

# Contents

1. Linear Algebra .....	5
1.1. Vectors .....	5
1.2. Real Coordinate Spaces .....	5
1.3. Vector Operations .....	5
1.3.1. Vector Addition .....	5
1.3.2. Vector Subtraction .....	6
1.3.3. Scalar Multiplication .....	7
1.3.4. Unit Vector .....	8
1.3.5. Parametric Representation of line .....	9
1.3.6. Vector Spaces .....	12
1.4. Matrices .....	13
1.4.1. Matrix-Vector Products .....	13
1.4.2. Null Space .....	14
1.4.3. Column Space .....	15
1.4.4. Dimension of a Subspace .....	16
1.4.5. Nullity .....	16
1.4.6. Rank .....	16
1.4.7. Matrix Representation of Systems of Equations .....	16
1.4.8. Matrix Multiplication .....	17
1.5. Eigenvectors .....	18
1.5.1. Transformation .....	18
1.6. Eigenvalues .....	19
2. Calculus .....	19
2.1. Fundamental Theorem of Calculus .....	19
2.2. Derivative .....	21
2.2.1. Power Rule .....	22
2.2.2. Product Rule .....	22
2.2.3. Quotient Rule .....	23
2.2.4. Chain Rule .....	23
2.2.5. Others .....	24
2.3. Integral .....	25
2.3.1. Power Rule .....	26
2.3.2. Sum & Difference Rule .....	26
2.3.3. Constant Multiple Rule .....	26
2.3.4. Integration by Substitution .....	26
2.3.5. Integration by Parts .....	26
2.3.6. Trigonometric Integrals .....	26
2.3.7. Exponential and Logarithmic Integrals .....	26
2.3.8. Definite Integrals of Even and Odd Functions .....	26
2.3.9. Improper Integrals .....	26
3. Causes of variation .....	26
3.1. Common .....	26
3.2. Special .....	26
4. Probability Theory .....	26
4.1. Probability Axioms .....	26
4.1.1. Non-Negativity .....	26
4.1.2. Normalization .....	26

4.1.3. Additivity .....	27
4.2. Rules .....	27
4.2.1. Complement Rule .....	27
4.2.2. Multiplication Rule .....	27
4.2.3. Addition Rule .....	27
4.2.4. Conditional Probability .....	28
4.2.5. Law of Total Probability .....	28
4.2.6. Law of Large Numbers .....	29
4.2.7. Central Limit Theorem .....	29
4.3. Bayes Theorem .....	29
5. Descriptive Statistics .....	29
5.1. Central Tendency .....	29
5.1.1. Mean .....	29
5.1.2. Median .....	29
5.1.3. Mode .....	30
5.2. Dispersion .....	30
5.2.1. Range .....	30
5.2.2. Variance .....	30
5.2.3. Standard deviation .....	31
5.2.4. Interquartile Range (IQR) .....	31
6. Probability Distributions .....	32
6.1. Gaussian (Normal) distribution .....	32
6.2. t-Distribution .....	32
6.3. Binomial distribution .....	33
6.4. Poisson distribution .....	33
6.5. Exponential distribution .....	33
7. Functions .....	34
7.1. PDF (Probability Density Function) .....	34
7.2. PMF (Probability Mass Function) .....	34
7.3. CDF (Cumulative Distribution Function) .....	34
7.4. PPF (Percent-Point Function) .....	35
7.5. SF (Survival Function) .....	35
8. Error Metrics .....	36
8.1. MAE (Mean Absolute Error) .....	36
8.2. MSE (Mean Squared Error) .....	36
8.3. RMSE (Root Mean Squared Error) .....	36
8.4. MAPE (Mean Absolute Percentage Error) .....	36
8.5. R-squared .....	36
8.6. Adj R-squared .....	37
8.7. MSLE (Mean Squared Logarithmic Error) .....	38
8.8. Cross-Entropy Loss (Log Loss) .....	38
9. Hypothesis Testing .....	38
9.1. Hypotheses .....	38
9.1.1. Null ( $H_0$ ) .....	38
9.1.2. Alternative ( $H_1$ or $H_a$ ) .....	38
9.2. Error Types .....	38
9.2.1. Type I ( $\alpha$ ) .....	38
9.2.2. Type II ( $\beta$ ) .....	38
9.3. t-Tests .....	38

9.3.1. One-sample .....	38
9.3.2. Independent .....	40
9.3.3. Paired .....	40
9.4. Chi-square tests .....	41
9.4.1. Goodness of Fit Test .....	41
9.4.2. Test of independence .....	41
9.5. ANOVA (Analysis of Variance) .....	41
9.5.1. One-way .....	41
9.5.2. Two-way .....	42
10. Regression Analysis: .....	42
10.1. Simple linear regression .....	42
10.2. Multiple regression .....	44
10.3. Logistic Regression .....	44
10.4. Model diagnostics .....	45
10.4.1. p-Values .....	45
10.4.2. F-Statistic .....	45
10.4.3. Confidence Intervals (CI) .....	46
11. Design of Experiments (DOE): .....	46
12. Control Charts: .....	46
12.1. P-charts (Proportion) .....	46
12.2. NP-charts (Number Proportion) .....	47
12.3. C-charts (Count) .....	47
12.4. U-charts (Unit) .....	48
12.5. $\bar{X}$ -chart .....	48
12.6. R-chart .....	48
13. Process Capability Analysis .....	48
13.1. $C_p$ (Process Capability Index) .....	48
13.2. $C_{pk}$ (Process Capability Index with Centering) .....	49
13.3. $C_{pm}$ (Taguchi Capability Index) .....	51
13.4. $P_p$ (Process Performance Index) .....	51
13.5. $P_{pk}$ (Process Performance Index with Centering) .....	52
14. Correlation .....	53
14.1. Pearson .....	53
14.2. Spearman's Rank .....	53
15. Non-Parametric Statistics .....	54
15.1. Mann-Whitney U .....	54
15.2. Wilcoxon Signed-Rank .....	54
15.3. Kolmogorov-Smirnov .....	54
15.4. Kruskal-Wallis .....	54
16. Time Series .....	54
16.1. SMA (Simple Moving Averages) .....	54
16.2. WMA (Weighted Moving Average) .....	55
16.3. Exponential Smoothing .....	56
16.4. Seasonal Decomposition .....	57
16.5. ARMA (AutoRegressive Moving Average) .....	58
16.6. ARIMA (AutoRegressive Integrated Moving Average) .....	58
17. Inventory Management .....	58
17.1. Newsvendor .....	58
17.2. ABC Analysis .....	60

17.3. Fill Rate .....	61
17.4. (OCT) Order Cycle Time .....	61
17.5. ROP (Reorder Point) .....	61
17.6. XYZ Analysis .....	62
17.7. EOQ (Economic Order Quantity) .....	63
17.7.1. Perfect Order Rate .....	63
17.8. Safety Stock .....	63
18. Queuing Theory .....	64
18.1. M/M/1 .....	64
19. Network Optimization .....	68
19.1. Shortest Path .....	68
19.2. Maximum Flow .....	68
19.3. Network Flow Optimization .....	68
19.4. Ford-Fulkerson .....	73
20. Optimization .....	76
20.1. LP (Linear Programming) .....	76
20.2. IP (Integer Programming) .....	78
20.3. Gradient Descent .....	80
20.4. Monte Carlo .....	81

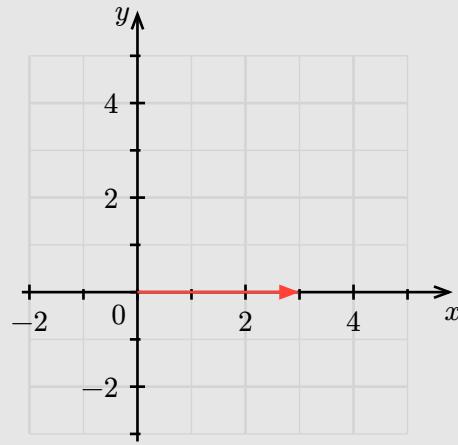
# 1. Linear Algebra

## 1.1. Vectors

Vector = Magnitude + Direction

A car is moving:

$\underbrace{\begin{array}{c} 3 \text{ MPH} \\ \text{Magnitude} \\ (\text{Speed} \rightarrow \text{Scalar}) \end{array}}_{\text{Velocity (Vector)}}$   $\underbrace{\text{East}}_{\text{Direction}}$



$$\vec{v} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

## 1.2. Real Coordinate Spaces

N-dimensional Real Coordinate Space

$$\mathbb{R}^n$$

$$\vec{x} \in \mathbb{R}^n$$

All possible real-valued n-tuples

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

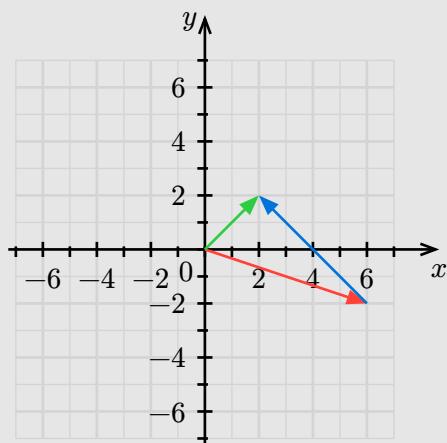
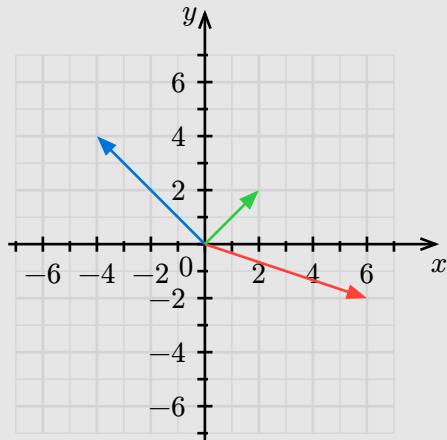
## 1.3. Vector Operations

### 1.3.1. Vector Addition

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{bmatrix}$$

$$\vec{a} = \begin{bmatrix} 6 \\ -2 \end{bmatrix} \quad \vec{b} = \begin{bmatrix} -4 \\ 4 \end{bmatrix}$$

$$\vec{a} + \vec{b} = \begin{bmatrix} 6 + -4 \\ -2 + 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

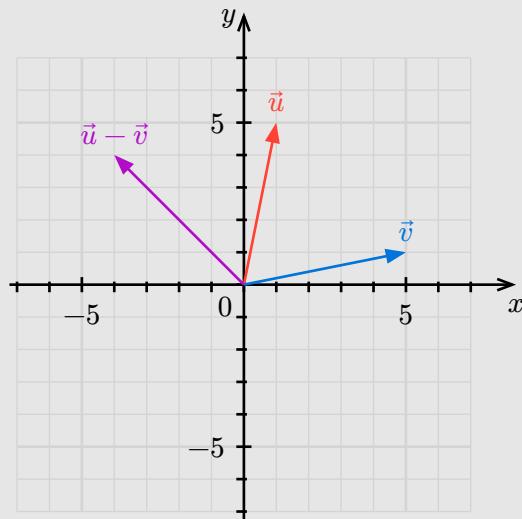


### 1.3.2. Vector Subtraction

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \\ x_3 - y_3 \end{bmatrix}$$

$$\vec{u} = \begin{bmatrix} 1 \\ 5 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

$$\vec{u} - \vec{v} = \begin{bmatrix} 1 - 5 \\ 5 - 1 \end{bmatrix} = \begin{bmatrix} -4 \\ 4 \end{bmatrix}$$

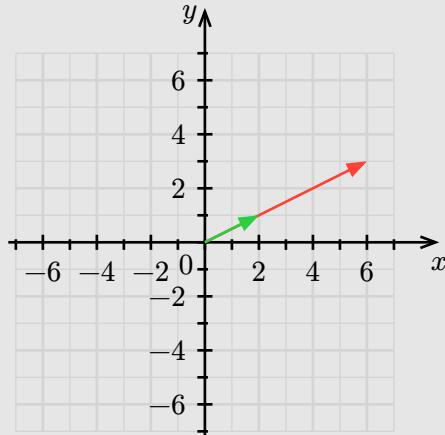


### 1.3.3. Scalar Multiplication

$$c \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} c \times x_1 \\ c \times x_2 \\ c \times x_3 \end{bmatrix}$$

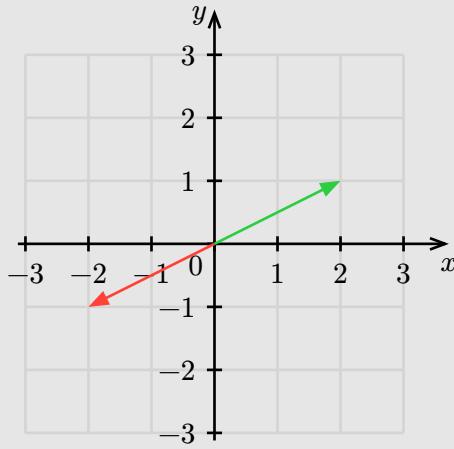
$$\vec{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$3\vec{a} = 3 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \cdot 2 \\ 3 \cdot 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$$



$$\vec{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$-1\vec{a} = -1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \cdot 2 \\ -1 \cdot 1 \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$$



#### 1.3.4. Unit Vector

A vector that has a magnitude (or length) of exactly 1

For a vector  $\vec{v}$  in  $n$ -dimensional space, a unit vector  $\hat{v}$  is defined as:

$$\hat{v} = \frac{\vec{v}}{\|\vec{v}\|}$$

Where:

- $\|\vec{v}\|$  is the **magnitude** (or **norm**) of the vector  $\vec{v}$ , computed as:

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$$

#### Key Properties

- **Magnitude:**

$$\|\hat{v}\| = 1$$

- **Direction:** A unit vector points in the same direction as the original vector  $\vec{v}$

Finding unit vector (vector of magnitude 1) with given direction

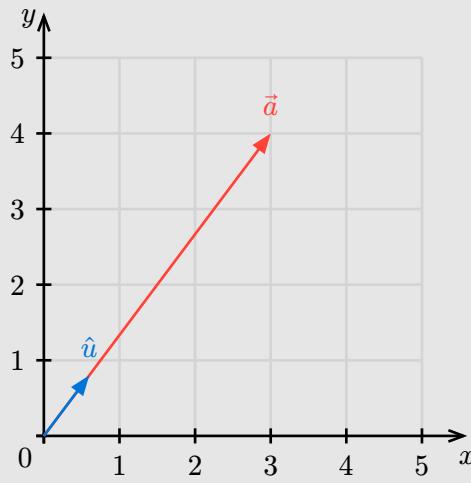
$$\vec{a} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

Magnitude

$$\|\vec{a}\| = \sqrt{3^2 + 4^2} = \sqrt{25} = 5$$

$$\hat{u} = \left( \frac{3}{\|\vec{a}\|}, \frac{4}{\|\vec{a}\|} \right) = \left( \frac{3}{5}, \frac{4}{5} \right)$$

$$\|\hat{u}\| = \sqrt{\left(\frac{3}{5}\right)^2 + \left(\frac{4}{5}\right)^2} = \sqrt{\frac{9}{25} + \frac{16}{25}} = \sqrt{\frac{25}{25}} = \sqrt{1} = 1$$



### 1.3.5. Parametric Representation of line

Set  $L$  of all points (i.e., line) equal to the set of all vectors  $\vec{x}$  plus some scalar  $t$  times the vector  $\vec{v}$  such that  $t$  can be any real number ( $\mathbb{R}$ )

$$L = \{\vec{x} + t\vec{v} \mid t \in \mathbb{R}\}$$

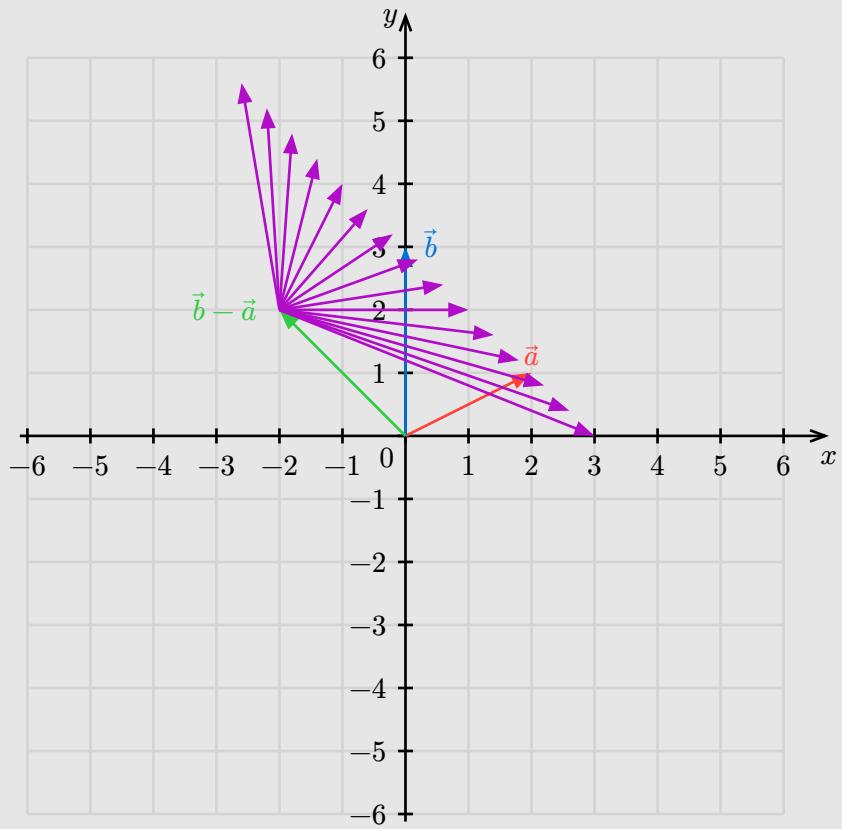
$$\vec{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\vec{b} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$t = 1$$

