Chapter 1

Useful Results on Linear Algebra, Distributions and Introduction to Regression Analysis

Some Results on Linear Algebra, Matrix Theory and Distributions:

We need some basic knowledge to understand the topics in the regression analysis and analysis of variance

Vectors:

A vector Y is an ordered n-tuple of real numbers. A vector can be expressed as a row vector or a column vector, as

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

is a column vector of order $n \times 1$ and

$$Y' = (y_1, y_2, ..., y_n)$$

is a row vector of order $1 \times n$.

If all $y_i = 0$ for all $i = 1,2,\ldots,n$ then $Y' = (0,0,...,0)$ is called the **null vector**.

If

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad Z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$$

then

$$X + Y = \begin{pmatrix} x_1 + y_2 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}, \quad kY = \begin{pmatrix} ky_1 \\ ky_2 \\ \vdots \\ ky_n \end{pmatrix}$$

$$X + (Y + Z) = (X + Y) + Z$$
$$X'(Y + Z) = X'Y + X'Z$$
$$k(X'Y) = (kX)'Y = X'(kY)$$
$$k(X + Y) = kX + kY$$
$$X'Y = x_1y_1 + x_2y_2 + ... + x_ny_n$$

where k is a scalar.

Orthogonal vectors:

Two vectors X and Y are said to be orthogonal if $X'Y = Y'X = 0$.

The null vector is orthogonal to every vector X and is the only such vector.

Linear combination:

If $x_1, x_2, ..., x_m$ are m vectors and $k_1, k_2, ..., k_m$ are m scalars, then

$$t = \sum_{i=1}^{m} k_i x_i$$

is called the linear combination of $x_1, x_2, ..., x_m$.

Linear independence:

If $X_1, X_2, ..., X_m$ are m vectors then they are said to be linearly independent if there exist scalars $k_1, k_2, ..., k_m$ such that

$$\sum_{i=1}^{m} k_i X_i = 0 \Rightarrow k_i = 0 \text{ for all } i = 1,2,\ldots,m.$$

If there exist $k_1, k_2, ..., k_m$ with at least one k_i to be nonzero, such that $\sum_{i=1}^{m} k_i x_i = 0$ then $x_1, x_2, ..., x_m$ are said to be **linearly dependent**.

- Any set of vectors containing the null vector is linearly dependent.
- Any set of non-null pair-wise orthogonal vectors is linearly independent.
- If $m > 1$ vectors are linearly dependent, it is always possible to express at least one of them as a linear combination of the others.

Linear function:

Let $K = (k_1, k_2, ..., k_m)'$ be a $m \times 1$ vector of scalars and $X = (x_1, x_2, ..., x_m)$ be a $m \times 1$ vector of variables, then $K'Y = \sum_{i=1}^{m} k_i y_i$ is called a linear function or linear form. The vector K is called the **coefficient vector**.

For example, the mean of $x_1, x_2, ..., x_m$ can be expressed as

$$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i = \frac{1}{m}(1,1,...,1)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \frac{1}{m}1_m^{'} X$$

where $1_m^{'}$ is a $m \times 1$ vector of all elements unity.

Contrast:

The linear function $K'X = \sum_{i=1}^{m} k_i x_i$ is called a contrast in $x_1, x_2, ..., x_m$ if $\sum_{i=1}^{m} k_i = 0.$

For example, the linear functions

$$x_1 - x_2, 2x_1 - 3x_2 + x_3, \frac{x_1}{2} - x_2 + \frac{x_3}{3}$$

are contrasts.

- A linear function $K'X$ is a contrast if and only if it is orthogonal to a linear function $\sum_{i=}^{m} x_i$ or to the linear function $\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_{.i}$.
- Contrasts $x_1 - x_2, x_1 - x_3, ..., x_1 - x_j$ are linearly independent for all $j = 2, 3, ..., m.$
- Every contrast in $x_1, x_2, ..., x_n$ can be written as a linear combination of (m - 1) contrasts $x_1 - x_2, x_1 - x_3, ..., x_1 - x_m$.

Matrix:

A matrix is a rectangular array of real numbers. For example

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

is a matrix of order $m \times n$ with m rows and n columns.

