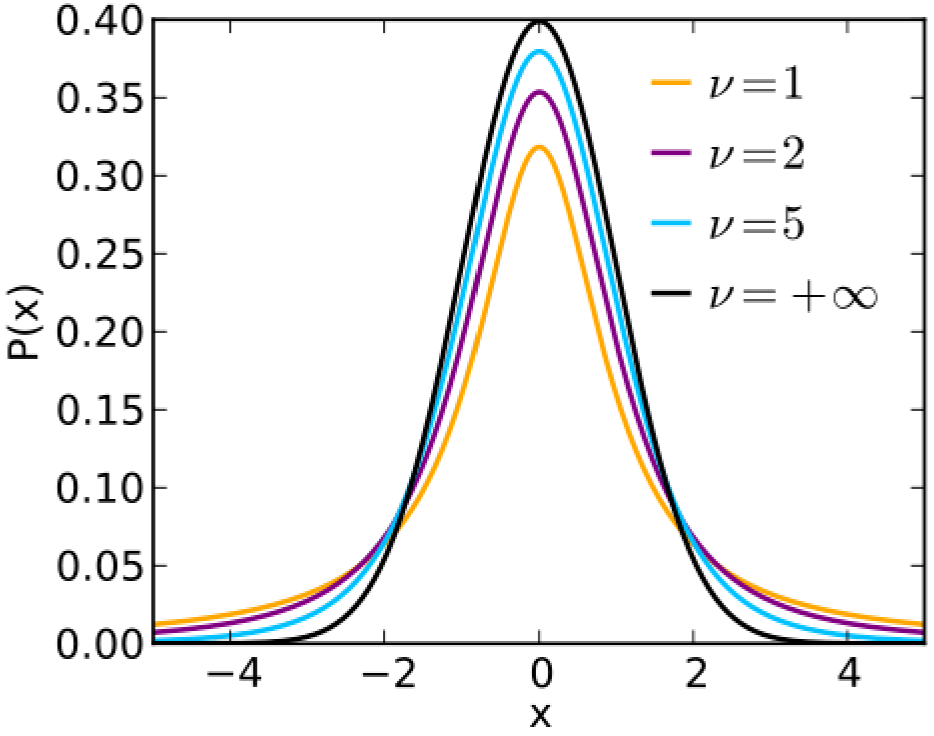
- What if we do not know the population standard deviation $\sigma$

*Student's t Distribution*

$$\boldsymbol{t} = \frac{(\bar{x} - \mu)}{\boldsymbol{s}/\sqrt{n}}$$

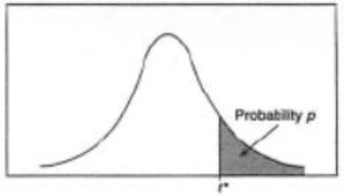
- This has a t distribution with n-1 degrees of freedom.
- s is the sample standard deviation.

*Student's t Distribution*




Reference: Wikipedia

# *Student's t Distribution*



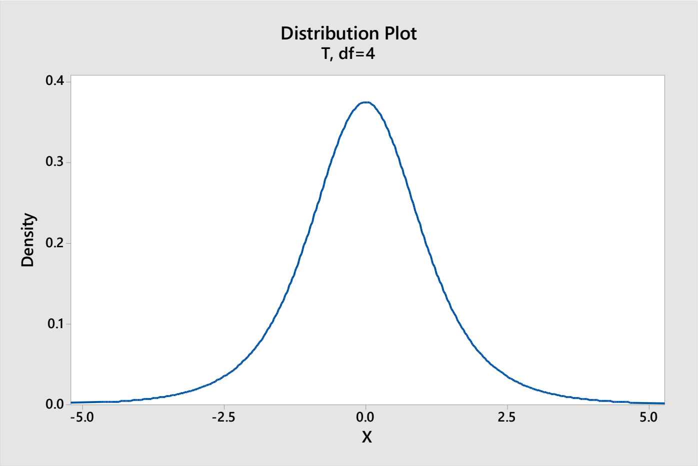




**TABLE B** ***t*** **Distribution Critical Values**

| | TAIL PROBABILITY *P* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| df | .25 | .20 | .15 | .10 | .05 | .025 | .02 | .01 | .005 | .0025 | .001 | .0005 |
| 1 | 1.000 | 1.376 | 1.963 | 3.078 | 6.314 | 12.71 | 15.89 | 31.82 | 63.66 | 127.3 | 318.3 | 636.6 |
| 2 | .816 | 1.061 | 1.386 | 1.886 | 2.920 | 4.303 | 4.849 | 6.965 | 9.925 | 14.09 | 22.33 | 31.60 |
| 3 | .765 | .978 | 1.250 | 1.638 | 2.353 | 3.182 | 3.482 | 4.541 | 5.841 | 7.453 | 10.21 | 12.92 |
| 4 | .741 | .941 | 1.190 | 1.533 | 2.132 | 2.776 | 2.999 | 3.747 | 4.604 | 5.598 | 7.173 | 8.610 |
| 5 | .727 | .920 | 1.156 | 1.476 | 2.015 | 2.571 | 2.757 | 3.365 | 4.032 | 4.773 | 5.893 | 6.869 |
| 6 | .718 | .906 | 1.134 | 1.440 | 1.943 | 2.447 | 2.612 | 3.143 | 3.707 | 4.317 | 5.208 | 5.959 |
| 7 | .711 | .896 | 1.119 | 1.415 | 1.895 | 2.365 | 2.517 | 2.998 | 3.499 | 4.029 | 4.785 | 5.408 |
| 8 | .706 | .889 | 1.108 | 1.397 | 1.860 | 2.306 | 2.449 | 2.896 | 3.355 | 3.833 | 4.501 | 5.041 |
| 9 | .703 | .883 | 1.100 | 1.383 | 1.833 | 2.262 | 2.398 | 2.821 | 3.250 | 3.690 | 4.297 | 4.781 |
| 10 | .700 | .879 | 1.093 | 1.372 | 1.812 | 2.228 | 2.359 | 2.764 | 3.169 | 3.581 | 4.144 | 4.587 |
| 11 | .697 | .876 | 1.088 | 1.363 | 1.796 | 2.201 | 2.328 | 2.718 | 3.106 | 3.497 | 4.025 | 4.437 |
| 12 | .695 | .873 | 1.083 | 1.356 | 1.782 | 2.179 | 2.303 | 2.681 | 3.055 | 3.428 | 3.930 | 4.318 |
| 13 | .694 | .870 | 1.079 | 1.350 | 1.771 | 2.160 | 2.282 | 2.650 | 3.012 | 3.372 | 3.852 | 4.221 |
| 14 | .692 | .868 | 1.076 | 1.345 | 1.761 | 2.145 | 2.264 | 2.624 | 2.977 | 3.326 | 3.787 | 4.140 |
| 15 | .691 | .866 | 1.074 | 1.341 | 1.753 | 2.131 | 2.249 | 2.602 | 2.947 | 3.286 | 3.733 | 4.073 |
| 16 | .690 | .865 | 1.071 | 1.337 | 1.746 | 2.120 | 2.235 | 2.583 | 2.921 | 3.252 | 3.686 | 4.015 |
| 17 | .689 | .863 | 1.069 | 1.333 | 1.740 | 2.110 | 2.224 | 2.567 | 2.898 | 3.222 | 3.646 | 3.965 |
| 18 | .688 | .862 | 1.067 | 1.330 | 1.734 | 2.101 | 2.214 | 2.552 | 2.878 | 3.197 | 3.611 | 3.922 |
| 19 | .688 | .861 | 1.066 | 1.328 | 1.729 | 2.093 | 2.205 | 2.539 | 2.861 | 3.174 | 3.579 | 3.883 |
| 20 | .687 | .860 | 1.064 | 1.325 | 1.725 | 2.086 | 2.197 | 2.528 | 2.845 | 3.153 | 3.552 | 3.850 |
| 21 | .686 | .859 | 1.063 | 1.323 | 1.721 | 2.080 | 2.189 | 2.518 | 2.831 | 3.135 | 3.527 | 3.819 |
| 22 | .686 | .858 | 1.061 | 1.321 | 1.717 | 2.074 | 2.183 | 2.508 | 2.819 | 3.119 | 3.505 | 3.792 |
| 23 | .685 | .858 | 1.060 | 1.319 | 1.714 | 2.069 | 2.177 | 2.500 | 2.807 | 3.104 | 3.485 | 3.768 |
| 24 | .685 | .857 | 1.059 | 1.318 | 1.711 | 2.064 | 2.172 | 2.492 | 2.797 | 3.091 | 3.467 | 3.745 |
| 25 | .684 | .856 | 1.058 | 1.316 | 1.708 | 2.060 | 2.167 | 2.485 | 2.787 | 3.078 | 3.450 | 3.725 |
| 26 | .684 | .856 | 1.058 | 1.315 | 1.706 | 2.056 | 2.162 | 2.479 | 2.779 | 3.067 | 3.435 | 3.707 |
| 27 | .684 | .855 | 1.057 | 1.314 | 1.703 | 2.052 | 2.158 | 2.473 | 2.771 | 3.057 | 3.421 | 3.690 |
| 28 | .683 | .855 | 1.056 | 1.313 | 1.701 | 2.048 | 2.154 | 2.467 | 2.763 | 3.047 | 3.408 | 3.674 |
| 29 | .683 | .854 | 1.055 | 1.311 | 1.699 | 2.045 | 2.150 | 2.462 | 2.756 | 3.038 | 3.396 | 3.659 |
| 30 | .683 | .854 | 1.055 | 1.310 | 1.697 | 2.042 | 2.147 | 2.457 | 2.750 | 3.030 | 3.385 | 3.646 |

# *Student's t Distribution*

=T.DIST(2.132, 4, TRUE) = 0.950009

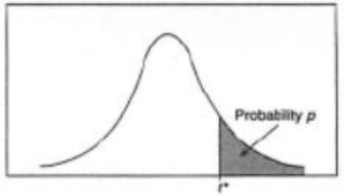



**TABLE B *t* Distribution Critical Values**

| | TAIL PROBABILITY *P* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| df | .25 | .20 | .15 | .10 | .05 | .025 | .02 | .01 | .005 | .0025 | .001 | .0005 |
| 1 | 1.000 | 1.376 | 1.963 | 3.078 | 6.314 | 12.71 | 15.89 | 31.82 | 63.66 | 127.3 | 318.3 | 636.6 |
| 2 | .816 | 1.061 | 1.386 | 1.886 | 2.920 | 4.303 | 4.849 | 6.965 | 9.925 | 14.09 | 22.33 | 31.60 |
| 3 | .765 | .978 | 1.250 | 1.638 | 2.353 | 3.182 | 3.482 | 4.541 | 5.841 | 7.453 | 10.21 | 12.92 |
| 4 | .741 | .941 | 1.190 | 1.533 | 2.132 | 2.776 | 2.999 | 3.747 | 4.604 | 5.598 | 7.173 | 8.610 |
| 5 | .727 | .920 | 1.156 | 1.476 | 2.015 | 2.571 | 2.757 | 3.365 | 4.032 | 4.773 | 5.893 | 6.869 |
| 6 | .718 | .906 | 1.134 | 1.440 | 1.943 | 2.447 | 2.612 | 3.143 | 3.707 | 4.317 | 5.208 | 5.959 |
| 7 | .711 | .896 | 1.119 | 1.415 | 1.895 | 2.365 | 2.517 | 2.998 | 3.499 | 4.029 | 4.785 | 5.408 |
| 8 | .706 | .889 | 1.108 | 1.397 | 1.860 | 2.306 | 2.449 | 2.896 | 3.355 | 3.833 | 4.501 | 5.041 |
| 9 | .703 | .883 | 1.100 | 1.383 | 1.833 | 2.262 | 2.398 | 2.821 | 3.250 | 3.690 | 4.297 | 4.781 |
| 10 | .700 | .879 | 1.093 | 1.372 | 1.812 | 2.228 | 2.359 | 2.764 | 3.169 | 3.581 | 4.144 | 4.587 |
| 11 | .697 | .876 | 1.088 | 1.363 | 1.796 | 2.201 | 2.328 | 2.718 | 3.106 | 3.497 | 4.025 | 4.437 |
| 12 | .695 | .873 | 1.083 | 1.356 | 1.782 | 2.179 | 2.303 | 2.681 | 3.055 | 3.428 | 3.930 | 4.318 |
| 13 | .694 | .870 | 1.079 | 1.350 | 1.771 | 2.160 | 2.282 | 2.650 | 3.012 | 3.372 | 3.852 | 4.221 |
| 14 | .692 | .868 | 1.076 | 1.345 | 1.761 | 2.145 | 2.264 | 2.624 | 2.977 | 3.326 | 3.787 | 4.140 |
| 15 | .691 | .866 | 1.074 | 1.341 | 1.753 | 2.131 | 2.249 | 2.602 | 2.947 | 3.286 | 3.733 | 4.073 |
| 16 | .690 | .865 | 1.071 | 1.337 | 1.746 | 2.120 | 2.235 | 2.583 | 2.921 | 3.252 | 3.686 | 4.015 |
| 17 | .689 | .863 | 1.069 | 1.333 | 1.740 | 2.110 | 2.224 | 2.567 | 2.898 | 3.222 | 3.646 | 3.965 |
| 18 | .688 | .862 | 1.067 | 1.330 | 1.734 | 2.101 | 2.214 | 2.552 | 2.878 | 3.197 | 3.611 | 3.922 |
| 19 | .688 | .861 | 1.066 | 1.328 | 1.729 | 2.093 | 2.205 | 2.539 | 2.861 | 3.174 | 3.579 | 3.883 |
| 20 | .687 | .860 | 1.064 | 1.325 | 1.725 | 2.086 | 2.197 | 2.528 | 2.845 | 3.153 | 3.552 | 3.850 |
| 21 | .686 | .859 | 1.063 | 1.323 | 1.721 | 2.080 | 2.189 | 2.518 | 2.831 | 3.135 | 3.527 | 3.819 |
| 22 | .686 | .858 | 1.061 | 1.321 | 1.717 | 2.074 | 2.183 | 2.508 | 2.819 | 3.119 | 3.505 | 3.792 |
| 23 | .685 | .858 | 1.060 | 1.319 | 1.714 | 2.069 | 2.177 | 2.500 | 2.807 | 3.104 | 3.485 | 3.768 |
| 24 | .685 | .857 | 1.059 | 1.318 | 1.711 | 2.064 | 2.172 | 2.492 | 2.797 | 3.091 | 3.467 | 3.745 |
| 25 | .684 | .856 | 1.058 | 1.316 | 1.708 | 2.060 | 2.167 | 2.485 | 2.787 | 3.078 | 3.450 | 3.725 |
| 26 | .684 | .856 | 1.058 | 1.315 | 1.706 | 2.056 | 2.162 | 2.479 | 2.779 | 3.067 | 3.435 | 3.707 |
| 27 | .684 | .855 | 1.057 | 1.314 | 1.703 | 2.052 | 2.158 | 2.473 | 2.771 | 3.057 | 3.421 | 3.690 |
| 28 | .683 | .855 | 1.056 | 1.313 | 1.701 | 2.048 | 2.154 | 2.467 | 2.763 | 3.047 | 3.408 | 3.674 |
| 29 | .683 | .854 | 1.055 | 1.311 | 1.699 | 2.045 | 2.150 | 2.462 | 2.756 | 3.038 | 3.396 | 3.659 |
| 30 | .683 | .854 | 1.055 | 1.310 | 1.697 | 2.042 | 2.147 | 2.457 | 2.750 | 3.030 | 3.385 | 3.646 |

# *Student's t Distribution*

# 3C
# Statistical Distributions

01 Normal distribution

02 Binominal distribution

03 Poisson distribution

04 Chi square distribution

05 Student's t distribution

06 F distribution

| CONTINUOUS | DISCRETE |
|---|---|
| Normal distribution | Binominal distribution |
| Chi square distribution | Poisson distribution |
| Student’s t distribution | |
| F distribution | |

# CONTINUOUS vs DISCRETE DATA

- F Distribution is the ratio of two chi square distributions divided by their respective degrees of freedom.

  - U1 has a $\chi^2$ distribution with $\nu 1$ degrees of freedom
  - U2 has a $\chi^2$ distribution with $\nu 2$ degrees of freedom
  - U1 and U2 are independent

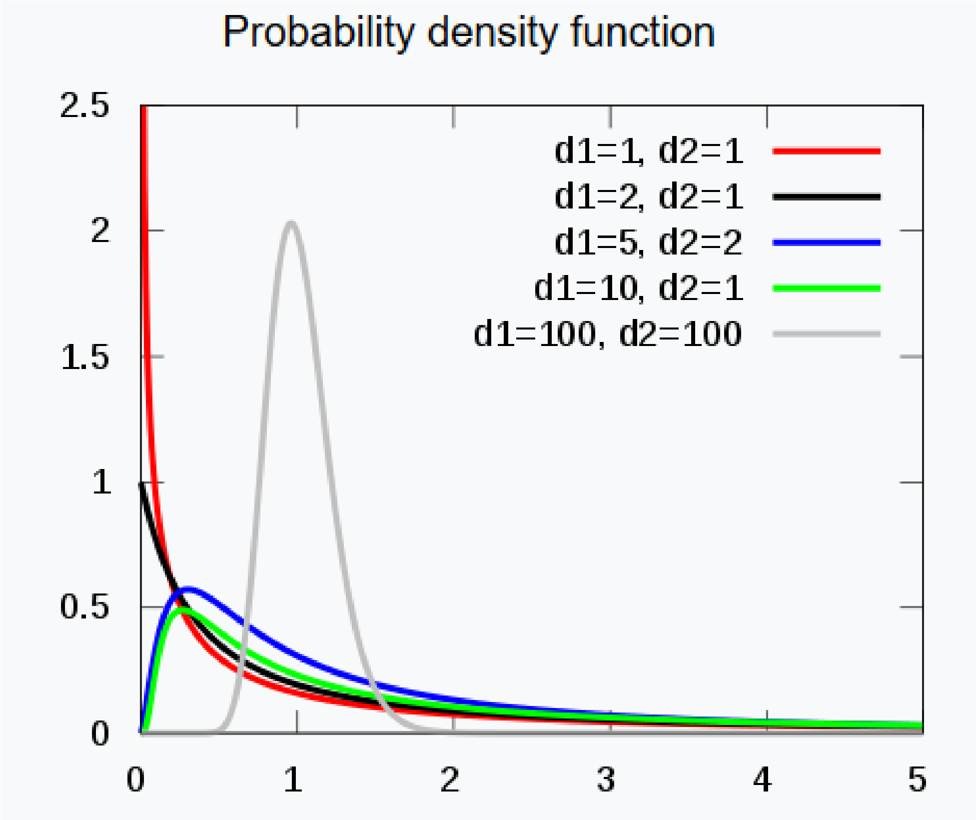
$$X = \frac{U_1/\nu_1}{U_2/\nu_2} \sim F(\nu_1, \nu_2)$$

# *F Distribution*

$$f(x; d_1, d_2) = \frac{\sqrt{\frac{(d_1\,x)^{d_1}\,\,d_2^{d_2}}{(d_1\,x+d_2)^{d_1+d_2}}}}{x\,\mathrm{B}\!\left(\frac{d_1}{2},\frac{d_2}{2}\right)}$$

$$= \frac{1}{\mathrm{B}\!\left(\frac{d_1}{2},\frac{d_2}{2}\right)} \left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}} x^{\frac{d_1}{2}-1} \left(1+\frac{d_1}{d_2}\,x\right)^{-\frac{d_1+d_2}{2}}$$

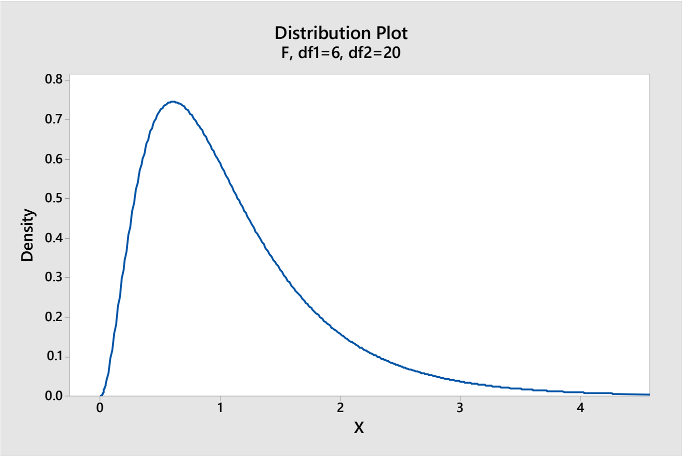
- Source: Wikipedia

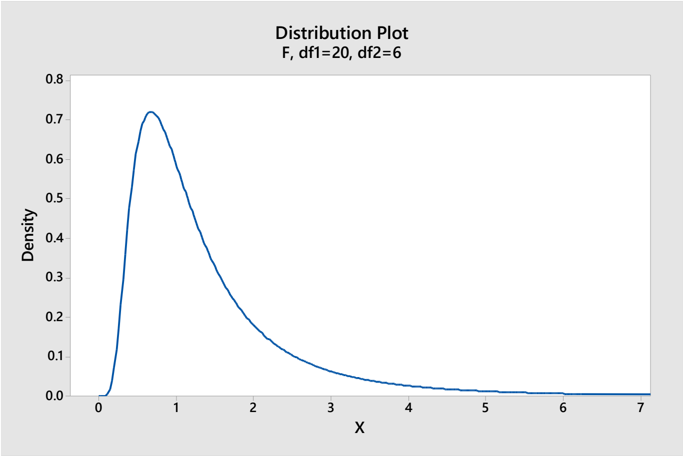



# *F Distribution*

$$F(\nu_1, \nu_2) \neq F(\nu_2, \nu_1)$$

$$F(\mathbf{6}, \mathbf{20}) \neq F(\mathbf{20}, \mathbf{6})$$




- F Distribution is used in ANOVA to test the equality or two or more means.

# *F Distribution*

# 3D Collecting and Summarizing Data

01 Types of data & measurement scales

02 Sampling and data collection methods

03 Descriptive statistics

04 Graphical methods

# Types of Data

- Continuous Data: (variable)
  - Measurements: Length, height, time
  - More information with less samples
  - More sensitive
  - Provide more information
  - More expensive to collect
- Discrete Data: (attribute)
  - Count: Number of students, Number of heads

# Measurement Scales

- Data
  - Nominal
  - Ordinal
  - Interval
  - Ratio

Data

- Nominal
- Ordinal
- Interval
- Ratio

Example:
Color: Blue, Green, Red

## *Measurement Scales*

Data

Nominal | Ordinal | Interval | Ratio

Example:
Pass/Fail
Good, Bad, Worst

# *Measurement Scales*

Data

Nominal | Ordinal | Interval | Ratio

Example:
Temperature: Celsius

# *Measurement Scales*

Data

Nominal | Ordinal | Interval | Ratio

Example:
Height, mass, volume

# *Measurement Scales*

| | Nominal | Ordinal | Interval | Ratio |
|---|---|---|---|---|
| Ordered | N | Y | Y | Y |
| Difference | N | N | Y | Y |
| Absolute Zero | N | N | N | Y |
| Example | Red, Blue | Good, Bad, Worst | Temperature: Degree C | Length, Weight |
| Central Tendency Measurement | Mode | Mode, Median | Mode, Median, Mean | Mode, Median, Mean |

# *Measurement Scales*

# 3D
# Collecting and Summarizing Data

01 Types of data & measurement scales

02 Sampling and data collection methods

03 Descriptive statistics

04 Graphical methods

- Because of the **cost and time** involved in studying the entire population.

# *Why Sampling?*

# Types of Sampling

- Probability Samples
  - Everyone in the population has an equal chance of being selected
- Non-Probability Samples
  - Where the probability of selection can't be accurately determined.
  - Sample may not be (generally isn’t) representative of the general population

# *Types of Sampling*

Probability Sampling

- Simple Random Sampling
- Systematic Random Sampling
- Stratified Random Sampling
- Cluster Sampling

Non Probability Sampling

- Accidental / Convenience Sampling
- Judgemental Sampling
- Quota Sampling

# Simple Random Sampling

- Each item in the population has an equal chance of being selected.
- Examples: Using random tables, Random draw of lot (lottery)

| Probability Sampling | |
|---|---|
| | Simple Random Sampling |
| | Systematic Random Sampling |
| | Stratified Random Sampling |
| | Cluster Sampling |

***Types of Sampling***

# Systematic Random Sampling

- Select elements at regular intervals through that ordered list.
- Example: Checking every $6^{th}$ piece produced by the machine.

Probability Sampling

- Simple Random Sampling
- Systematic Random Sampling
- Stratified Random Sampling
- Cluster Sampling

*Types of Sampling*

# Stratified Random Sampling

- Used to ensure that sub-groups within a population are represented proportionally in the sample.
- Example: If 10 people are drawn to represent a country, 5 of them are male and 5 females to avoid the sex bias.

Probability Sampling

- Simple Random Sampling
- Systematic Random Sampling
- Stratified Random Sampling
- Cluster Sampling

*Types of Sampling*

# Cluster Sampling

- Sometimes it is more cost-effective to select respondents in groups ('clusters'). Sampling is often clustered by geography, or by time periods.
- Example: Survey all customers visiting particular stores on particular days.

Probability Sampling

- Simple Random Sampling
- Systematic Random Sampling
- Stratified Random Sampling
- Cluster Sampling

*Types of Sampling*

# Convenience Sampling

- The researcher selects whomever is convenient. The samples are being drawn from that part of the population which is close to hand
- Example: A researcher at the mall selects the first five people who walk by to get their opinion of a product.