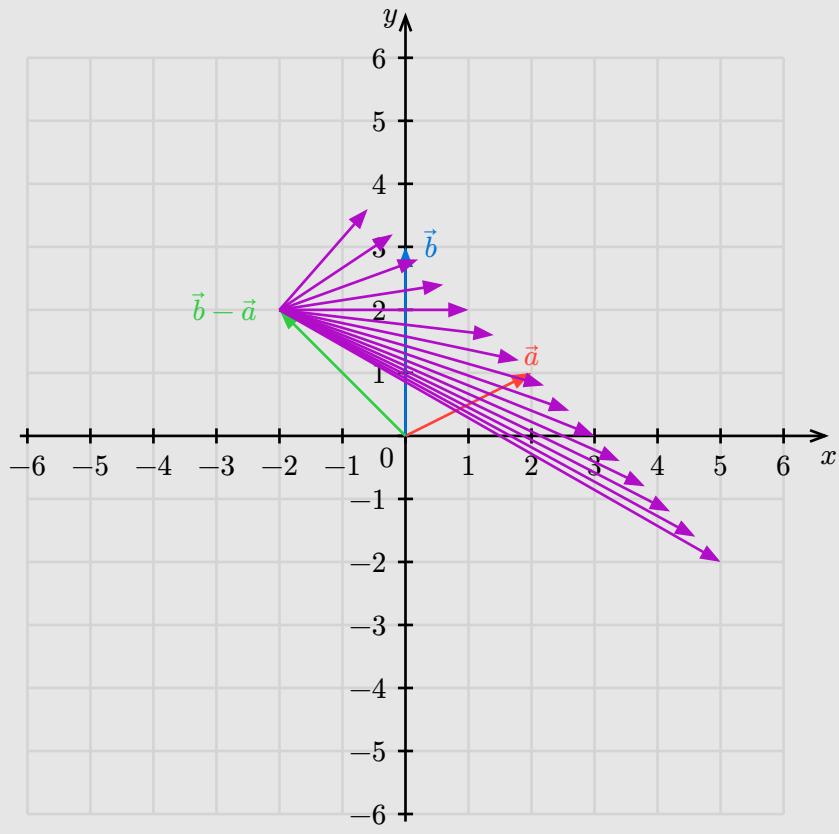
The line  $L$  can be defined as:

$$\begin{aligned} \vec{b} + t(\vec{b} - \vec{a}) &= \begin{bmatrix} 0 \\ 3 \end{bmatrix} + 1 \left( \begin{bmatrix} 0 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right) \\ &= \begin{bmatrix} 0 \\ 3 \end{bmatrix} + 1 \begin{bmatrix} -2 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} -2 \\ 5 \end{bmatrix} \end{aligned}$$



The line  $L$  can also be defined as:

$$\begin{aligned}
 \vec{a} + t(\vec{b} - \vec{a}) &= \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 1 \left( \begin{bmatrix} 0 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right) \\
 &= \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} -2 \\ 2 \end{bmatrix} \\
 &= \begin{bmatrix} 0 \\ 3 \end{bmatrix}
 \end{aligned}$$



Generalization

$$P_1 = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad P_2 = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$L = P_1 + t(P_1 - P_2) \mid t \in \mathbb{R}$$

$$\overrightarrow{P_1} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \quad \overrightarrow{P_2} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$\overrightarrow{P_1} - \overrightarrow{P_2} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

- $L$  starts at  $P_1$  and moves toward  $P_2$  as  $t$  decreases

$$L = P_1 + t(P_1 - P_2) = \begin{bmatrix} -1 \\ 2 \end{bmatrix} + t \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$x = -1 + t(-1)$$

$$y = 2 + t(-1)$$

- $L$  starts at  $P_2$  and moves toward  $P_1$  as  $t$  increases

$$L = P_2 + t(P_1 - P_2) = \begin{bmatrix} 0 \\ 3 \end{bmatrix} + t \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$x = 0 + t(-1)$$

$$y = 3 + t(-1)$$

- $L$  starts at  $P_1$  and moves toward  $P_2$  as  $t$  increases

$$L = P_1 + t(P_2 - P_1) = \begin{bmatrix} -1 \\ 2 \end{bmatrix} + t \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$x = -1 + t(1)$$

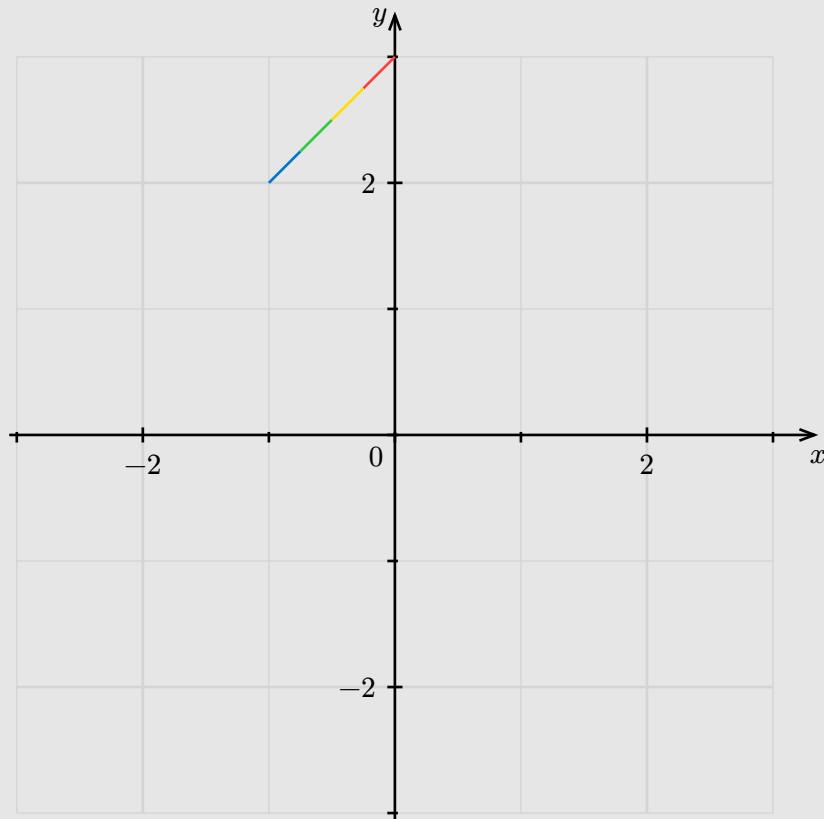
$$y = 2 + t(1)$$

- $L$  starts at  $P_2$  and moves toward  $P_1$  as  $t$  decreases

$$L = P_2 + t(P_1 - P_2) = \begin{bmatrix} 0 \\ 3 \end{bmatrix} + t \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$x = 0 + t(1)$$

$$y = 3 + t(1)$$



### 1.3.6. Vector Spaces

Let's say your factory can produce up to 300 units of product 1, 500 units of product 2, and 400 units of product 3. The set of all possible production combinations forms a vector space:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$0 \leq x_1 \leq 300$$

$$0 \leq x_2 \leq 500$$

$$0 \leq x_3 \leq 400$$

## 1.4. Matrices

$m \times n$  matrix  $A$

- $m$ : rows
- $n$ : columns

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

### 1.4.1. Matrix-Vector Products

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$A\vec{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

For the dot product to be defined, the number of columns in the matrix  $A$  (which is  $n$ ) must match the number of elements in the vector  $\vec{x}$  (also  $n$ ).

The result of multiplying matrix  $A$  and vector  $\vec{x}$  will be a column vector with dimensions  $m \times 1$ , where  $m$  is the number of rows in the matrix  $A$

$$(m \times n) \cdot (n \times 1) = m \times 1$$

#### 1. As Row vectors

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$$\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\vec{a}^T = [a_1, a_2, \dots, a_n]$$

$$\vec{b}^T = [b_1, b_2, \dots, b_n]$$

$$A = \begin{bmatrix} [a_1, a_2, \dots, a_n] \\ [b_1, b_2, \dots, b_n] \end{bmatrix}$$

$$A = \begin{bmatrix} \vec{a} \\ \vec{b} \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\begin{bmatrix} \vec{a}^T \\ \vec{b}^T \end{bmatrix} \cdot \vec{x} = \begin{bmatrix} \vec{a} \cdot \vec{x} \\ \vec{b} \cdot \vec{x} \end{bmatrix}$$

## 2. As Column Vectors

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$$\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$A = \begin{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} & \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \end{bmatrix}$$

$$A = \begin{bmatrix} \vec{a} & \vec{b} \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$A\vec{x} = x_1\vec{a} + x_2\vec{b}$$

### 1.4.2. Null Space

The null space (or kernel) of a matrix  $A$  is the set of all vectors  $x$  that satisfy the equation:

$$A\vec{x} = \mathbf{0}$$

Where:

- $A$ :  $m \times n$  matrix
- $\vec{x}$ :  $n$ -dimensional vector
- $\mathbf{0}$ : zero vector in  $\mathbb{R}^m$

$$N(A) = N(\text{rref}(A)) = \text{span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$$

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{bmatrix}$$

We want to find the null space of  $A$ , which consists of all vectors  $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$  that satisfy:

$$A\vec{x} = \mathbf{0}$$

This expands to the following system of linear equations:

$$\begin{cases} 1x_1 + 1x_2 + 1x_3 + 1x_4 = 0 \\ 1x_1 + 2x_2 + 3x_3 + 4x_4 = 0 \\ 4x_1 + 3x_2 + 2x_3 + 1x_4 = 0 \end{cases}$$

This can be represented as the augmented matrix:

$$\left[ \begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 4 & 0 \\ 4 & 3 & 2 & 1 & 0 \end{array} \right]$$

### 1.4.3. Column Space

The **columns space** (or range) of matrix  $A$  is span of its columns vectors

If the matrix  $A$  has columns  $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n$ , then the column space of  $A$  is defined as:

$$\text{Col}(A) = \{ \vec{y} \in \mathbb{R}^m \mid \vec{y} = A\vec{x} \text{ for some } \vec{x} \in \mathbb{R}^n \}$$

or equivalently,

$$\text{Col}(A) = \text{span}(\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n\})$$

Consider the simple example of a  $2 \times 2$  matrix:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$$

The matrix has two columns:

$$\vec{a}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \text{and} \quad \vec{a}_2 = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$$

The column space, denoted  $\text{Col}(A)$ , is the span of these two vectors:

$$\text{Col}(A) = \text{span}\left(\left\{\begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \end{bmatrix}\right\}\right)$$

### Finding the Column Space

We observe that the two columns  $\vec{a}_1$  and  $\vec{a}_2$  are **linearly dependent**:

$$\vec{a}_2 = k\vec{a}_1$$

This means that  $\vec{a}_2$  is a scalar multiple of  $\vec{a}_1$ , the the two columns are **linearly dependent**. As a result, the column space is spanned by just one vector,  $\vec{a}_1$ , because any linear combination of  $\vec{a}_1$  and  $\vec{a}_2$  can be reduced to a multiple of  $\vec{a}_1$ .

Therefore, the column space of  $A$  is:

$$\text{Col}(A) = \text{span}\left(\left\{\begin{bmatrix} 1 \\ 3 \end{bmatrix}\right\}\right)$$

which represents all vectors of the form:

$$c \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} c \\ 3c \end{bmatrix} \quad \text{for any scalar } c$$

In other words, the column space is a line in  $\mathbb{R}^2$  through the origin in the direction of

### Rank of $A$

The rank of  $A$ , which is the **dimension of its column space**, is 1 because there is only one linearly independent column

This means the column space is the span of the columns of  $A$ , or all vectors that can be formed by taking linear combinations of the columns of  $A$ .

#### 1.4.4. Dimension of a Subspace

Number of elements in a basis for the subspace

#### 1.4.5. Nullity

##### Dimension of the Null Space

$$\dim(N(A))$$

The nullity of  $A$ : number of non-pivot columns (i.e., free variables) in the rref of  $A$

#### 1.4.6. Rank

##### Dimension of the column space

$$\text{rank}(A) = \dim(C(A))$$

#### 1.4.7. Matrix Representation of Systems of Equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1m}x_m &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2m}x_m &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nm}x_m &= b_n \end{aligned}$$

Coefficient Matrix ( $A$ ):

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

Variable Vector (x):

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

Constant Vector (b):

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\mathbf{Ax} = \mathbf{b}$$

The system of equations:

$$\begin{aligned} 2x_1 + 3x_2 + 5x_3 &= 100 \\ 4x_1 + 2x_2 + 1x_3 &= 80 \\ 1x_1 + 5x_2 + 2x_3 &= 60 \end{aligned}$$

Can be represented as a matrix equation:

$$\begin{bmatrix} 2 & 3 & 5 \\ 4 & 2 & 1 \\ 1 & 5 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 100 \\ 80 \\ 60 \end{bmatrix}$$

#### 1.4.8. Matrix Multiplication

$m \times n$  matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

$n \times p$  matrix:

$$B = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix}$$

Compute Each Element of Result Matrix  $C$

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$$

Let  $A$  be an  $n \times m$  matrix:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

Let  $B$  an  $p \times n$  matrix:

$$B = \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}$$

Calculate Each Element of  $C$

$$c_{11} = (1 \cdot 7) + (2 \cdot 9) + (3 \cdot 11) = 58$$

$$c_{12} = (1 \cdot 8) + (2 \cdot 10) + (3 \cdot 12) = 64$$

$$c_{21} = (4 \cdot 7) + (5 \cdot 9) + (6 \cdot 11) = 138$$

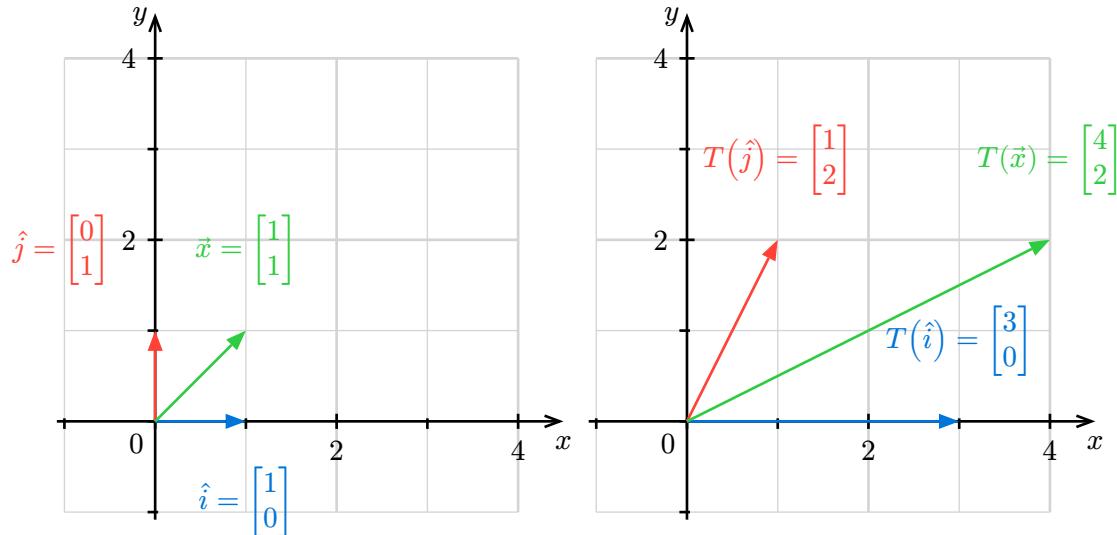
$$c_{22} = (4 \cdot 8) + (5 \cdot 10) + (6 \cdot 12) = 154$$

$C$  is a  $m \times p$  matrix

$$C = \begin{bmatrix} 58 & 64 \\ 139 & 154 \end{bmatrix}$$

## 1.5. Eigenvectors

### 1.5.1. Transformation



$$A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$$

$$\begin{aligned}
T(\hat{i}) &= A\hat{i} \\
&= \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 3 \\ 0 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
T(\hat{j}) &= A\hat{j} \\
&= \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ 2 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
T(\vec{x}) &= A\vec{x} \\
&= \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} 4 \\ 2 \end{bmatrix}
\end{aligned}$$

## 1.6. Eigenvalues

## 2. Calculus

### 2.1. Fundamental Theorem of Calculus

1. If  $F(x)$  is the antiderivative of  $f(x)$  (meaning  $F'(x) = f(x)$ ), then the definite integral of  $f(x)$  from  $a$  to  $b$  is the difference in the values of the antiderivative:

$$\int_a^b f(x)dx = F(b) - F(a)$$

If you take the derivative of a function and then integrate it, you recover the original function

2. If  $f(x)$  is continuous on an interval, then the derivative of the integral of  $f(x)$  from a constant  $a$  to  $x$  is the original function  $f(x)$ :

$$\frac{d}{dx} \left( \int_a^x f(t)dt \right) = f(x)$$

Differentiating the integral of a function recovers the original function

Suppose

$$f(x) = 3x^2$$

We want to:

- Find the definite integral of  $f(x)$  from 1 to 2
- Verify that the derivative of the integral function recovers  $f(x)$

#### 1. Using the First Part of the FTC:

We need to find the definite integral:

$$\int_1^2 3x^2 dx$$

- The antiderivative of  $3x^2$  is  $F(x) = x^3$  (since the derivative of  $x^3$  is  $3x^2$ )
- Apply the formula:

$$\int_1^2 3x^2 dx = F(2) - F(1) = 2^3 - 1^3 = 8 - 1 = 7$$

So,

$$\int_1^2 3x^2 dx = 7$$

## 2. Using the Second Part of the FTC:

Let's define the integral function as:

$$F(x) = \int_1^x 3t^2 dt$$

Now, we take the derivative of  $F(x)$ . By the second part of the FTC:

$$F'(x) = 3x^2$$

This shows that the derivative of the integral function gives us back the original function  $f(x) = 3x^2$ , verifying the relationship between differentiation and integration.

```
import sympy as sp

# Define the variable
x = sp.Symbol('x')

# Define the function f(x)
f = 3 * x**2

# 1. Compute the definite integral of f(x) from 1 to 2

# Find the antiderivative of f(x)
F = sp.integrate(f, x)

# Compute the definite integral using the antiderivative
a, b = 1, 2
definite_integral = F.subs(x, b) - F.subs(x, a)

# Print the result
print("Definite Integral of 3x^2 from 1 to 2:", definite_integral)

# 2. Define the integral function F(x) with limits from 1 to x
F_x = sp.integrate(f, (x, 1, x))

# Compute the derivative of F(x)
```

```

F_prime = sp.diff(F_x, x)

# Print the results
print("Integral Function F(x):", F_x)
print("Derivative of the Integral Function F'(x):", F_prime)

```

## 2.2. Derivative

Measures the instantaneous rate of change of a function

Derivative of a function  $f(x)$

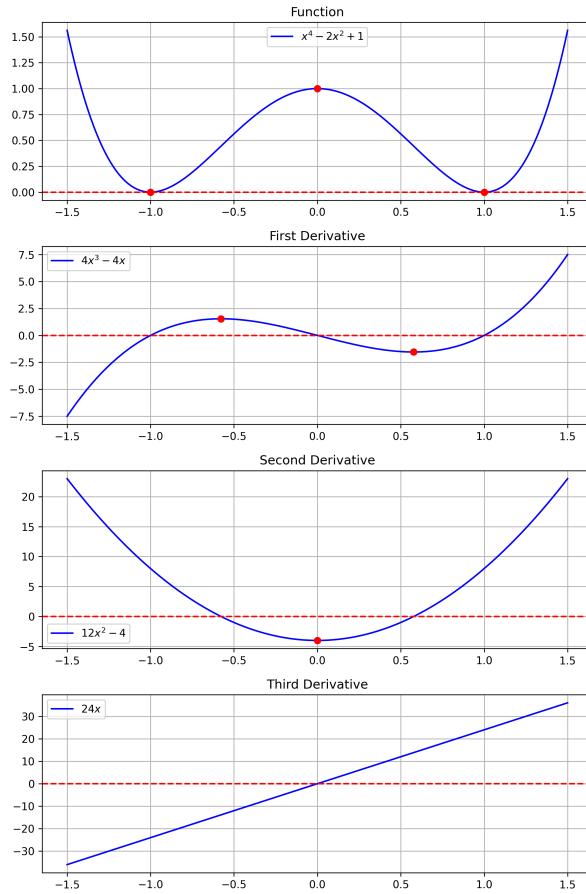
$$f'(x) = \frac{d}{dx} f(x)$$

General Pattern:

- **Linear term:** If a function is linear (e.g.,  $ax + b$ ), the first derivative will be a constant  $a$ .
- **Constant term:** The derivative of a constant is zero.
- **Higher-degree polynomials:** For polynomials of higher degrees, each derivative decreases the degree of the polynomial until you get a constant, after which all further derivatives are zero.