- If $m = n,$ then A is called a square matrix.
- If $a_{ij} = 0, i \neq j, m = n,$ then A is a diagonal matrix and is denoted as

$$A = diag(a_{11}, a_{22}, ..., a_{mm}).$$

- If $m = n$ (square matrix) and $a_{ij} = 0$ for $i > j$, then A is called an **upper triangular matrix**. On the other hand if $m = n$ and $a_{ij} = 0$ for $i < j$ then A is called a **lower triangular matrix.**
- If A is a $m \times n$ matrix, then the matrix obtained by writing the rows of A and columns of A as columns of A and rows of A respectively is called the **transpose of a matrix** A and is denoted as A'.
- If $A = A'$ then A is a **symmetric matrix**.
- If $A = -A'$ then A is **skew-symmetric matrix**.
- A matrix whose all elements are equal to zero is called a **null matrix**.
- An **identity matrix** is a square matrix of order p whose diagonal elements are unity (ones) and all the off-diagonal elements are zero. It is denoted as I_p.
- If A and B are matrices of order $m \times n$ then $(A+B)' = A' + B'$.
- If A and B are the matrices of order m x n and n x p respectively and k is any scalar, then

  $(AB)' = B'A'$

  $(kA)B = A(kB) = k(AB) = kAB.$
- If the orders of matrices A is m x n, B is n x p and C is n x p then $A(B+C) = AB + AC$.
- If the orders of matrices A is m x n, B is n x p and C is p x q then $(AB)C = A(BC)$.
- If A is the matrix of order m x n then $I_m A = AI_n = A$..

Trace of a matrix:

The trace of $n \times n$ matrix A, denoted as $tr(A)$ or $trace(A)$ is defined to be the sum of all the diagonal elements of A, i.e. $tr(A) = \sum_{i=1}^{n} a_{ii}$.

- If A is of order $m \times n$ and B is of order $n \times m$, then
  $tr(AB) = tr(BA)$.
- If A is $n \times n$ matrix and P is any nonsingular $n \times n$ matrix then
  $tr(A) = tr(P^{-1}AP)$.

  If P is an orthogonal matrix than $tr(A) = tr(P'AP)$.

* If A and B are $n \times n$ matrices, a and b are scalars then
  $tr(aA + bB) = a\,tr(A) + b\,tr(B)$.

- If A is an $m \times n$ matrix, then

$$tr(A'A) = tr(AA') = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij}^2$$

and

$$tr(A'A) = tr(AA') = 0 \text{ if and only if } A = 0.$$

- If A is $n \times n$ matrix then

$$tr(A') = trA.$$

Rank of matrices:

The rank of a matrix A of $m \times n$ is the number of linearly independent rows in A.

Let B be any other matrix of order $n \times q$.

- A square matrix of order m is called **non-singular** if it has full rank.
- $rank(AB) \le \min(rank(A), rank(B))$
- $rank(A+B) \le rank(A) + rank(B)$
- *Rank A* is equal to the maximum order of all nonsingular square submatrices of A.
- $rank(AA') = rank(A'A) = rank(A) = rank(A')$.
- A is of full row rank if $rank(A) = m < n$.
- A is of full column rank if $rank(A) = n < m$.

Inverse of a matrix:

The inverse of a square matrix A of order m, is a square matrix of order m, denoted as A^{-1}, such that $A^{-1}A = AA^{-1} = I_m$.

The inverse of A exists if and only if A is nonsingular.

- $(A^{-1})^{-1} = A$.
- If A is non-singular, then $(A')^{-1} = (A^{-1})'$
- If A and B are non-singular matrices of the same order, then their product, if defined, is also nonsingular and $(AB)^{-1} = B^{-1}A^{-1}$.

Idempotent matrix:

A square matrix A is called idempotent if $A^2 = AA = A.$.

If A is an $n \times n$ idempotent matrix with $rank(A) = r \le n$. Then

- eigenvalues of A are 1 or 0.
- $trace(A) = rank(A) = r.$
- If A is of full rank n, then $A = I_n$.
- If A and B are idempotent and $AB = BA$, then AB is also idempotent.
- If A is idempotent then $(I - A)$ is also idempotent and $A(I - A) = (I - A)A = 0$.

Quadratic forms:

If A is a given matrix of order $m \times n$ and X and Y are two given vectors of order $m \times 1$ and $n \times 1$ respectively

$$X'AY = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j$$

where a_{ij} are the nonstochastic elements of A.

If A is a square matrix of order m and $X = Y$, then

$$X'AX = a_{11}x_1^2 + ... + a_{mm}x_m^2 + (a_{12} + a_{21})x_1x_2 + ... + (a_{m-1,m} + a_{m,m-1})x_{m-1}x_m.$$

If A is symmetric also, then

$$\begin{aligned} X'AX &= a_{11}x_1^2 + ... + a_{mm}x_m^2 + 2a_{12}x_1x_2 + ... + 2a_{m-1,m}x_{m-1}x_m \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i x_j \end{aligned}$$

is called a quadratic form in m variables $x_1, x_2, ..., x_m$ or a quadratic form in X.