Non Probability Sampling

- Accidental / Convenience Sampling
- Judgemental Sampling
- Quota Sampling

*Types of Sampling*

# Judgmental Sampling

- The researcher chooses the sample based on who they think would be appropriate for the study.
- Example: Auditor selects a sample based on the concerns he/she had in the earlier audit

Non Probability Sampling

- Accidental / Convenience Sampling
- Judgemental Sampling
- Quota Sampling

*Types of Sampling*

# Quota Sampling

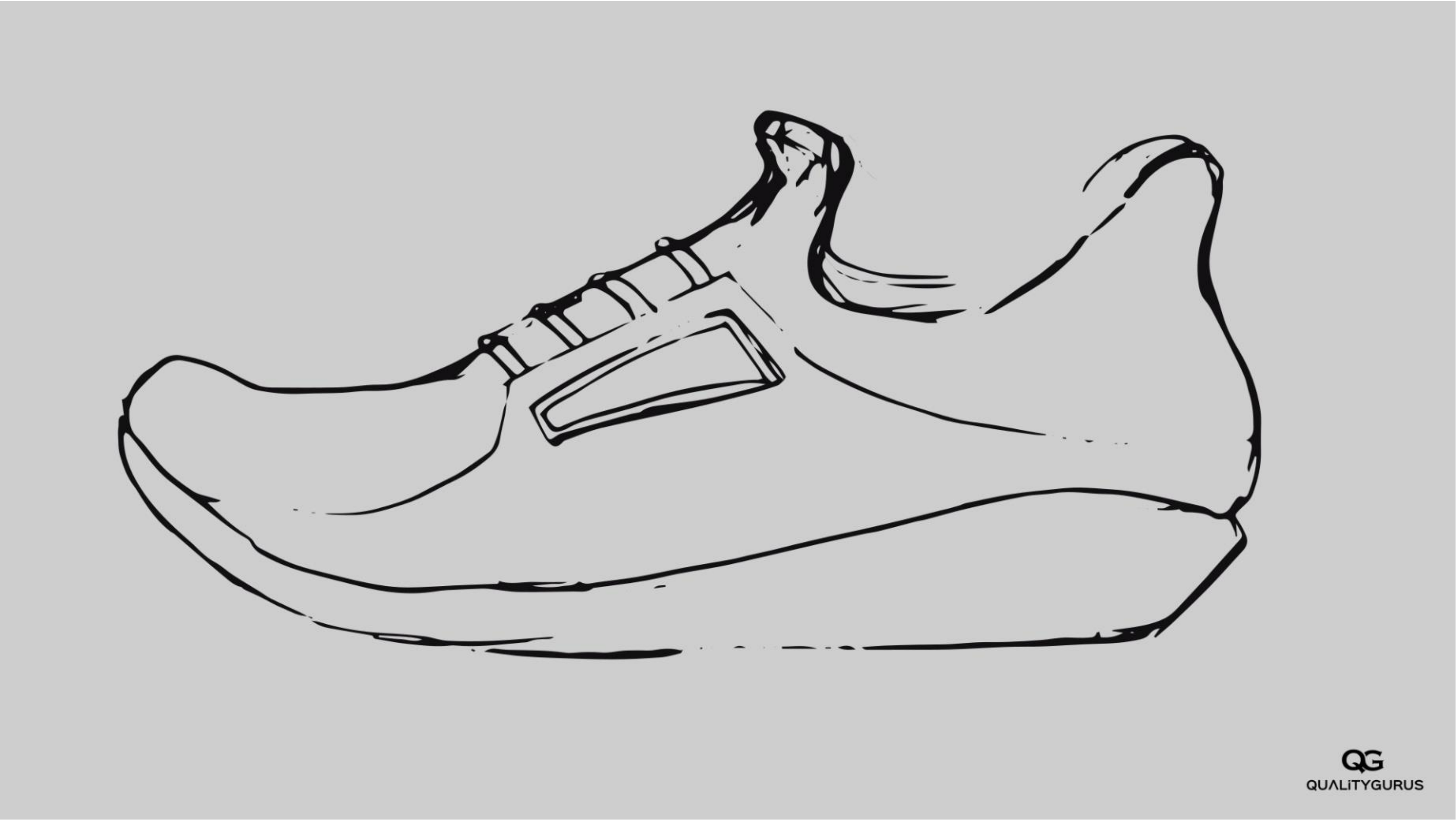
- A quota is established and auditor are free to choose any sample they wish as long as the quota is met.
- Example:  2% of the calibration records.

Non Probability Sampling

- Accidental / Convenience Sampling
- Judgemental Sampling
- Quota Sampling

**_Types of Sampling_**

| Defect | Count |
|---|---|
| Excess glue | |
| Weak joint | |
| Abrasion marks | |
| Asymmetry | |
| Nail / sharp points | |




# Check sheets

- Adding, Subtracting
  - Example: -95, -97, -98, -90
  - Add 100 to each: 5, 3, 2, 10
  - Coded mean: 5
  - Un-coded mean: 5-100 = -95
  - Standard deviation remains same and is not affected by addition and subtraction.
  - s = 3.559

# Data Coding

- Multiplying or dividing
  - Example: 1.05, 1.03, 1.02, 1.10
  - Multiply 100 to each: 105, 103, 102, 110
  - Coded mean: 105
  - Un-coded mean: 105 / 100 = 1.05
  - Standard deviation need to divided by you multiplied for coding.
  - For coded data s = 3.559
  - For original data s = 3.559/100 = 0.03559

## *Data Coding*

- By truncation of repetitive terms
  - Example: 0.555, 0.553, 0.552, 0.550
  - Truncate 0.55 from all: 5,3,2,0
  - This means we multiplied it by 1000 and subtracted 550
  - Coded mean: 2.5
  - Un-coded mean: (2.5+550)/ 1000 =.5525
  - Standard deviation need to divided by you multiplied for coding.
  - For coded data s = 2.0816
  - For original data s = 2.0816/1000 = 0.0020816

# *Data Coding*

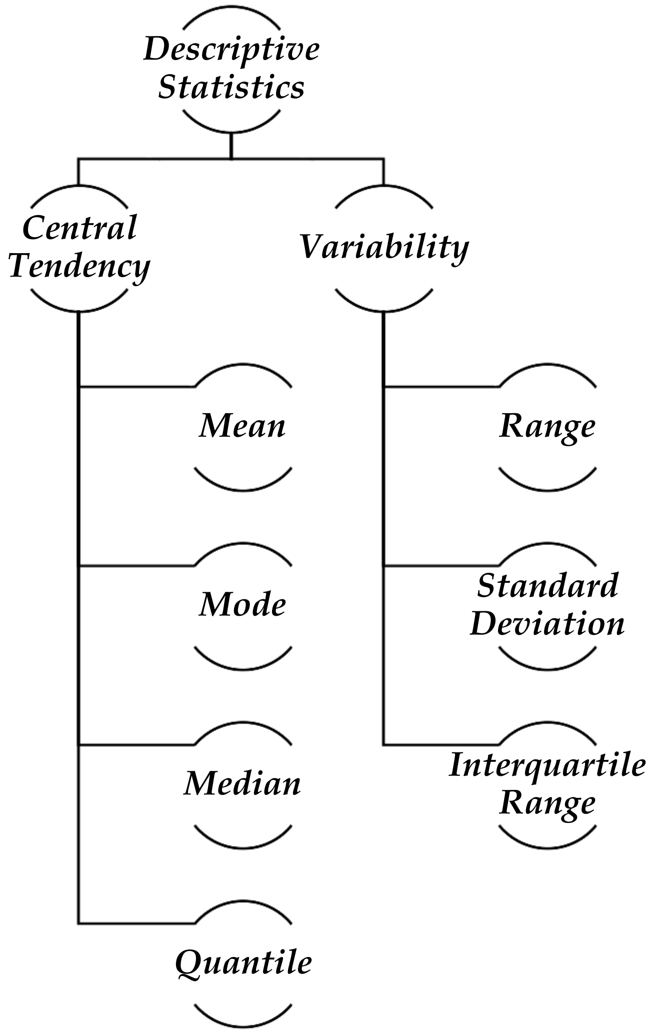
# 3D
# Collecting and Summarizing Data

01 Types of data & measurement scales

02 Sampling and data collection methods

03 Descriptive statistics

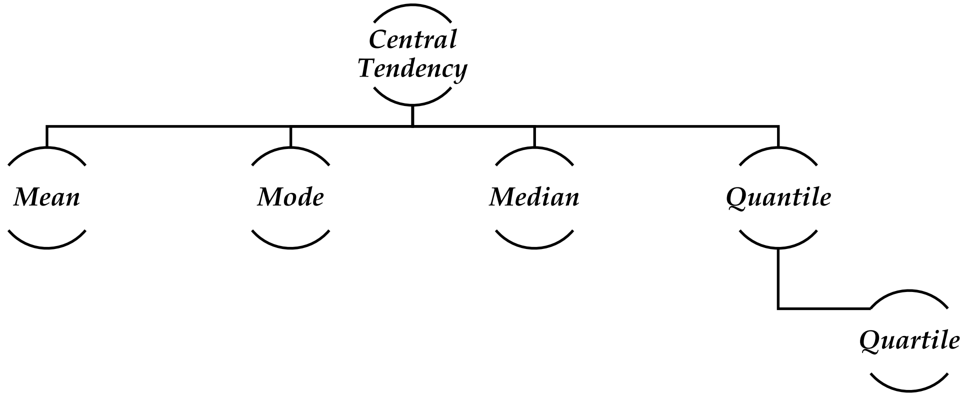
04 Graphical methods




# Descriptive Statistics

# Mean

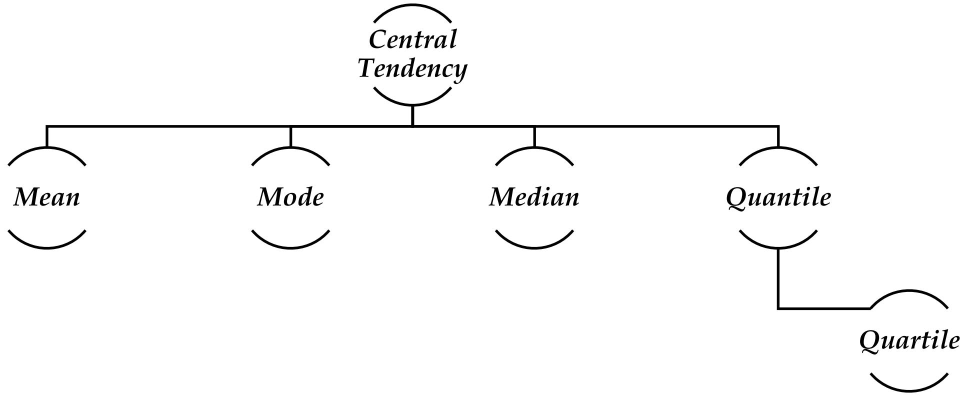
- Also known as Average
- Affected by extreme values
- Example: 10, 11, 14, 9, 6
- Mean = (10+11+14+9+6)/5 = 50/5 = 10




**Descriptive Statistics**

# Mode

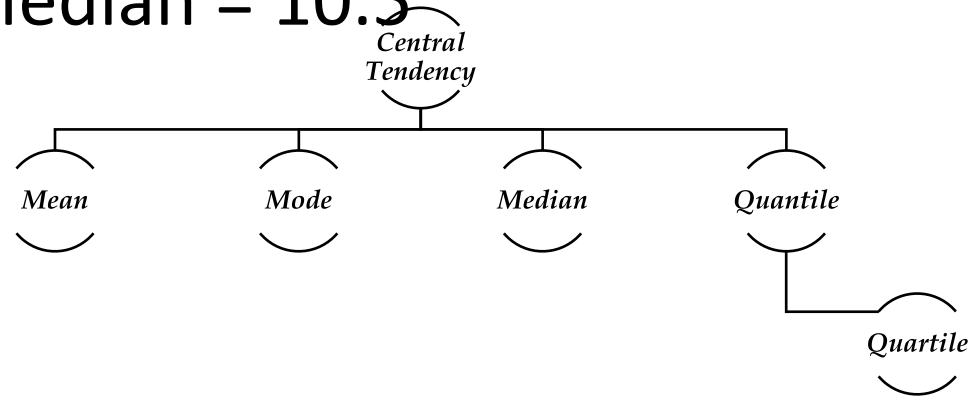
- Most occurring item
- Example: 10, 11, 14, 9, 6, 10
- Mode = 10




**Descriptive Statistics**

# Median

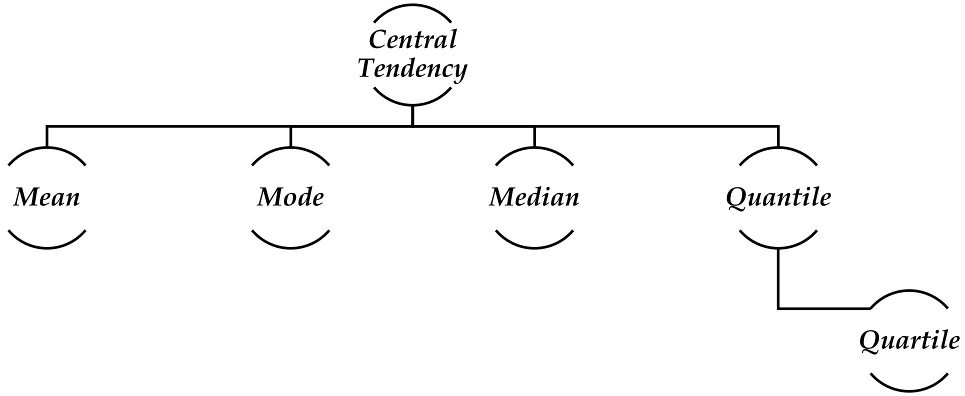
- Middle value when put in ascending or descending order.
- Example 1: 10, 11, 14, 9, 6
- In ascending order - 6,9,10,11,14
- Median = 10
- Example 2: 10, 11, 14, 9, 6, 11
- In order - 6,9,10,11, 11,14
- Median = 10.5




*Descriptive Statistics*

# Quartile

- Arrange in ascending or descending order
- Divide the data into 4 parts (just like we divided it into 2 parts for median)
- Example: 6,9,10,11, 11,14
- Q1=9, Q2=10.5, Q3=11




**Descriptive Statistics**

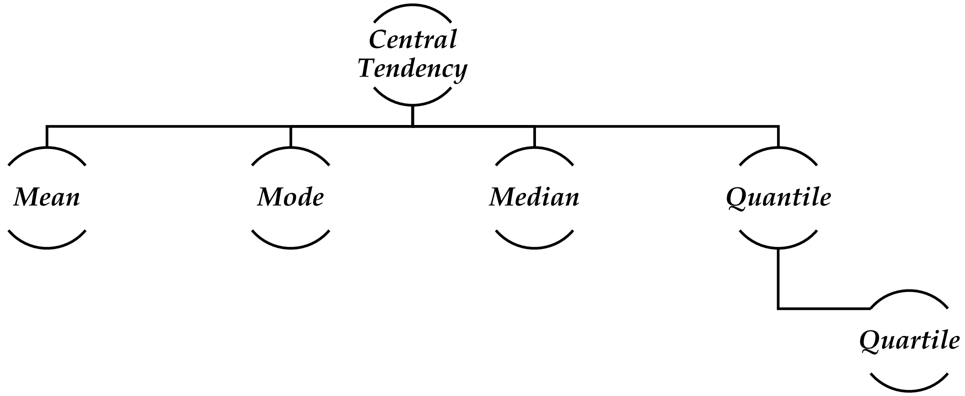
# Percentile

- Arrange in ascending or descending order
- Calculate location(i) = P.(n)/100
- P=percentile, n=numbers in data set
- If i is whole number – Percentile is average of (i)th and (i+1)th location
- If i is “not” a whole number – Percentile is located at (i+1)th whole-num.
- Example: 6,9,10,11, 11,14
- Q1=9, Q2=10.5, Q3=11

*Descriptive Statistics*

# Quantile

- Quantiles can go from anything to anything.
- Percentile divides data in 100 parts
- Quartile divides data in 4 parts
- Percentiles and quartiles are examples of quantiles.



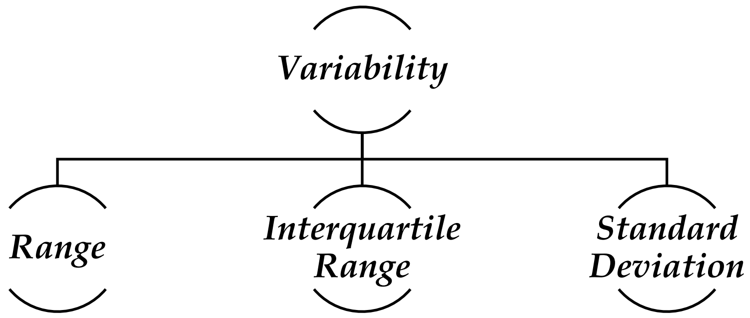

**Descriptive Statistics**

# Descriptive Statistics

- Variability
  - Range
  - Interquartile Range
  - Standard Deviation

# Range

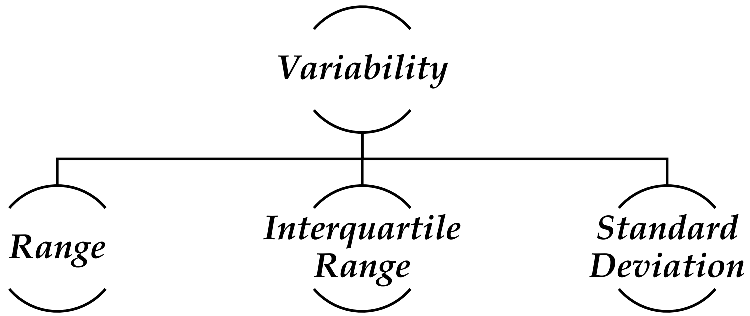
- Difference between lowest and the highest value.
- Example: 6,9,10,11, 11,14
- Range = 14-6 = 8




*Descriptive Statistics*

# Interquartile Range

- Range of middle 50% data
- IQR = Q3-Q1
- Example: 6,9,10,11, 11,14
- Q1=9, Q2=10.5, Q3=11
- IQR = 11-9 = 2
- Box-and-Whisker Plot

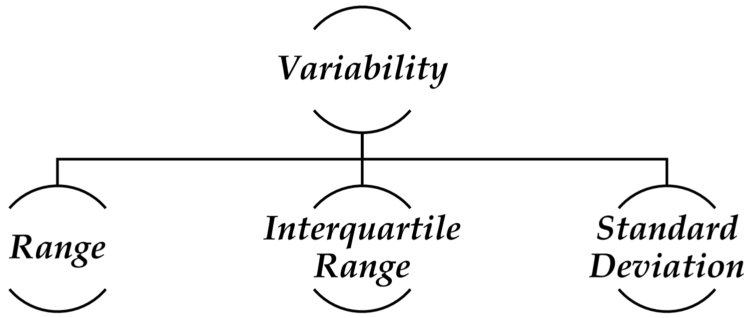



***Descriptive Statistics***

# Standard Deviation

- Variance = average of squared deviation about the arithmetic mean.
- Square root of variance is standard deviation

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}} \qquad s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$




**Descriptive Statistics**

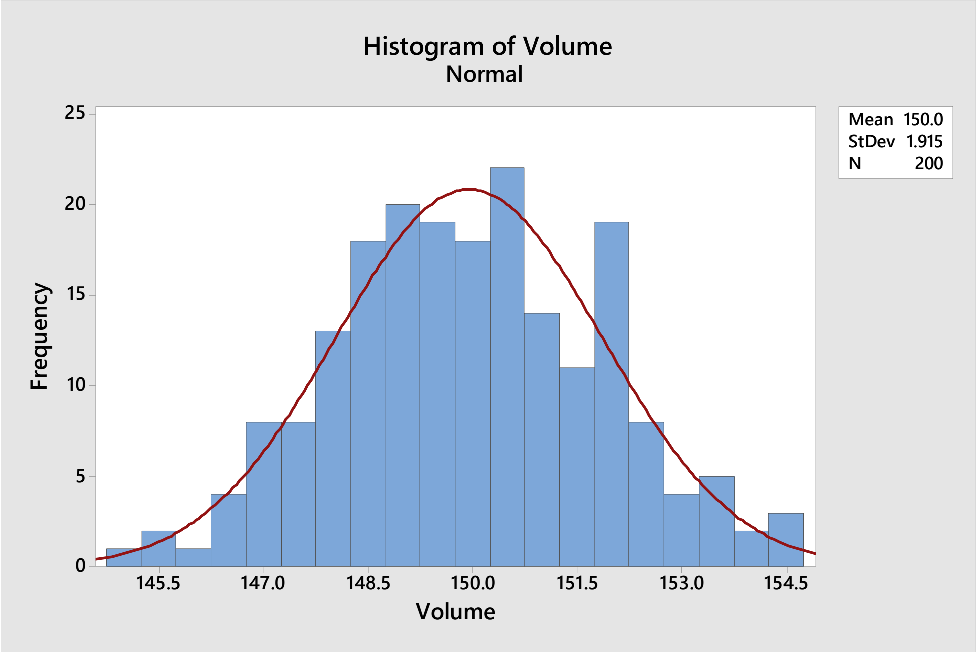
| x | x-$\bar{x}$ | (x-$\bar{x}$ )$^2$ |
|---|---|---|
| 100 | 0 | 0 |
| 101 | 1 | 1 |
| 99 | -1 | 1 |
| 102 | 2 | 4 |
| 98 | -2 | 4 |
| 100 | 0 | 0 |
| $\bar{x}$ =100 | ∑(x-$\bar{x}$ )=0 | ∑(x-$\bar{x}$ )$^2$=10 |

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

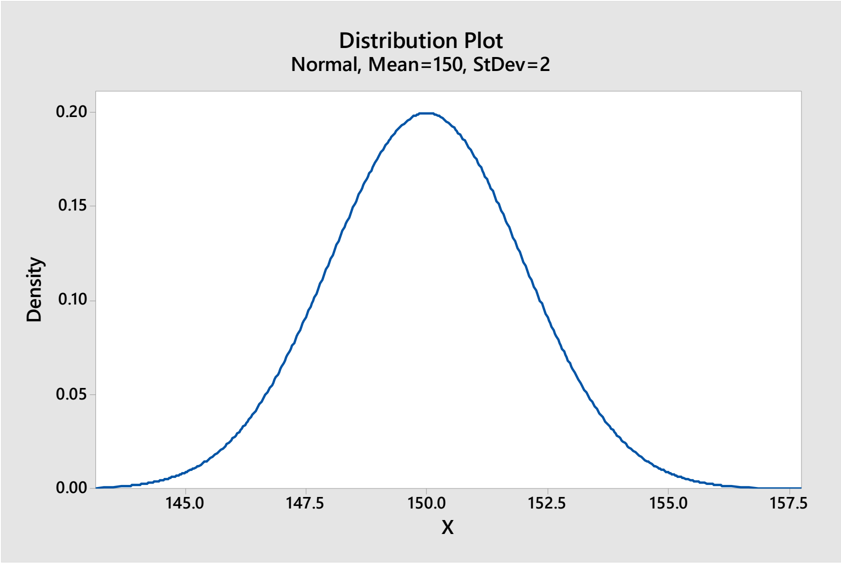
$$s = \sqrt{\frac{10}{(6-1)}} \qquad s = \sqrt{2} = 1.414$$

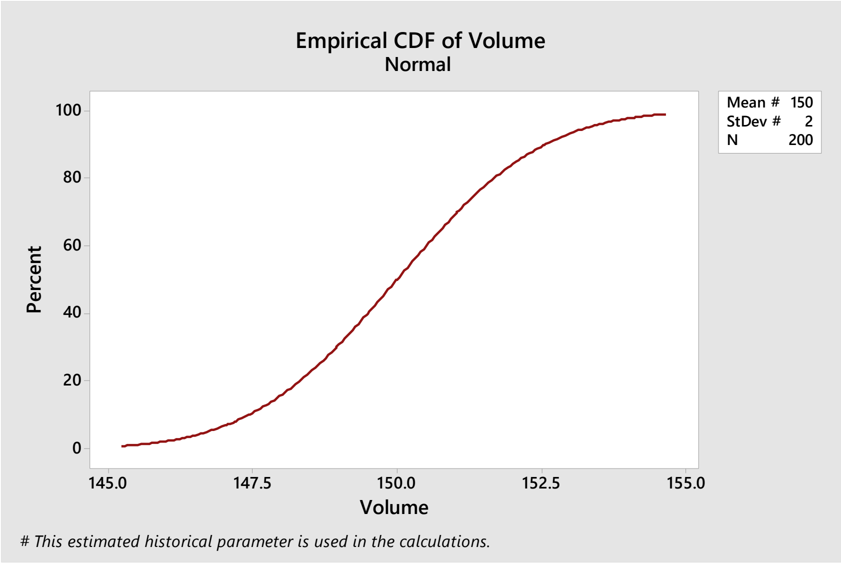
*Descriptive Statistics*

Volume
150.270
150.552
148.448
145.241
147.832
152.614
149.804
149.920
153.426
153.780
…..
…..