Consider the quartic function:  $f(x) = x^4 - 2x^2 + 1$

- First derivative:  $f'(x) = 4x^3 - 4x$  (**cubic**)
- Second derivative:  $f''(x) = 12x^2 - 4$  (**quadratic**)
- Third derivative:  $f'''(x) = 24x$  (**linear**)
- Fourth derivative:  $f''''(x) = 24$  (**constant**)
- Fifth derivative:  $f'''''(x) = 0$  (**zero**)



### 2.2.1. Power Rule

$$f(x) = x^n$$

$$f'(x) = nx^{n-1}$$

$$f(x) = x^3$$

$$f'(x) = 3x^2$$

### 2.2.2. Product Rule

$$f(x) = u(x) \cdot v(x)$$

$$f'(x) = u'(x)v(x) + u(x)v'(x)$$

$$f(x) = x^2 \cdot x^3$$

$$u(x) = x^2$$

$$v(x) = x^3$$

1. Compute the derivatives of  $u(x)$  and  $v(x)$

$$u'(x) = 2x$$

$$v'(x) = 3x^2$$

2. Apply rule

$$\begin{aligned}f'(x) &= (2x)(x^3) + (x^2)(3x^2) \\&= 2x^2 + 3x^4 \\&= 5x^4\end{aligned}$$

### 2.2.3. Quotient Rule

$$\begin{aligned}f(x) &= \frac{u(x)}{v(x)} \\f'(x) &= \frac{u'(x)v(x) - u(x)v'(x)}{v(x)^2}\end{aligned}$$

$$f(x) = \frac{x^3}{x^2}$$

$$u(x) = x^3$$

$$v(x) = x^2$$

1. Compute the derivatives of  $u(x)$  and  $v(x)$

$$u'(x) = 3x^2$$

$$v'(x) = 2x$$

2. Apply rule

$$\begin{aligned}f'(x) &= \frac{(3x^2)(x^2) - (x^3)(2x)}{(x^2)^2} \\&= \frac{3x^4 - 2x^4}{x^4} \\&= \frac{x^4}{x^4} \\&= 1\end{aligned}$$

### 2.2.4. Chain Rule

$$\begin{aligned}f(g(x)) \\f'(x) = f'(g(x)) \cdot g'(x)\end{aligned}$$

$$f(x) = (2x + 1)^3$$

• Outer function:

$$f(u) = u^3$$

• Inner function

$$g(x) = 2x + 1$$

1. Compute the derivatives of  $f'(u)$  and  $g'(x)$

$$f'(u) = 3u^2$$

$$g'(x) = 2$$

2. Apply rule

$$\begin{aligned}f'(x) &= 3(2x+1)^2 \cdot 2 \\&= 6(2x+1)^2\end{aligned}$$

### 2.2.5. Others

#### 1. Trigonometric Functions

$$\frac{d}{dx}[\sin(x)] = \cos(x)$$

$$\frac{d}{dx}[\cos(x)] = -\sin(x)$$

$$\frac{d}{dx}[\tan(x)] = \sec^2(x)$$

$$\frac{d}{dx}[\cot(x)] = \csc^2(x)$$

$$\frac{d}{dx}[\sec(x)] = \sec(x) \tan(x)$$

$$\frac{d}{dx}[\csc(x)] = -\csc(x) \cot(x)$$

#### 2. Inverse Trigonometric Functions

$$\frac{d}{dx}[\arcsin(x)] = \frac{1}{\sqrt{1-x^2}}$$

$$\frac{d}{dx}[\arccos(x)] = -\frac{1}{\sqrt{1-x^2}}$$

$$\frac{d}{dx}[\arctan(x)] = \frac{1}{1+x^2}$$

$$\frac{d}{dx}[\text{arccot}(x)] = -\frac{1}{1+x^2}$$

$$\frac{d}{dx}[\text{arcsec}(x)] = \frac{1}{|x| \sqrt{x^2-1}}$$

$$\frac{d}{dx}[\text{arccsc}(x)] = -\frac{1}{|x| \sqrt{x^2-1}}$$

#### 3. Exponential and Logarithmic Functions

$$\frac{d}{dx}[e^x] = e^x$$

$$\frac{d}{dx}[a^x] = a^x \ln(a)$$

$$\frac{d}{dx}[\ln(x)] = \frac{1}{x}$$

$$\frac{d}{dx}[\log_{a(x)}] = \frac{1}{x \ln(a)}$$

```
import sympy as sp

# Define the symbolic variable x
x = sp.symbols('x')

# Define the function f(x) = 3x^2 + 2x + 1
f = 3 * x**2 + 2 * x + 1

# Compute the derivative of f with respect to x
f_prime = sp.diff(f, x)

# Print the derivative of f
print(f_prime) # Output: 6x + 2

# Create a numerical function from the symbolic derivative
f_prime_func = sp.lambdify(x, f_prime)

# Evaluate the derivative at x = 2
result = f_prime_func(2)

# Print the result of f'(2)
print(result) # Output: 14

# Solve f'(x) = 0 to find the critical values (where the derivative is 0)
critical_values = sp.solve(f_prime, x)

# Print the critical values (roots of the derivative)
print(critical_values) # Output: [-1/3]
```

## 2.3. Integral

Measures the accumulation of quantities

Integral (antiderivative) of a function  $f(x)$

$$F(x) = \int f(x)dx$$

Area under a curve between two points

- 2.3.1. Power Rule
- 2.3.2. Sum & Difference Rule
- 2.3.3. Constant Multiple Rule
- 2.3.4. Integration by Substitution
- 2.3.5. Integration by Parts
- 2.3.6. Trigonometric Integrals
- 2.3.7. Exponential and Logarithmic Integrals
- 2.3.8. Definite Integrals of Even and Odd Functions
- 2.3.9. Improper Integrals

### 3. Causes of variation

#### 3.1. Common

Common Cause Variation: Normal, expected variation inherent to the process; predictable and stable.

- Temperature
- Humidity
- Material properties

#### 3.2. Special

Special Cause Variation: Abnormal, unexpected variation due to specific causes; unpredictable and requires immediate action.

External disruptions

- Machine malfunction
- Faulty raw materials
- Staff shortages
- Weather events

### 4. Probability Theory

#### 4.1. Probability Axioms

##### 4.1.1. Non-Negativity

Probability of any event cannot be negative

$$P(A) \geq 0$$

##### 4.1.2. Normalization

The probability of the sample space is always 1

$$P(S) = 1$$

Where:

- $S$  is the sample space (the set of all possible outcomes)

### 4.1.3. Additivity

If two events are mutually exclusive (cannot happen at the same time), the probability of either occurring is the sum of their individual probabilities

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

## 4.2. Rules

### 4.2.1. Complement Rule

The probability of the complement of an event  $A$  is  $P(A^c) = 1 - P(A)$

How likely it is that the event does not occur.

Consider a fair die:

- Let  $A$  be the event “rolling a 4”. Then  $P(A) = \frac{1}{6}$
- The complement of  $A$ , denoted  $P(A^c)$ , is “not rolling 4”
- $P(A^c) = 1 - P(A) = 1 - \frac{1}{6} = \frac{5}{6}$

### 4.2.2. Multiplication Rule

- Independent Events

If  $A$  and  $B$  are independent, then  $P(A \cap B) = P(A) \times P(B)$ . The probability of both events occurring is the product of their individual probabilities.

Consider flipping two fair coins. Let  $A$  be the event “the first coin is heads” and  $B$  be the event “the second coin is heads”

- Since the flips are independent,  $P(A \cap B) = P(A) \times P(B)$
- $P(A) = \frac{1}{2}$  and  $P(B) = \frac{1}{2}$
- Thus,  $P(A \cap B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

- Dependent Events

If  $A$  and  $B$  are dependent,  $P(A \cap B) = P(A) \times P(B | A)$ , where  $P(B | A)$  is the conditional probability of  $B$  given  $A$ .

Draw two cards from a standard deck without replacement. Let  $A$  be the event “drawing an Ace on the first draw” and  $B$  be the event “drawing an Ace on the second draw”

- $P(A) = \frac{4}{52} = \frac{1}{13}$
- If  $A$  occurs (i.e., an Ace is drawn first), there are 3 Aces left out of 51 cards. So,  $P(B | A) = \frac{3}{51} = \frac{1}{17}$
- Thus,  $P(A \cap B) = P(A) \times P(B | A) = \frac{1}{13} \times \frac{1}{17} = \frac{1}{221}$

### 4.2.3. Addition Rule

For Any Two Events:  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ . This accounts for the overlap between the two events to avoid double counting.

Suppose you roll a die, and you want to find the probability of rolling a 2 or a 4.

- Let  $A$  be the event “rolling a 2,” and  $B$  be the event “rolling a 4”
- $P(A) = \frac{1}{6}$  and  $P(B) = \frac{1}{6}$
- Since  $A$  and  $B$  are mutually exclusive,  $P(A \cap B) = 0$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{6} + \frac{1}{6} - 0 = \frac{2}{6} = \frac{1}{3}$

#### 4.2.4. Conditional Probability

The probability of an event  $A$  given that  $B$  has occurred is  $P(A | B) = \frac{P(A \cap B)}{P(B)}$ , provided  $P(B) > 0$ .

In a deck of 52 cards, if you know a card is a spade, what is the probability that it is an Ace?

- Let  $A$  be the event “drawing an Ace,” and  $B$  be the event “drawing a spade”
- $P(B) = \frac{12}{52} = \frac{1}{4}$
- There is 1 Ace of Spades out of 13 spades, so  $P(A \cap B) = \frac{1}{32}$
- Thus,  $P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{32}}{\frac{1}{4}} = \frac{1}{13}$

#### 4.2.5. Law of Total Probability

If  $\{B_i\}$  is a partition of the sample space, then for any event  $A$ :

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A | B_i) \times P(B_i)$$

Suppose you want to calculate the probability of raining on a given day. You know that it's either sunny or cloudy, and the probability of rain is different in each condition.

- Let  $B_1$  be the event “sunny” and  $B_2$  be the event “cloudy”
- $P(B_1) = \frac{3}{4}$  and  $P(B_2) = \frac{1}{4}$
- The probability of rain given sunny is  $P(A | B_1) = \frac{1}{10}$ , and given cloudy is  $P(A | B_2) = \frac{2}{5}$
- The total probability of rain is:

$$\begin{aligned} P(A) &= P(A | B_1) \times P(B_1) + P(A | B_2) \times P(B_2) \\ &= \frac{1}{10} \times \frac{3}{4} + \frac{2}{5} \times \frac{1}{4} \\ &= \frac{3}{40} + \frac{2}{20} \\ &= \frac{3}{40} + \frac{4}{40} \\ &= \frac{7}{40} \end{aligned}$$

#### 4.2.6. Law of Large Numbers

As the number of trials of a random experiment increases, the sample mean will converge to the expected value (mean) of the random variable

You flip a fair coin 1000 times. As you increase the number of flips, the proportion of heads should get closer to the probability of heads (which is 0.5).

#### 4.2.7. Central Limit Theorem

For a sufficiently large number of trials, the distribution of the sample mean approaches a normal distribution, regardless of the distribution of the population

Suppose you are rolling a fair die repeatedly and calculating the average of the results. If you take 30 samples, each being the average of 10 die rolls, the distribution of those sample means will be approximately normal with a mean of 3.5 (the mean of a fair die) and a standard deviation that can be computed based on the original die's distribution.

### 4.3. Bayes Theorem

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}$$

- $P(A)$ : prior probability (initial belief about event  $A$ )
- $P(B | A)$ : likelihood (probability of observing  $B$  given that  $A$  is true)
- $P(B)$ : marginal likelihood (overall probability of observing  $B$  under all possible conditions)
- $P(A | B)$ : posterior probability (updated belief after considering the evidence  $B$ )

## 5. Descriptive Statistics

### 5.1. Central Tendency

#### 5.1.1. Mean

Sum of all the values divided by the number of values

$$\mu = \frac{\sum_{i=1}^n x_i}{n}$$

[1, 2, 3]

$$\bar{x} = \frac{1 + 2 + 3}{3} = 150$$

#### 5.1.2. Median

Middle value in a set of values when they are arranged in ascending or descending order

##### 1. Odd Number of Values

[1, 2, 3]

- **Step 1:** Arrange the Data in Ascending Order

[1, 2, 3]

- **Step 2:** Identify the Median

Median = 2

## 2. Even Number of Values

[1, 2, 3, 4]

- **Step 1:** Arrange the Data in Ascending Order

[1, 2, 3, 4]

- **Step 2:** Identify the Median

$$\text{Median} = \frac{2+3}{2} = \frac{5}{2} = 2.5$$

### 5.1.3. Mode

Value that appears most frequently

[1, 1, 2, 3]

- **Step 1:** Identify the Most Frequent Number

1 : 2

2 : 1

3 : 1

- **Step 2:** Determine the Mode

Mode = 1

## 5.2. Dispersion

### 5.2.1. Range

Range = max – min

[1, 2, 3, 4, 5]

- **Step 1:** Identify the Maximum and Minimum Values

max = 5

min = 1

- **Step 2:** Calculate the Range

Range = 5 – 1 = 4

### 5.2.2. Variance

Quantifies the spread or dispersion of a set of data points in a dataset

- Population

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^n (x_i - \mu)^2$$

- Sample

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

$$[70, 75, 80, 85, 90]$$

**Step 1:** Find mean

$$\bar{x} = \frac{70 + 75 + 80 + 85 + 90}{5} = \frac{400}{5} = 80$$

**Step 2:** Subtract the Mean and Square the result

$$\begin{aligned} (70 - 80)^2 &= (-10)^2 = 100 \\ (75 - 80)^2 &= (-5)^2 = 25 \\ (80 - 80)^2 &= 0^2 = 0 \\ (85 - 80)^2 &= 5^2 = 25 \\ (90 - 80)^2 &= 10^2 = 100 \end{aligned}$$

**Step 3:** Calculate variance

$$s^2 = \frac{100 + 25 + 0 + 25 + 100}{5-1} = \frac{250}{4} = 62.5$$

### 5.2.3. Standard deviation

Quantifies the amount of variation or dispersion in a set of data points

- Population

$$\sigma = \sqrt{\sigma^2}$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

- Sample

$$s = \sqrt{s^2}$$

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

### 5.2.4. Interquartile Range (IQR)

$$IQR = Q3 - Q1$$

$$[1, 2, 3, 4, 5, 6, 7]$$

Step 1: Arrange the Data in Ascending Order

[1, 2, 3, 4, 5, 6, 7]

Step 2: Find the Quartiles

1. Calculate the Median (Q2)

$$\text{Median (Q2)} = 4$$

2. Find First Quartile (Q1)

Q1 is the median of the first half of the dataset

$$Q1 = 2$$

3. Find Third Quartile (Q3)

Q3 is the median of the second half of the dataset

$$Q3 = 5$$

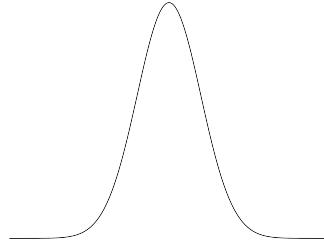
Step 3: Calculate the Interquartile Range (IQR)

$$IQR = 5 - 2 = 3$$

## 6. Probability Distributions

### 6.1. Gaussian (Normal) distribution

$$f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

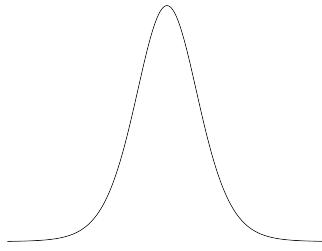


### 6.2. t-Distribution

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

Where:

- $t$ : t-statistic
- $\nu$  (or  $df$ ): degrees of freedom
- $\Gamma$ : Gamma function (generalizes the factorial function)



Continuous probability distribution for estimating the mean of a normally distributed population in situations where:

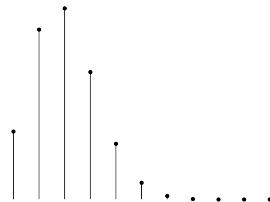
- Sample size is small
- Population standard deviation is unknown

Similar in shape to the normal distribution but has heavier tails, which means it gives more probability to values further from the mean

### 6.3. Binomial distribution

Discrete probability distribution that describes the number of successes in a fixed number of independent Bernoulli trials, each with the same probability of success

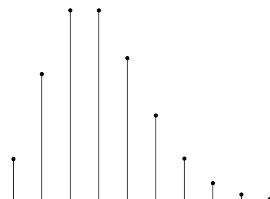
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$



### 6.4. Poisson distribution

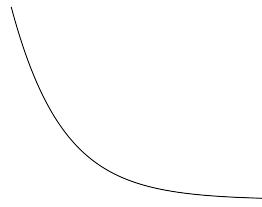
Used to model the number of events that occur within a fixed interval of time or space, given a constant mean rate and assuming that these events occur independently of each other

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$



### 6.5. Exponential distribution

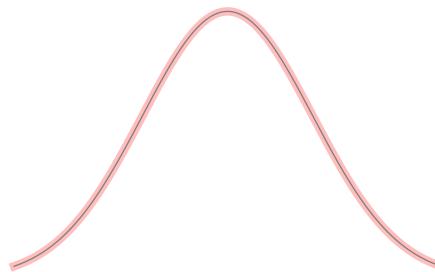
$$f(x \mid \lambda) = \lambda e^{-\lambda x}$$