- To every quadratic form corresponds a symmetric matrix and vice versa.
- The matrix A is called the matrix of the quadratic form.
- The quadratic form $X'AX$ and the matrix A of the form is called.
  - ❖ Positive definite if $X'AX > 0$ for all $x \ne 0$.
  - ❖ Positive semidefinite if $X'AX \ge 0$ for all $x \ne 0$.
  - ❖ Negative definite if $X'AX < 0$ for all $x \ne 0$.
  - ❖ Negative semidefinite if $X'AX \le 0$ for all $x \ne 0$.

- If A is positive semi-definite matrix then $a_{ii} \geq 0$ and if $a_{ii} = 0$ then $a_{ij} = 0$ for all j, and $a_{ji} = 0$ for all j.
- If P is any nonsingular matrix and A is any positive definite matrix (or positive semi-definite matrix) then $P'AP$ is also a positive definite matrix (or positive semi-definite matrix).
- A matrix A is positive definite if and only if there exists a non-singular matrix P such that $A = P'P$.
- A positive definite matrix is a nonsingular matrix.
- If A is $m \times n$ matrix and $rank(A) = m < n$ then AA' is positive definite and $A'A$ is positive semidefinite.
- If A $m \times n$ matrix and $rank(A) = k < m < n$, then both $A'A$ and AA' are positive semidefinite.

Simultaneous linear equations:

The set of m linear equations in n unknowns $x_1, x_2, ..., x_n$ and scalars a_{ij} and b_i, $i = 1, 2, ..., m$, $j = 1, 2, ..., n$ of the form

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

can be formulated as $AX = b$

where A is a real matrix of known scalars of order $m \times n$ called as a coefficient matrix, X is $n \times 1$ real vector and b is $n \times 1$ real vector of known scalars given by

$$A = \begin{pmatrix} a_{11} & a_{12} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2n} \\ \vdots & \vdots \ \ddots \ \vdots \\ a_{m1} & a_{m2} \ldots a_{mn} \end{pmatrix},$$ is an $m \times n$ real matrix called as coefficient matrix,

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$ is an $n \times 1$ vector of variables and $$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$ is an $m \times 1$ real vector.

- If A is a $n \times n$ nonsingular matrix, then $AX = b$ has a unique solution.
- Let $B = [A, b]$ is an augmented matrix. A solution to $AX = b$ exist if and only if $rank(A) = rank(B)$.
- If A is a $m \times n$ matrix of rank m, then $AX = b$ has a solution.

- Linear homogeneous system $AX = 0$ has a solution other than $X = 0$ if and only if $rank(A) < n$.
- If $AX = b$ is consistent then $AX = b$ has a unique solution if and only if $rank(A) = n$
- If a_{ii} is the i^{th} diagonal element of an orthogonal matrix, then $-1 \le a_{ii} \le 1$.
- Let the $n \times n$ matrix be partitioned as $A = [a_1, a_2, ..., a_n]$ where a_i is an $n \times 1$ vector of the elements of i^{th} column of A. A necessary and sufficient condition that A is an orthogonal matrix is given by the following:

  (i) $a_i' a_i = 1$ for $i = 1, 2, ..., n$

  (ii) $a_i' a_j = 0$ for $i \ne j = 1, 2, ..., n.$

Orthogonal matrix:

A square matrix A is called an orthogonal matrix if $A'A = AA' = I$ or equivalently if $A^{-1} = A'$.

- An orthogonal matrix is non-singular.
- If A is orthogonal, then AA' is also orthogonal.
- If A is an $n \times n$ matrix and let P is an $n \times n$ orthogonal matrix, then the determinants of A and $P'AP$ are the same.

Random vectors:

Let $Y_1, Y_2, ..., Y_n$ be n random variables then $Y = (Y_1, Y_2, ..., Y_n)'$ is called a random vector.

- The **mean vector** Y is

  $E(Y) = ((E(Y_1), E(Y_2), ..., E(Y_n))'$

- The **covariance matrix** or **dispersion matrix** of Y is

$$Var(Y) = \begin{pmatrix} Var(Y_1) & Cov(Y_1, Y_2) & \cdots & Cov(Y_1, Y_n) \\ Cov(Y_2, Y_1) & Var(Y_2) & \cdots & Cov(Y_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Y_n, Y_1) & Cov(Y_n, Y_2) & \cdots & Var(Y_n) \end{pmatrix}$$

  which is a symmetric matrix

- If $Y_1, Y_2, ..., Y_n$ are independently distributed, then the covariance matrix is a diagonal matrix.
- If $Var(Y_i) = \sigma^2$ for all i = 1, 2,…,n then $Var(Y) = \sigma^2 I_n$.