# *Frequency Distribution*




# *Cumulative Frequency Distribution*

# 3D
# Collecting and Summarizing Data

01 Types of data & measurement scales

02 Sampling and data collection methods

03 Descriptive statistics

04 Graphical methods

# 3D-4 Graphical Methods

01 Scatter diagrams

02 Histograms

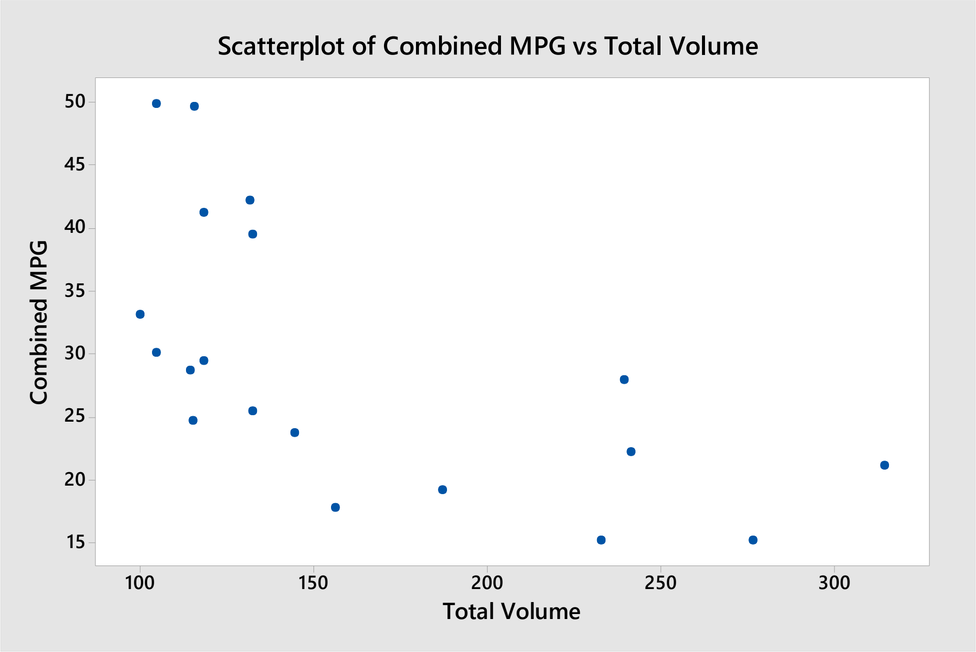
03 Box-and-whisker plots

04 Stem-and-leaf plots

05 Normal probability plots

# Scatter diagrams

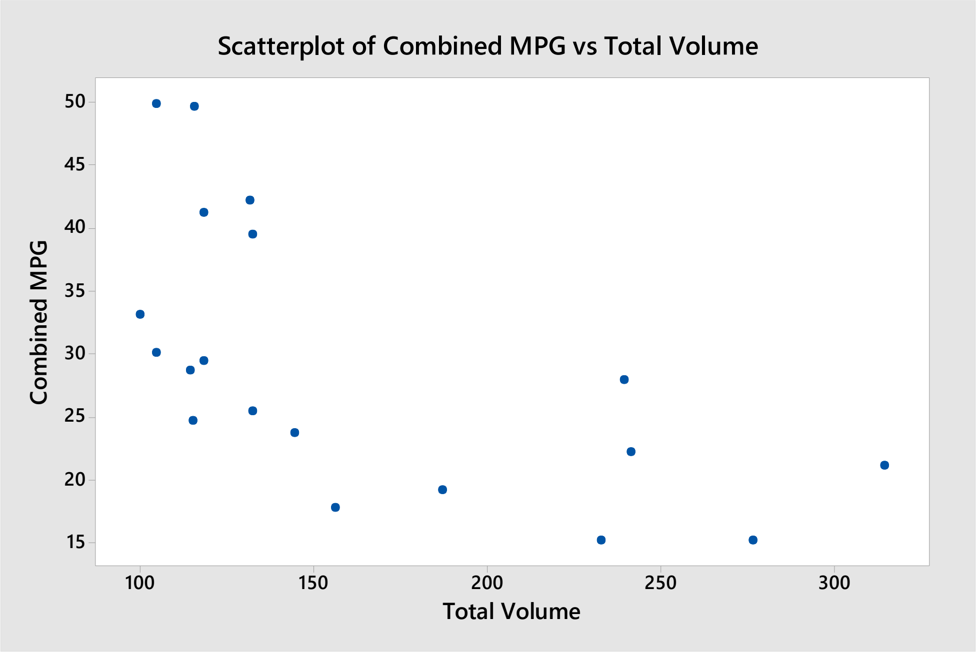
- One of seven basic quality tools
- To see relationship between two variables
- Relationship should make practical sense
- Temperature(X) vs Ice cream sale (Y)
- Some times relationship between two variables is because of a third variable. (ice cream sale vs heat stroke cases)
- Correlation/Regression is covered in the Analyze Phase

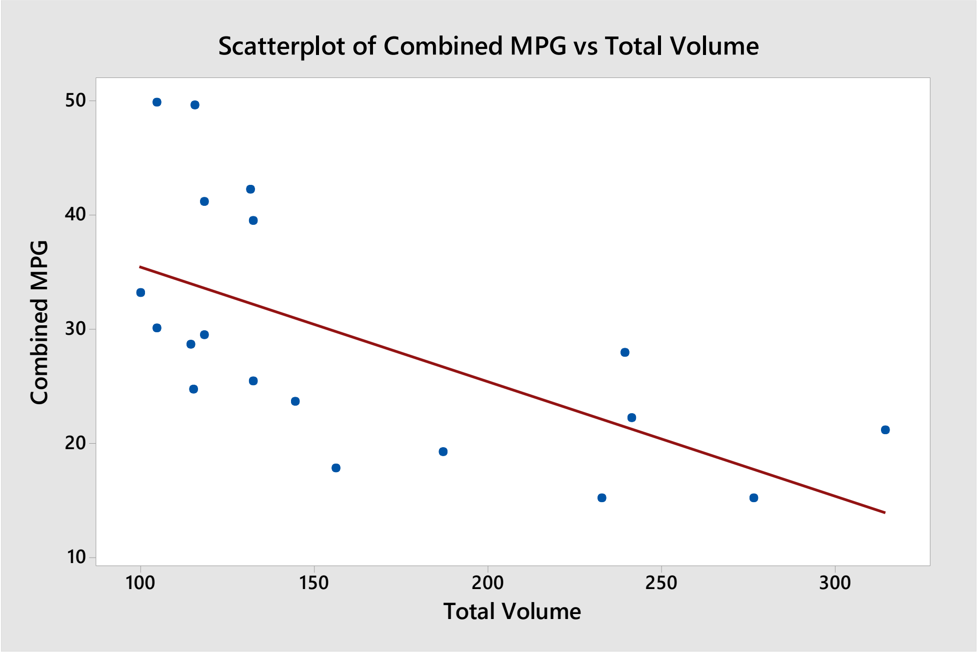



Minitab: Graph > Scatter plot …

| Worksheet column | Description |
|---|---|
| Vehicle | The model identifier |
| Type | The body type: SUV, Sedan, Hatchback, Wagon, or Minivan |
| Fuel | The type of fuel used: Gas or Hybrid |
| City MPG | Fuel economy for city driving in miles per gallon |
| Hwy MPG | Fuel economy for highway driving in miles per gallon |
| Combined MPG | Overall fuel economy in miles per gallon |
| Retail ($1000) | Retail price in units of $1000 |
| Safety (0-5) | Safety rating (0 to 5), where 0 is poor and 5 is excellent |
| Interior Volume | Passenger volume in cubic feet |
| Cargo Volume | Cargo volume in cubic feet |
| Total Volume | Total volume in cubic feet |

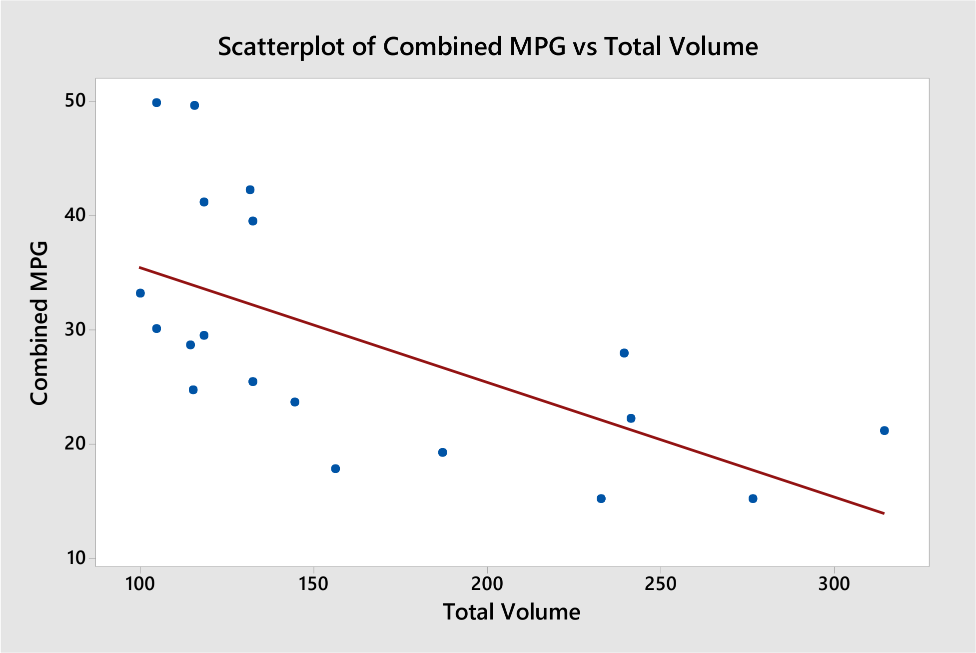
# *Scatter diagrams*



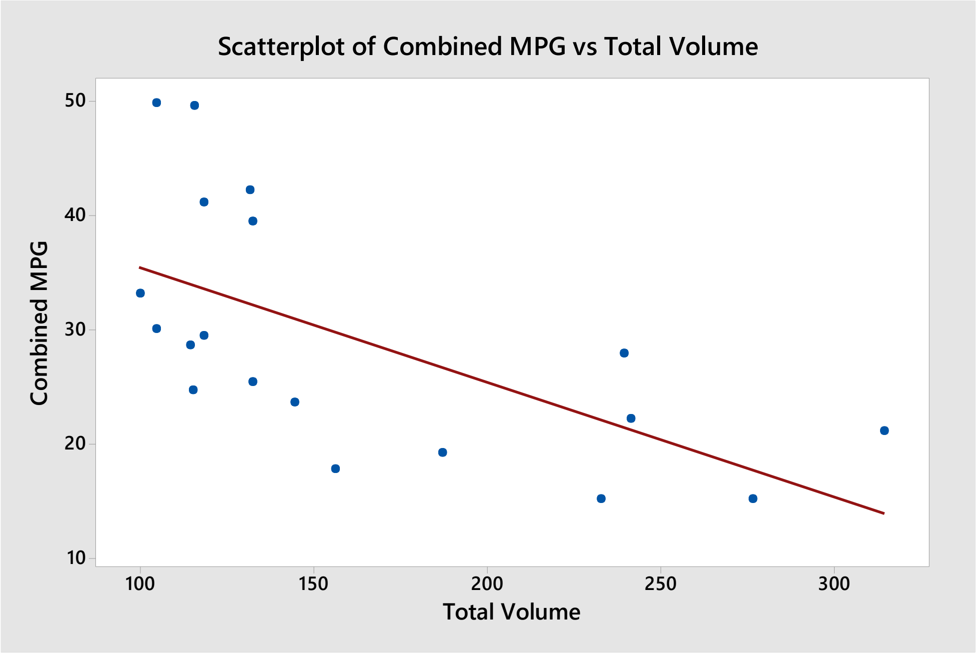

With Linear Regression Line

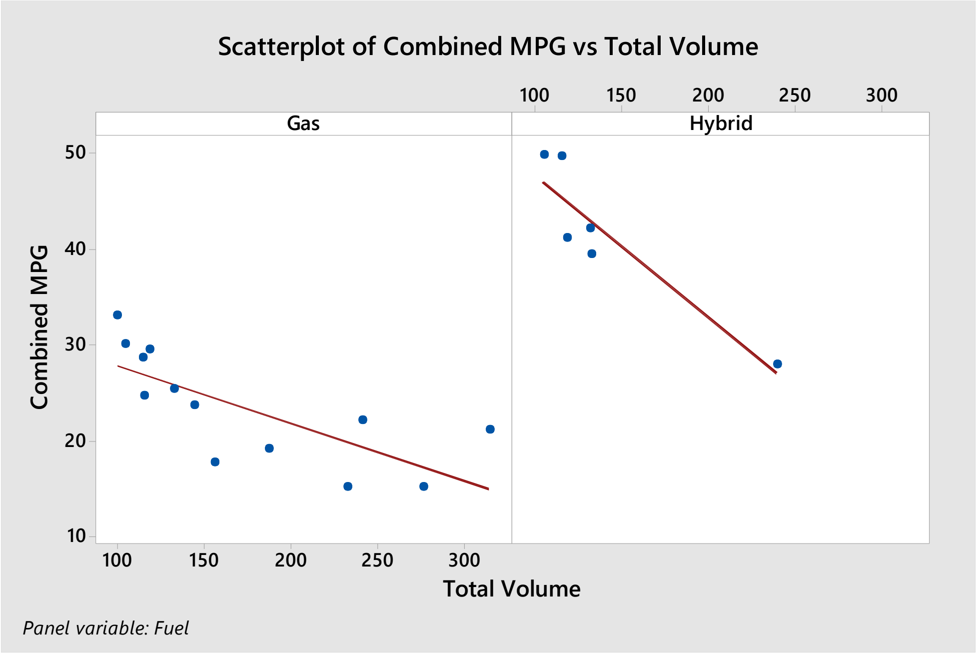
# *Scatter diagrams*




| Worksheet column | Description |
|---|---|
| Vehicle | The model identifier |
| Type | The body type: SUV, Sedan, Hatchback, Wagon, or Minivan |
| Fuel | The type of fuel used: Gas or Hybrid |
| City MPG | Fuel economy for city driving in miles per gallon |
| Hwy MPG | Fuel economy for highway driving in miles per gallon |
| Combined MPG | Overall fuel economy in miles per gallon |
| Retail ($1000) | Retail price in units of $1000 |
| Safety (0-5) | Safety rating (0 to 5), where 0 is poor and 5 is excellent |
| Interior Volume | Passenger volume in cubic feet |
| Cargo Volume | Cargo volume in cubic feet |
| Total Volume | Total volume in cubic feet |

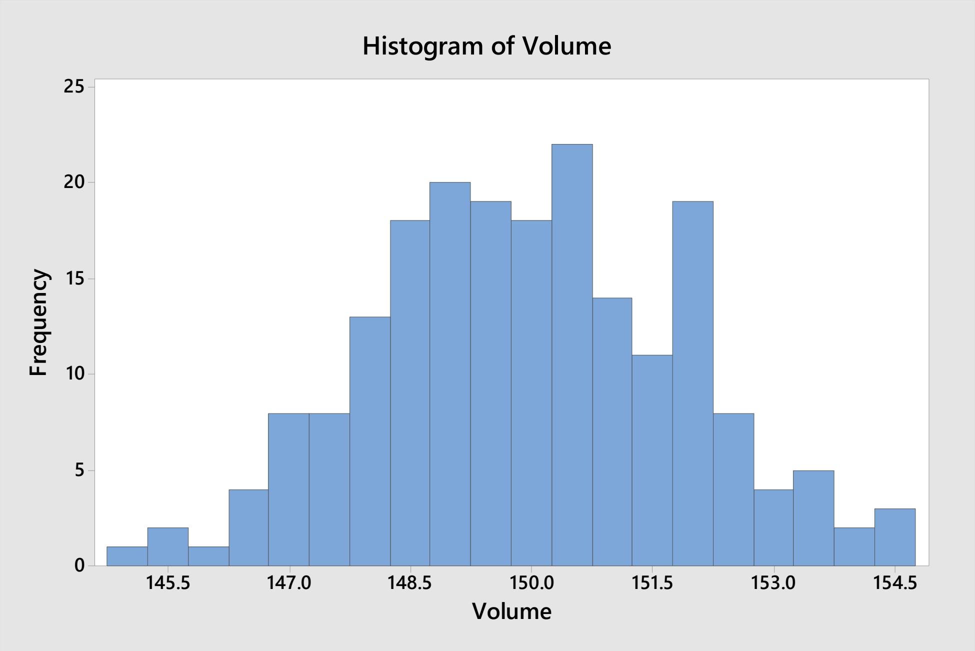
# *Scatter diagrams*




# *Scatter diagrams*

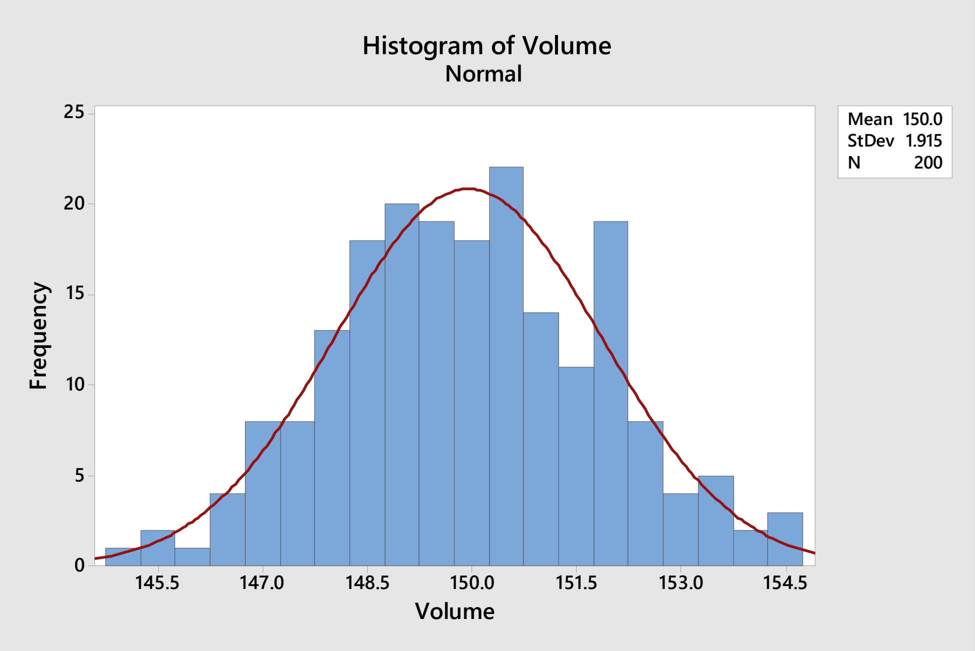
**Volume**
150.270
150.552
148.448
145.241
147.832
152.614
149.804
149.920
153.426
153.780
….
….




Minitab: Graph > Histogram… (Simple)

# *Histogram*

**Volume**
150.270
150.552
148.448
145.241
147.832
152.614
149.804
149.920
153.426
153.780
….
….

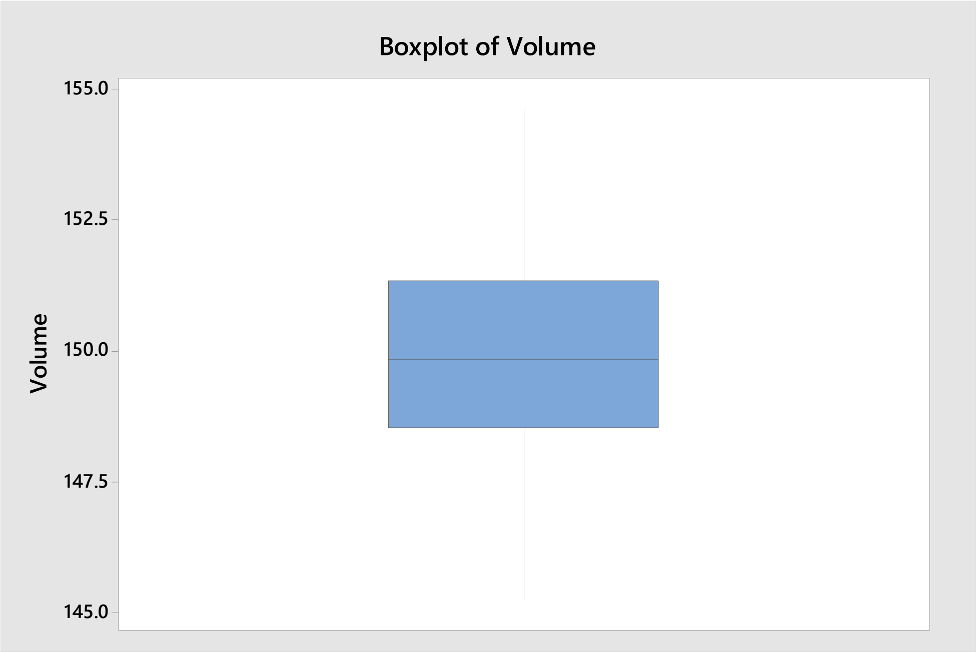



Minitab: Graph > Histogram… (with fit)

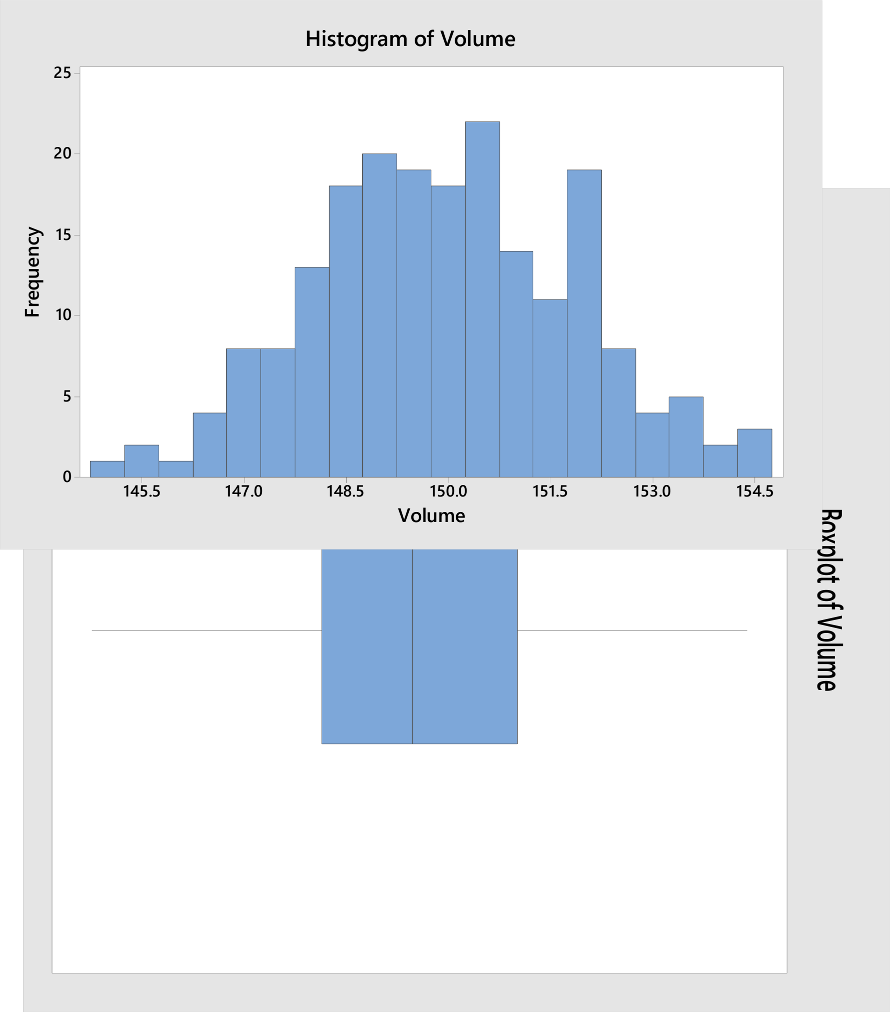
# *Histogram*

- Also known as Box Plot
- Shows the median
- Shows Q1, Q3 and IQR

| Volume |
|---|
| 150.270 |
| 150.552 |
| 148.448 |
| 145.241 |
| 147.832 |
| 152.614 |
| 149.804 |
| 149.920 |
| 153.426 |
| 153.780 |
| …. |
| …. |