## 7. Functions

### 7.1. PDF (Probability Density Function)

Function that describes the likelihood of a continuous random variable taking on a particular value



Properties:

- The area under the curve of a PDF over the entire range of possible values equals 1
- The PDF itself is non-negative everywhere
- The probability that the variable falls within a certain range is given by the integral of the PDF over that range

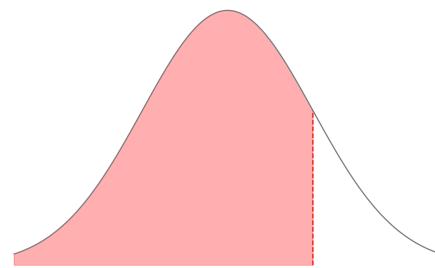
### 7.2. PMF (Probability Mass Function)

Frequency distribution that provides the probability that a categorical variable takes on each of its possible values

### 7.3. CDF (Cumulative Distribution Function)

Gives the probability that  $X$  will take a value less than or equal to  $x$

$$F(x) = P(X \leq x)$$



1. Categorical

$$F(x) = \int_{-\infty}^x f(t)dt$$

## 2. Continuous

$$F(x) = \sum_{t \leq x} P(X = t)$$

```
from scipy.stats import norm

x = 1
mu = 0
sigma = 1

norm.cdf(x, loc=mu, scale=sigma)
```

## 7.4. PPF (Percent-Point Function)

Gives the value  $x$  such that the probability of a random variable being less than or equal to  $x$  is equal to a given probability  $p$

```
from scipy.stats import norm

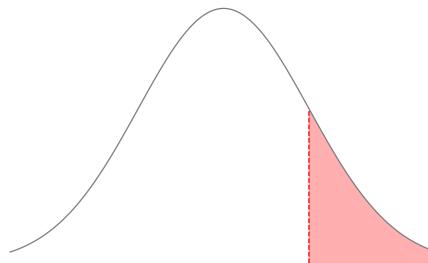
p = 0.95
mu = 0
sigma = 1

x_value = norm.ppf(p, loc=mu, scale=sigma)
```

## 7.5. SF (Survival Function)

Probability that a certain event has not occurred by a certain time

$$S(t) = P(T > t)$$



Relationship to PDF:

$$S(t) = 1 - F(t)$$

```
from scipy.stats import norm

z = 3.4
mu = 0
```

```
sigma = 1

norm.pdf(z, loc=mu, scale=sigma)
```

## 8. Error Metrics

### 8.1. MAE (Mean Absolute Error)

Average of squared differences between predicted ( $\hat{y}_i$ ) and actual values ( $y_i$ )

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

### 8.2. MSE (Mean Squared Error)

Average of squared differences between predicted ( $\hat{y}_i$ ) and actual ( $y_i$ ) values

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

### 8.3. RMSE (Root Mean Squared Error)

square root of the average squared differences between predicted ( $\hat{y}_i$ ) and actual ( $y_i$ ) values

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

### 8.4. MAPE (Mean Absolute Percentage Error)

Average percentage difference between predicted ( $\hat{y}_i$ ) and actual ( $y_i$ ) values

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

### 8.5. R-squared

Proportion of variance explained by the model

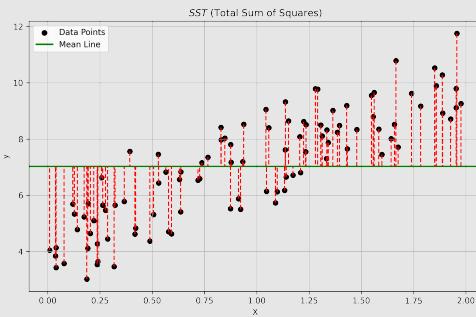
$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

- 1: model explains all the variance in the dependent variable
- 0: model explains none of the variance in the dependent variable

**Step 1:** Fit the Regression Model

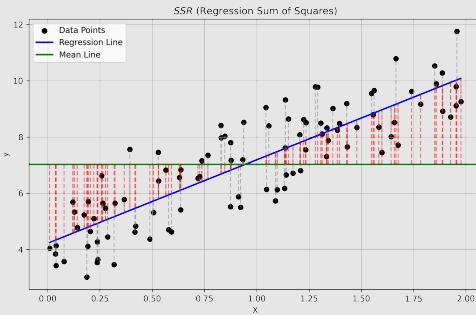
**Step 2:** Compute **Total Sum of Squares (SST)**

$$\text{SST} = \sum_{i=1}^n (y_i - \bar{y})^2$$



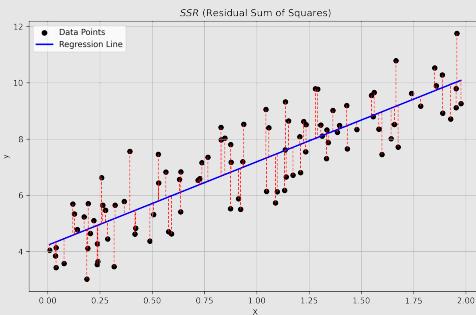
### Step 3: Compute Regression Sum of Squares (SSR)

$$SSR = \sum_{i=1}^n = (\hat{y}_i - \bar{y})^2$$



### Step 4: Compute Residual Sum of Squares (SSE)

$$SSE = \sum_{i=1}^n = (y_i - \hat{y}_i)^2$$



### Step 5: Calculate $R^2$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

## 8.6. Adj R-squared

Adjusts the  $R^2$  value based on the number of predictors (penalty for adding non-informative variables)

$$\text{Adj } R^2 = 1 - (1 - R^2) \frac{n - 1}{n - p - 1}$$

## 8.7. MSLE (Mean Squared Logarithmic Error)

When predictions and actual values span several orders of magnitude

$$\text{MSLE} = \frac{1}{n} \sum_{i=1}^n (\log(1 + y_i) - \log(1 + \hat{y}_i))^2$$

## 8.8. Cross-Entropy Loss (Log Loss)

Binary and multi-class classification

$$\text{Log Loss} = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

# 9. Hypothesis Testing

## 9.1. Hypotheses

### 9.1.1. Null ( $H_0$ )

No effect or difference (observed effect is due to sampling variability)

$$H_0 : \mu = \mu_0$$

### 9.1.2. Alternative ( $H_1$ or $H_a$ )

Presence of an effect or a difference

- Two-Tailed

$$H_1 : \mu \neq \mu_0$$

- Right-Tailed

$$H_1 : \mu > \mu_0$$

- Left-Tailed

$$H_1 : \mu < \mu_0$$

## 9.2. Error Types

### 9.2.1. Type I ( $\alpha$ )

$$\alpha = P(\text{Reject } H_0 \mid H_0 \text{ is true})$$

### 9.2.2. Type II ( $\beta$ )

$$\beta = P(\text{Fail to Reject } H_0 \mid H_1 \text{ is true})$$

## 9.3. t-Tests

### 9.3.1. One-sample

Tests if the mean of a single sample differs from a known or hypothesized population mean.

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

- $\bar{x}$ : sample mean
- $\mu_0$ : hypothesized population mean
- $s$ : sample standard deviation
- $n$ : sample size

$$[78, 82, 89]$$

Step 1: State Hypotheses

$$H_0 : \mu = 85 \text{ cm}$$

$$H_1 : \mu \neq 85 \text{ cm}$$

Step 2: Summarize Data

- Sample values:

$$x_1 = 78, x_2 = 82, x_3 = 89$$

- Sample size:

$$n = 3$$

Step 3: Calculate Sample Mean ( $\bar{x}$ )

$$\bar{x} = \frac{x_1 + x_2 + x_3}{n} = \frac{78 + 82 + 89}{3} = 83$$

Step 4: Calculate Sample Standard Deviation:

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

- Find the deviations from the mean and square them

$$(x_1 - \bar{x})^2 = (78 - 83)^2 = (-5)^2 = 25$$

$$(x_2 - \bar{x})^2 = (82 - 83)^2 = (-1)^2 = 1$$

$$(x_3 - \bar{x})^2 = (89 - 83)^2 = (6)^2 = 36$$

- Sum of squared deviations

$$SSD = 25 + 1 + 36$$

- Calculate variance of sample

$$s^2 = \frac{SSD}{n-1} = \frac{62}{3-1} = \frac{62}{2} = 31$$

- Calculate Sample Standard Deviation

$$s = \sqrt{s^2} = \sqrt{32} = 5.57$$

Step 5: Calculate the Test Statistic

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{83 - 85}{\frac{5.57}{\sqrt{3}}} = -\frac{3}{3.22} = -0.62$$

Step 6: Determine the Degrees of Freedom

$$df = n - 1 = 15 - 1 = 14$$

Step 7: Find Critical t-value

- For a two-tailed test at a significance level ( $\alpha$ ) of 0.05 and 2 degrees of freedom ( $df$ )

4.303

Step 8: Compare the t-Value to the Critical t-Value

- If the absolute value of the test statistic is greater than the critical t-value, reject the null hypothesis.
- If the absolute value of the test statistic is less than the critical t-value, fail to reject the null hypothesis.

Step 8: Find the p-Value

$$t = 4.303 \text{ and } df = 3$$

$$\text{p-value} = 0.58$$

```
from scipy import stats
rvs = stats.uniform.rvs(size=50)
stats.ttest_1samp(rvs, popmean=0.5)
```

### 9.3.2. Independent

Compares the means of two independent samples.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)}}$$

- $\bar{x}_1$  and  $\bar{x}_2$ : sample means
- $s_1^2$  and  $s_2^2$ : sample variances
- $n_1$  and  $n_2$ : sample sizes

```
from scipy import stats
rvs1 = stats.norm.rvs(loc=5, scale=10, size=500)
rvs2 = stats.norm.rvs(loc=5, scale=10, size=500)
stats.ttest_ind(rvs1, rvs2)
```

### 9.3.3. Paired

$$t = \frac{\bar{D}}{\frac{s_D}{\sqrt{n}}}$$

- $\bar{D}$ : mean of the differences between paired observations
- $s_D$ : standard deviation of the differences
- $n$ : number of pairs

## 9.4. Chi-square tests

### 9.4.1. Goodness of Fit Test

Compares an **observed** categorical distribution to a **theoretical** categorical distribution.

$$\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$$

- $o_i$ : observed frequency in category  $i$
- $e_i$ : expected frequency in category  $i$

```
from scipy import stats
f_obs = np.array([43, 52, 54, 40])
f_exp = np.array([47, 47, 47, 47])
stats.chisquare(f_obs=f_obs, f_exp=f_exp)
```

### 9.4.2. Test of independence

Compares two **observed** categorical distributions.

$$\chi^2 = \sum_{i=1}^k \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

- $o_{ij}$ : observed frequency in cell  $(i, j)$
- $e_{ij}$ : expected frequency in category  $(i, j)$  calculated as:

$$E_{ij} = \frac{R_i \times C_j}{N}$$

- $R_i$ : Row total for row  $i$
- $C_j$ : Column total for column  $j$
- $N$ : Total number of observations

## 9.5. ANOVA (Analysis of Variance)

### 9.5.1. One-way

Compares the means of three or more groups based on one independent variable

- $H_0: \mu_1 = \mu_2 = \dots = \mu_k$
- $H_1: \text{At least one } \mu_i \text{ differs from the others}$

**Step 1:** Calculate Between-Group Variation ( $SS_{\text{between}}$ )

$$SS_{\text{between}} = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X}_{\text{overall}})^2$$

- $n_i$ : Number of observations in group  $i$
- $\bar{X}_i$ : Mean of group  $i$
- $\bar{X}_{\text{overall}}$ : Overall mean of all groups

**Step 2:** Calculate Within-Group Variation ( $SS_{\text{within}}$ )

$$SS_{\text{within}} = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

- $X_{ij}$ : Observation  $j$  in group  $i$

**Step 3:** Calculate Total Variation ( $SS_{\text{total}}$ )

$$SS_{\text{total}} = SS_{\text{between}} + SS_{\text{within}}$$

**Step 4:** Calculate Mean Squares

- Mean Square Between ( $MS_{\text{between}}$ )

$$MS_{\text{between}} = \frac{SS_{\text{between}}}{k - 1}$$

- Mean Square Within ( $MS_{\text{within}}$ )

$$MS_{\text{within}} = \frac{SS_{\text{within}}}{N - k}$$

- $N$ : total number of observations
- $k$ : number of groups

**Step 5:** Calculate the F-statistic

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

**Step 6:** Decision Rule

- Compare the F-statistic to the critical value from the F-distribution table (based on chosen significance level  $\alpha$ ).
- Alternatively, compare the p-value to the significance level  $\alpha$ .
- Reject  $H_0$  if the F-statistic is greater than the critical value or if the p-value is less than  $\alpha$ , indicating that at least one group mean is significantly different.

#### 9.5.2. Two-way

## 10. Regression Analysis:

### 10.1. Simple linear regression

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

- $y$ : dependent variable
- $x$ : independent variables
- $\beta_0$ : intercept (value of  $y$  when  $x = 0$ )
- $\beta_1$ : slope (change in  $y$  for a one-unit change in  $x$ )
- $\varepsilon$ : error term (difference between the actual data points and the predicted values)

Data

$x$ (Hours Studied)	$y$ (Test Score)
1	2
2	3

$x$ (Hours Studied)	$y$ (Test Score)
3	5
4	4
5	6

**Step 1:** Calculate Means

$$\bar{x} = 3$$

$$\bar{y} = 4$$

**Step 2:** Calculating Slope  $\beta_1$

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

- The numerator

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 9$$

- The denominator

$$\sum_{i=1}^n (x_i - \bar{x})^2 = 10$$

- The slope  $\beta_1$  is

$$\beta_1 = \frac{9}{10} = 0.9$$

**Step 3:** Calculate Intercept  $\beta_0$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} = 1.3$$

**Step 4:** Calculate p-Value

- Calculate Standard Error of the Slope ( $SE_{\beta_1}$ )

$$SE_{\beta_1} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{(n-2) \cdot \sum_{i=1}^n (x_i - \bar{x})^2}}$$

- Calculate the Residual Sum of Squares (RSS)

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

- Calculate the t-statistic for the Slope

$$t = \frac{\beta_1}{SE_{\beta_1}}$$

- Determine Degrees of Freedom

$$df = n - 2$$

- Look up the p-value corresponding to  $t$  with 3 degrees of freedom

**Step 5:** Calculate  $R^2$  ( $R_{\text{adj}}^2$ )

$$R^2 = \frac{SS_{\text{reg}}}{SS_{\text{total}}}$$

$SS_{\text{reg}}$  (Regression Sum of Squares): sum of the squared differences between the predicted  $\hat{y}$  values and the mean of the observed  $y$  values.

$$SS_{\text{reg}} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$SS_{\text{total}}$  (Total Sum of Squares): sum of the squared differences between the observed  $y$  values and the mean of the observed  $y$  values.

$$SS_{\text{total}} = \sum_{i=1}^n (y_i - \bar{y})^2$$

Adjusting for the number of independent variables

$$R_{\text{adj}}^2 = 1 - \left( \frac{(1 - R^2)(n - 1)}{n - k - 1} \right)$$

- $n$ : number of observations
- $k$ : number of independent variables

## 10.2. Multiple regression

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$$

Matrix form:

$$Y = X\beta + \varepsilon$$

Estimating Coefficients (OLS):

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Where:

- $\hat{\beta}$ : vector of estimated coefficients

## 10.3. Logistic Regression

Binary classification (1 or 0, true or false, yes or no) based on one or more predictor variables

Sigmoid function:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Where:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

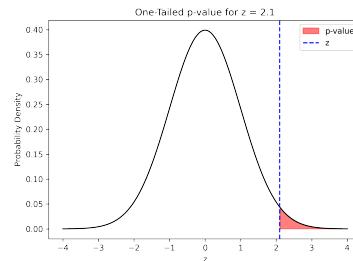
## 10.4. Model diagnostics

### 10.4.1. p-Values

Probability of obtaining results at least as extreme as the observed results

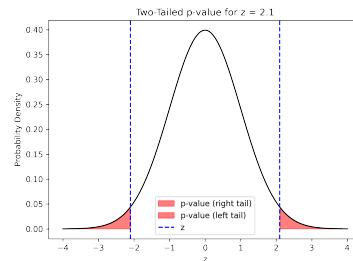
#### 1. One-Tailed

$$p = P(Z > z_{\text{observed}})$$



#### 2. Two-Tailed

$$p = 2 \cdot P(Z > |z_{\text{observed}}|)$$



```
z = 2.1
df = 3
scipy.stats.norm.sf(abs(z), df=df)
```

### 10.4.2. F-Statistic

$$F = \frac{\text{MSR}}{\text{MSE}}$$

Where:

- Mean Square Regression (MSR):

$$\text{MSR} = \frac{\text{SSR}}{\text{df}_{\text{regression}}}$$

- $\text{df}_{\text{regression}} = p$

Where:

- $p$ : Number of independent
- Mean Square Error (MSE):

$$MSE = \frac{SSE}{df_{\text{error}}}$$

- $df_{\text{error}} = n - p - 1$

Where:

- $p$ : Number of independent
- $n$ : Number of observations
- 1: Constant (for intercept)

#### 10.4.3. Confidence Intervals (CI)

Range within which we can be confident that the true value (population parameter) lies, based on the sample data

1. Known population standard deviation ( $\sigma$ ):

$$CI = \bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

Where:

- $\bar{x}$ : sample mean
- $z$ : z-score corresponding to the desired confidence level
- $\sigma$ : population standard deviation
- $n$ : sample size

2. Unknown population standard deviation ( $\sigma$ ):

$$CI = \bar{x} \pm t \frac{s}{\sqrt{n}}$$

Where:

- $\bar{x}$ : sample mean
- $t$ : critical value from t-distribution
- $s$ : sample standard deviation
- $n$ : sample size

## 11. Design of Experiments (DOE):

## 12. Control Charts:

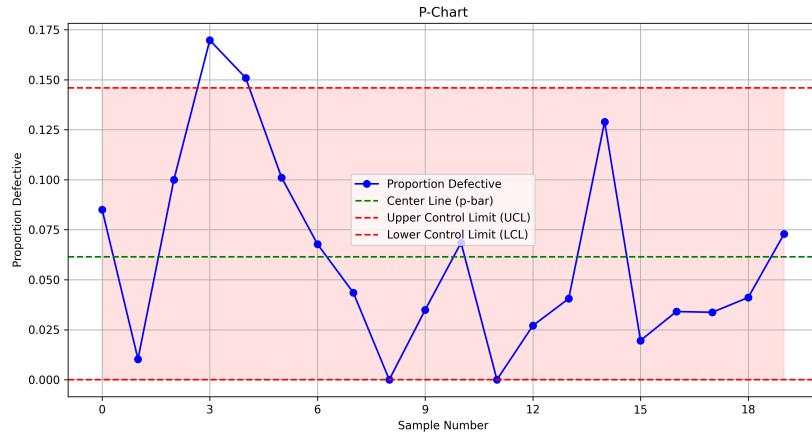
### 12.1. P-charts (Proportion)