Linear function of random variable:

If $Y_1, Y_2, ..., Y_n$ are *n* random variables, and $k_1, k_2, .., k_n$ are scalars, then $\sum_{i=1}^{n} k_i Y_i$ is called a linear function of random variables $Y_1, Y_2, ..., Y_n$.

If $Y = (Y_1, Y_2, ..., Y_n)'$, $K = (k_1, k_2, ..., k_n)'$ then $K'Y = \sum_{i=1}^{n} k_i Y_i$.

- the mean $K'Y$ is $E(K'Y) = K'E(Y) = \sum_{i=1}^{n} k_i E(Y_i)$ and
- the variance of $K'Y$ is $Var\left(K'Y\right) = K'Var\left(Y\right)K.$

Multivariate normal distribution:

A random vector $Y = (Y_1, Y_2, ..., Y_n)'$ has a multivariate normal distribution with mean vector $\mu = (\mu_1, \mu_2, ..., \mu_n)$ and dispersion matrix Σ if its probability density function is

$$f(Y / \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(Y - \mu)'\Sigma^{-1}(Y - \mu) \right]$$

assuming Σ is a nonsingular matrix.

Chi-square distribution:

- If $Y_1, Y_2, ..., Y_k$ are independently distributed following the normal distribution random variables with common mean 0 and common variance 1, then the distribution of $\sum_{i=1}^{k} Y_i^2$ is called the χ^2-distribution with k degrees of freedom.
- The probability density function of χ^2-distribution with k degrees of freedom is given as

$$f_{\chi^2}(x) = \frac{1}{\Gamma(k/2) 2^{k/2}} x^{\frac{k}{2}-1} \exp\left(-\frac{x}{2} \right); \quad 0 < x < \infty$$

- If $Y_1, Y_2, ..., Y_k$ are independently distributed following the normal distribution with common mean 0 and common variance σ^2, then $\frac{1}{\sigma^2}\sum_{i=1}^{k} Y_i^2$ has χ^2 distribution with *k* degrees of freedom.

- If the random variables $Y_1, Y_2, ..., Y_k$ are normally distributed with non-null means $\mu_1, \mu_2, ..., \mu_k$ but common variance 1, then the distribution of $\sum_{i=1}^{k} Y_i^2$ has noncentral χ^2 distribution with *k* degrees of freedom and noncentrality parameter $\lambda = \sum_{i=1}^{k} \mu_i^2$.
- If $Y_1, Y_2, ..., Y_k$ are independently and normally distributed following the normal distribution with means $\mu_1, \mu_2, ..., \mu_k$ but common variance σ^2 then $\frac{1}{\sigma^2}\sum_{i=1}^{k} Y_i^2$ has non-central χ^2 distribution with *k* degrees of freedom and noncentrality parameter $\lambda = \frac{1}{\sigma^2}\sum_{i=1}^{k} \mu_i^2$.
- If *U* has a Chi-square distribution with *k* degrees of freedom then $E(U) = k$ and $Var(U) = 2k$.
- If *U* has a noncentral Chi-square distribution with *k* degrees of freedom and noncentrality parameter λ then $E\left(U\right) = k + \lambda$ and $Var(U) = 2k + 4\lambda$.
- If $U_1, U_2, ..., U_k$ are independently distributed random variables with each U_i having a noncentral Chi-square distribution with n_i degrees of freedom and non-centrality parameter $\lambda_i, \ i = 1, 2, ..., k$ then $\sum_{i=1}^{k} U_i$ has noncentral Chi-square distribution with $\sum_{i=1}^{k} n_i$ degrees of freedom and noncentrality parameter $\sum_{i=1}^{k} \lambda_i$.
- Let $X = (X_1, X_2, ..., X_n)'$ has a multivariate distribution with mean vector μ and positive definite covariance matrix Σ. Then $X'AX$ is distributed as noncentral χ^2 with *k* degrees of freedom if and only if ΣA is an idempotent matrix of rank *k*.
- Let $X = (X_1, X_2, ..., X_n)$ has a multivariate normal distribution with mean vector μ and positive definite covariance matrix Σ. Let the two quadratic forms-
  - $X'A_1X$ is distributed as χ^2 with n_1 degrees of freedom and noncentrality parameter $\mu' A_1 \mu$ and
  - $X'A_2X$ is distributed as χ^2 with μ_2 degrees of freedom and noncentrality parameter $\mu' A_2 \mu$.

  Then $X'A_1X$ and $X'A_2X$ are independently distributed if $A_1 \Sigma A_2 = 0$

t-distribution:

- If
  - X has a normal distribution with mean 0 and variance 1,
  - Y has a χ^2 distribution with n degrees of freedom, and
  - X and Y are independent random variables,

  then the distribution of the statistic $T = \frac{X}{\sqrt{Y/n}}$ is called the *t*-distribution with n degrees of freedom.