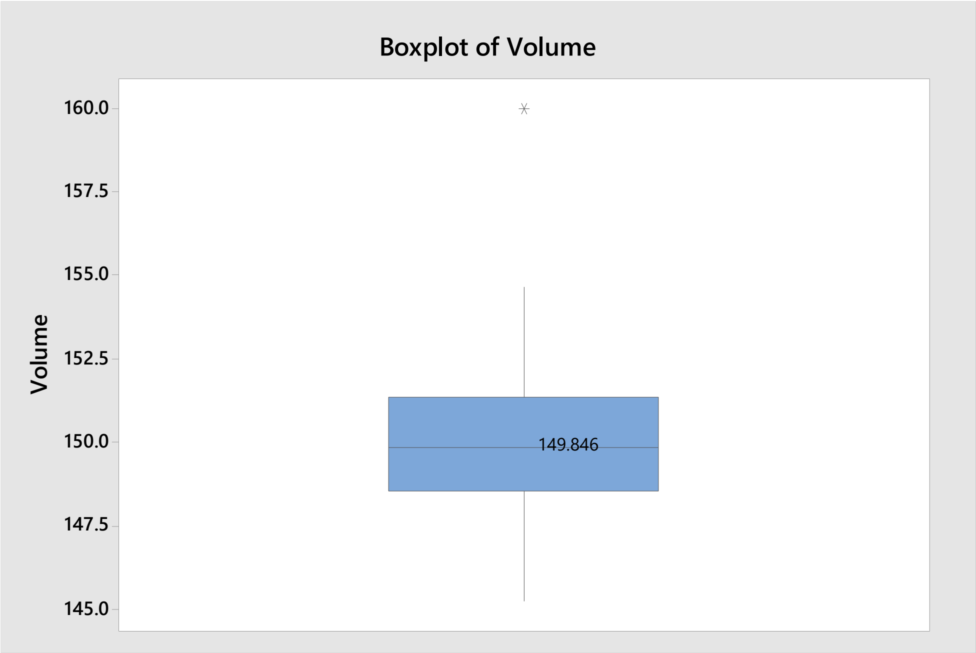



# *Box-and-Whisker Plot*

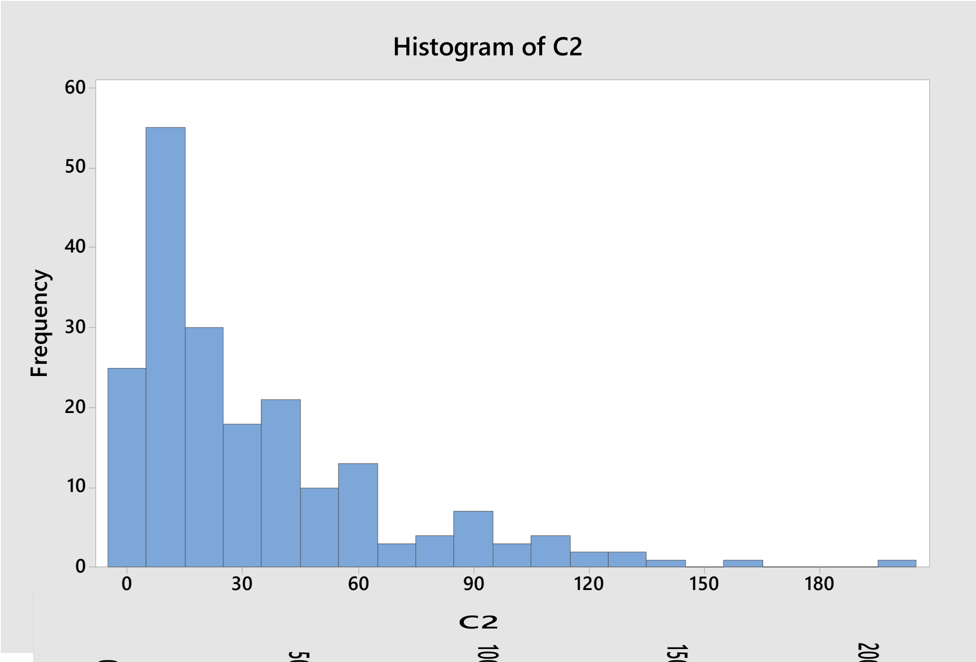


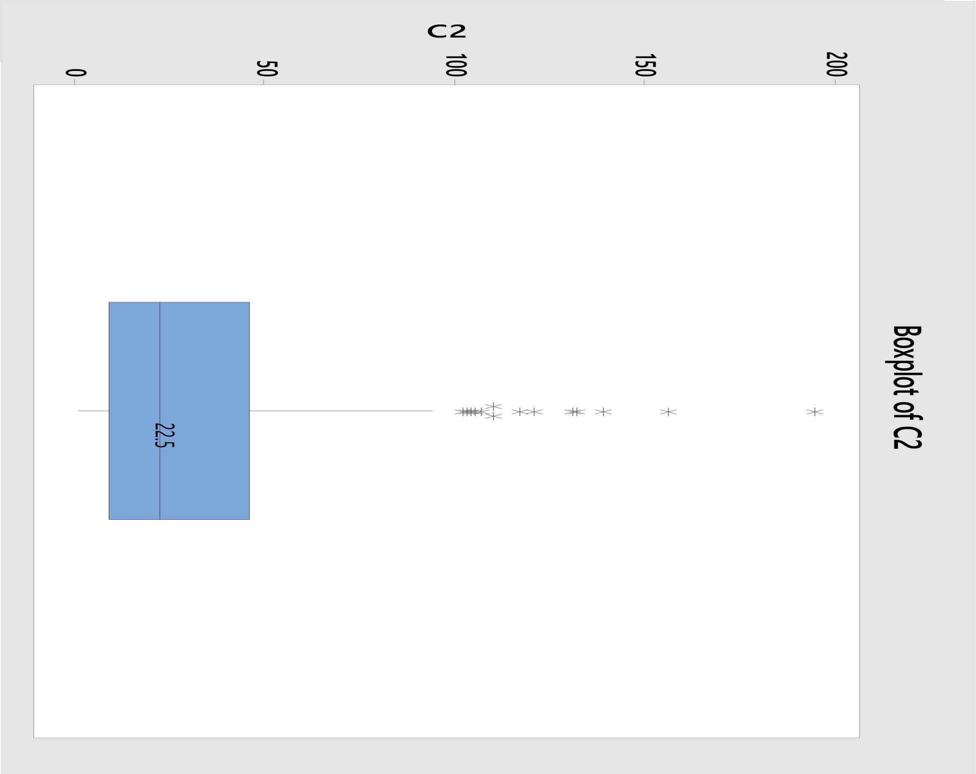

# *Box-and-Whisker Plot*

❖ Outlier – A data point which is more than 1.5 times IQR from the box.

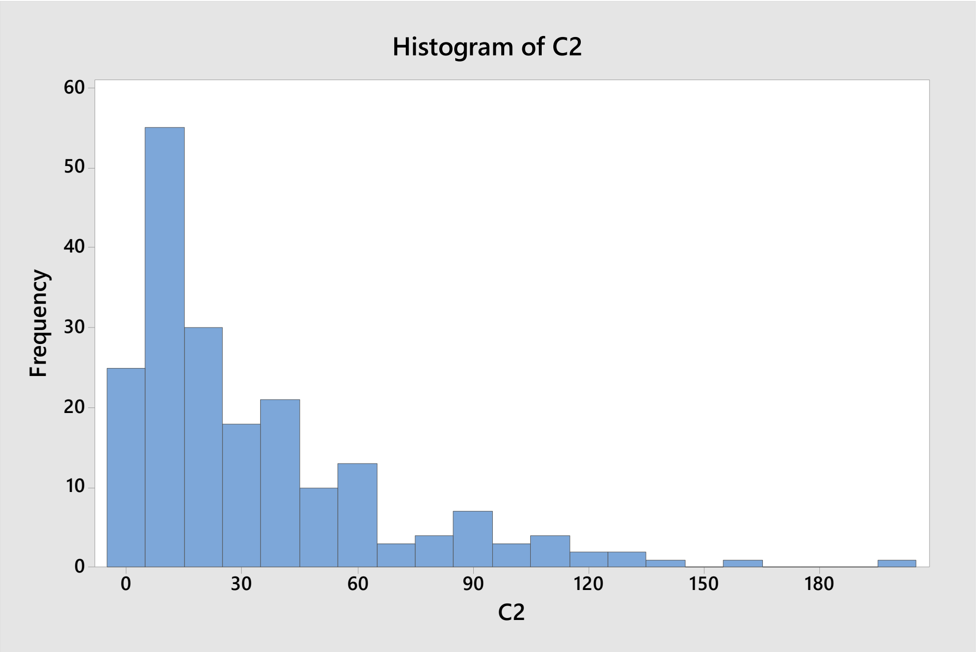


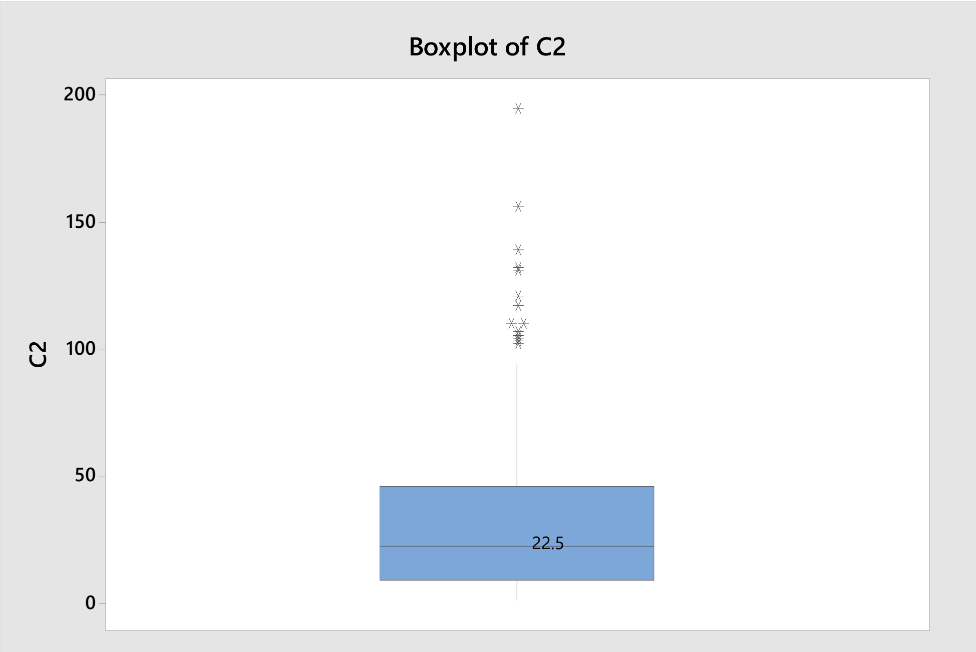

# Box-and-Whisker Plot




# *Box-and-Whisker Plot*




# *Box-and-Whisker Plot*

❖ 11, 22, 55, 13, 45, 14, 19, 10, 33, 52, 13

| Stem | Leaf |
| --- | --- |
| 1 | 0 1 3 3 4 9 |
| 2 | 2 |
| 3 | 3 |
| 4 | 5 |
| 5 | 2 5 |

***Stem-and-leaf Plot***

**Volume**
150.270
150.552
148.448
145.241
147.832
152.614
149.804
149.920
153.426
153.780
….
….

## Stem-and-Leaf Display: Volume

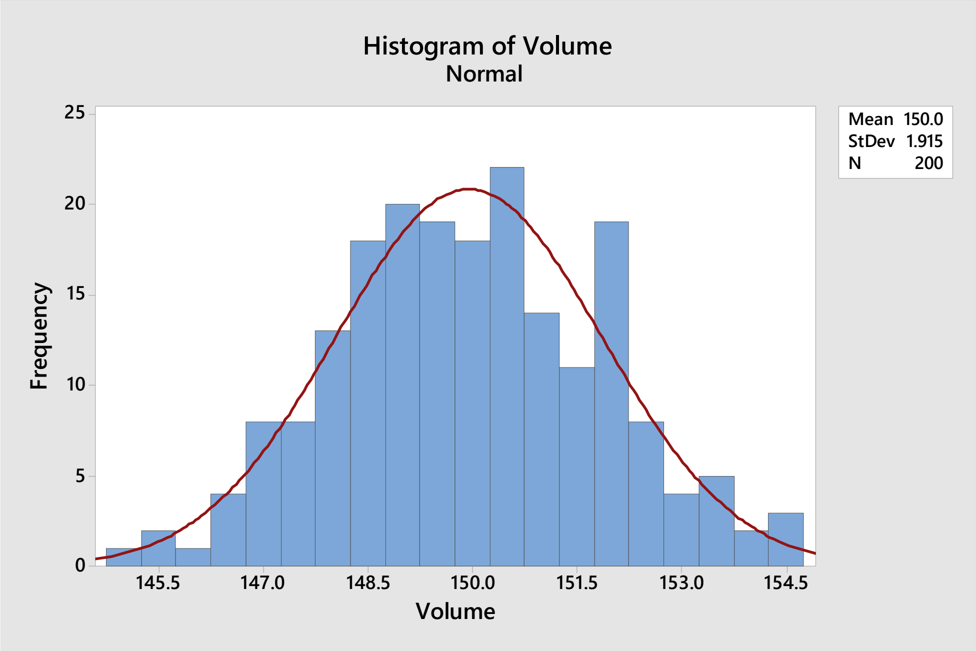
Stem-and-leaf of Volume N = 200

| | | |
|---|---|---|
| 3 | 145 | 266 |
| 13 | 146 | 0333488899 |
| 31 | 147 | 012223555667778899 |
| 62 | 148 | 0011122333334444555555677799999 |
| (43) | 149 | 0000011111112222333344445566667778888888999 |
| 95 | 150 | 00011122333344444444555556666778899 9 |
| 60 | 151 | 0111222223334444455788899 |
| 35 | 152 | 00000000111222234446677 |
| 12 | 153 | 002446777 |
| 3 | 154 | 456 |

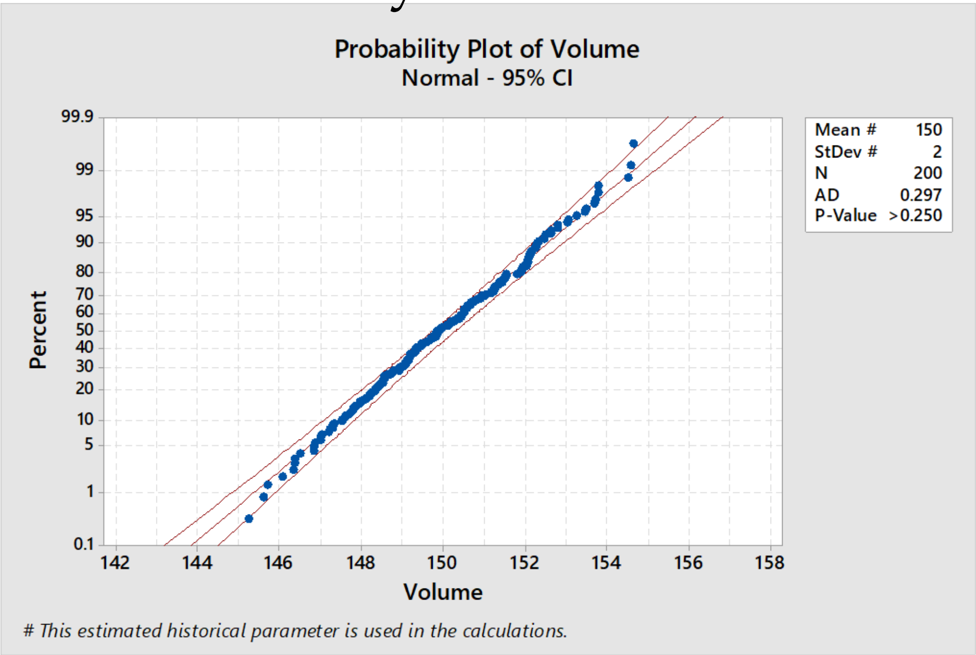
*Leaf Unit = 0.1*

Minitab: Graph > Stem-and-Leaf…

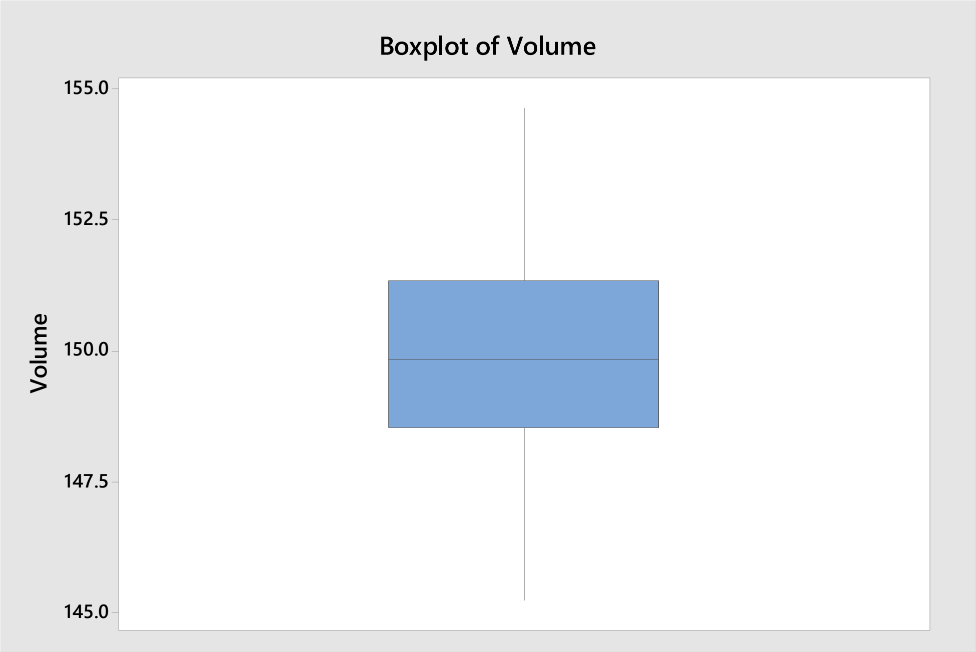
***Stem-and-Leaf Plot***



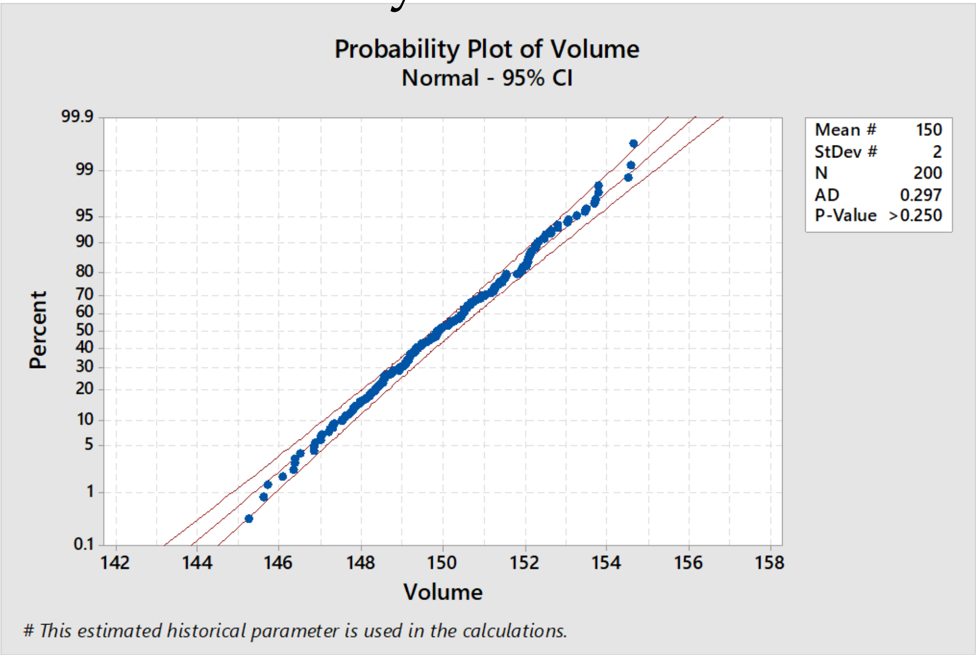

Minitab: Graph > Probability Plot …




# *Normal Probability Plots*

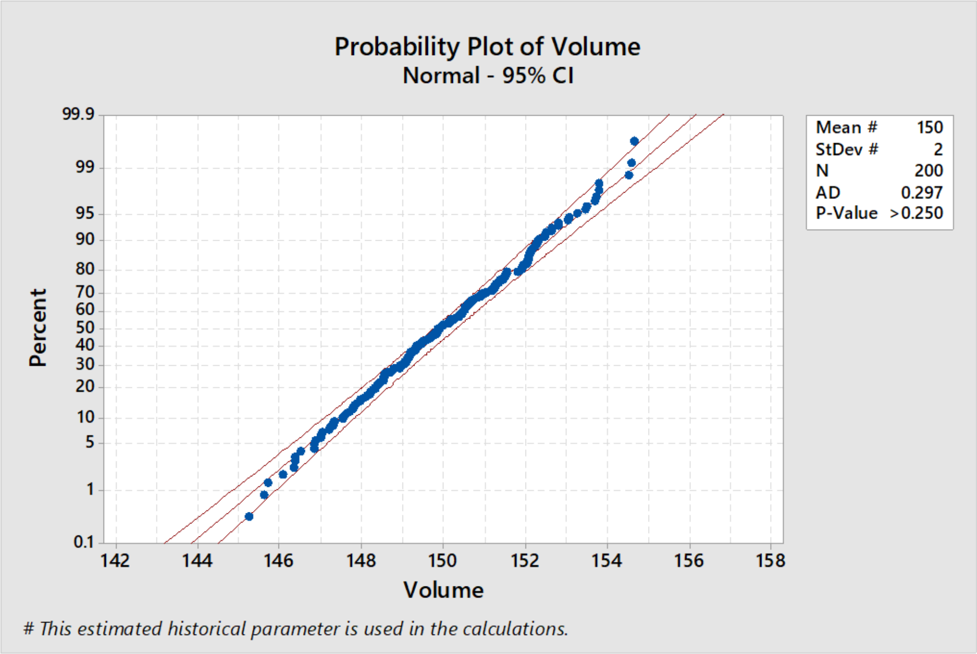



Minitab: Graph > Probability Plot …




# *Normal Probability Plots*

- The p value > 0.05 and hence data is normal with 95% confidence level.




***Normal Probability Plots***

# 2022 – Changes in the BoK – 3D

| 2014 BoK | 2022 BoK Details | Notes |
| --- | --- | --- |
| III.D | **D. Collecting and summarizing data** | |
| III.D.1 | 1. Types of data and measurement scales<br>Identify and classify continuous (variables) and discrete (attributes) data. Describe and define nominal, ordinal, interval, and ratio measurement scales. (Analyze) | |
| III.D.2 | 2. Sampling and data collection plans and methods<br>Define and apply various sampling methods (random and stratified) and data collection methods (check sheets and data coding). Prepare data collection plans that include gathering data and performing quality checks (e.g., minimum/maximum values, erroneous data, null values). (Apply) | Revised subtopic name and added data collection plans and quality checks |
| III.D.3 | 3. Descriptive statistics<br>Define, calculate, and interpret measures of dispersion and central tendency. Develop and interpret frequency distributions and cumulative frequency distributions. (Evaluate) | |
| III.D.4 | 4. Graphical methods<br>Construct and interpret diagrams and charts that are designed to communicate numerical analysis efficiently, including scatter diagrams, normal probability plots, histograms, stem-and-leaf plots, box-and-whisker plots. (Create) | |

*Moved from BB to GB*

# Data Collection Plan

- Why you need to collect data?
  - Goal and Objective
- Operational Definition
  - How much? How? Where? When? Etc.
  - Type of data – NOIR
  - Manual or Automatic
  - Past data vs Future
- Is data reliable?

# Data Collection Plan

| Measurement | Operational Definition | How is it measured? | Type of Data | Sample size | Who? | Data Recording Form | Comments |
|---|---|---|---|---|---|---|---|
| Time to assemble | Time from picking up the first piece to placing the assembled item in tray | Using a stop watch | Continuous<br><br>Ratio | Every $10^{th}$ piece | Operator | Assembly Record F-0156 | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

# Quality Checks on Data

- Minimum/maximum values
  - Range Check (e.g. Employee age between 18 to 65)
- Erroneous data
  - List of Options (e.g. US, CA, IN for country codes)
- Null values or Missing value (discussed on the next slide)*
- Duplicate check (e.g. Employee ID)

# Data Cleaning – Missing Data

- In statistics, imputation is the process of replacing missing data with substituted values.
- Missing data can introduce bias.
  - Missing randomly
  - Reason for missing
- Options in case the data is missing
  - Delete the row
  - Replace with the average value

# 3E Measurement System Analysis (MSA)

01 Definitions

02 Gage R & R studies

03 Precision to Tolerance Ratio

- You need measurement to see how the process is performing.
- Process has variation.
- What about measurement error / variation?

*Measurement System Analysis*

- Measurement System :
  - Operator
  - Measuring instrument
  - Procedure

Measurement System Analysis

- True Value – Actual value, which is unknown
- Reference Value – Accepted value or substitute of true value.

*Measurement System Analysis*

# Resolution

- Resolution/ Discrimination
  - Smallest readable unit of the measuring instrument.
  - 10 to 1 Rule of Thumb:
  - “Rule of Ten” or “one to ten” is that the discrimination (resolution) of the measuring instrument should divide the tolerance of the characteristic to be measured into ten parts.

*Measurement System Analysis*

# Resolution

- 10 to 1 Rule of Thumb:
  - Which of these two would you use if the part tolerance is 52.00 +/- 0.05 (51.95 – 52.05)



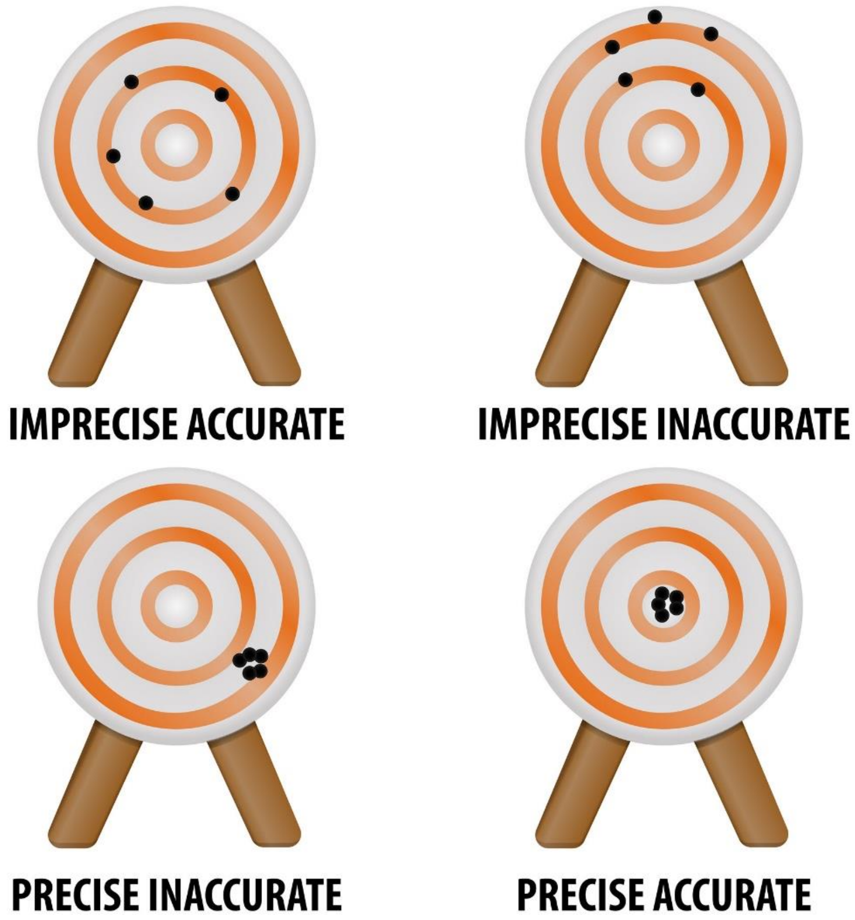
***Measurement System Analysis***

# Resolution

- 10 to 1 Rule of Thumb:
  - Which of these two would you use if the part tolerance is 52.00 +/- 0.05 (51.95 – 52.05)
  - Tolerance Range = 0.10
  - Minimum Reading of Digital Vernier = 0.01
  - Digital Vernier divides the tolerance into 10 parts, hence acceptable.