Proportion of defective items

$$\hat{p} = \frac{D}{n}$$

Where:

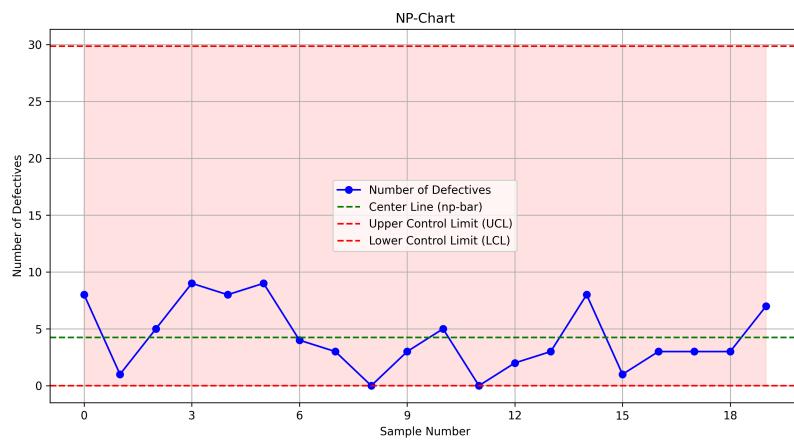
- $D$ : number of defective items
- $n$ : sample size



## 12.2. NP-charts (Number Proportion)

Number of defective items (constant sample size)

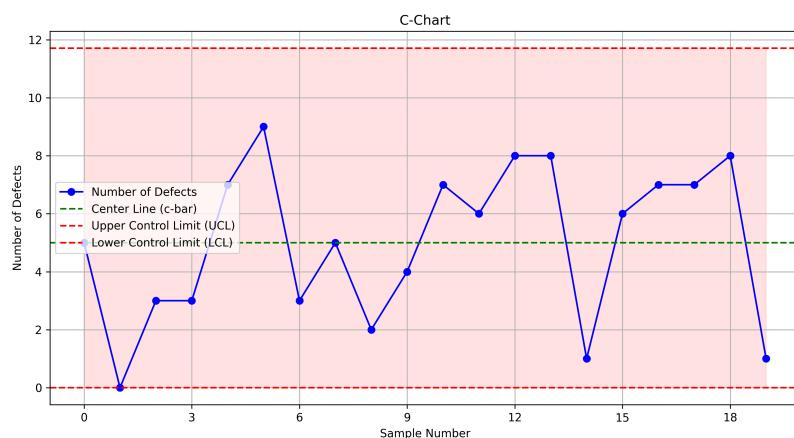
- $N$ : count of defective items in each sample



## 12.3. C-charts (Count)

Count of defects (fixed unit size)

Number of defects observed in each sample or inspection unit



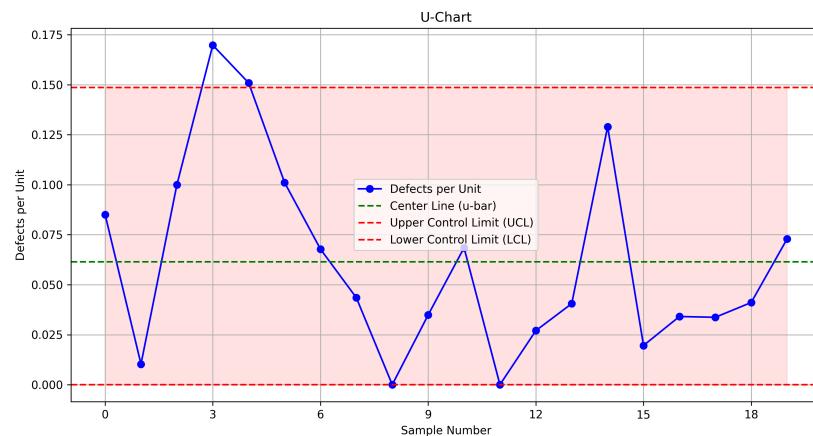
## 12.4. U-charts (Unit)

Defects per unit (variable unit size)

$$\hat{u} = \frac{C}{n}$$

Where:

- $C$ : number of defects
- $n$ : size of unit



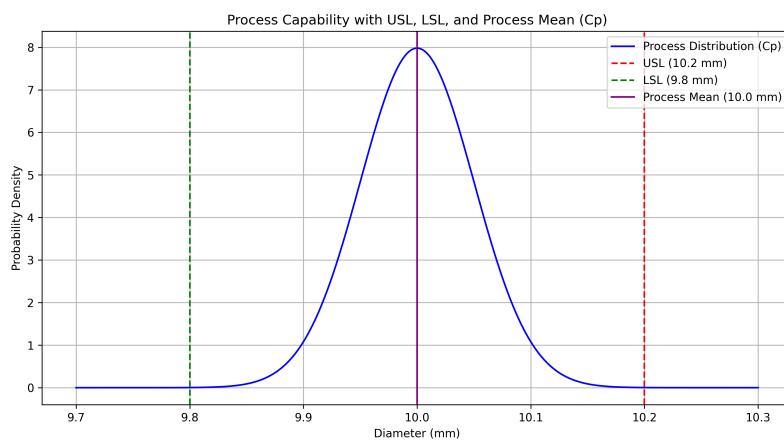
## 12.5. $\bar{X}$ -chart

## 12.6. R-chart

# 13. Process Capability Analysis

## 13.1. $C_p$ (Process Capability Index)

$$C_p = \frac{USL - LSL}{6\sigma}$$



Suppose a company manufactures metal rods, and the specification limits for the diameter of the rods are:

- Upper Specification Limit (**USL**): 10.2 mm
- Lower Specification Limit (**LSL**): 9.8 mm

The process has a standard deviation  $\sigma$  of 0.05 mm.

### Step 1: Determine the Specification Width

The specification width is the difference between the USL and LSL.

$$\text{Specification Width} = \text{USL} - \text{LSL} = 10.2 \text{ mm} - 9.8 \text{ mm}$$

### Step 2: Calculate the Process Capability Index $C_p$

The formula for  $C_p$  is:

$$C_p = \frac{\text{Specification Width}}{6\sigma} = \frac{\text{USL} - \text{LSL}}{6\sigma}$$

Substitute the values:

$$C_p = \frac{0.4 \text{ mm}}{6 \times 0.05 \text{ mm}} = \frac{0.4 \text{ mm}}{0.3 \text{ mm}} = 1.33$$

Interpretation:

$C_p = 1.33$  means the process spread ( $6\sigma$ ) fits 1.33 times within the tolerance range (the distance between the Upper Specification Limit and Lower Specification Limit).

- $C_p = 1.00$ : Process variation fits exactly within the specification limits. 99.73% of the output will be within specifications **if the process is centered** (3 sigma process).
- $C_p > 1.00$ : Process variation is narrower than the specification limits. The higher the  $C_p$ , the more capable the process is, meaning it can produce parts within the tolerance more consistently.
- $C_p < 1$ : Process variation is wider than the specification limits. Significant portion of the output will fall outside the specification limits.

Limitations:

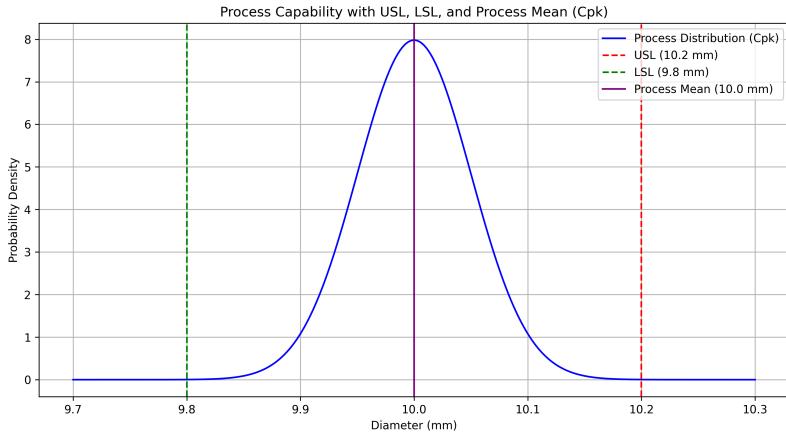
Since  $C_p$  **does not account for the centering of the process**, it may give a false sense of security if the process mean is off-center (see  $C_{pk}$ ).

Use:

When you are interested in understanding the **potential capability** of a process under ideal conditions, typically in a short-term study where the process is stable and controlled.

## 13.2. $C_{pk}$ (Process Capability Index with Centering)

$$C_{pk} = \text{Min} \left( \frac{\text{USL} - \bar{x}}{3\sigma}, \frac{\bar{x} - \text{LSL}}{3\sigma} \right)$$



Suppose a company manufactures metal rods, and the specification limits for the diameter of the rods are:

- Upper Specification Limit (**USL**): 10.2 mm
- Lower Specification Limit (**LSL**): 9.8 mm

The process has:

- A standard deviation  $\sigma$  of 0.05 mm.
- A process mean  $\mu$  of 10.1 mm.

**Step 1:** Calculate the distance from the mean to the USL:

$$\frac{USL - \mu}{3\sigma} = \frac{10.2 \text{ mm} - 10.1 \text{ mm}}{3 \times 0.05 \text{ mm}} = \frac{0.1 \text{ mm}}{0.15 \text{ mm}} = 0.67$$

**Step 2:** Calculate the distance from the mean to the LSL:

$$\frac{\mu - LSL}{3\sigma} = \frac{10.1 \text{ mm} - 9.8 \text{ mm}}{3 \times 0.05 \text{ mm}} = \frac{0.3 \text{ mm}}{0.15 \text{ mm}} = 2.00$$

**Step 3:** Determine  $C_{pk}$ :

$$C_{pk} = \min(0.67, 2.00) = 0.67$$

Interpretation:

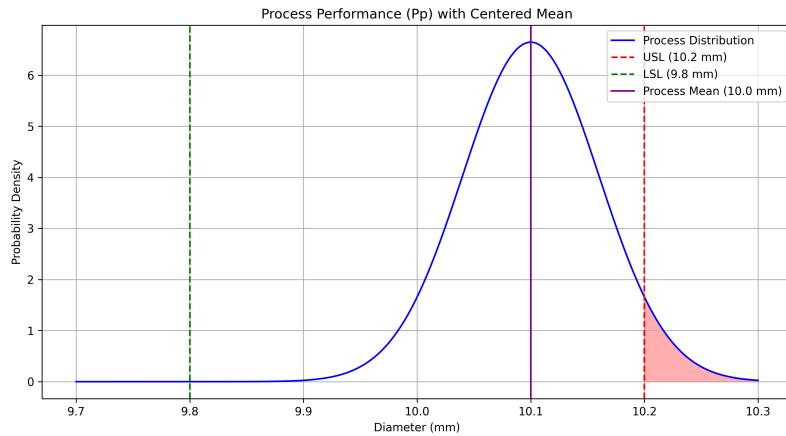
$C_p = 1.33$  means the process spread ( $6\sigma$ ) fits 1.33 times within the tolerance range (the distance between the Upper Specification Limit and Lower Specification Limit).

- $C_p = 1.00$ : Process mean is exactly at the midpoint of the specification limits, and the process variation fits exactly within these limits. 99.73% of the output will be within specifications, indicating a capable process (3 sigma process).
- $C_p > 1.00$ : The higher the  $C_{pk}$ , the more capable and stable the process is, meaning it can consistently produce parts within tolerance with minimal risk of defects.
- $C_p < 1$ : The process mean is off-center or the variation is wider than the specification limits, or both. A significant portion of the output may fall outside the specification limits.

### 13.3. $C_{\text{pm}}$ (Taguchi Capability Index)

### 13.4. $P_p$ (Process Performance Index)

$$P_p = \frac{\text{USL} - \text{LSL}}{6\sigma_{\text{overall}}}$$



Suppose a company manufactures metal rods, and the specification limits for the diameter of the rods are:

- Upper Specification Limit (**USL**): 10.2 mm
- Lower Specification Limit (**LSL**): 9.8 mm
- Overall standard deviation ( $\sigma_{\text{overall}}$ ): 0.06 mm

#### Step 1: Determine the Specification Width

The specification width is the difference between the USL and LSL.

$$\text{Specification Width} = \text{USL} - \text{LSL} = 10.2 \text{ mm} - 9.8 \text{ mm} = 0.4$$

#### Step 2: Calculate the Process Performance Index $P_p$

The formula for  $P_p$  is:

$$P_p = \frac{\text{Specification Width}}{6\sigma_{\text{overall}}} = \frac{\text{USL} - \text{LSL}}{6\sigma_{\text{overall}}}$$

Substitute the values:

$$P_p = \frac{0.4 \text{ mm}}{6 \times 0.06 \text{ mm}} = \frac{0.4 \text{ mm}}{0.36 \text{ mm}} = 1.11$$

Interpretation:

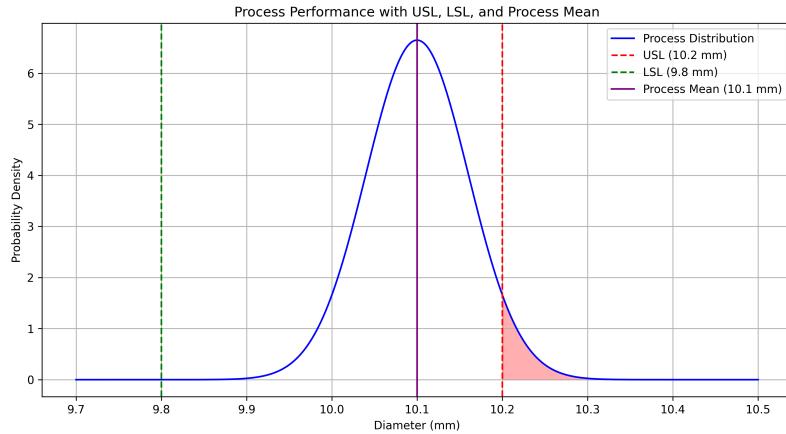
- A  $P_p$  of 1.11 indicates that the process performance, considering all sources of variation over time, is capable but less so than the potential capability indicated by  $C_p$ . The value being slightly above 1 suggests that the process can generally produce rods within specifications, but there might be more variability in the process compared to the short-term capability measured by  $C_p$
- Decrease from the  $C_p$  value (1.33 to 1.11) reflects the impact of additional variability when evaluating the process over a longer time or under different conditions.

Use:

When you need to evaluate the **actual performance** of a process over a longer period, considering all sources of variation, including shifts, drifts, and other long-term factors.

### 13.5. $P_{pk}$ (Process Performance Index with Centering)

$$P_{pk} = \text{Min} \left( \frac{\text{USL} - \mu_{\text{overall}}}{3\sigma}, \frac{\mu_{\text{overall}} - \text{LSL}}{3\sigma} \right)$$



Suppose a company manufactures metal rods, and the specification limits for the diameter of the rods are:

- Upper Specification Limit (**USL**): 10.2 mm
- Lower Specification Limit (**LSL**): 9.8 mm
- Overall standard deviation ( $\sigma_{\text{overall}}$ ): 0.06 mm
- Overall process mean ( $\mu_{\text{overall}}$ ): 10.1 mm

**Step 1:** Calculate the Distance from the Process Mean to the Specification Limits

Calculate the distance from the overall process mean to both the USL and LSL:

$$\text{USL} - \mu_{\text{overall}} = 10.2 \text{ mm} - 10.1 \text{ mm} = 0.1 \text{ mm}$$

$$\mu_{\text{overall}} - \text{LSL} = 10.1 \text{ mm} - 9.8 \text{ mm} = 0.3 \text{ mm}$$

**Step 2:** Calculate the Process Performance Index  $P_k$

The formula for  $P_{pk}$  is:

$$P_{pk} = \text{Min} \left( \frac{\text{USL} - \mu_{\text{overall}}}{3\sigma}, \frac{\mu_{\text{overall}} - \text{LSL}}{3\sigma} \right)$$

Substitute the values:

$$P_{pk} = \text{Min} \left( \frac{0.1 \text{ mm}}{3 \times 0.06 \text{ mm}}, \frac{0.3 \text{ mm}}{3 \times 0.06 \text{ mm}} \right)$$

$$P_{pk} = \text{Min} \left( \frac{0.1 \text{ mm}}{0.18 \text{ mm}}, \frac{0.3 \text{ mm}}{0.18 \text{ mm}} \right)$$

$$P_{pk} = \min(0.56, 1.67) = 0.56$$

Interpretation:

- A  $P_p$  of 1.11 indicates that the process performance, considering all sources of variation over time, is capable but less so than the potential capability indicated by  $C_p$ . The value being slightly above 1 suggests that the process can generally produce rods within specifications, but there might be more variability in the process compared to the short-term capability measured by  $C_p$
- Decrease from the  $C_p$  value (1.33 to 1.11) reflects the impact of additional variability when evaluating the process over a longer time or under different conditions.

Use:

When you need to evaluate the **actual performance** of a process over a longer period, considering all sources of variation, including shifts, drifts, and other long-term factors.

## 14. Correlation

### 14.1. Pearson

Linear relationships

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \times (y_i - \bar{y})^2}}$$

Interpretation:

- 1: perfect positive linear relationship
- -1: perfect negative linear relationship
- 0: no linear relationship

$$X : [1, 2, 3, 4, 5]$$

$$Y : [2, 4, 6, 8, 10]$$

**Step 1:** Calculate the Means of X and Y

$$\bar{X} = 3$$

$$\bar{Y} = 6$$

**Step 2:** Calculate the Differences from the Mean

### 14.2. Spearman's Rank

Non-linear relationships

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

Where

- $\rho$ : Spearman rank correlation coefficient
- $d_i$ : difference between the ranks of corresponding values
- $n$ : number of observations

Rank: position of a value within a data set when the values are ordered in ascending order

Observation $i$	$X$	Rank $X$	$Y$	Rank $Y$	$d_i = \text{Rank } X - \text{Rank } Y$	$d_i^2$
1	3	3	8	5	-2	4
2	1	1	6	3	-2	4
3	4	4	7	4	0	0
4	2	2	4	1	1	1
5	5	5	5	2	3	9

Interpretation

- $\rho = 1$ : Perfect positive correlation
- $\rho = -1$ : Perfect negative correlation
- $\rho = 0$ : No correlation

## 15. Non-Parametric Statistics

### 15.1. Mann-Whitney U

Determine whether there is a significant difference between the distributions of two independent samples (used as an alternative to the independent samples t-test when the assumptions of normality are not met)

### 15.2. Wilcoxon Signed-Rank

Compare two paired samples or to assess whether the median of a single sample is different from a specified value (used as a non-parametric alternative to the paired t-test when the data does not meet the assumptions of normality)

### 15.3. Kolmogorov-Smirnov

Determine if a sample is drawn from a population with a specific distribution.