  The probability density function of t is

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1+\frac{t^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}; \quad -\infty < t < \infty$$

- If the mean of X is nonzero then the distribution of $\frac{X}{\sqrt{Y/n}}$ is called the noncentral *t*- distribution with n degrees of freedom and noncentrality parameter μ.

F-distribution:

- If X and Y are independent random variables with χ^2-distribution with m and n degrees of freedom respectively, then the distribution of the statistic $F = \frac{X/m}{Y/n}$ is called the *F*-distribution with m and n degrees of freedom. The probability density function of F is

$$f_F(f) = \frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{m/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} f^{\left(\frac{m-2}{2}\right)}\left(1+\left(\frac{m}{n}\right)f\right)^{-\left(\frac{m+n}{2}\right)}; \quad 0 < f < \infty$$

- If X has a noncentral Chi-square distribution with m degrees of freedom and noncentrality parameter $\lambda; Y$ has a χ^2 distribution with n degrees of freedom, and X and Y are independent random variables, then the distribution of $F = \frac{X/m}{Y/n}$ is the noncentral F distribution with m and n degrees of freedom and noncentrality parameter λ.

Estimable functions:

A linear parametric function $\lambda'\beta$ of the parameter is said to be an estimable parametric function or estimable if there exists a linear function of random variables $\ell' y$ of Y where $Y = (Y_1, Y_2, ..., Y_n)'$ such that

$$E(\ell' y) = \lambda' \beta$$

with $\ell = (\ell_1, \ell_2, ..., \ell_n)'$ and $\lambda = (\lambda_1, \lambda_2, ..., \lambda_n)'$ being vectors of known scalars.

Fisher-Cochran theorem:

If $X = (X_1, X_2, ..., X_n)$ has a multivariate normal distribution with mean vector μ and positive definite covariance matrix Σ and let

$$X'\Sigma^{-1}X = Q_1 + Q_2 + ... + Q_k$$

where $Q_i = X' A_i X$ with rank $(A_i) = N_i$, $i = 1, 2, ..., k$. Then Q_i's are independently distributed noncentral Chi-square distribution with N_i degrees of freedom and noncentrality parameter $\mu' A_i \mu$ if and only if $\sum_{i=1}^{k} N_i = N$ is which case

$$\mu'\Sigma^{-1}\mu = \sum_{i=1}^{k} \mu' A_i \mu.$$

Derivatives of quadratic and linear forms:

Let $X = (x_1, x_2, ..., x_n)'$ and $f(X)$ be any function of n independent variables $x_1, x_2, ..., x_n$, then

$$\frac{\partial f(X)}{\partial X} = \begin{pmatrix} \frac{\partial f(X)}{\partial x_1} \\ \frac{\partial f(X)}{\partial x_2} \\ \vdots \\ \frac{\partial f(X)}{\partial x_n} \end{pmatrix}.$$

If $K = (k_1, k_2, ..., k_n)'$ is a vector of constants, then

$$\frac{\partial K'X}{\partial X} = K$$

If A is a $m \times n$ matrix, then

$$\frac{\partial X'AX}{\partial X} = 2(A + A')X\ .$$

Independence of linear and quadratic forms:

- Let Y be an $n\times 1$ vector having multivariate normal distribution $N(\mu, I)$ and B be a $m\times n$ matrix. Then the $m\times 1$ vector linear form BY is independent of the quadratic form $Y'AY$ if $BA = 0$ where A is a symmetric matrix of known elements.
- Let Y be a $n\times 1$ vector having multivariate normal distribution $N(\mu, \Sigma)$ with $rank(\Sigma) = n$. If $B\Sigma A = 0$, then the quadratic form $Y'AY$ is independent of linear form BY where B is a $m\times n$ matrix.

Linear Models and Regression Analysis:

Linear models play a central part in modern statistical methods. On the one hand, these models can approximate a large amount of metric data structures over their entire range of definition, or at least piecewise.

Suppose the outcome of any process is denoted by a random variable y, called as dependent (or study) variable, depends on k independent (or explanatory) variables denoted by $X_1, X_2, ..., X_k$. Suppose the behaviour of y can be explained by a relationship given by

$$y = f(X_1, X_2, ..., X_k, \beta_1, \beta_2, ..., \beta_k) + \varepsilon$$

where f is some well-defined function and $\beta_1, \beta_2, ..., \beta_k$ are the parameters which characterize the role and contribution of $X_1, X_2, ..., X_k$, respectively. The term ε reflects the stochastic nature of the relationship between y and $X_1, X_2, ..., X_k$ indicates that such a relationship is not exact in nature. When $\varepsilon = 0$, then the relationship is called the mathematical model otherwise the statistical model. The term "**model**" is broadly used to represent any phenomenon in a mathematical framework.