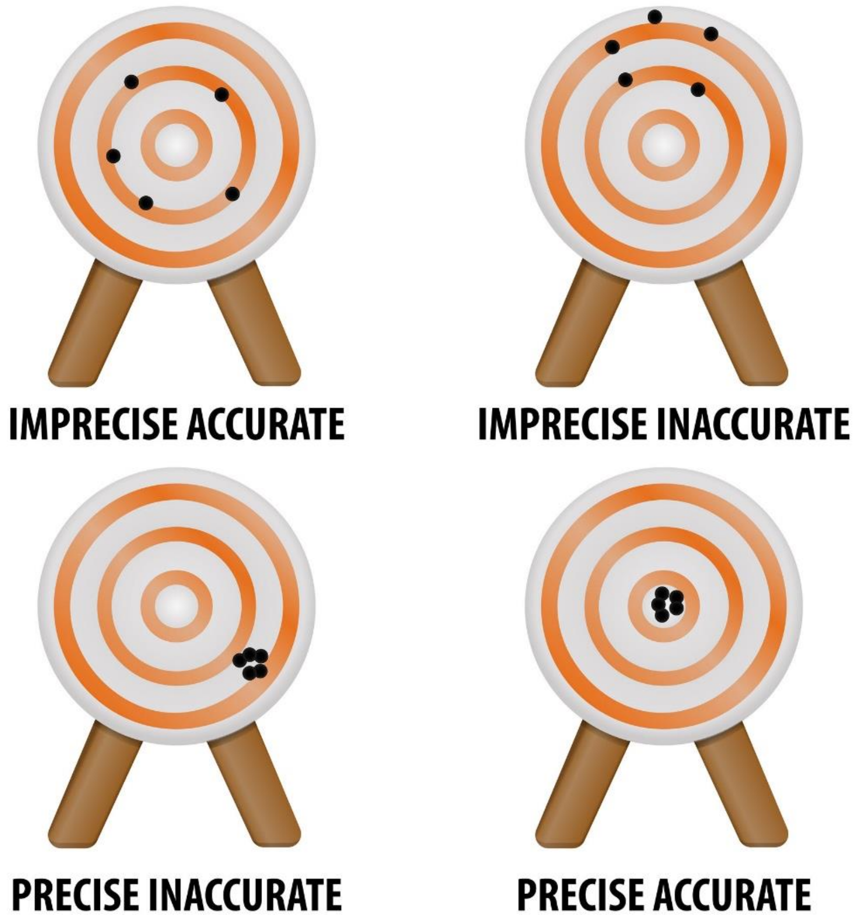

***Measurement System Analysis***

- ❖ Resolution
- ❖ Accuracy
    - ❖ Bias
    - ❖ Linearity
    - ❖ Stability
- ❖ Precision
    - ❖ Repeatability
    - ❖ Reproducibility




# *Measurement System Analysis*

- Accuracy
  - "Closeness" to the true value, or to an accepted reference value.
    - Bias
    - Linearity
    - Stability
- Precision
  - "Closeness" of repeated readings to each other
    - Repeatability
    - Reproducibility




***Measurement System Analysis***

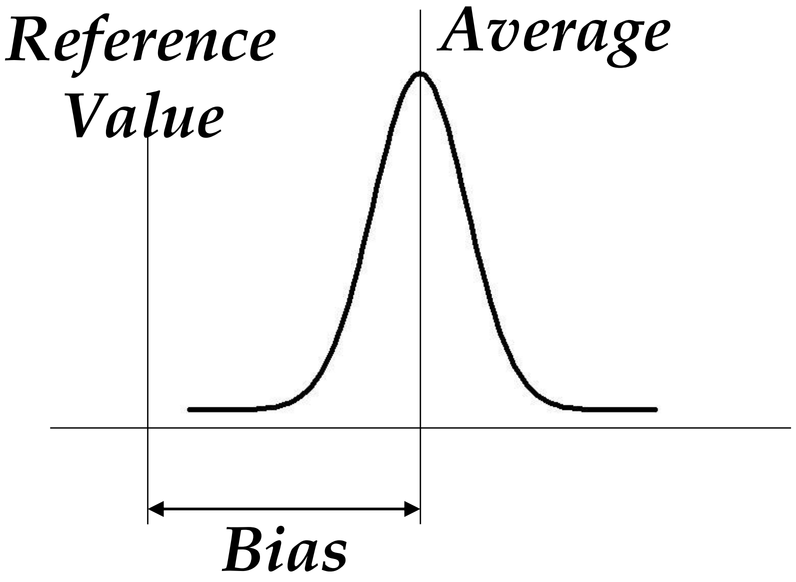
# Bias

- Bias is the difference between the observed average of measurements and the reference value.

| | Reference Value (psi) | Measured Value (psi) |
|---|---|---|
| | 100 | 100 |
| | 100 | 101 |
| | 100 | 102 |
| | 100 | 102 |
| | 100 | 101 |
| | 100 | 100 |
| Average | | 101 |
| Bias = 101 psi -100 psi = 1 psi | | |

***Measurement System Analysis***

| | Reference Value (psi) | Measured Value (psi) |
|---|---|---|
| | 100 | 100 |
| | 100 | 101 |
| | 100 | 102 |
| | 100 | 102 |
| | 100 | 101 |
| | 100 | 100 |
| Average | | 101 |
| Bias = 101 psi -100 psi = 1 psi | | |




# *Measurement System Analysis*

# Bias

- ❖ Bias is the systematic error.
- ❖ Bias is addressed by calibration.

*Measurement System Analysis*

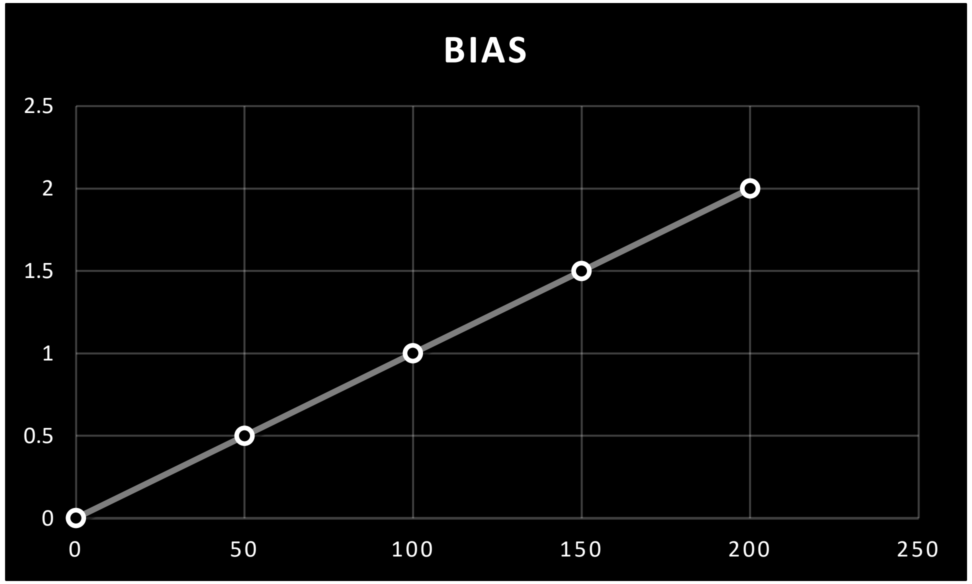
# Linearity

- Linearity measures the bias across the operating range of a tool or instrument.

| Reference Value (psi) | Average Measured Value (psi) | Bias |
|---|---|---|
| 0 | 0 | 0 |
| 50 | 50.5 | 0.5 |
| 100 | 101 | 1 |
| 150 | 151.5 | 1.5 |
| 200 | 202 | 2 |

*Measurement System Analysis*

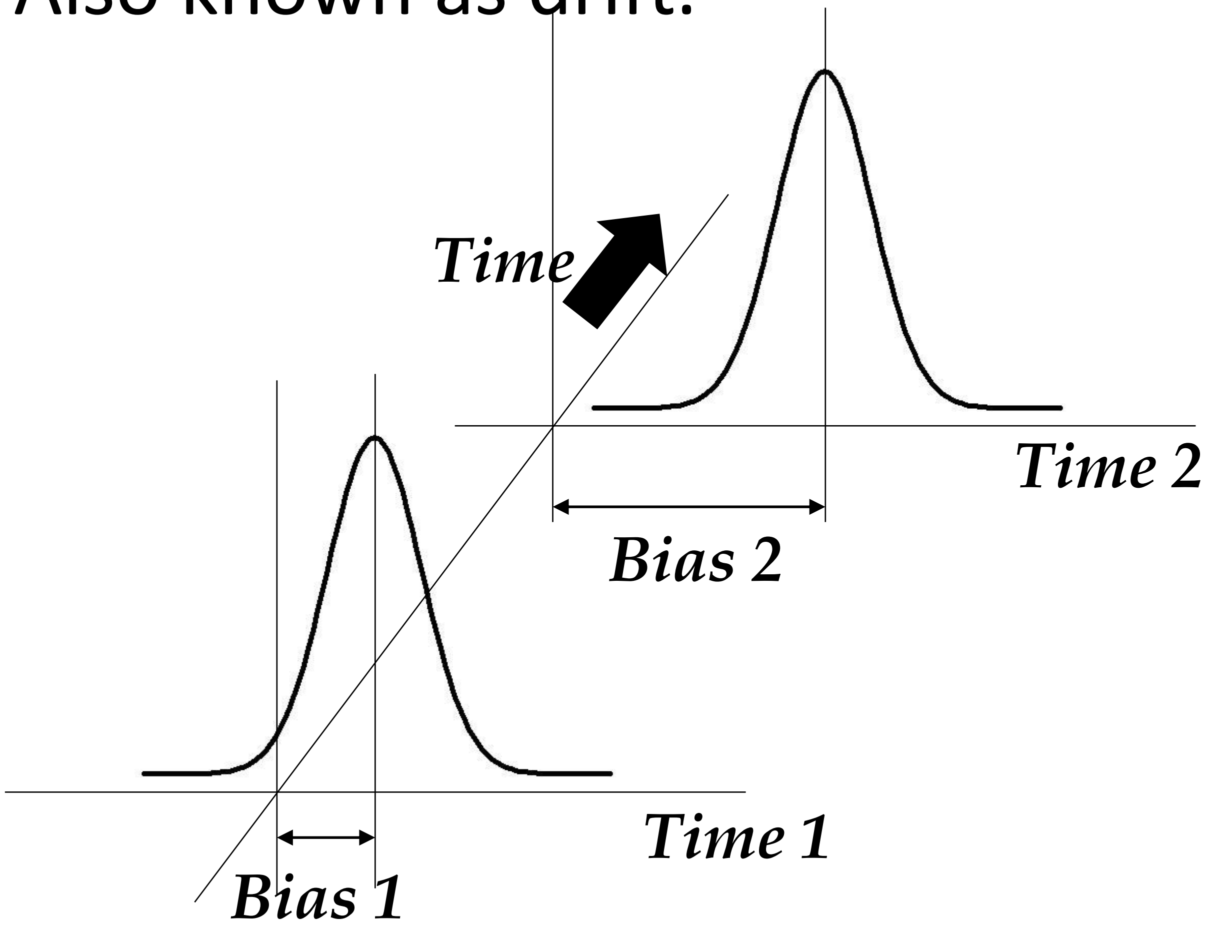
| Reference Value (psi) | Average Measured Value (psi) | Bias |
|---|---|---|
| 0 | 0 | 0 |
| 50 | 50.5 | 0.5 |
| 100 | 101 | 1 |
| 150 | 151.5 | 1.5 |
| 200 | 202 | 2 |




# Measurement System Analysis

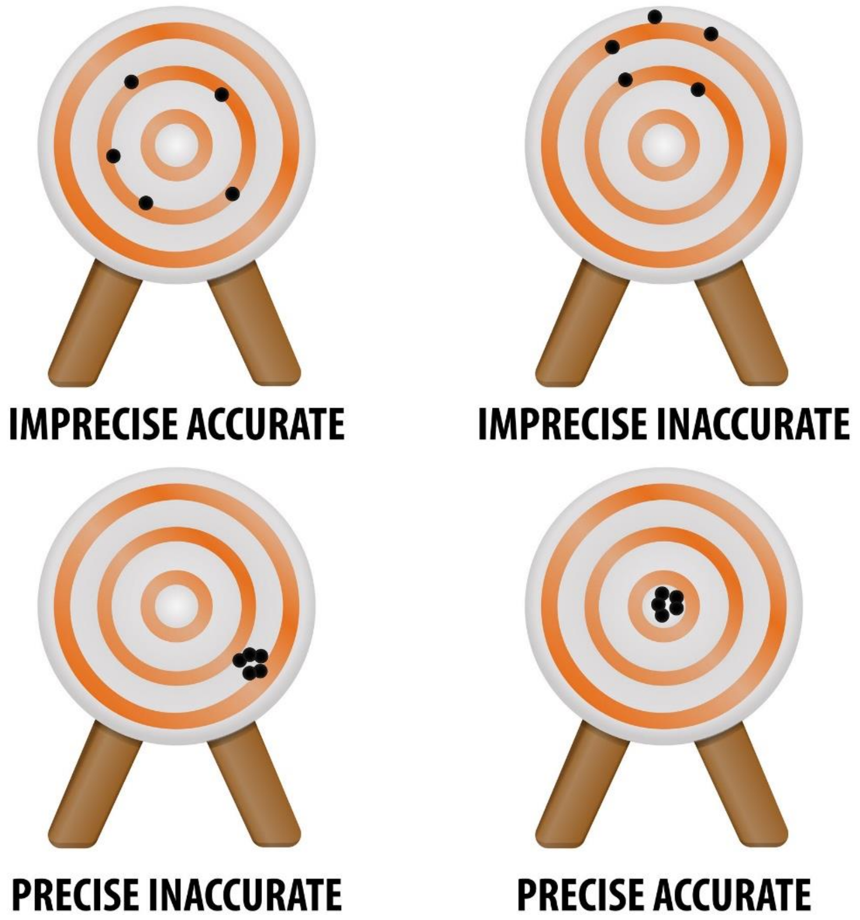
# Stability

- ❖ Stability measures the bias over time. Also known as drift.



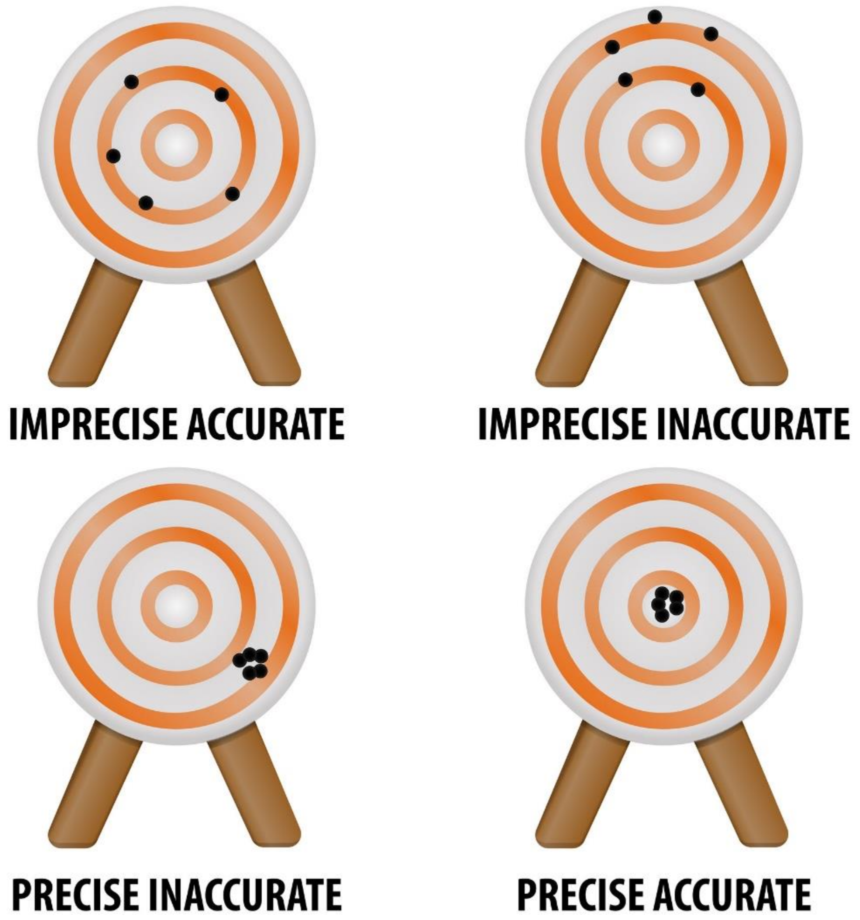

**Measurement System Analysis**

- Accuracy
  - "Closeness" to the true value, or to an accepted reference value.
    - Bias
    - Linearity
    - Stability
- Precision
  - "Closeness" of repeated readings to each other
    - Repeatability
    - Reproducibility




# *Measurement System Analysis*

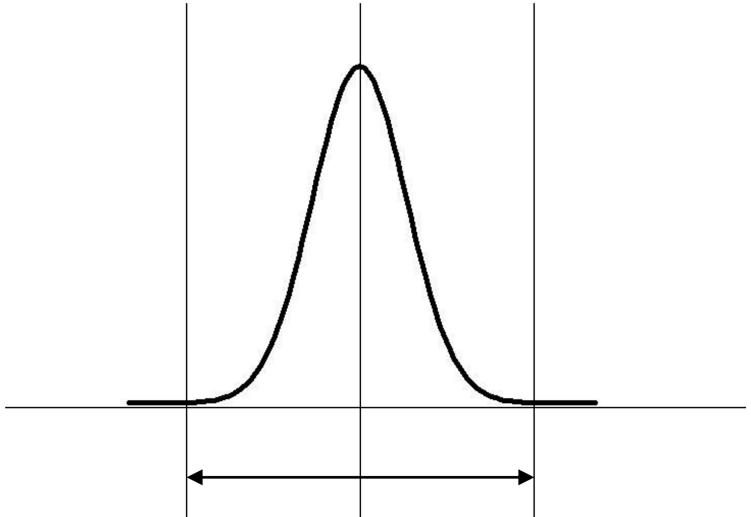
- Accuracy
  - "Closeness" to the true value, or to an accepted reference value.
    - Bias
    - Linearity
    - Stability
- Precision
  - "Closeness" of repeated readings to each other
    - Repeatability
    - Reproducibility




# *Measurement System Analysis*

# Repeatability

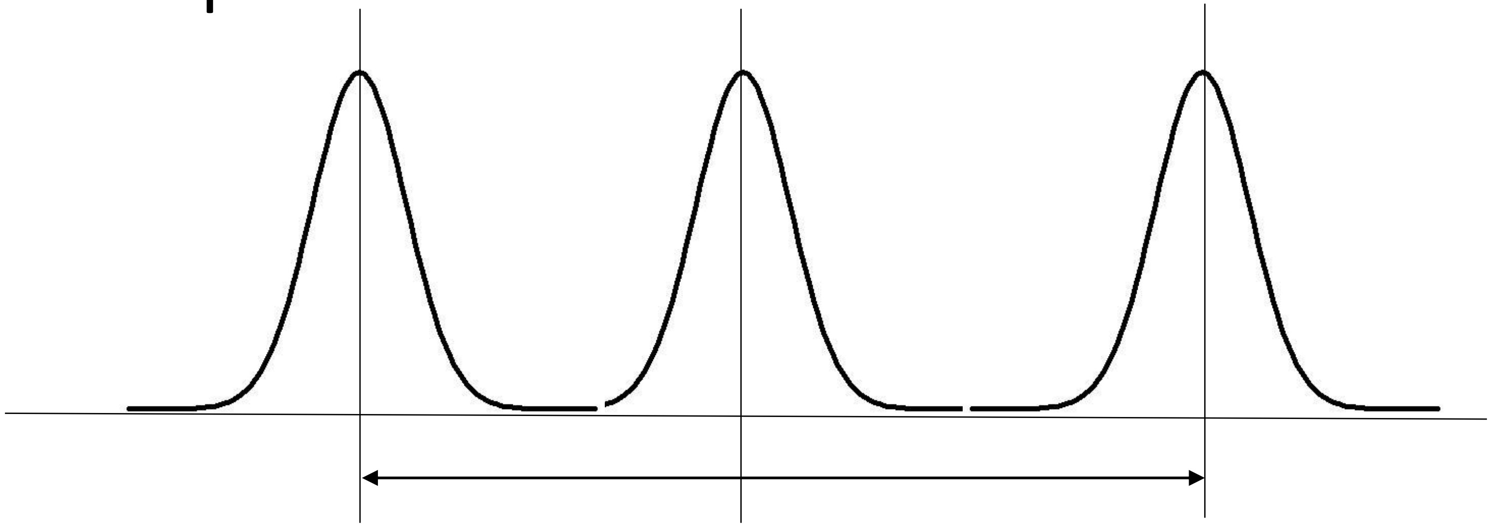
- Variation in measurements obtained with **one measuring instrument when used several times by an appraiser**.
- Also called Equipment Variation (EV)
- It’s the capability of the **gauge** to produce consistent results.



***Measurement System Analysis***

# Reproducibility

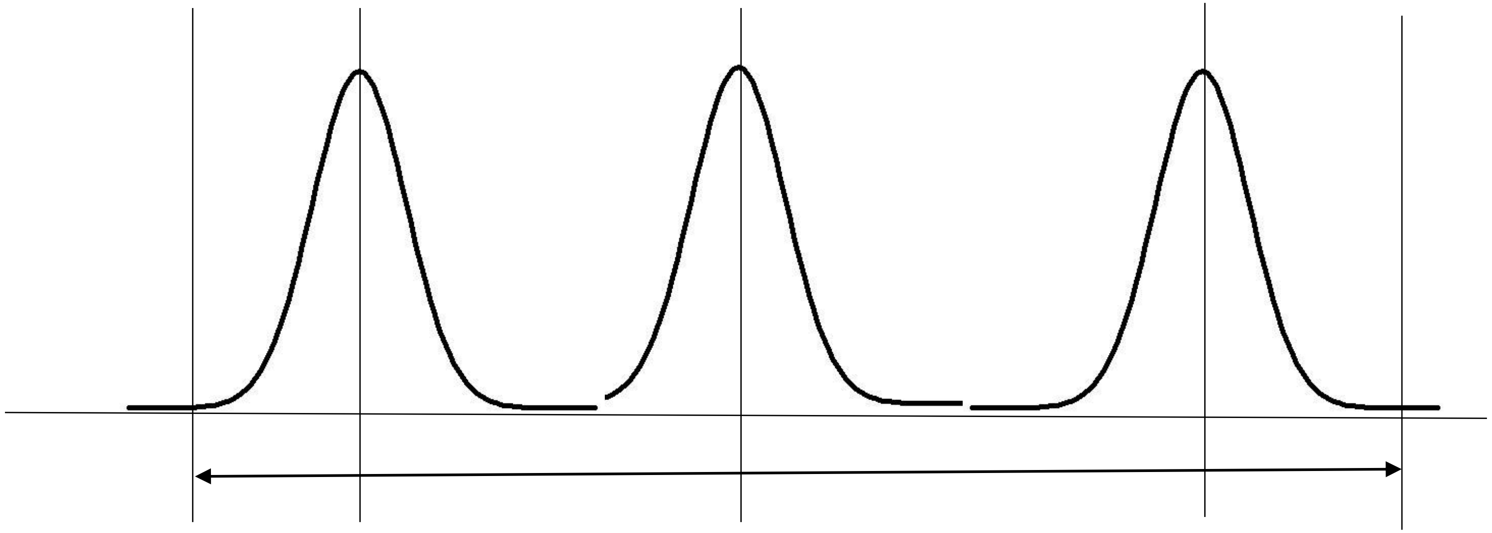
- Variation in the average of the measurements made **by different appraisers using the same gage**
- Also called Appraiser Variation (AV)
- It's the capability of the **appraiser** to produce consistent results.



*Measurement System Analysis*

# Gage R&R (GRR)

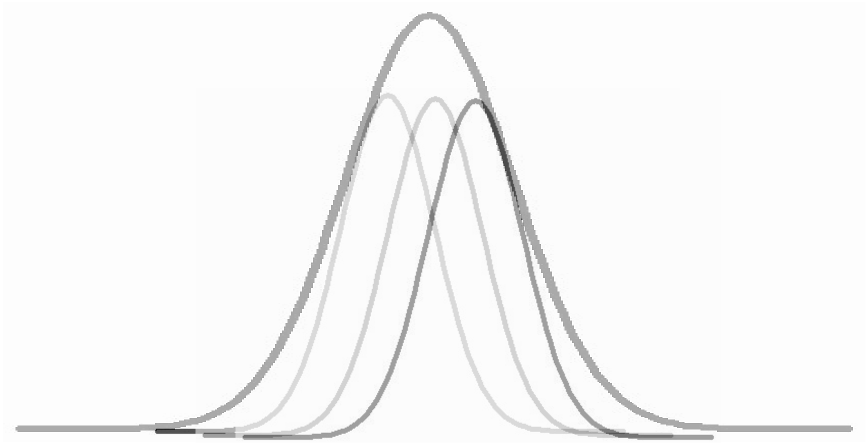
- Combined estimate of repeatability and reproducibility.



Measurement System Analysis

# Gage R&R (GRR)

❖ Combined estimate of repeatability and reproducibility.