It compares the empirical distribution function (EDF) of the sample with the cumulative distribution function (CDF) of the reference distribution.

### 15.4. Kruskal-Wallis

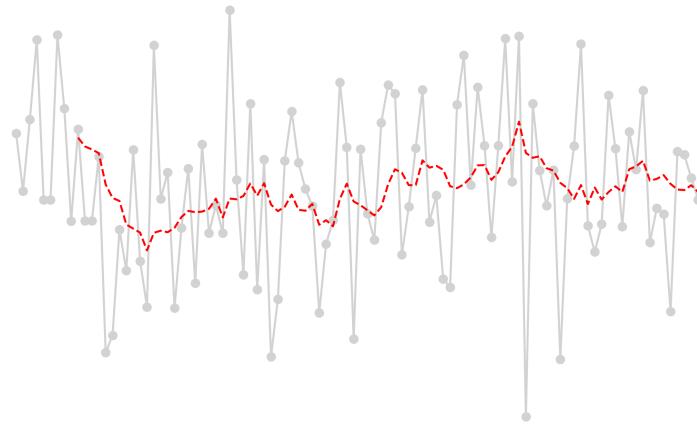
Determine if there are significant differences between the medians of three or more independent groups (extension of the Mann-Whitney U test, which is used for comparing two groups)

## 16. Time Series

### 16.1. SMA (Simple Moving Averages)

Creating a series of averages of different subsets (i.e., window) of the full data set

Minimize impact of short-term fluctuations



### 3-Day Simple Moving Average for the 7 Day Time Series

#### Data

- Day 1: \$10
- Day 2: \$12
- Day 3: \$14
- Day 4: \$16
- Day 5: \$18
- Day 6: \$20
- Day 7: \$22

$$\text{SMA for Day 3} = \frac{\text{Day 1} + \text{Day 2} + \text{Day 3}}{3}$$

$$\text{SMA for Day 4} = \frac{\text{Day 2} + \text{Day 3} + \text{Day 4}}{3}$$

$$\text{SMA for Day 5} = \frac{\text{Day 3} + \text{Day 4} + \text{Day 5}}{3}$$

$$\text{SMA for Day 6} = \frac{\text{Day 4} + \text{Day 5} + \text{Day 6}}{3}$$

$$\text{SMA for Day 7} = \frac{\text{Day 5} + \text{Day 6} + \text{Day 7}}{3}$$

```
pl.col('X').rolling_mean(window_size)
```

## 16.2. WMA (Weighted Moving Average)

Each data point in the window is assigned a specific weight (usually decrease linearly)

$$\text{WMA} = \frac{\sum_{i=1}^n (x_i w_i)}{\sum_{i=1}^n w_i}$$

### Data

- Day 1: \$10
- Day 2: \$12
- Day 3: \$14
- Day 4: \$13
- Day 5: \$15

### Weights

- 1st most recent: 3
- 2nd most recent: 2
- 3rd most recent: 1

$$\text{WMA for Day 3} = \frac{(14 \times 3) + (12 \times 2) + (10 \times 1)}{3 + 2 + 1} = 12.67$$

$$\text{WMA for Day 4} = \frac{(13 \times 3) + (14 \times 2) + (12 \times 1)}{3 + 2 + 1} = 13.17$$

$$\text{WMA for Day 5} = \frac{(15 \times 3) + (13 \times 2) + (14 \times 1)}{3 + 2 + 1} = 14.7$$

### 16.3. Exponential Smoothing

Gives more weight to recent data points (more responsive to new information)

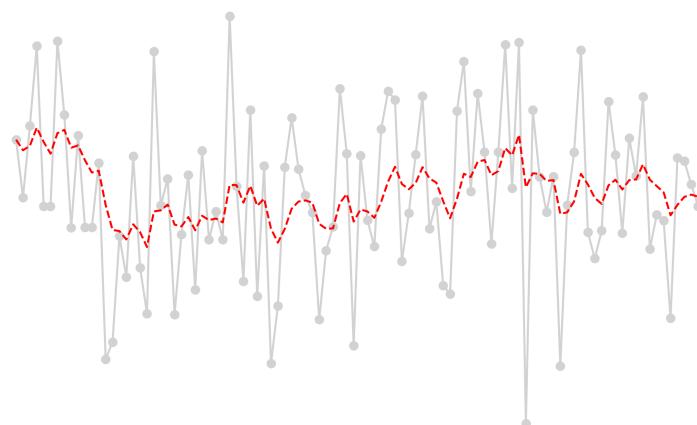
$$\text{EMA}_t = \alpha X_t + (1 - \alpha)\text{EMA}_{t-1}$$

Where

- $\alpha$ : smoothing factor

$$\alpha = \frac{2}{N + 1}$$

- $N$ : Period (i.e., window size) of the EMA



## Data

- Day 1: \$10
- Day 2: \$12
- Day 3: \$14
- Day 4: \$13
- Day 5: \$15

## Calculate Smoothing Factor

$$\alpha = \frac{2}{3+1} = 0.5$$

## Calculate SMA for First Value

$$\text{SMA for Day 3} = \frac{10 + 12 + 14}{3}$$

## Calculate EMA

$$\text{EMA for Day 4} = (0.5 \times 13) + (12 \times 0.5) = 12.5$$

$$\text{EMA for Day 5} = (0.5 \times 15) + (12.5 \times 0.5) = 13.75$$

```
pl.col("X").ewm_mean(span=window_size, adjust=False)
```

## 16.4. Seasonal Decomposition

- Level: baseline value around which the time series fluctuates
- Trend: long-term progression or direction of the time series
- Seasonality: regular, repeating patterns or cycles in the time series
- Noise: random fluctuations or irregular variations (cannot be explained by the trend, seasonality, or other components)

### Additive Decomposition

When the magnitude of seasonal fluctuations and trend does not change over time

$$Y_t = L_t + T_t + S_t + N_t$$

Where

- $L_t$ : Level at time  $t$
- $T_t$ : Trend at time  $t$
- $S_t$ : Seasonal component at time  $t$
- $N_t$ : Noise (residuals) at time  $t$

### Multiplicative Decomposition

When the magnitude of seasonal fluctuations and trend change proportionally over time

$$Y_t = L_t \times T_t \times S_t \times N_t$$

Where

- $L_t$ : Level at time  $t$
- $T_t$ : Trend at time  $t$
- $S_t$ : Seasonal component at time  $t$
- $N_t$ : Noise (residuals) at time  $t$

## 16.5. ARMA (AutoRegressive Moving Average)

## 16.6. ARIMA (AutoRegressive Integrated Moving Average)

# 17. Inventory Management

## 17.1. Newsvendor

determine the optimal order quantity  $Q^*$  that minimizes the total expected cost or maximizes the expected profit, based on the trade-off between the overage and underage costs

Assumptions

- Products are separable
- Planning is done for a single period
- Demand is random
- Deliveries are made in advance of demand
- Costs of overage or underage are linear

### 1. Parameters

- $P$ : Sale Price
- $C$ : Purchase Cost
- $S$ : Unsold Value
- $\mu$ : Mean Demand
- $\sigma$ : Standard Deviation Demand

### 2. Calculate Underage and Overage Costs:

- Underage Cost ( $C_u$ ): Profit lost for each unit of demand not met

$$C_u = P - C$$

- Overage Cost ( $C_o$ ): This is the cost of holding an unsold newspaper.

$$C_o = C - S$$

### 3. Calculate Critical Ratio ( $CR$ )

$$CR = \frac{C_u}{C_u - C_o}$$

### 4. Find z-score: Find the number of standard deviations away from the mean corresponding to the critical ratio:

$$z^* = \Phi^{-1}(CR)$$

Where:

- $\Phi^{-1}$ : Inverse of the CDF of the standard normal distribution (PPF)

### 5. Calculate Optimal Order Quantity ( $Q^*$ )

$$Q^* = \mu + z^* \sigma$$

Where:

- $z^*$ : z-score corresponding to the critical ratio  $CR$  from the standard normal distribution

Consider a newsvendor selling newspapers:

1. Parameters

- Sale Price ( $P$ ): \$3 (per unit)
- Purchase Cost ( $C$ ): \$1 (per unit)
- Unsold Value ( $S$ ): \$0 (per unit)
- Mean Demand ( $\mu$ ): 100 (units)
- Standard Deviation Demand ( $\sigma$ ): 20 (units)

2. Calculate Underage and Overage Costs:

- Underage Cost ( $C_u$ ): Profit lost for each unit of demand not met

$$C_u = P - C = 3 - 1 = 2$$

- Overage Cost ( $C_o$ ): This is the cost of holding an unsold newspaper.

$$C_o = C - S = 1 - 0 = 1$$

3. Calculate Critical Ratio ( $CR$ )

$$CR = \frac{C_u}{C_u - C_o} = \frac{2}{2+1} = \frac{2}{3} = 0.67$$

4. Find z-score: Find the number of standard deviations away from the mean corresponding to the critical ratio:

$$z^* = \Phi^{-1}(CR) = 0.44$$

5. Calculate Optimal Order Quantity ( $Q^*$ )

$$Q^* = \mu + z^* \sigma = 100 + 0.44 \cdot 20 = 108.8$$

```

import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt

# Parameters for the newsvendor example
selling_price = 3 # Selling price per newspaper
purchase_cost = 1 # Purchase cost per newspaper
unsold_value = 0 # Value of unsold newspapers
mu = 100 # Mean of demand
sigma = 20 # Standard deviation of demand

# Calculate underage and overage costs
C_u = selling_price - purchase_cost # Underage cost
C_o = purchase_cost - unsold_value # Overage cost

# Calculate the critical ratio
CR = C_u / (C_u + C_o)

```

```

# Assume a normal distribution for demand
mu = 50      # Mean demand
sigma = 10    # Standard deviation of demand

# Calculate the optimal order quantity (Q*)
z_star = stats.norm.ppf(CR)  # z-score corresponding to the critical ratio
Q_star = mu + z_star * sigma # Optimal order quantity

```

## 17.2. ABC Analysis

categorizes inventory items into three groups (A, B, and C) based on their importance

- A: Top 70-80% of the total annual consumption value
- B: Next 15-25% of the total annual consumption value
- C: Remaining 5-10% of the total annual consumption value

Item	Usage Quantity (Annual)	Unit Cost	Consumption Value (Annual)
$I_1$	50	\$100	\$5000
$I_2$	150	\$20	\$3000
$I_3$	300	\$10	\$3000
$I_4$	400	\$5	\$2000
$I_5$	500	\$1	\$500

**Step 1:** Calculate Annual Consumption Values

**Step 2:** Sort Items by Annual Consumption Value (Descending)

**Step 3:** Calculate Total Annual Consumption Value

$$\text{Total} = \$5,000(I_1) + \$3,000(I_2) + \$3,000(I_3) + \$2,000(I_4) + \$500(I_5) = \$13500$$

**Step 4:** Calculate Cumulative Consumption Value Percentages

- $I_1 : \frac{5000}{13500} \times 100\% = 37.04\%$
- $I_2 : \frac{3000}{13500} \times 100\% = 22.22\%$
- $I_3 : \frac{3000}{13500} \times 100\% = 22.22\%$
- $I_4 : \frac{2000}{13500} \times 100\% = 14.81\%$
- $I_5 : \frac{500}{13500} \times 100\% = 3.7\%$

**Step 5:** Cumulative Percentages

- $I_1 : 37.04\%$
- $I_1 + I_2 : 59.26\%$
- $I_1 + I_2 + I_3 : 81.48\%$

- $I_1 + I_2 + I_3 + I_4$ : 96.3%
- $I_1 + I_2 + I_3 + I_4 + I_5$ : 100%

#### Step 6: Categorize Items

- A:  $I_1, I_2, I_3$
- B:  $I_4$
- C:  $I_5$

### 17.3. Fill Rate

Percentage of customer demand that is satisfied from available inventory

$$F = \frac{U_f}{U_o} \times 100\%$$

Where:

- $F$ : Fill Rate
- $U_f$ : Number of units fulfilled
- $U_o$ : Total number of units ordered

Customers order 100 units of a product and 90 units are fulfilled from stock:

$$F = \frac{90}{100} \times 100\% = 90\%$$

### 17.4. (OCT) Order Cycle Time

Measures the total time taken from when a customer places an order to when the order is delivered

$$OCT = T_{\text{order}} + T_{\text{processing}} + T_{\text{production}} + T_{\text{shipping}}$$

Where:

- $T_{\text{order}}$ : Order Entry Time (time it takes to receive and log the order)
- $T_{\text{processing}}$ : Order Processing Time (time to check inventory, verify details, and prepare for production or shipment)
- $T_{\text{production}}$ : Production Time (time to manufacture or prepare the product)
- $T_{\text{shipping}}$ : Shipping Time (time it takes to deliver the product from the warehouse to the customer)

### 17.5. ROP (Reorder Point)

The inventory level at which a new order should be placed to avoid stockouts

$$ROP = (\text{Average Demand per Period} \times \text{Lead Time})$$

Suppose your business sells 50 units per week, and the lead time for a new order is 2 weeks.  
Using the formula:

$$ROP = 50(\text{units per week}) \times 2(\text{weeks}) = 100 \text{ units}$$

This means that when your inventory level drops to 100 units, you should place a new order to avoid running out of stock.

## 17.6. XYZ Analysis

Categorization based on variability

- X: Low variability
- Y: Moderate variability
- Z: High variability

Coefficient of Variation:

$$CV = \frac{\sigma}{\mu}$$

- X Items: Low CV
  - $CV < k_1$ , where  $k_1$  is the threshold value indicating low variability
- Y Items: Moderate CV
  - $CV < k_2$ , where  $k_2$  is the moderate value indicating low variability
- Z Items: High CV
  - $CV < k_3$ , where  $k_3$  is the threshold value indicating high variability

### Step 1: Collect Historical Data (12 months)

- Product A: [100, 105, 98, 102, 101, 104, 103, 100, 99, 100, 101, 102]
- Product B: [150, 155, 145, 160, 140, 150, 155, 150, 165, 155, 150, 140]
- Product C: [200, 180, 220, 190, 210, 240, 180, 230, 220, 210, 250, 190]

### Step 2: Calculate the Mean and Standard Deviation

- Product A:
  - $\mu = 101$
  - $\sigma = 2$
- Product B:
  - $\mu = 150$
  - $\sigma = 8$
- Product C:
  - $\mu = 210$
  - $\sigma = 25$

### Step 3: Calculate the Coefficient of Variation (CV)

- $CV = \frac{2}{101} = 0.0198$
- $CV = \frac{8}{150} = 0.0533$
- $CV = \frac{25}{210} = 0.1190$

### Step 4: Categorization

- X: Product A (Low variability)
- Y: Product B (Moderate variability)

- Z: Product C (High variability)

## 17.7. EOQ (Economic Order Quantity)

Optimal order quantity that **minimizes the total cost**, which includes both **holding costs** and **ordering costs**

$$\text{EOQ} = \sqrt{\frac{2DS}{H}}$$

Where:

- $D$ : Demand per time period
- $S$ : Ordering cost per order
- $H$ : Holding cost per unit, per time period

A company sells widgets and wants to determine the optimal order quantity for inventory.

- The annual demand for widgets ( $D$ ) is 12,000 units.
- The cost to place an order ( $S$ ) is \$50.

The holding cost per unit per year ( $H$ ) is \$2.

$$\text{EOQ} = \sqrt{\frac{2 \times 12000 \times 50}{2}} = 775$$

The company should order 775 widgets each time they place an order to minimize the total cost, which includes both ordering and holding costs

### 17.7.1. Perfect Order Rate

Measures the percentage of orders delivered to customers in full, on time, and without any damage

$$\text{Perfect Order Rate} = \frac{\text{Number of Perfect Orders}}{\text{Total Number of Orders}} \times 100\%$$

Where:

- **Number of Perfect Orders**: The number of orders that are delivered on time, complete, and undamaged.
- **Total Number of Orders**: The total number of orders fulfilled within a specific period.

Suppose you received 1,000 orders over a quarter, and 900 of those orders were delivered on time, complete, and without damage.

$$\text{Perfect Order Rate} = \frac{900}{1000} \times 100\% = 90\%$$

## 17.8. Safety Stock

Additional quantity of inventory kept on hand to protect against uncertainties in demand or supply. buffer to prevent stockouts due to unexpected variations in demand or delays in supply.