A model or relationship is termed as linear if it is linear in parameters and nonlinear if it is not linear in parameters. In other words, if all the partial derivatives of y with respect to each of the parameters $\beta_1, \beta_2, ..., \beta_k$, are independent of the parameters, then the model is called a **linear model**. If any of the partial derivatives of y with respect to any of the $\beta_1, \beta_2, ..., \beta_k$ is not independent of the parameters, then the model is called as nonlinear. Note that the linearity or non-linearity of the model is not described by the linearity or nonlinearity of explanatory variables in the model.

For example

$$y = \beta_1 X_1^2 + \beta_2 \sqrt{X_2} + \beta_3 \log X_3 + \varepsilon$$

is a linear model because $\partial y / \partial \beta_i$, $(i = 1,2,3)$ are independent of the parameters $\beta_i, (i = 1,2,3)$. On the other hand,

$$y = \beta_1^2 X_1 + \beta_2 X_2 + \beta_3 \log X + \varepsilon$$

is a nonlinear model because $\partial y / \partial \beta_1 = 2\beta_1 X_1$ depends on β_1 although $\partial y / \partial \beta_2$ and $\partial y / \partial \beta_3$ are independent of any of the β_1, β_2 or β_3.

When the function f is linear in parameters, then $y = f(X_1, X_2, ..., X_k, \beta_1, \beta_2, ..., \beta_k) + \varepsilon$ is called a linear model and when the function f is nonlinear in parameters, then it is called a nonlinear model. In general, the function f is chosen as

$$f(X_1, X_2, ..., X_k, \beta_1, \beta_2 ..., \beta_k) = \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k$$

to describe a linear model. Since $X_1, X_2, ..., X_k$ are pre-determined variables and y is the outcome, so both are known. Thus the knowledge of the model depends on the knowledge of the parameters $\beta_1, \beta_2, ..., \beta_k$.

The linear statistical modelling essentially consists of developing approaches and tools to determine $\beta_1, \beta_2, ..., \beta_k$ in the linear model

$$y = \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon$$

given the observations on y and $X_1, X_2, ..., X_k$.

Different statistical estimation procedures, e.g., method of maximum likelihood, principal of least squares, method of moments etc. can be employed to estimate the parameters of the model. The method of maximum likelihood needs further knowledge of the distribution of y. In contrast, the method of moments and the principle of least squares do not need any knowledge about the distribution of y.

The regression analysis is a tool to determine the values of the parameters given the data on y and $X_1, X_2, ..., X_k$. The literal meaning of regression is "to move in the backward direction". Before discussing and understanding the meaning of "backward direction", let us find which of the following statement is correct:

$S1$: model generates data or

$S2$: data generates model.

Obviously, $S1$ is correct. It can be broadly thought that the model exists in nature but is unknown to the experimenter. When some values to the explanatory variables are provided, then the values for the output or study variable are generated accordingly, depending on the form of the function f and the nature of the phenomenon. So ideally, the pre-existing model gives rise to the data. Our objective is to determine the functional form of this model. Now we move in the backward direction. We propose to first collect the data on study and explanatory variables. Then we employ some statistical techniques and use this data to know the form of function f. Equivalently, the data from the model is recorded first and then used to determine the parameters of the model. The regression analysis is a technique which helps in determining the statistical model by using the data on study and explanatory variables. The classification of linear and nonlinear regression analysis is based on the determination of linear and nonlinear models, respectively.

Consider a simple example to understand the meaning of "regression". Suppose the yield of the crop (y) depends linearly on two explanatory variables, viz., the quality of fertilizer (X_1) and level of irrigation (X_2) as

$$y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon.$$

There exist the true values of β_1 and β_2 in nature but are unknown to the experimenter. Some values on y are recorded by providing different values to X_1 and X_2. There exists some relationship between y and X_1, X_2 which gives rise to a systematically behaved data on y, X_1 and X_2. Such a relationship is unknown to the experimenter. To determine the model, we move in the backward direction in the sense that the collected data is used to determine the parameters β_1 and β_2 of the model. In this sense, such an approach is termed as regression analysis.

The theory and fundamentals of linear models lay the foundation for developing regression analysis tools based on valid statistical theory and concepts.

Steps in regression analysis:

Regression analysis includes the following steps:

- Statement of the problem under consideration
- Choice of relevant variables
- Collection of data on relevant variables
- Specification of model
- Choice of method for fitting the data
- Fitting of model
- Model validation and criticism
- Using the chosen model(s) for the solution of the posed problem.