**_Measurement System Analysis_**

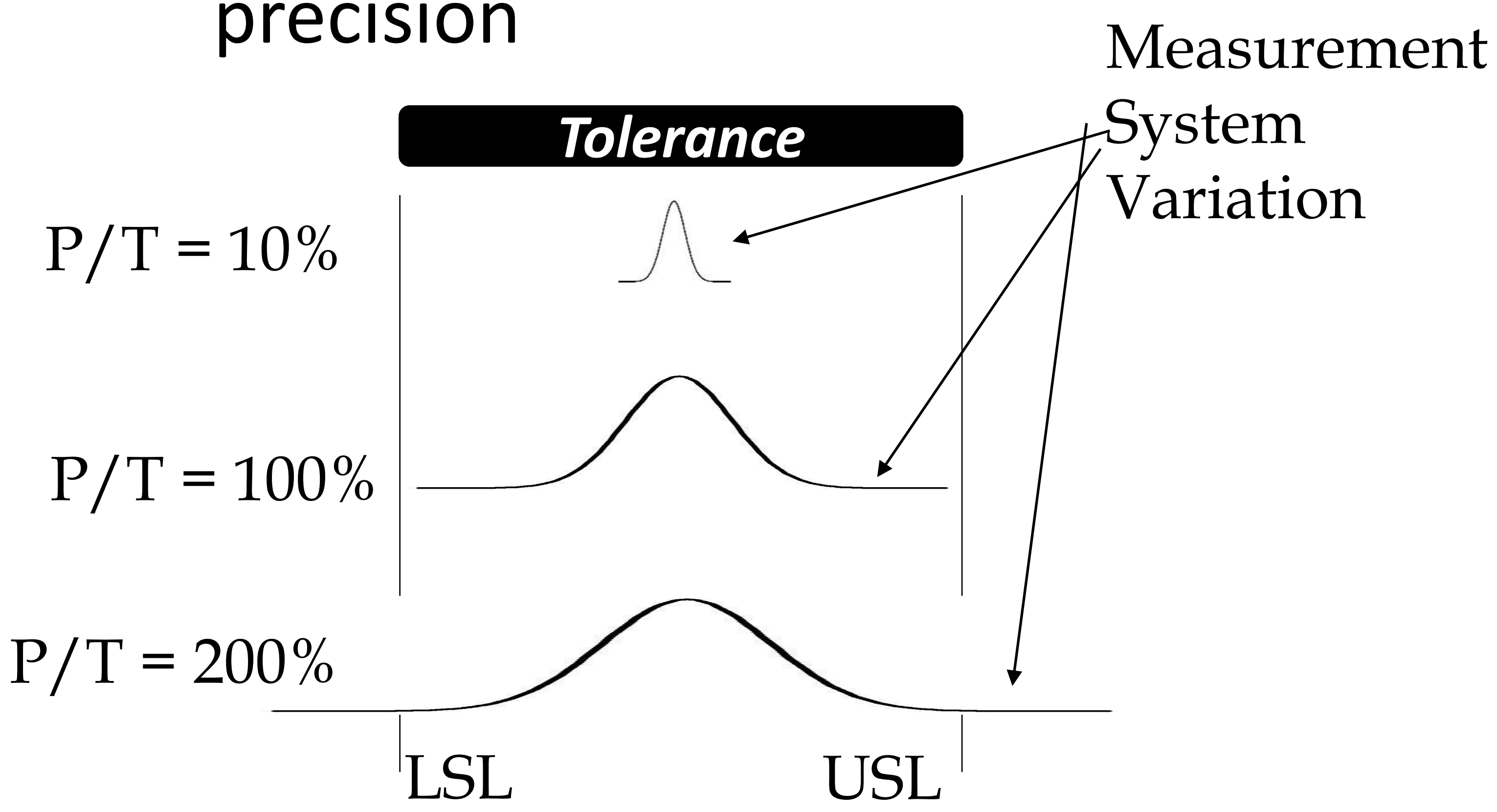
# Precision to Tolerance Ratio

- How capable your measurement system is?
- Precision/Tolerance (P/T) is the ratio between the estimated measurement error (precision) and the tolerance of the characteristic being measured.

*Measurement System Analysis*

# Precision to Tolerance Ratio

- P/T ratio is the most common estimate of measurement system precision




***Measurement System Analysis***

# Precision to Tolerance Ratio

PTR

- PTR $= \dfrac{6\ \sigma_{ms}}{USL-LSL}$

Instead of $6\sigma_{ms}$ some use $5.15\ \sigma_{ms}$

6 sigma includes 99.73% area

5.15 sigma includes 99% area

**_Measurement System Analysis_**

# 3F
# Process and Performance Capability

01 Process performance vs process specification

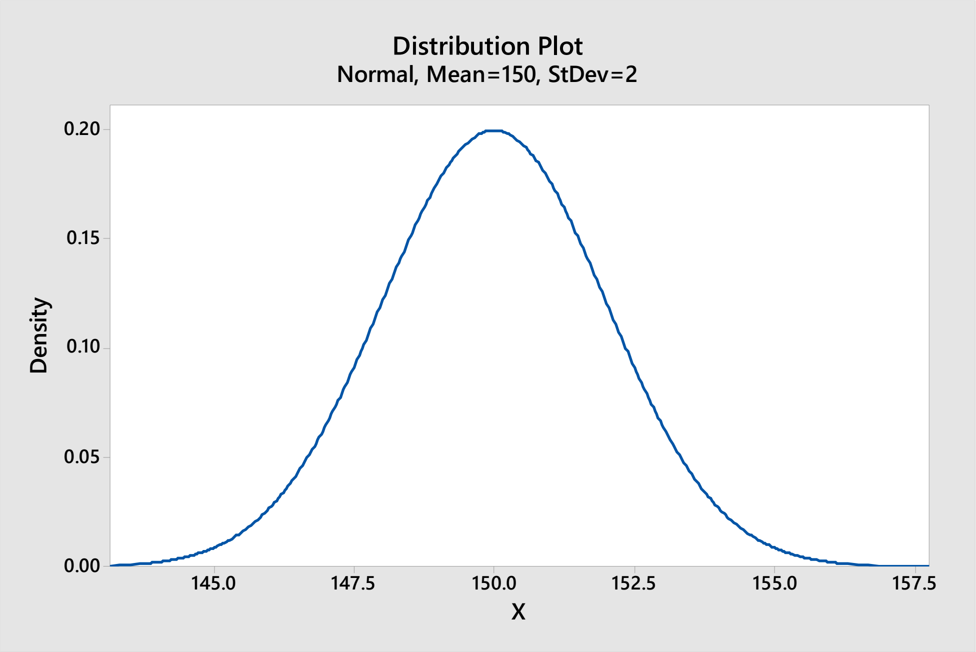
02 Process capability studies

03 Process capability and process performance

04 Short-term vs long-term capability and sigma shift

# Process Performance

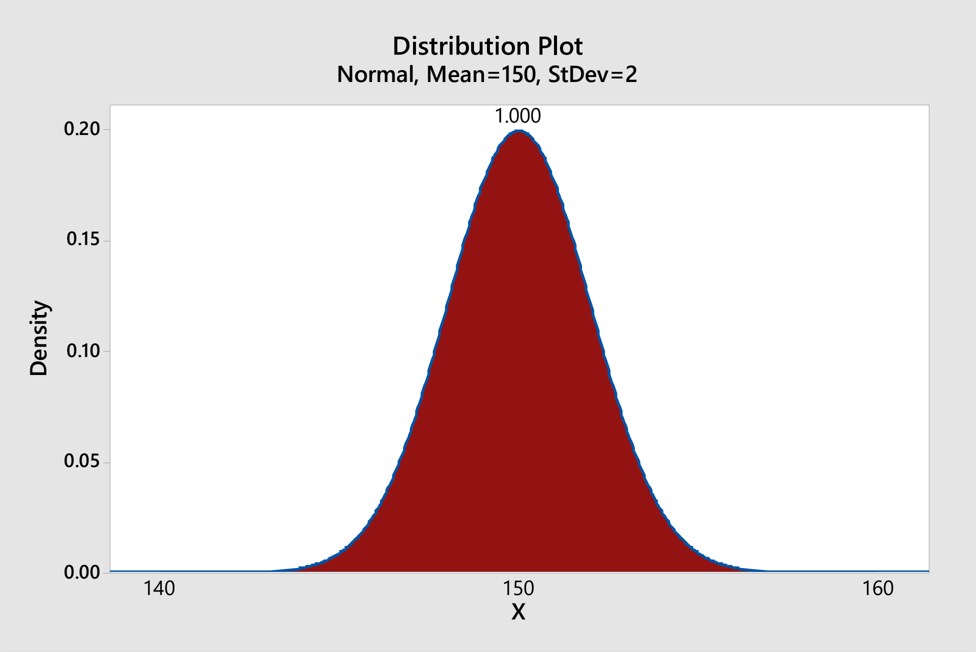
Natural process limits




**_Process Performance vs Specifications_**

# Process Performance

Specification: 140 to 160

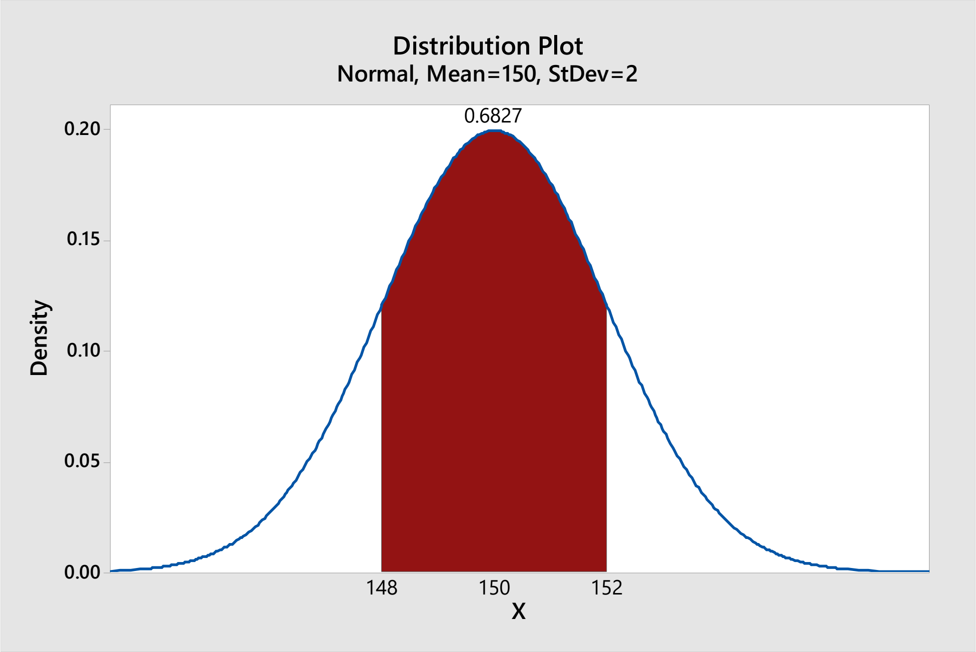



***Process Performance vs Specifications***

# Process Performance

Specification: 148 to 152

Rejections: 1-0.6827 = 0.3173

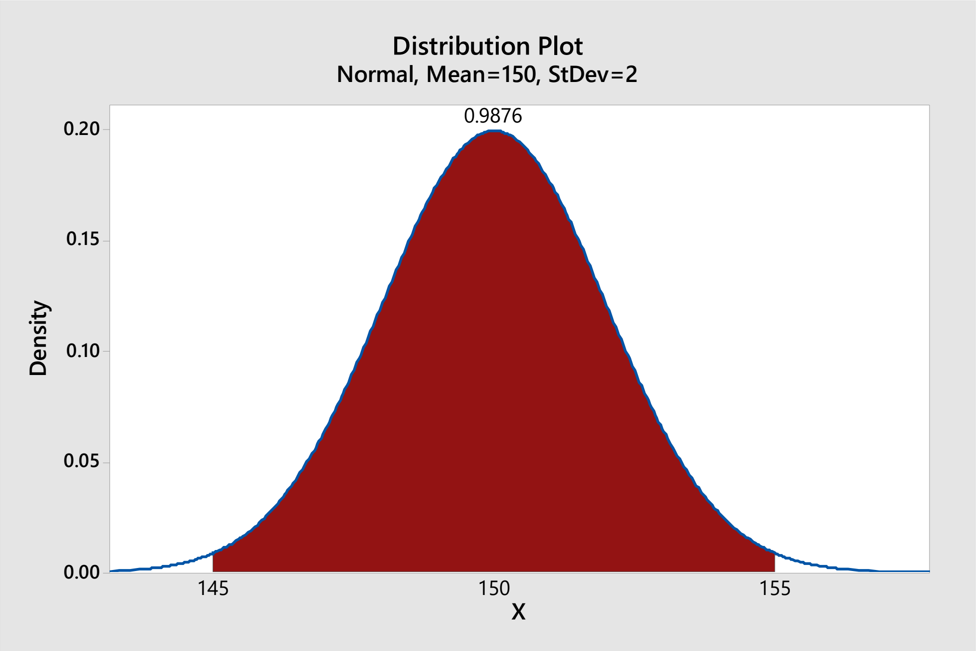



***Process Performance vs Specifications***

# Process Performance

Specification: 145 to 155

Rejections: 1-0.9876 = 0.0124

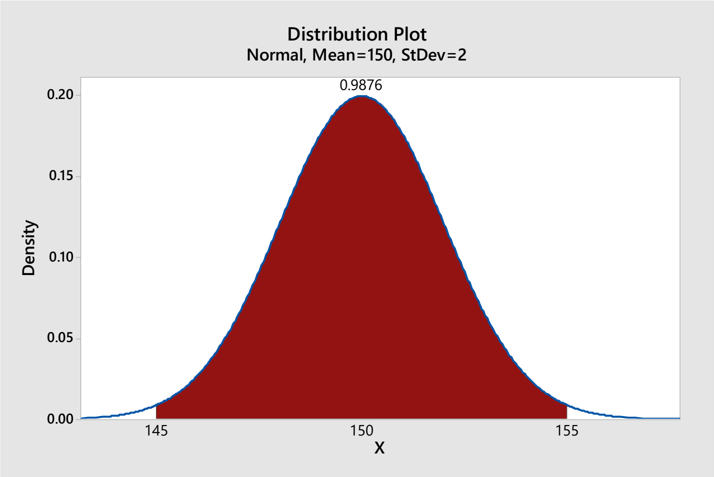



***Process Performance vs Specifications***

# Process Performance

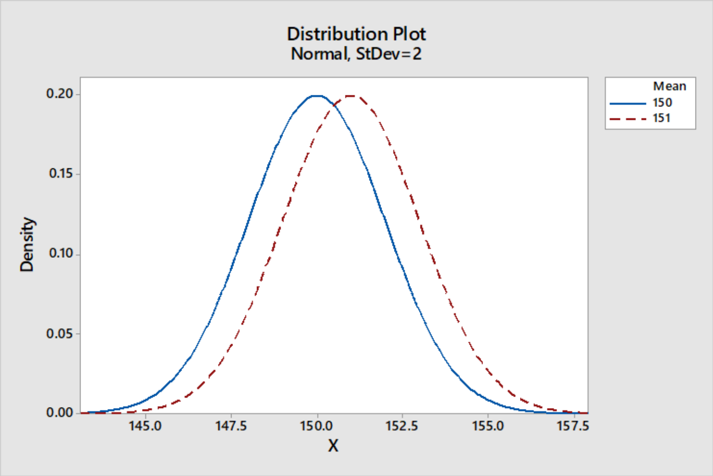
Specification: 145 to 155

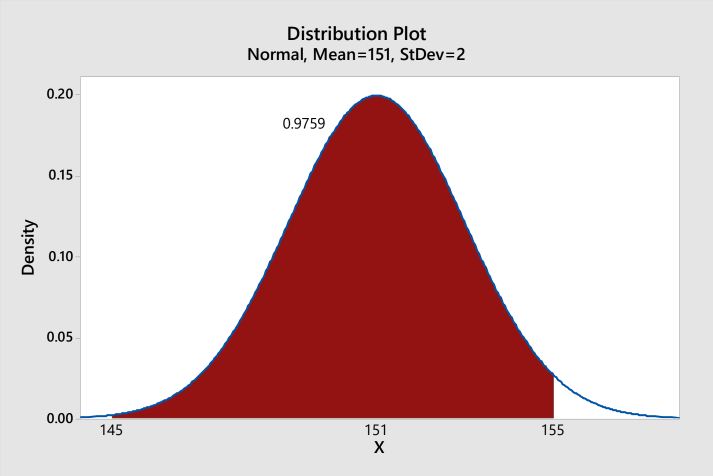
Rejections: 1-0.9875 = 0.0124




Specification: 148 to 152

Rejections: 1-0.9759 = 0.0241



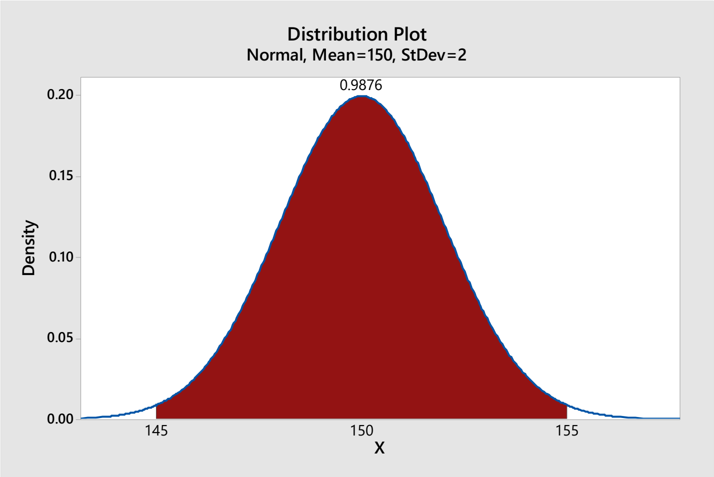

***Process Performance vs Specifications***

# Process Performance

Process sd = 2

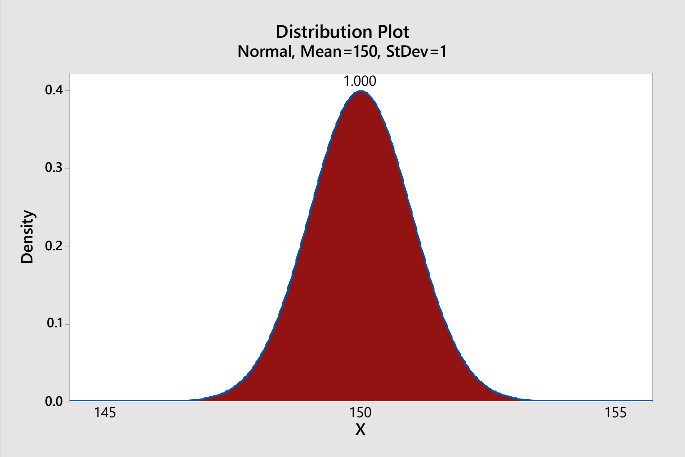
Specification: 145 to 155

Rejections: 1-0.9875 = 0.0124



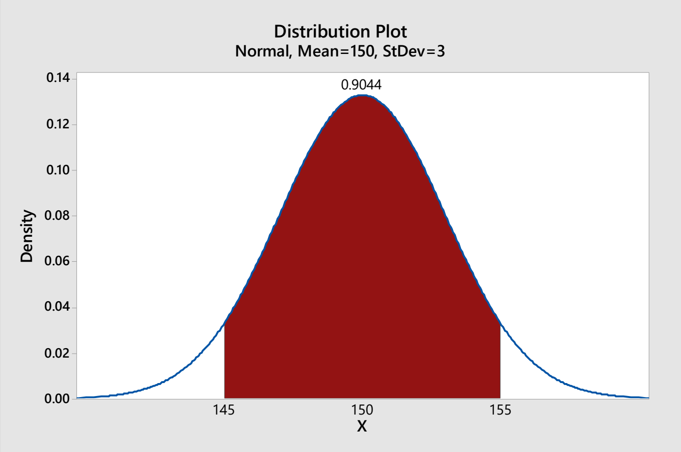

Process sd = 1

Rejections: 1-1 = almost zero



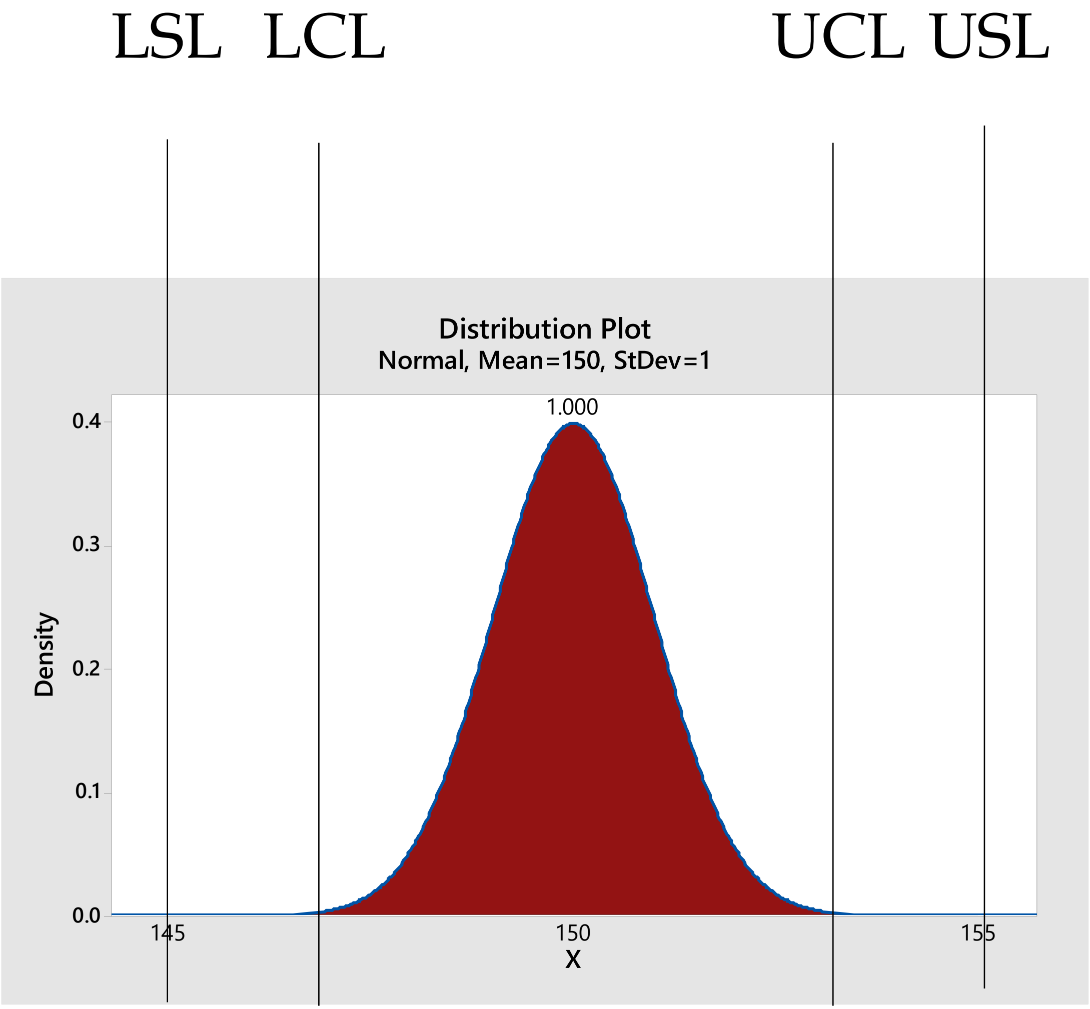

Process sd = 3

Rejections: 1-0.9044 = 0.0956



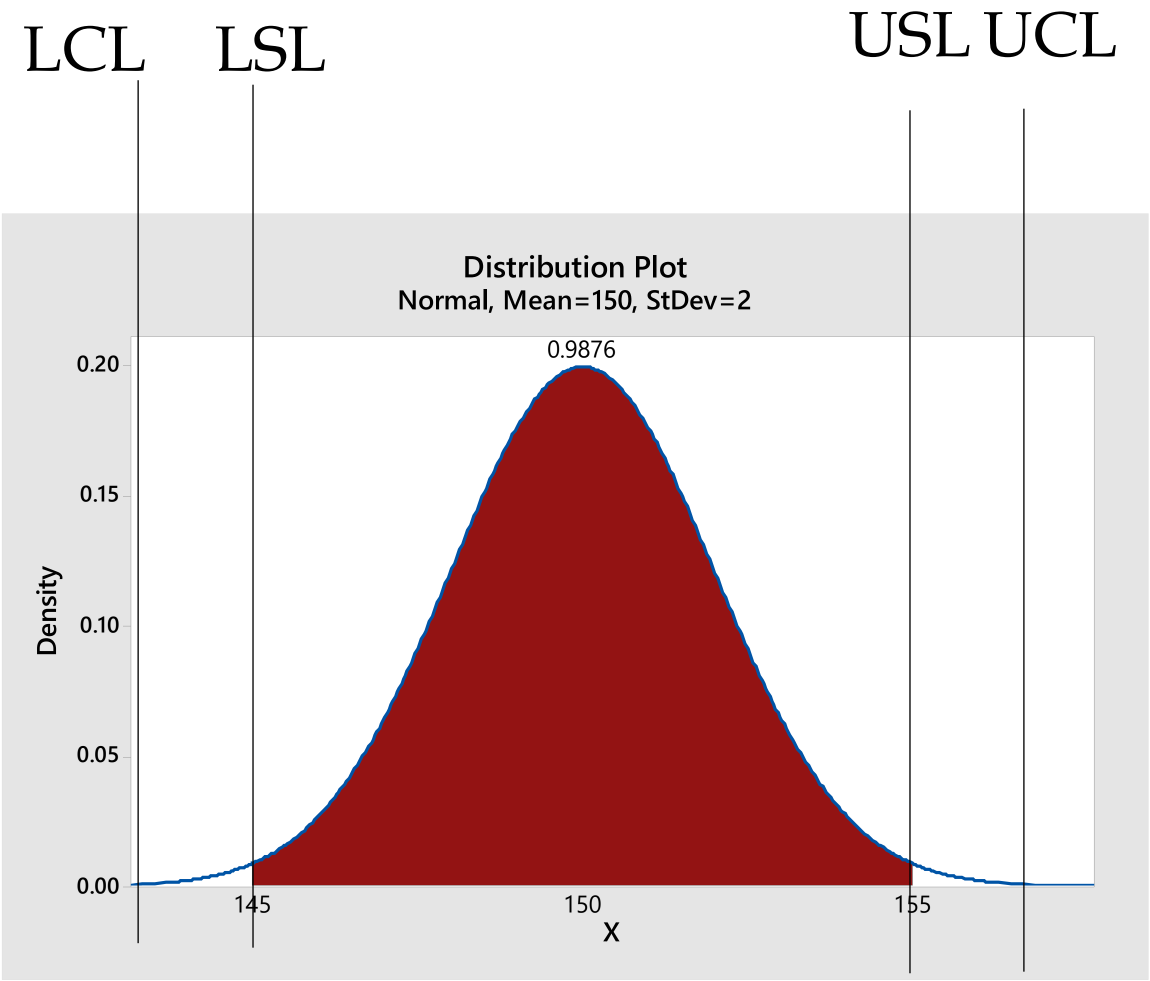

***Process Performance vs Specifications***

# Capable Process




***Process Performance vs Specifications***

# Is this process capable? .. No




***Process Performance vs Specifications***

# Process Performance Matrices

- ❖ Percent Defectives
- ❖ PPM
- ❖ DPMO
- ❖ DPU
- ❖ Rolled Throughput Yield

(already covered in 2E-1)

***Process Performance vs Specifications***

# 3F
# Process and Performance Capability

01 Process performance vs process specification

02 Process capability studies

03 Process capability and process performance

04 Short-term vs long-term capability and sigma shift

# Process Capability Studies

- Select the process
- Data Collection Plan
- Measurement System Analysis
- Gather data
- Confirm normality of data
- Confirm that the process is in control
- Estimate the process capability (Cp, Cpk)
- Continually improve process

*Process Capability Studies*

# 3F
# Process and Performance Capability

01 Process performance vs process specification

02 Process capability studies

03 Process capability and process performance

04 Short-term vs long-term capability and sigma shift

# Process Capability Indices

- Ratio of the spread between the process specifications to the spread of the process values, (6 process standard deviations) .