### 1. Constant Demand & Constant Lead Time

$$\text{SS} = Z \times \sigma_D$$

Where:

- $Z$ : Z-score corresponding to the desired service level
- $\sigma$ : standard deviation of demand

## 2. Variable Demand & Constant Lead Time

$$SS = Z \times \sigma_D \times \sqrt{L}$$

Where:

- $Z$ : Z-score corresponding to the desired service level
- $\sigma_D$ : Standard deviation of demand per unit of time
- $L$ : Lead time

## 3. Variable Lead Time & Constant Demand

$$SS = Z \times \bar{D} \times \sigma_L$$

Where:

- $Z$ : Z-score corresponding to the desired service level
- $\bar{D}$ : Average demand
- $\sigma_L$ : Standard deviation of lead time

## 4. Variable Demand & Variable Lead Time

$$SS = Z \times \sqrt{(\bar{D}^2 \times \sigma_L^2) + (L \times \sigma_D^2)}$$

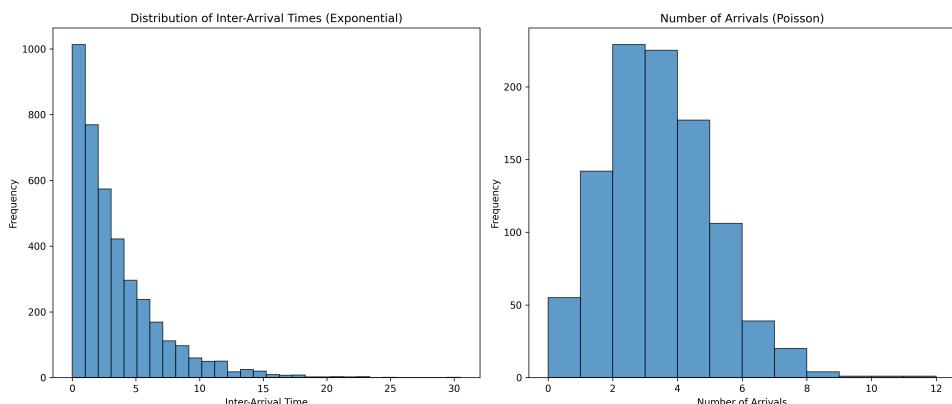
Where:

- $\bar{L}$ : Average Lead Time (average time it takes to receive inventory after placing an order)
- $\sigma_L^2$ : Demand Variance (variability in demand during the lead time)
- $\bar{D}$ : Average Demand (mean quantity of demand per time period)
- $\sigma_D^2$ : Lead Time Variance (variability in lead time)

# 18. Queuing Theory

## 18.1. M/M/1

- Arrival rate ( $\lambda$ ): Average number of customers arriving per unit of time
- Service rate ( $\mu$ ): Average number of customers served per unit of time



### 1. Utilization ( $\rho$ )

Fraction of time the server is busy

$$\rho = \frac{\lambda}{\mu}$$

2. Average Number of Customers in the System ( $L$ )

Average number of customers (both waiting and being served) in the system

$$L = \frac{\rho}{1 - \rho}$$

3. Average Number of Customers in the Queue ( $L_q$ )

Average number of customers waiting in the queue

$$L_q = \frac{\rho^2}{1 - \rho}$$

4. Average Time a Customer Spends in the System ( $W$ )

Average time a customer spends in the system (from arrival until they are done being served)

$$W = \frac{1}{\mu - \lambda}$$

5. Average Waiting Time in the Queue ( $W_q$ )

Average time a customer spends just waiting in line before being served

$$W_q = \frac{\rho}{\mu - \lambda}$$

6. Probability that the System is Empty ( $P_0$ )

Probability that there are zero customers in the system (no one is being served and no one is waiting)

$$P_0 = 1 - \rho$$

7. Probability that  $n$  Customers are in the System ( $P_n$ )

Probability that there are  $n$  customers in the system (either waiting or being served)

$$P_n = (1 - \rho) \cdot \rho^n$$

8. Probability the Queue is Full (if the Queue has Limited Capacity) ( $P_{n_{\max}}$ )

Probability that the system is at full capacity

$$P_{n_{\max}} = (1 - \rho) \cdot \rho^{n_{\max}}$$

9. System Throughput

Rate at which customers are served and leave the system

$$\text{Throughput} = \lambda$$

10. Expected Time in Service ( $W_s$ )

Average time a customer spends actually being served (not including waiting time)

$$W_s = \frac{1}{\mu}$$

### 11. Idle Time ( $1 - \rho$ )

Fraction of time that the server is idle (i.e., not serving any customers)

$$\text{Idle Time} = 1 - \rho$$

### 12. Probability of Having to Wait in the Queue ( $P_w$ )

Probability that an arriving customer will have to wait before being served, i.e., that the server is busy when the customer arrives

$$P_w = \rho$$

### 13. Variance of the Number of Customers in the System ( $\text{Var}(L)$ )

Variance of the number of customers in the system

$$\text{Var}(L) = \frac{\rho}{(1 - \rho)^2}$$

A bank with a single teller

- **Arrival rate ( $\lambda$ ):** On average, 4 customers arrive every 10 minutes ( $\lambda = 4$  customers per 10 minutes)
- **Service Rate ( $\mu$ ):** The teller can serve 6 customers every 10 minutes ( $\mu = 6$  customers per 10 minutes)

#### 1. Utilization ( $\rho$ )

$$\rho = \frac{\lambda}{\mu} = \frac{4}{6} = 0.67$$

The teller is busy 67% of the time. The remaining 33% of the time, the teller is idle, waiting for the next customer

#### 2. Average Number of Customers in the System ( $L$ )

$$L = \frac{\rho}{1 - \rho} = \frac{0.67}{1 - 0.67} = 2$$

On average, there are 2 customers in the coffee shop at any given time, either being served or waiting in line

#### 3. Average Number of Customers in the Queue ( $L_q$ )

$$L_q = \frac{\rho^2}{1 - \rho} = \frac{0.67^2}{1 - 0.67} = 1.33$$

On average, about 1.33 customers are waiting in line at any time

#### 4. Average Time a Customer Spends in the System ( $W$ )

$$W = \frac{1}{\mu - \lambda} = \frac{1}{6} - 4 = 0.5$$

On average, a customer spends 5 minutes ( $0.5 \times 10$  minutes) in the shop (including both waiting in line and getting served)

#### 5. Average Waiting Time in the Queue ( $W_q$ )

$$W_q = \frac{\rho}{\mu - \lambda} = \frac{0.67}{6 - 4} = 0.33$$

On average, a customer waits 3.3 minutes ( $0.33 \times 10$  minutes) in line before being served by the teller

6. Probability that the System is Empty ( $P_0$ )

$$P_0 = 1 - \rho = 1 - 0.67 = 0.33$$

There is a 33% chance that the coffee shop is empty, meaning there is no customer in the queue or being served

7. Probability that  $n$  Customers are in the System ( $P_n$ )

$$P_n = (1 - \rho) \cdot \rho^n = (1 - 0.67) \cdot 0.67^2 = 0.148$$

There is a 14.8% chance that exactly 2 customers are either in line or being served

8. Probability the Queue is Full (if the Queue has Limited Capacity) ( $P_{n_{\max}}$ )

$$P_{n_{\max}} = (1 - \rho) \cdot \rho^{n_{\max}} = (1 - 0.67) \cdot 0.67^5 = 0.028$$

There is a 2.8% chance that the system is full, and no new customers can enter

9. System Throughput

$$\text{Throughput} = \lambda = 4$$

The coffee shop serves 4 customers every 10 minutes, on average

10. Expected Time in Service ( $W_s$ )

$$W_s = \frac{1}{\mu} = \frac{1}{6} = 0.167$$

On average, a customer spends 1.67 minutes being served by the teller

11. Idle Time ( $1 - \rho$ )

$$\text{Idle Time} = 1 - \rho = 1 - 0.67 = 0.33$$

The teller is idle 33% of the time

12. Probability of Having to Wait in the Queue ( $P_w$ )

$$P_w = \rho = 0.67$$

There is a 67% chance that a customer will have to wait when they arrive

13. Variance of the Number of Customers in the System ( $\text{Var}(L)$ )

$$\text{Var}(L) = \frac{\rho}{(1 - \rho)^2} = \frac{0.67}{(1 - 0.67)^2} = 6.12^2$$

The queue length varies significantly, with a variance of 6.12 customers

Costs

- **Cost per Waiting Customer per Unit Time ( $C_w$ )**: cost incurred per customer for each unit of time they spend waiting in the queue

$$\text{Total Waiting Cost} = L_q \times C_w \times \text{Unit Time}$$

- **Cost per Idle Server per Unit Time ( $C_s$ ):** cost incurred per unit time when the server is not serving customers

$$\text{Total Idle Cost} = (1 - \rho) \times C_s \times \text{Unit Time}$$

Total Cost

$$\text{Total Cost} = \text{Total Waiting Cost} + \text{Total Idle Cost}$$

- Arrival Rate ( $\lambda$ ): 4 customers per 10 minutes
- Service Rate ( $\mu$ ): 6 customers per 10 minutes
- Cost per Waiting Customer per Hour ( $C_w$ ): \$10
- Cost per Idle Server per Hour ( $C_s$ ): \$20
- Operational Time: 1 hour

1. Utilization

$$\rho = \frac{\lambda}{\mu} = \frac{4}{6} = 0.67$$

2. Average Number of Customers in Queue ( $L_q$ )

$$L_q = \frac{\rho^2}{1 - \rho} = \frac{0.67^2}{1 - 0.67} = 1.33 \text{ customers}$$

3. Total Waiting Cost

$$\text{Total Waiting Cost} = L_q \times C_w \times \text{Unit Time} = 1.33 \times 10 \times 1 = \$13.33$$

4. Idle Time

$$\text{Idle Time} = 1 - \rho = 1 - 0.67 = 0.33$$

5. Total Idle Cost

$$\text{Total Idle Cost} = (1 - \rho) \times C_s \times \text{Unit Time} = 0.33 \times 20 \times 1 = \$6.67$$

6. Total Cost

$$\text{Total Cost} = \text{Total Waiting Cost} + \text{Total Idle Cost} = 13.33 + 6.67 = \$20$$

## 19. Network Optimization

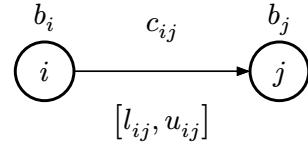
### 19.1. Shortest Path

### 19.2. Maximum Flow

### 19.3. Network Flow Optimization

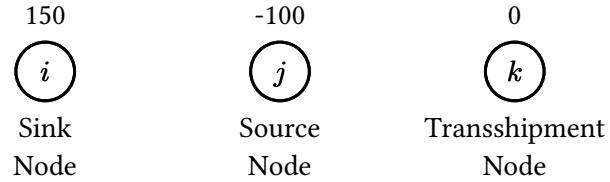
Types of Problems:

- Maximum Flow
- Minimum Cost Flow
- Multi-Commodity Flow



Node  $i$

- Sink node:  $b_i > 0$  Has demand of  $b_i$  units
- Source node:  $b_i < 0$  Has supply of  $-b_i$  units
- Transshipment node:  $b_i = 0$  Neither supply or demand



Units shipped from node  $i$  to node  $j$ :

$$x_{ij}$$

Minimize:

$$Z = \sum_{i,j} c_{ij} x_{ij}$$

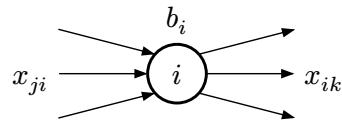
s.t.

$$\sum_k x_{ki} - \sum_l x_{il} = \text{or } \leq \text{ or } \geq b_1$$

$$l_{ij} \leq x_{ij} \leq u_{ij}$$

Where:

- $c_{ij}$ : Unit cost of flow from node  $i$  to node  $j$
- $b_i$ : Demand on node  $i$
- $l_{ij}$ : flow lower bound from  $i$  to  $j$
- $u_{ij}$ : Flow upper bound from  $i$  to  $j$
- $\sum_k x_{ki}$ : Inflow to  $i$
- $\sum_l x_{il}$ : Outflow from  $j$

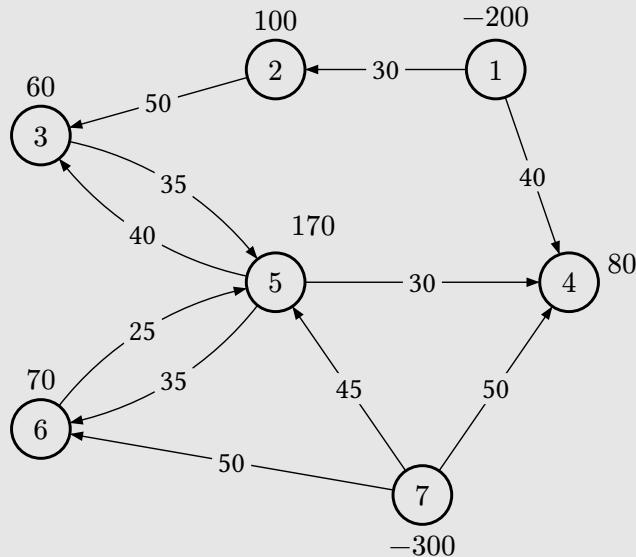


If:

- Total Supply = Total Demand  $[Inflow to i] - [Outflow from i] = b_i$
- Total Supply > Total Demand  $[Inflow to i] - [Outflow from i] \geq b_i$
- Total Supply < Total Demand  $[Inflow to i] - [Outflow from i] \leq b_i$

Important:

- One decision variable  $x_{ij}$  for each edge  $(i, j)$
- One flow balancing constraint for each node  $i$



Minimize

$$Z = 30x_{12} + 40x_{14} + 50x_{23} + 35x_{35} + 40x_{53} + 30x_{54} + 35x_{56} + 25x_{65} + 50x_{74} + 45x_{75} + 50x_{76}$$

s.t.

$$x_{12} + x_{14} \leq 200 \quad (\text{Node 1})$$

$$x_{12} + x_{23} \geq 100 \quad (\text{Node 2})$$

$$x_{23} + x_{53} - x_{35} \geq 60 \quad (\text{Node 3})$$

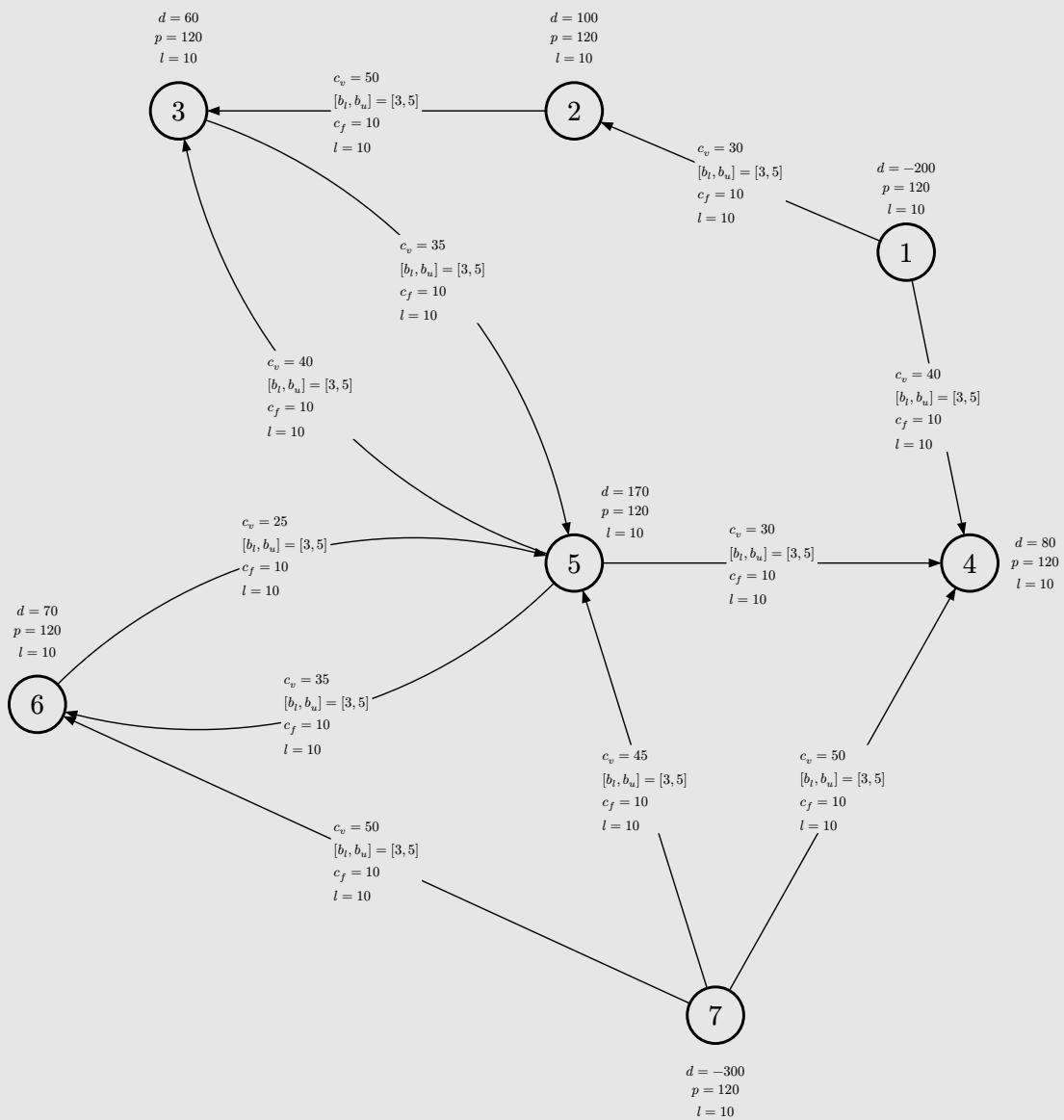
$$x_{14} + x_{54} + x_{74} \geq 80 \quad (\text{Node 4})$$

$$x_{35} + x_{65} + x_{75} - x_{53} - x_{54} - x_{56} \geq 170 \quad (\text{Node 5})$$

$$x_{56} + x_{76} - x_{65} \geq 70 \quad (\text{Node 6})$$

$$x_{76} + x_{75} + x_{74} \leq 300 \quad (\text{Node 7})$$

$$x_{ij} \geq 0 \quad \forall (i, j) \in E$$



Minimize:

$$\sum_{ij} (c_{ij}^f \cdot y_{ij}) + \sum_{ij} (c_{ij}^v \cdot x_{ij}) + \sum_i (c_i^f \cdot y_i) + \sum_i (p_i \cdot s_i) +$$

Where:

- $\sum_{ij} (c_{ij}^f \cdot y_{ij})$ : Edge fixed cost contribution
- $\sum_{ij} (c_{ij}^v \cdot x_{ij})$ : Variable cost contribution
- $\sum_i (c_i^f \cdot y_i)$ : Node fixed cost contribution
- $\sum_i (p_i \cdot s_i)$ : Penalty contribution
- $\sum_{ij} (l_{ij} \cdot x_{ij})$ : Edge lead time weighted by flow
- $\sum_i l_i \cdot \sum_j x_{ij}$ : Node service time weighted by flow

s.t.

$$\begin{aligned}
 \sum_j x_{ji} - \sum_j x_{ij} &= \text{or } \leq \text{ or } \geq d_i && \text{Flow Conservation} \\
 b_{ij}^l \leq x_{ij} \leq b_{ij}^u & && \text{Lower & Upper Flow Bound} \\
 x_{ij} \leq M \cdot y_{ij} & && \text{Fixed Cost Route} \\
 \sum_i (x_{ji} + x_{ij}) \leq M \cdot y_i & && \text{Fixed Cost Node} \\
 \sum_i x_{ij} + s_j \geq d_j & && \text{Unmet Demand Penalty} \\
 x_{ij} \geq 0 \quad \forall (i, j) \in E & && \text{Non Negative Flow}
 \end{aligned}$$

Node Properties:

- **Node Type**: Source, sink, or intermediary.
- **Supply (Source)**: Flow capacity of source nodes.
- **Holding Cost (Intermediary)**: Inventory cost for stored goods.
- **Service Time**: Processing time at the node.
- **Demand**: Required flow at sink nodes.
- **Storage Capacity (Intermediary)**: Maximum amount of goods that can be held at a node
- **Penalty for Unfulfilled Demand (Sink)**: Cost for unmet demand.
- **Disruption Risk (All Nodes)**: Probability of a node being unavailable due to unforeseen circumstances

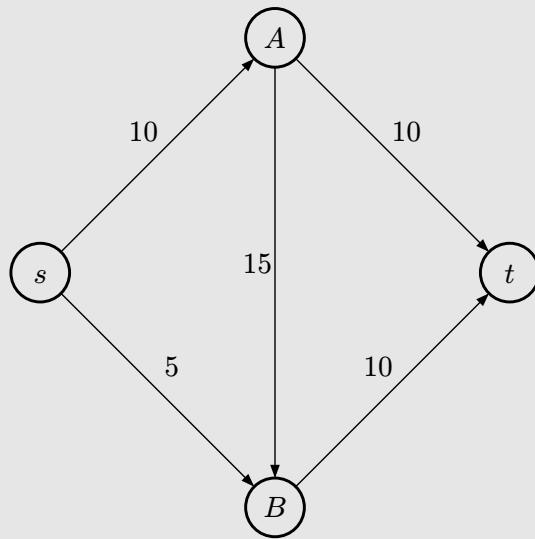
Edge Properties:

- **Edge Type**: Transport mode (air, water, road, rail).
- **Fixed Cost**: Cost incurred for using the edge, regardless of flow.
- **Reliability**: Probability of edge availability.
- **Flow Bounds**: Minimum and maximum allowable flow.
- **Unit Cost**: Cost per unit of flow.
- **Lead Time**: Time it takes for flow to travel along the edge.