These steps are examined below.

1. Statement of the problem under consideration

The first important step in conducting any regression analysis is to specify the problem and the objectives to be addressed by the regression analysis. The wrong formulation or understanding of the problem will yield erroneous statistical inferences. The choice of variables depends upon the objectives of the study and the understanding of the problem. For example, the height and weight of children are related. Now, there are two issues to be addressed.

(i) Determination of height for a given weight, or

(ii) determination of weight for a given height.

In case 1, the height is a response variable, whereas weight is a response variable is case 2. The roles of the explanatory variables are also interchanged in cases 1 and 2.

2. Choice of relevant variables

Once the problem is carefully formulated and objectives have been decided, the next question is to choose the relevant variables. It has to be kept in mind that the correct choice of variables will determine the validity of statistical inferences. For example, in an agricultural experiment, the yield depends on explanatory variables like quantity of fertiliser, rainfall, irrigation, temperature, etc. These variables are denoted by $X_1, X_2, ..., X_k$ as a set of k explanatory variables.

3. Collection of data on relevant variables

Once the objective of the study is clearly stated and the variables are chosen, the next question that arises is to collect data on such relevant variables. The data is essentially the measurement of these variables. For example, it is important to know how to record the data on age. For example, the data collection is on age in total complete years, or the date of birth is to be recorded, which can give the exact age at a specific time. Moreover, it is important to decide whether the data will be collected on quantitative or qualitative variables. For example, if the ages (in years) are 15, 17, 19, 21, and 23, then these are quantitative values. If the ages are defined by a variable that takes the value 1 if the ages are less than 18 years and 0 if the ages are more than 18 years, then the earlier recorded data is converted to 1, 1, 0, 0, 0. Note that information is lost when quantitative data are converted to qualitative data. The methods and approaches for qualitative and quantitative data are also different. If the study variable is binary, then **logistic regression** is used. If all explanatory variables are qualitative, then the **analysis of variance** technique is used. If some explanatory variables are qualitative and others are quantitative, then the **analysis of covariance** technique is used. The techniques of analysis of variance and analysis of covariance are special cases of regression analysis.

Generally, the data is collected on n subjects, then y on data, then y denotes the response or study variable and $y_1, y_2, ..., y_n$ are the n values. If there are k explanatory variables, $X_1, X_2, .., X_k$ then x_{ij} denotes the i^{th} value of j^{th} variable. The observation can be presented in the following table:

Notations for the data used in regression analysis

Observation Number	**Response** y	**Explanatory variables**			
		X_1	X_2	$\cdots$	X_k
1	y_1	x_{11}	x_{12}	$\cdots$	x_{1k}
2	y_2	x_{21}	x_{22}	$\cdots$	x_{2k}
3	y_3	x_{31}	x_{32}	$\cdots$	x_{3k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_n	x_{n1}	x_{n2}	$\cdots$	x_{nk}

4. Specification of the model

The experimenter or the person working on the subject usually helps in determining the form of the model. Only the form of the tentative model can be determined; it will depend on unknown parameters. For example, a general form will be like

$$y = f(X_1, X_2, ..., X_k; \beta_1, \beta_2, ..., \beta_k) + \varepsilon$$

where ε is the random error reflecting mainly the difference in the observed value of y and the value of y obtained through the model. The form of $f(X_1, X_2, ..., X_k; \beta_1, \beta_2, ..., \beta_k)$ can be linear as well as nonlinear depending on the form of parameters $\beta_1, \beta_2, ..., \beta_k$. A model is said to be linear if it is linear in parameters.

For example,

$$y = \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \varepsilon$$
$$y = \beta_1 + \beta_2 \ln X_2 + \varepsilon$$

are linear models whereas

$$y = \beta_1 X_1 + \beta_2^2 X_2 + \beta_3 X_2 + \varepsilon$$
$$y = \ln \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

are the non-linear models. Often, nonlinear models can be converted to linear models via appropriate transformations. So the class of linear models is wider than it initially appears.

A model containing only one explanatory variable is called a **simple regression model.** When there is more than one independent variable, it is called a **multiple regression model.** When there is only one study variable, the regression is termed as **univariate regression.** When there is more than one study variable, the regression is termed as **multivariate regression.** Note that the simple and multiple regressions are not the same as univariate and multivariate regressions. The choice between simple and multiple regression is determined by the number of explanatory variables, whereas the choice between univariate and multivariate regressions is determined by the number of study variables.