- Voice of customer / Voice of process > 1

*Cp, Cpk*
*Pp, Ppk*

# Process Capability Indices

- Voice of Customer:
  - LSL – Lower Specification Limit
  - USL - Upper Specification Limit

- Voice of Process:
  - LCL – Lower Control Limit
  - UCL - Upper Control Limit

***Cp, Cpk***
***Pp, Ppk***

# Cp

$$C_p = \frac{(\boldsymbol{U}SL - LSL)}{6 \times \sigma_{\boldsymbol{within}}}$$

***Cp, Cpk***
***Pp, Ppk***

# Cpk

$$C_{pL} = \frac{(\textbf{Process Mean} - LSL)}{\mathbf{3} \times \sigma_{\boldsymbol{within}}}$$

$$C_{pU} = \frac{(\boldsymbol{USL} - \textbf{Process Mea}\boldsymbol{n})}{\mathbf{3} \times \sigma_{\boldsymbol{within}}}$$

$$C_{\mathrm{pk}} = \mathrm{Min}\ (C_{\mathrm{pU},}\ C_{\mathrm{pL}}\ )$$

***Cp, Cpk***
***Pp, Ppk***

QG

## σ within

- Short term standard deviation
- It is an estimate of the variation within the subgroups.
- It should not be influenced by changes to process inputs, such as tool wear or different lots of material.

## σ overall

- Long term standard deviation

***Cp, Cpk***
***Pp, Ppk***

piston ring diameter are 74.0 mm ± 0.05 mm.

Every 5 rows represent a subgroup.

Total 25 subgroups

| Diameter |
|---|
| 74.030 |
| 74.002 |
| 74.019 |
| 73.992 |
| 74.008 |
| 73.995 |
| 73.992 |
| 74.001 |
| 74.011 |
| 74.004 |
| 73.988 |

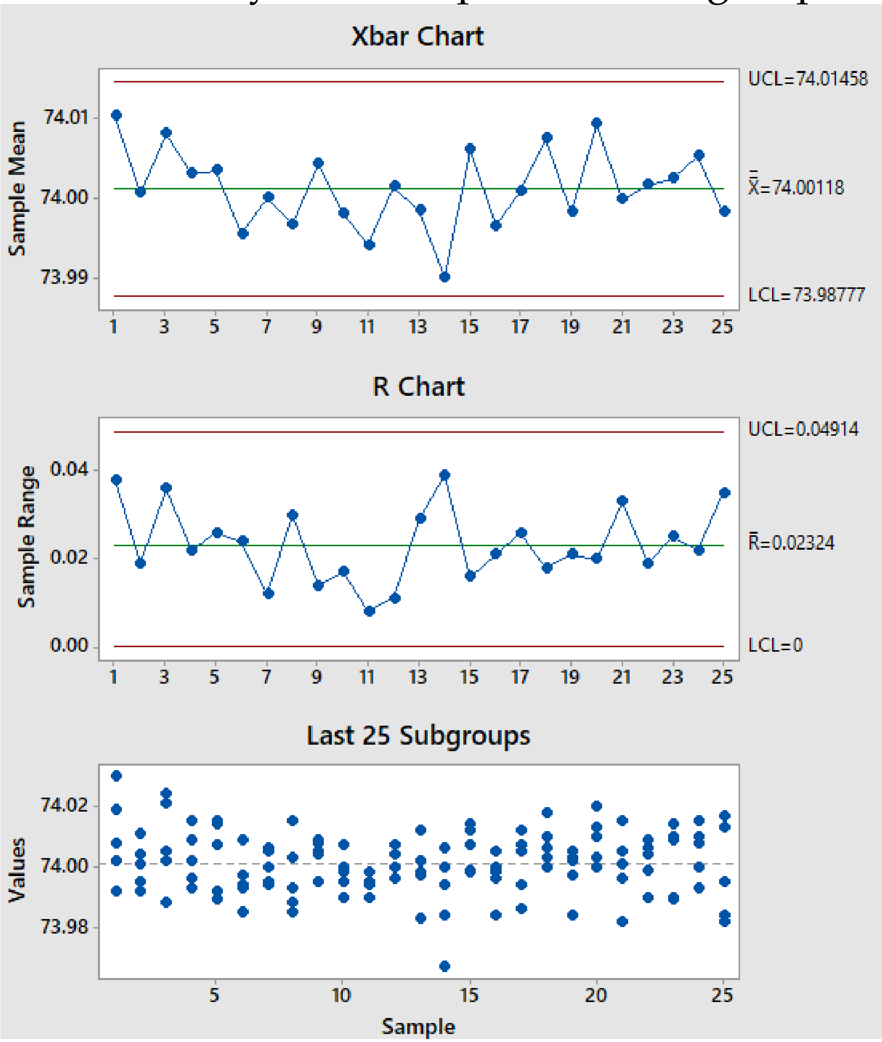
# Cp, Cpk

- Conditions to be met:
  - Sample to represent the population
  - Normal distribution of data *
  - The process must be in statistical control *
  - Sample size must be sufficient

## *Short term vs Long term SD*

piston ring diameter are 74.0 mm ± 0.05 mm.
Every 5 rows represent a subgroup.

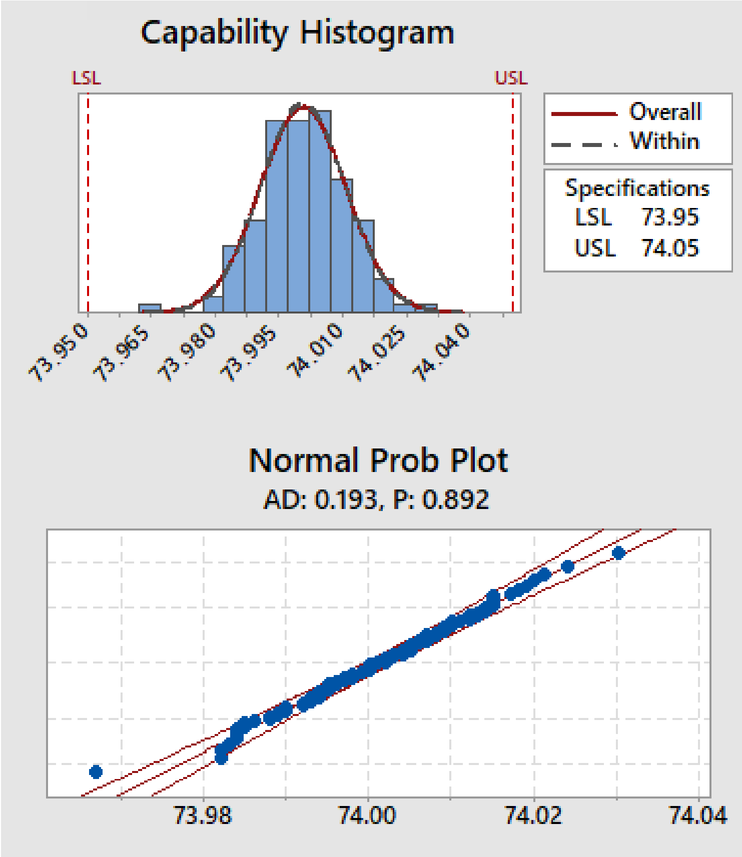
# Cp, Cpk




| PistonRingDiameter |
| --- |
| Diameter |
| 74.030 |
| 74.002 |
| 74.019 |
| 73.992 |
| 74.008 |
| 73.995 |
| 73.992 |
| 74.001 |
| 74.011 |
| 74.004 |
| 73.988 |

$$S = \frac{\bar{R}}{d_2}$$

**$d_2$ = 2.326**




**Stat > Quality Tools > Capability Sixpack > Normal …**

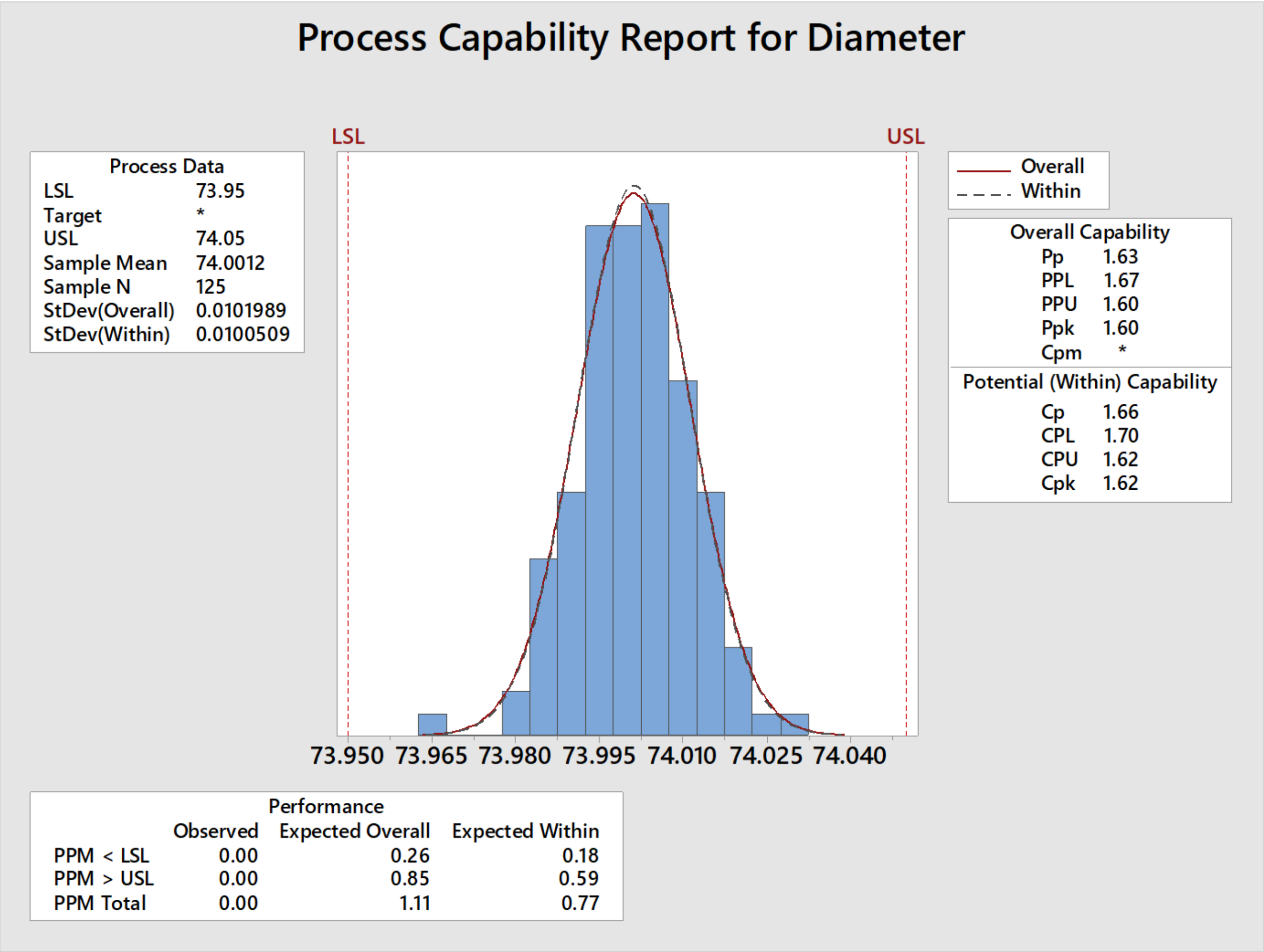
***Short term vs Long term SD***

Piston ring diameter are 74.0 mm ± 0.05 mm.

$$C_p = \frac{(USL - LSL)}{6 \times \sigma_{within}}$$

$$C_p = \frac{(74.05 - 73.95)}{6 \times 0.0100509}$$

$$C_p = \frac{(74.5 - 73.5)}{6 \times 0.0100509} = 1.658$$




**Stat > Quality Tools > Capability Analysis > Normal …**

# *Short term vs Long term SD*

Piston ring diameter are 74.0 mm ± 0.05 mm.

$$C_{pL} = \frac{(\textbf{Process Mean} - LSL)}{3 \times \sigma_{within}}$$

$$C_{pL} = \frac{(74.0012 - 73.95)}{3 \times 0.0100509} = 1.698$$

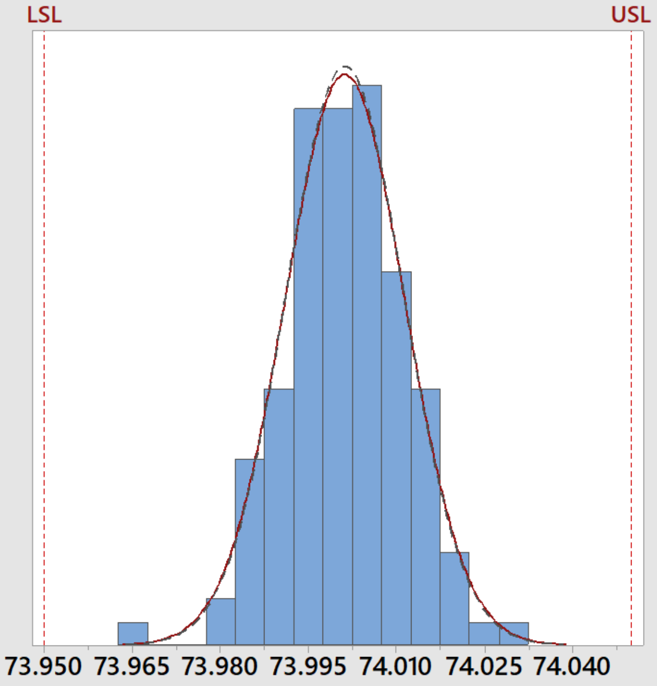
$$C_{pU} = \frac{(USL - \textbf{Process Mean})}{3 \times \sigma_{within}}$$

$$C_{pU} = \frac{(74.05 - 74.0012)}{3 \times 0.0100509} = 1.618$$

$$C_{pk} = \text{Min}\,(C_{pU}, C_{pL}) = 1.618$$

## Process Capability Report for Diameter

| Process Data | |
|---|---|
| LSL | 73.95 |
| Target | * |
| USL | 74.05 |
| Sample Mean | 74.0012 |
| Sample N | 125 |
| StDev(Overall) | 0.0101989 |
| StDev(Within) | 0.0100509 |




| Overall Capability | |
|---|---|
| Pp | 1.63 |
| PPL | 1.67 |
| PPU | 1.60 |
| Ppk | 1.60 |
| Cpm | * |

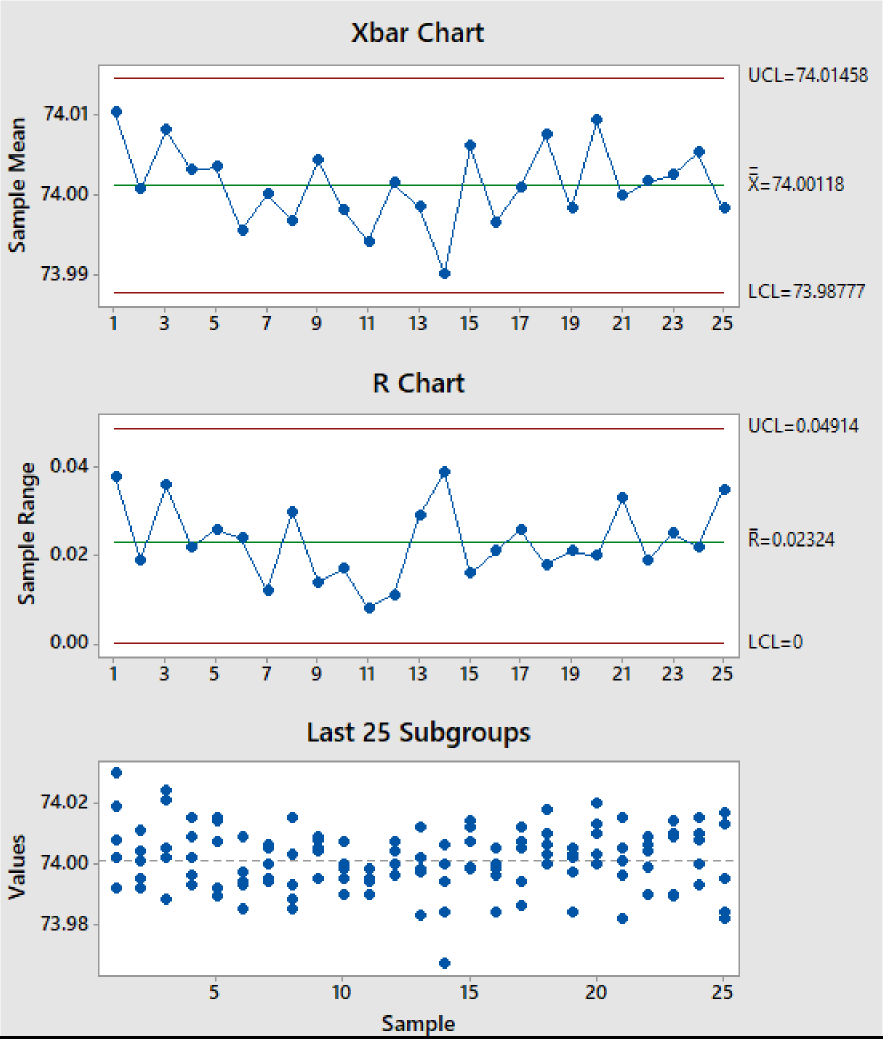
| Potential (Within) Capability | |
|---|---|
| Cp | 1.66 |
| CPL | 1.70 |
| CPU | 1.62 |
| Cpk | 1.62 |

| Performance | Observed | Expected Overall | Expected Within |
|---|---|---|---|
| PPM < LSL | 0.00 | 0.26 | 0.18 |
| PPM > USL | 0.00 | 0.85 | 0.59 |
| PPM Total | 0.00 | 1.11 | 0.77 |

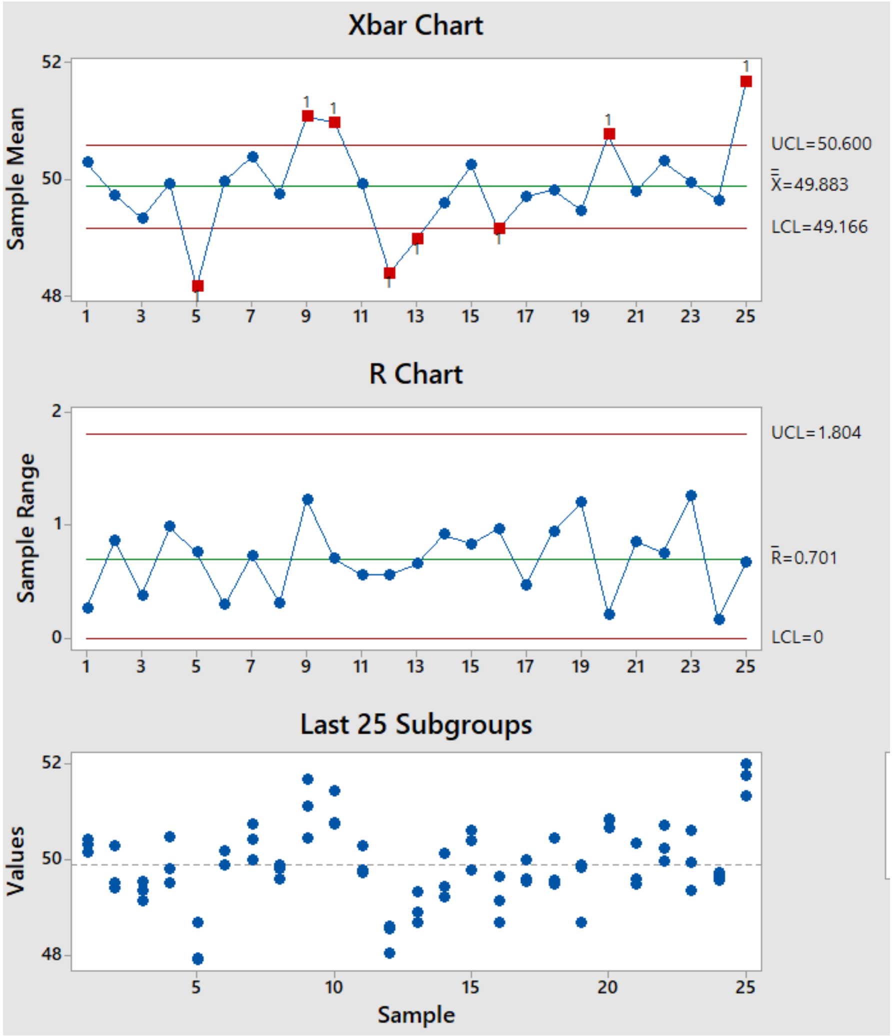
**Stat > Quality Tools > Capability Analysis > Normal ...**

# *Short term vs Long term SD*

piston ring diameter are 74.0 mm ± 0.05 mm.
Every 5 rows represent a subgroup.



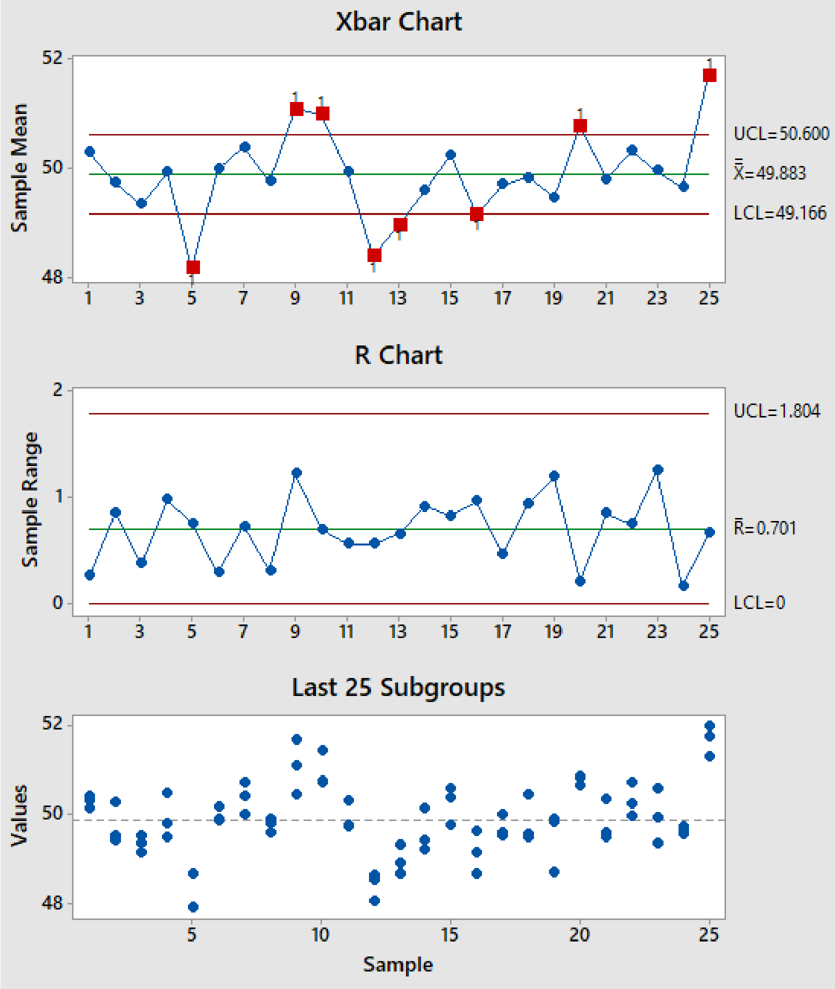

The film thickness must be 50 ± 3 microns
Subgroup size of 3.