- **Environmental Impact:** Account for the carbon footprint of using certain transport modes.

#### 19.4. Ford-Fulkerson

Find augmenting paths in the network and increase the flow until no more augmenting paths can be found



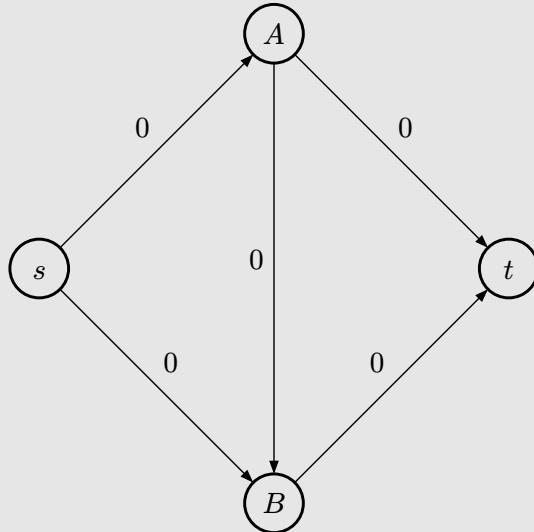
- $s$ : Source
- $A$  &  $B$ : Intermediate nodes
- $t$ : Sink

Capacities:

- $s \rightarrow t$ : 10
- $s \rightarrow B$ : 5
- $A \rightarrow B$ : 15
- $A \rightarrow t$ : 10
- $B \rightarrow t$ : 10

**Step 1:** Initialize flow to 0

All flows through the edges are initially set to 0.



**Step 2:** Find an augmenting path

Find an augmenting path using Depth-First Search (DFS). Start from the source  $s$

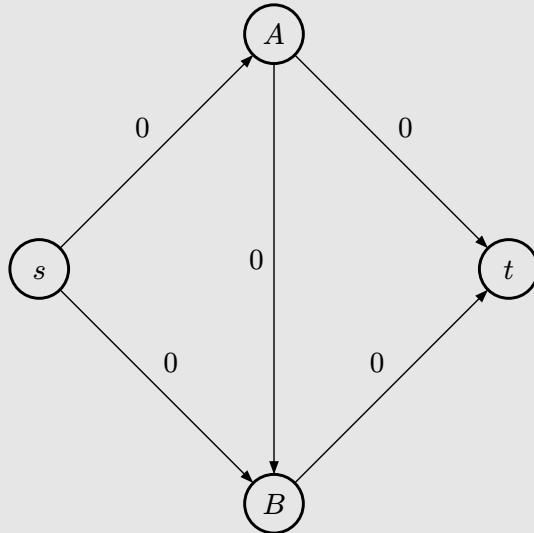
$$s \rightarrow A \rightarrow t$$

The minimum capacity along this path is 10 (bottleneck on edge  $A \rightarrow t$ )

We can push a flow of 10 units along this path.

**Step 3:** Update the residual graph

- $s \rightarrow A$ : Capacity becomes  $10 - 10 = 0$  (no residual capacity)
- $A \rightarrow t$ : Capacity becomes  $10 - 10 = 0$  (no residual capacity)



**Step 4:** Find another augmenting path

Find another augmenting path.

We cannot use  $s \rightarrow A$  or  $A \rightarrow t$  because their capacities are 0.

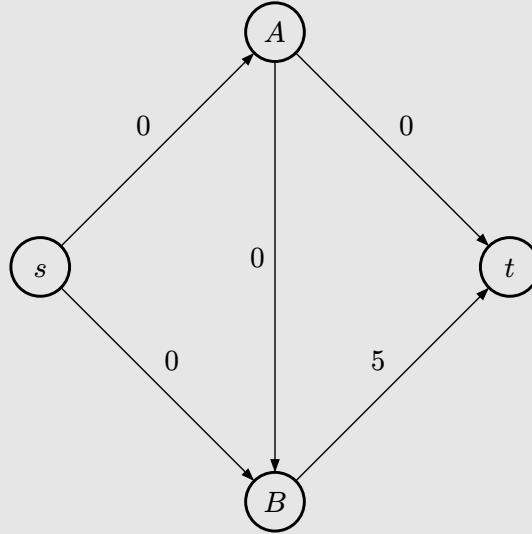
$$s \rightarrow B \rightarrow t$$

The minimum capacity along this path is 5 (bottleneck on edge  $s \rightarrow B$ )

We can push a flow of 5 units along this path

**Step 5:** Update the residual graph

- $s \rightarrow B$ : Capacity becomes  $5 - 5 = 0$
- $B \rightarrow t$ : Capacity becomes  $10 - 5 = 5$



**Step 6:** Termination

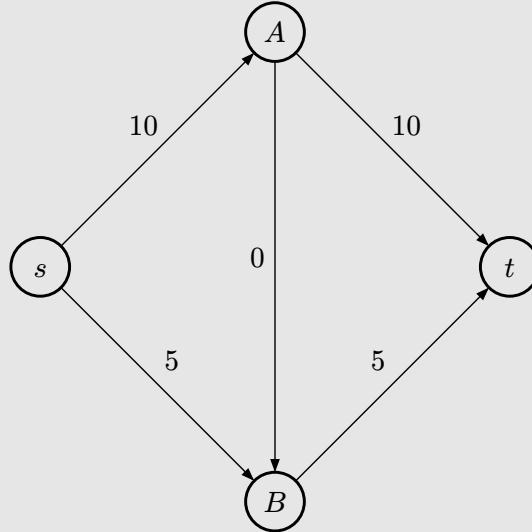
No more augmenting paths from  $s$  to  $t$  can be found, as all edges from  $s$  are fully saturated

Maximum flow:

- $s \rightarrow A \rightarrow t$ : 10 units
- $s \rightarrow B \rightarrow t$ : 5 units

Thus, the maximum flow from source  $s$  to sink  $t$  is 15 units

**Step 7:** Final Flow Distribution



## 20. Optimization

### 20.1. LP (Linear Programming)

Optimizing (maximizing or minimizing) a linear **objective function** subject to linear equality or inequality **constraints**. **Decision variables** can take any continuous real values.

#### 1. Objective Function

Maximize:

$$Z = c_1x_1 + c_2x_2 + \dots + c_nx_n$$

Or, equivalently:

$$Z = c^T x = \sum_{i=1}^n c_i x_i$$

Where:

- $Z$ : Objective Function
- $c_1, c_2, \dots, c_n$ : Coefficients
- $x_1, x_2, \dots, x_n$ : Decision Variables

#### 2. Constraints

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &\leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &\leq b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &\leq b_m \end{aligned}$$

Or, in matrix form:

$$Ax \leq b$$

Where:

- $A$ :  $m \times n$  matrix of coefficients  $a_{ij}$
- $x = (x_1, x_2, \dots, x_n)^T$ : vector of decision variables
- $b = (b_1, b_2, \dots, b_m)^T$ : vector of known constants

#### 3. Non-Negativity Constraints

$$x_i \geq 0 \text{ for } i = 1, 2, \dots, n$$

#### 1. Problem

A company produces two products:  $x_1$  (Product A) and  $x_2$  (Product B). The company wants to maximize profit, where:

- Each unit of Product A gives a profit of \$40.
- Each unit of Product B gives a profit of \$30.

The company has constraints on the production process:

- It takes 2 hours of labor to produce one unit of Product A and 1 hour to produce one unit of Product B. The company has a maximum of 100 labor hours available.

- The company can only use up to 80 units of raw material, and each unit of Product A uses 1 unit of material, while Product B uses 2 units of material.

The goal is to decide how many units of Product A ( $x_1$ ) and Product B ( $x_2$ ) to produce to maximize profit.

2. Formulation:

**Objective Function** (maximize the profit):

$$Z = 40x_1 + 30x_2$$

### Constraints

1. Labor (maximum 100 hours):

$$2x_1 + x_2 \leq 100$$

2. Raw material (maximum 60 units):

$$x_1 + 2x_2 \leq 80$$

3. Non-negativity (can't produce negative quantities):

$$x_1 \geq 0$$

$$x_2 \geq 0$$

4. Summary

Maximize:

$$Z = 40x_1 + 30x_2$$

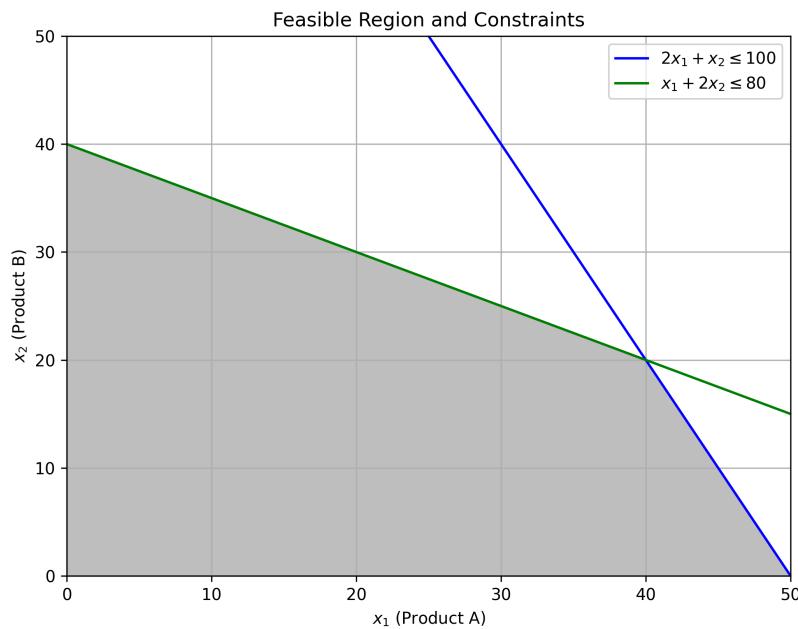
s.t.

$$2x_1 + x_2 \leq 100$$

$$x_1 + 2x_2 \leq 80$$

$$x_1 \geq 0$$

$$x_2 \geq 0$$



```

import pulp

# Initialize the problem
prob = pulp.LpProblem("Maximize Profit", pulp.LpMaximize)

# Define decision variables
x1 = pulp.LpVariable('x1', lowBound=0, cat='Continuous') # Product A
x2 = pulp.LpVariable('x2', lowBound=0, cat='Continuous') # Product B

# Objective function: Maximize 40*x1 + 30*x2
prob += 40 * x1 + 30 * x2, "Total Profit"

# Constraints
prob += 2 * x1 + x2 <= 100, "Labor Constraint"
prob += x1 + 2 * x2 <= 80, "Material Constraint"

# Solve the problem
prob.solve()

# Print the results
print("Status:", pulp.LpStatus[prob.status])
print(f"Optimal x1 (Product A): {pulp.value(x1)}")
print(f"Optimal x2 (Product B): {pulp.value(x2)}")
print(f"Maximum Profit: {pulp.value(prob.objective)}")

```

## 20.2. IP (Integer Programming)

Optimizing (maximizing or minimizing) a linear **objective function** subject to linear equality or inequality **constraints**. **Decision variables** can take any integer real values.

Minimize or Maximize

$$c^T x$$

s.t.

$$Ax \leq b$$

and

$$x \in \mathbb{Z}^n$$

Where

- $x$ : vector of decision variables
- $c$ : vector of coefficients for the objective function
- $A$ : matrix of constraint coefficients
- $b$ : vector of constraint constants
- $x \in \mathbb{Z}^n$ : each  $x_i$  of  $x$  must be integer values

You are organizing a small event and want to minimize costs. You have to decide how many chairs and tables to rent.

- Each chair ( $x_1$ ) costs \$5, each table ( $x_2$ ) costs \$20.
- You need at least 3 tables and 10 chairs.
- Your budget is \$100.

You can only rent whole numbers of chairs and tables.

Minimize

$$5x_1 + 20x_2$$

s.t.

$$\begin{aligned}x_1 &\geq 10 \\x_2 &\geq 3 \\5x_1 + 20x_2 &\leq 100\end{aligned}$$

$$x_1, x_2 \in \mathbb{Z}^+$$

```
import pulp

# Create a linear programming problem instance
# We are minimizing the cost
prob = pulp.LpProblem("Minimize_Cost", pulp.LpMinimize)

# Define decision variables
# x1 is the number of chairs
# x2 is the number of tables
x1 = pulp.LpVariable("x1", lowBound=10, cat='Integer')
x2 = pulp.LpVariable("x2", lowBound=3, cat='Integer')

# Objective function: Minimize 5*x1 + 20*x2
prob += 5 * x1 + 20 * x2, "Total_Cost"

# No additional constraints in this example, as the bounds cover the requirements

# Solve the problem
prob.solve()
```

```

# Print the results
print(f"Status: {pulp.LpStatus[prob.status]}")
print(f"Number of chairs (x1): {x1.varValue}")
print(f"Number of tables (x2): {x2.varValue}")
print(f"Total cost: {pulp.value(prob.objective)}")

```

### 20.3. Gradient Descent

Find the minimum of a function

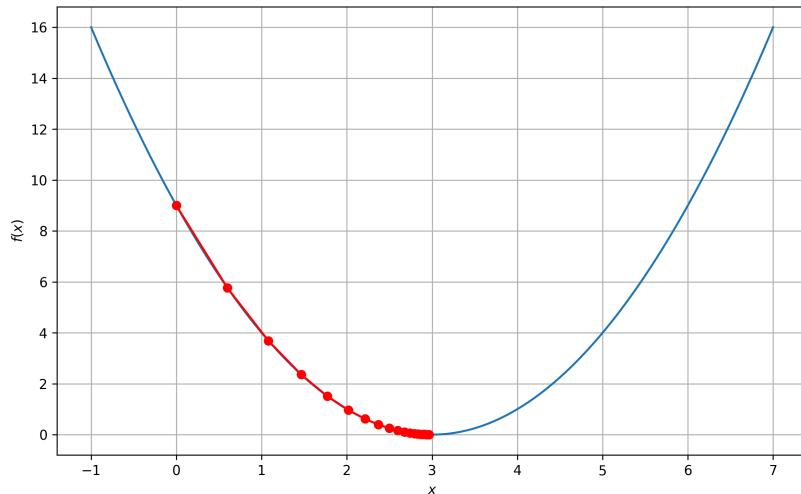
1. **Initialize:** Start with an initial guess for the parameters
2. **Compute Gradient:** Find the gradient of the function at the current parameters
3. **Update Parameters:** Adjust the parameters by moving in the opposite direction of the gradient, scaled by the learning rate
4. **Repeat:** Continue the process until the parameters converge to a minimum or the changes are minimal

**Update Rule:**

$$\theta \leftarrow \theta - \alpha \nabla f(\theta)$$

Where:

- $\theta$ : parameter being optimized
- $\alpha$ : learning rate
- $\nabla f(\theta)$ : gradient of the function  $f$  with respect to  $\theta$



#### 1. Function and Gradient:

- Function:  $\theta_1^2 + \theta_2^2$
- Gradient:  $\nabla f(\theta) = \left( \frac{\partial f}{\partial \theta_1}, \frac{\partial f}{\partial \theta_2} \right) = (2\theta_1, 2\theta_2)$

#### 2. Initial Values:

$$\theta_1 = 1$$

$$\theta_2 = 2$$

3. Learning Rate:

$$\alpha = 0.1$$

4. Gradient Calculation

- For  $\theta_1 = 1$  and  $\theta_2 = 2$ :

$$\nabla f(\theta) = (2 \cdot 1, 2 \cdot 2) = (2, 4)$$

5. Parameter Update:

- Update  $\theta_1$  and  $\theta_2$  using the rule  $\theta \leftarrow \theta - \alpha \nabla f(\theta)$ :

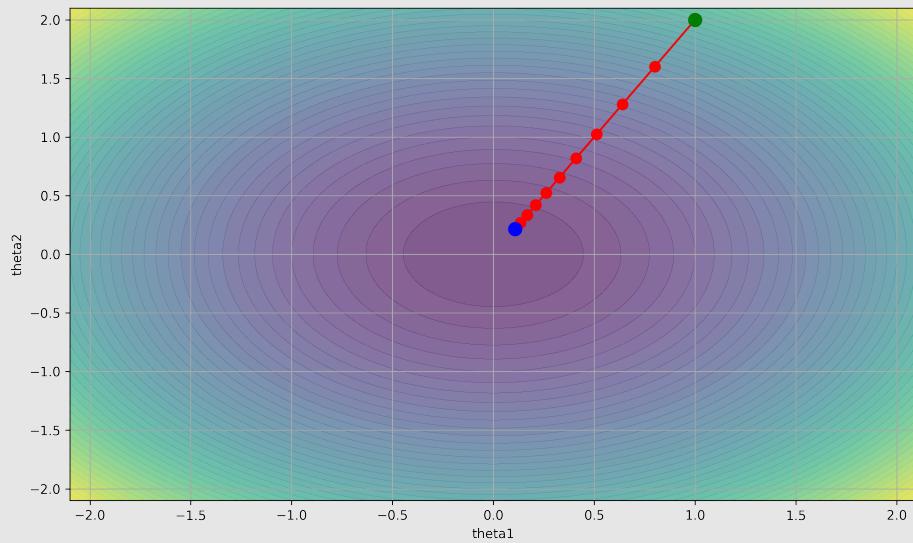
$$\theta_1 \leftarrow 1 - 0.1 \cdot 2 = 1 - 0.2 = 0.8$$

$$\theta_2 \leftarrow 2 - 0.1 \cdot 4 = 2 - 0.4 = 1.6$$

6. New Values:

$$\theta_1 = 0.8$$

$$\theta_2 = 1.6$$



## 20.4. Monte Carlo