5. Choice of method for fitting the data

After the model has been defined and the data have been collected, the next step is to estimate the parameters of the model based on the collected data. This is also referred to as **parameter estimation** or **model fitting**. The most commonly used method of estimation is called the least-squares method. Under certain assumptions, the least-squares method produces estimators with desirable properties. The other estimation methods are the maximum likelihood method, the ridge method, the principal components method, etc.

6. Fitting of the model

Estimating unknown parameters using an appropriate method yields parameter values. Substituting these values into the equation yields a usable model. This is termed as model fitting. The estimates of parameters $\beta_1, \beta_2, ..., \beta_k$ in the model

$$y = f(X_1, X_2, ..., X_k, \beta_1, \beta_2, ..., \beta_k) + \varepsilon$$

are denotes as $\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$ which gives the fitted model as

$$y = f(X_1, X_2, ..., X_k, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k).$$

When the value of y is obtained for the given values of $X_1, X_2, ..., X_k,$ it is denoted as $\hat{y}$ and called as fitted value.

The fitted equation is used for prediction. In this case, $\hat{y}$ is termed as the **predicted value.** Note that the fitted value is where the values used for explanatory variables correspond to one of the n observations in the data, whereas the predicted value is the one obtained for any set of values of explanatory variables. It is not generally recommended to predict the y-values for the set of those values of explanatory variables which lie outside the range of data. When the values of explanatory variables are the future values of explanatory variables, the predicted values are called as forecasted values.

There are different methodologies based on regression analysis. They are described in the following table:

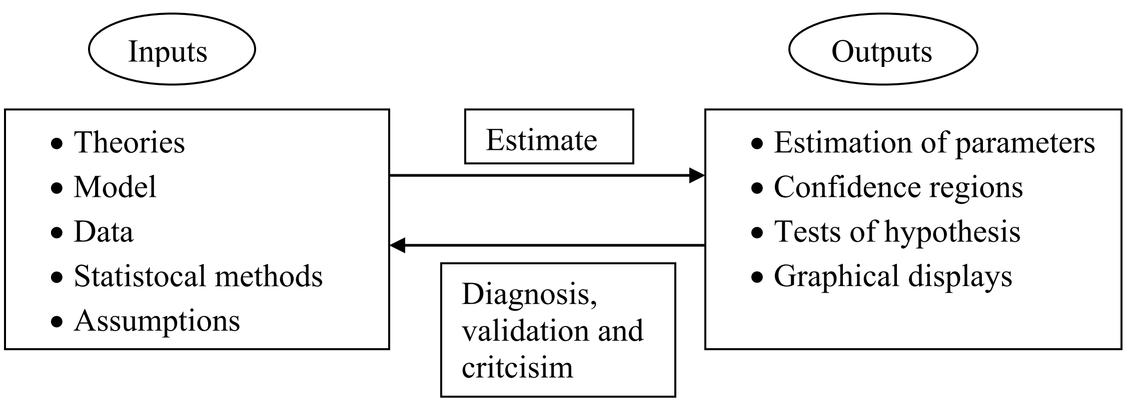
Type of Regression	**Conditions**
Univariate	Only one quantitative response variable
Multivariate	Two or more quantitative response variables
Simple	Only one explanatory variable
Multiple	Two or more explanatory variables
Linear	All parameters enter the equation linearly, possibly after transformation of the data
Nonlinear	The relationship between the response and some of the explanatory variables is nonlinear or some of the parameters appear nonlinearly, but no transformation is possible to make the parameters appear linearly
Analysis of variance	All explanatory variables are qualitative variables

Analysis of Covariance	Some explanatory variables are quantitative variables and others are qualitative variables
Logistic	The response variable is qualitative

7. Model criticism and selection

The validity of the statistical method used for regression analysis depends on several assumptions. These assumptions essentially become the assumptions for the model and the data. The quality of statistical inferences depends heavily on whether these assumptions are satisfied. For these assumptions to be valid and satisfied, care is needed from the beginning of the experiment. One has to be careful in choosing the required assumptions and decide whether the assumptions are valid for the given experimental conditions. It is also important to identify the situations in which the assumptions may not hold.

The assumptions must be validated before drawing any statistical conclusions. Any departure from the validity of assumptions will be reflected in the statistical inferences. In fact, the regression analysis is an iterative process where the outputs are used to diagnose, validate, criticize, and modify the inputs. The iterative process is illustrated in the following figure.



8. Objectives of regression analysis

Determining the explicit form of the regression equation is the ultimate objective of regression analysis. It is finally a good, valid relationship between the study variable and the explanatory variables. Such a regression equation can be used for several purposes. For example, to determine the role of any explanatory variable in the joint relationship in any policy formulation, or to forecast the values of the response variable for a given set of values of explanatory variables. The regression equation helps in understanding the interrelationships among the variables.