# *Short term vs Long term SD*

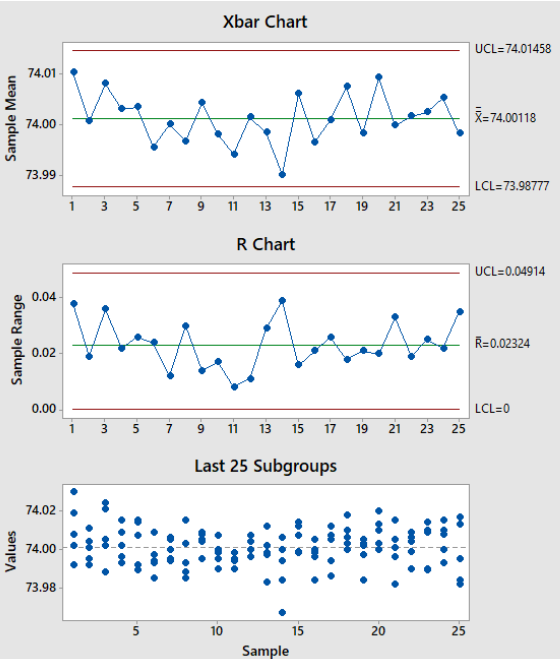
The film thickness must be 50 ± 3 microns
Subgroup size of 3.

# Cp, Cpk, Pp, Ppk




**FilmThickness**

| Coating | Roll |
|---|---|
| 50.3175 | 1 |
| 50.1432 | 1 |
| 50.4081 | 1 |
| 49.5077 | 2 |
| 50.2715 | 2 |
| 49.4091 | 2 |
| 49.3587 | 3 |
| 49.1436 | 3 |
| 49.5229 | 3 |




$$S = \frac{\bar{R}}{d_2}$$

**For Cp, Cpk use Short-term sd**

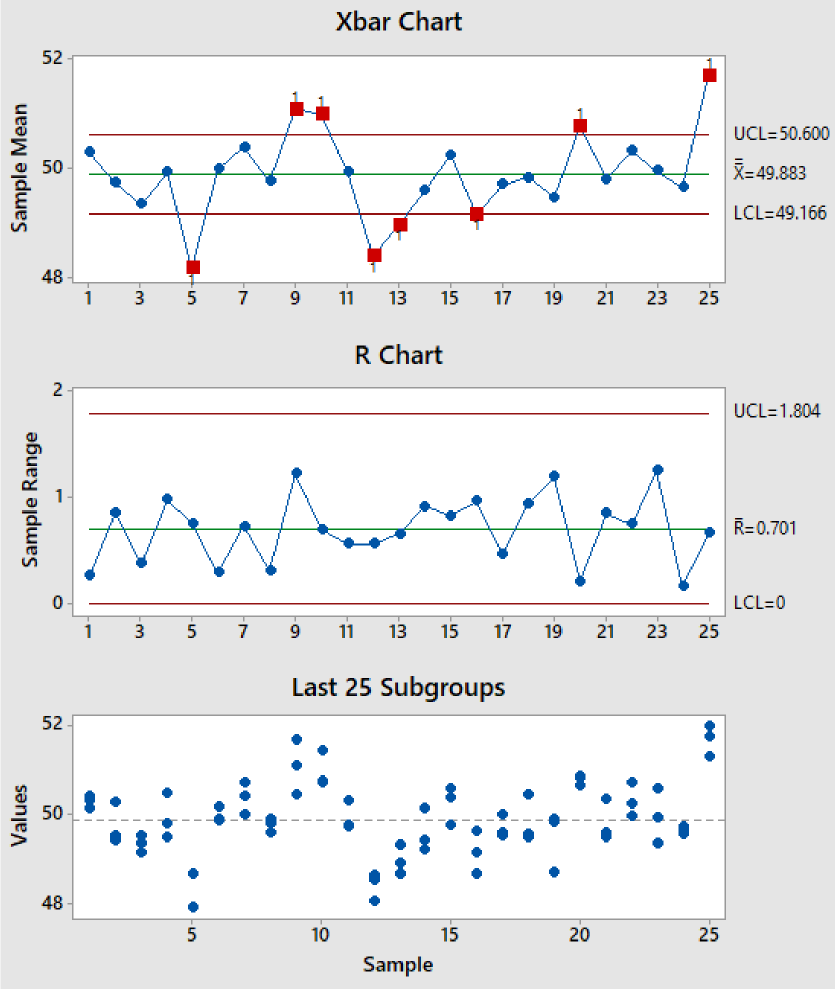
$$s = \sqrt{\frac{(x - \bar{x})^2}{n - 1}}$$

**For Pp, Ppk use Long-term sd**

# *Short term vs Long term SD*

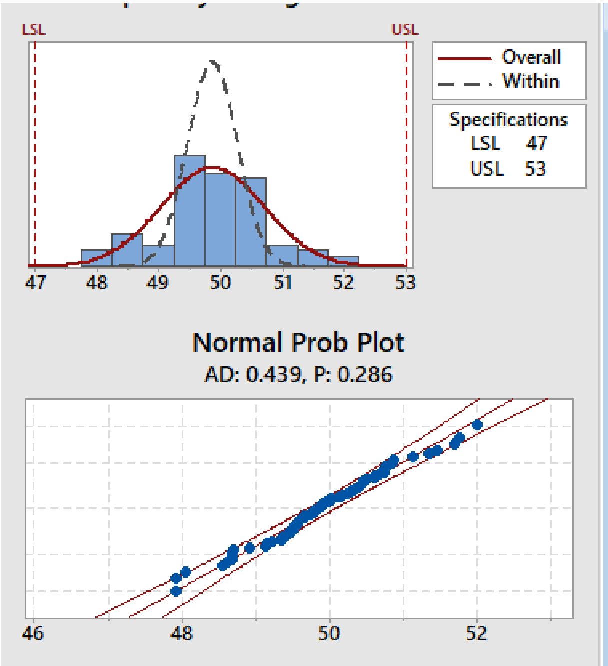
The film thickness must be 50 ± 3 microns
Subgroup size of 3.

# Cp, Cpk, Pp, Ppk




**FilmThickness**

| Coating | Roll |
|---|---|
| 50.3175 | 1 |
| 50.1432 | 1 |
| 50.4081 | 1 |
| 49.5077 | 2 |
| 50.2715 | 2 |
| 49.4091 | 2 |
| 49.3587 | 3 |
| 49.1436 | 3 |
| 49.5229 | 3 |




$$S = \frac{\bar{R}}{d_2}$$

**For Cp, Cpk use Short-term sd**

$$s = \sqrt{\frac{(x - \bar{x})^2}{n-1}}$$

**For Pp, Ppk use Long-term sd**

***Short term vs Long term SD***

$$C_p = \frac{(USL - LSL)}{6 \times \sigma_{within}}$$

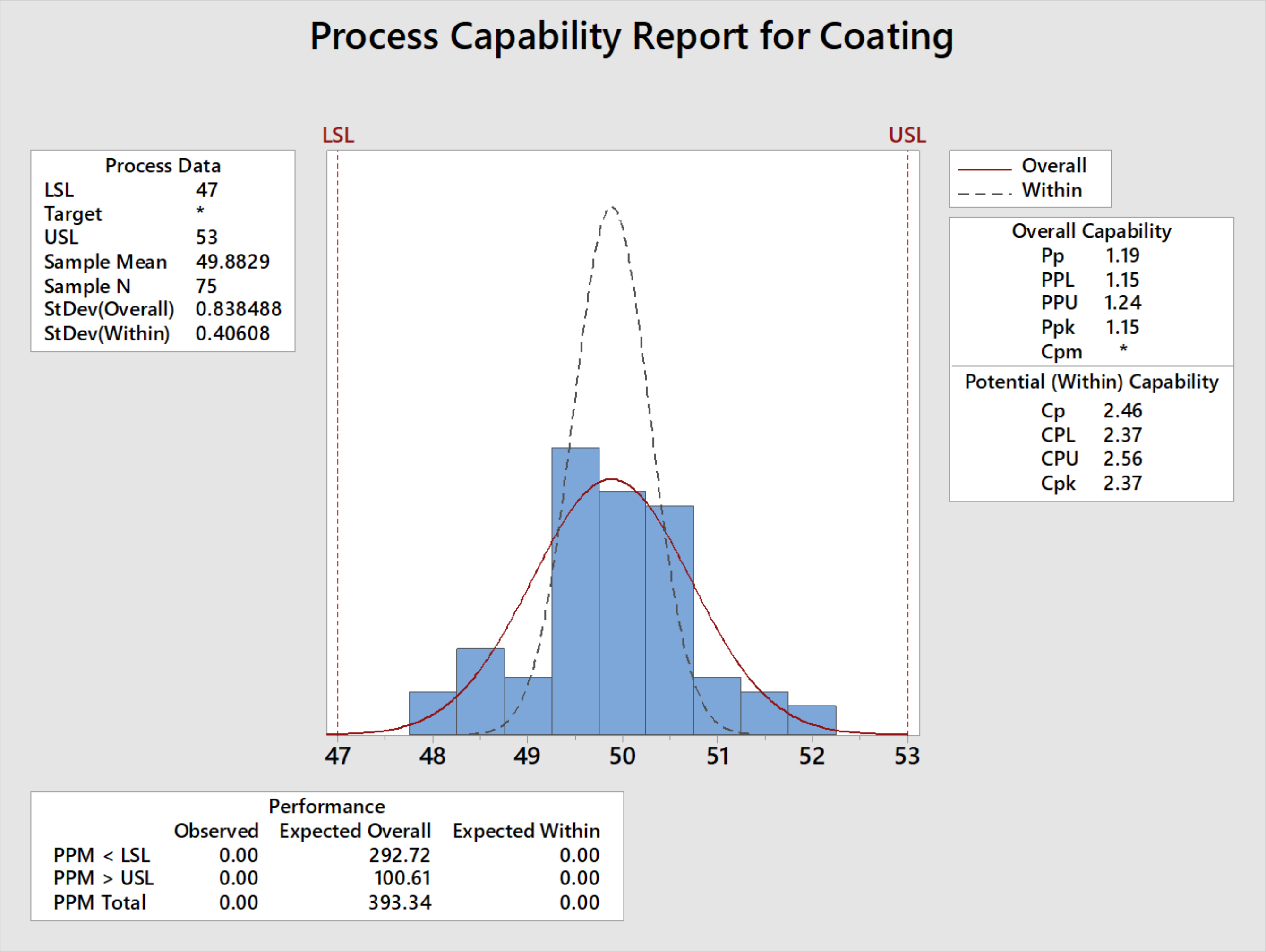
$$C_{pL} = \frac{(\textbf{Process Mean} - LSL)}{3 \times \sigma_{within}}$$

$$C_{pU} = \frac{(USL\ -\ \textbf{Process Mean})}{3 \times \sigma_{within}}$$

$$C_{pk} = \mathrm{Min}\ (C_{pU,}\ C_{pL}\ )$$

Process capability

Potential capability




**Stat > Quality Tools > Capability Analysis > Normal …**

$$P_p = \frac{(USL - LSL)}{6 \times \sigma_{overall}}$$

$$P_{pL} = \frac{(\textbf{Process Mean} - LSL)}{3 \times \sigma_{overall}}$$

$$P_{pU} = \frac{(USL\ -\ \textbf{Process Mean})}{3 \times \sigma_{overall}}$$

$$P_{pk} = \mathrm{Min}\ (P_{pU,}\ P_{pL}\ )$$

Process Performance

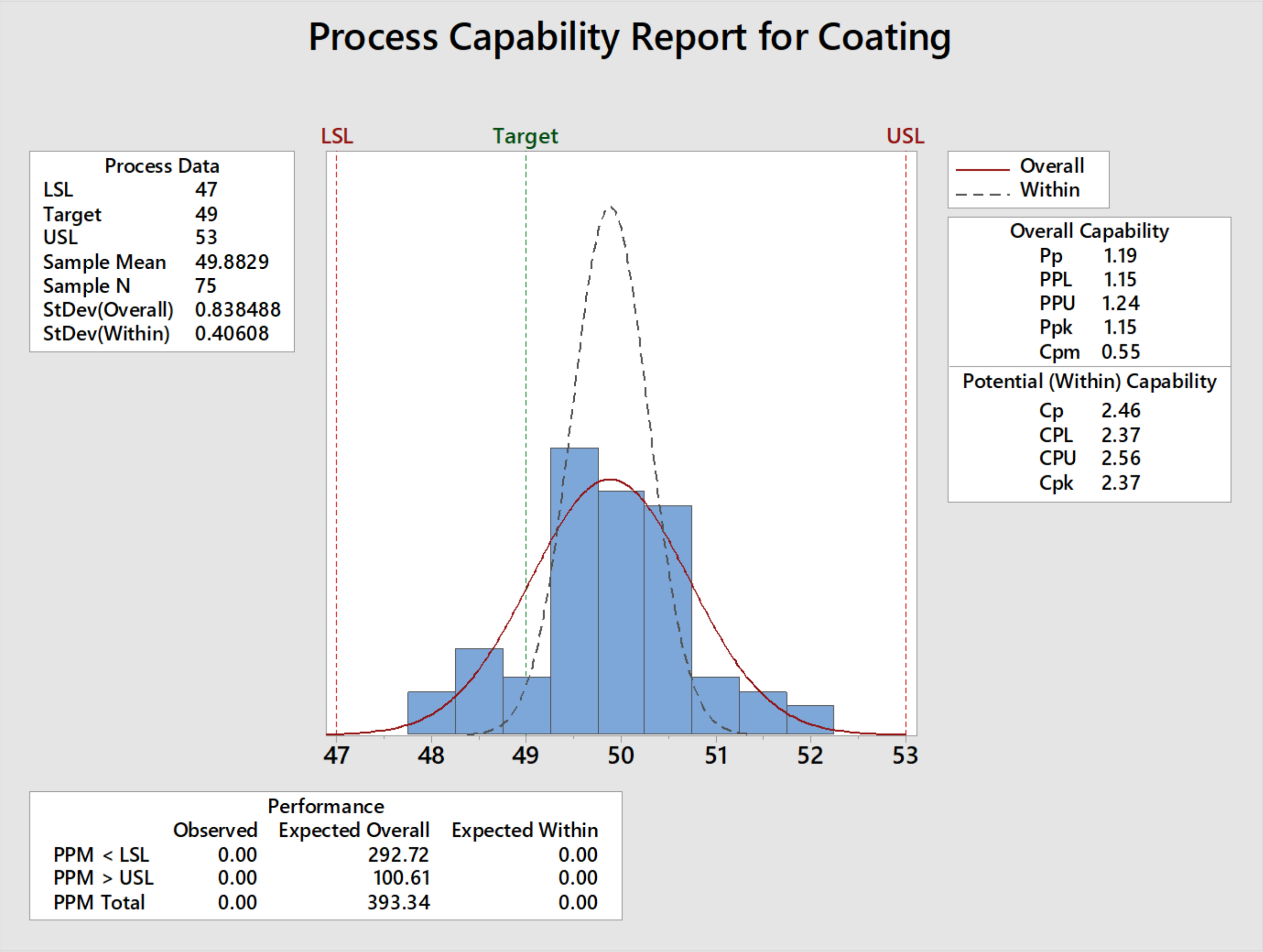
Overall capability

# *Short term vs Long term SD*

# Cpm

$$P_p = \frac{(USL - LSL)}{6 \times \sigma_{overall}}$$

$$C_{pm} = \frac{(USL - LSL)}{6\sqrt{\sigma^2 + (\mu - T)^2}}$$




*Cpm*

# 3F
# Process and Performance Capability

01 Process performance vs process specification

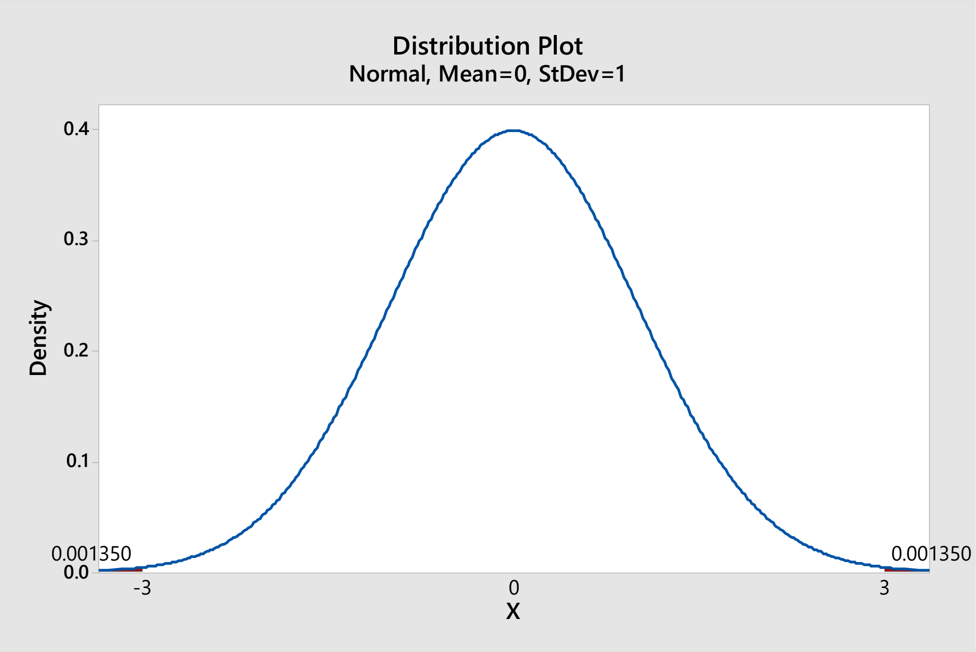
02 Process capability studies

03 Process capability and process performance

04 Short-term vs long-term capability and sigma shift

# 6 Sigma = 3.4 DPMO … Why?

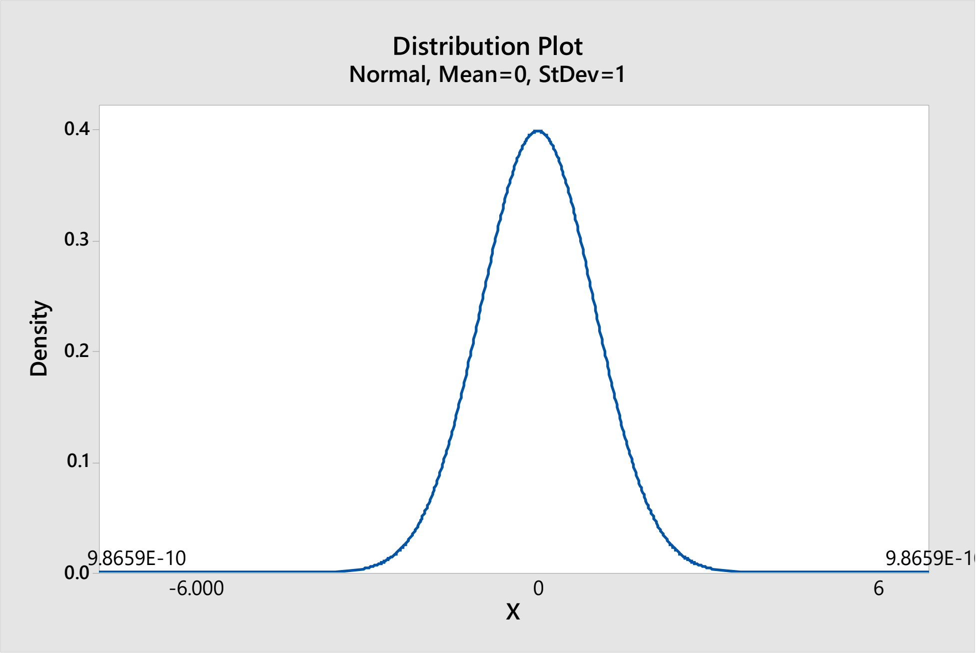
- SPC required processes to be controlled within plus minus 3 sigma.
- This results in 2.7 defects per thousand or 2700 DPMO. (with no long-term shift)




***Long-term Sigma Shift***

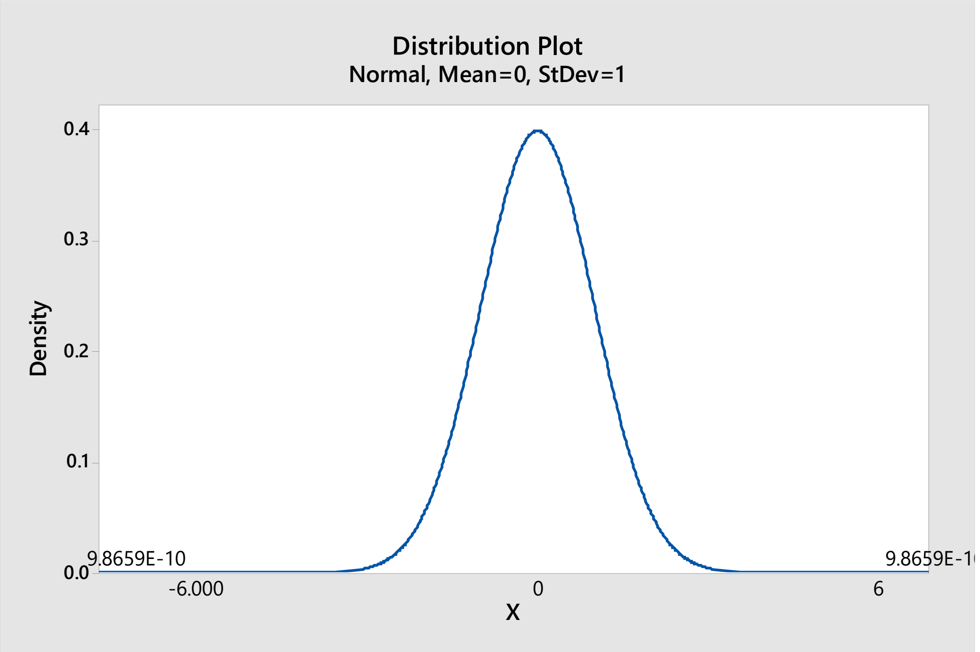
# 6 Sigma = 3.4 DPMO … Why?

- Instead of plus minus 3 sigma, if the process is controlled in plus minus 6 sigma, what will be the DPMO?



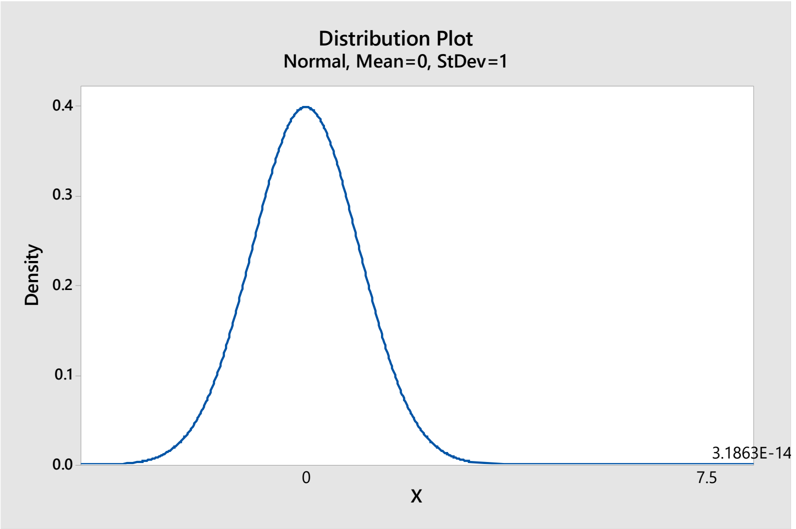

***Long-term Sigma Shift***

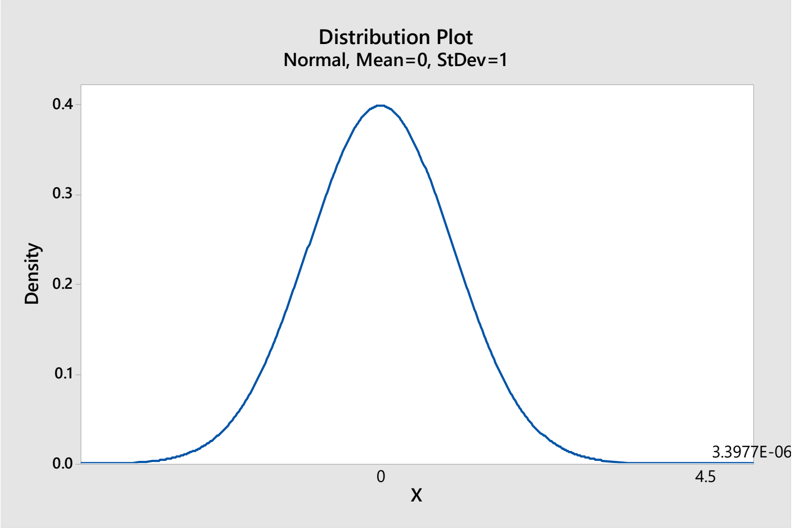
- $2 \times 9.86 \times 10^{-10}$
- $2 \times 9.86 \times 10^{-4}$ per million
- 0.001972 DPMO
- But this should have been 3.4 DPMO .. Why it is so low?
- The difference is because of 1.5 sigma shift allowed on long-term




# *Long-term Sigma Shift*

- After 1.5 sigma shift, one side will be in 4.5 sigma and other side in 7.5 sigma.
- Rejections will be sum of
  - 3.397 DPMO
  - 3.186 x $10^{-8}$ DPMO
- Hence 3.4 DPMO




# *Long-term Sigma Shift*