Chapter 1

Some Results on Linear Algebra, Matrix Theory and Distributions

We need some basic knowledge to understand the topics in the 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

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

then

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

$$\begin{aligned} &X + (Y + Z) = (X + Y) + Z \\ &X'(Y + Z) = X'Y + X'Z \\ &k(X'Y) = (kX)'Y = X'(kY) \\ &k(X + Y) = kX + kY \\ &X'Y = x_1y_1 + x_2y_2 + ... + x_ny_n \end{aligned}$$

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

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

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

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

- 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 sub-matrices 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 non-singular.

- $(A^{-1})^{-1} = A.$
- 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

$$X'AX = a_{11} x_1^2 + ... + a_{mm} x_m^2 + (a_{12} + a_{21}) x_1 x_2 + ... + (a_{m-1,m} + a_{m,m-1}) x_{m-1} x_m .$$

If A is symmetric also, then

$$\begin{aligned} X'AX &= a_{11} x_1^2 + ... + a_{mm} x_m^2 + 2a_{12} x_1 x_2 + ... + 2a_{m-1,m} x_{m-1} x_m \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i x_j \end{aligned}$$

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 $x \neq 0$.
    - ❖ Positive semidefinite if $X'AX \geq 0$ for all $x \neq 0$.
    - ❖ Negative definite if $X'AX < 0$ for all $x \neq 0$.
    - ❖ Negative semidefinite if $X'AX \leq 0$ for all $x \neq 0$.
- 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

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

  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 $i = 1, 2, \ldots, n$ 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 $Var(K'Y) = K'Var(Y)K$.

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 $E(U) = k + \lambda$ 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.

Linear model:

Suppose there are n observations. In the linear model, we assume that these observations are the values taken by n random variables $Y_1, Y_2, .., Y_n$ satisfying the following conditions:

1. $E(Y_i)$ is a linear combination of p unknown parameters $\beta_1, \beta_2, ..., \beta_p,$
   $E(Y_i) = x_{i1}\beta_1 + x_{i2}\beta_2 + ... + x_{ip}\beta_p, \; i = 1, 2, ..., n$
   where $x_{ij}'s$ are known constants.
2. $Y_1, Y_2, ..., Y_n$ are uncorrelated and normality distributed with variance $Var(Y_i) = \sigma^2$.

The linear model can be rewritten by introducing independent normal random variables following $N(0, \sigma^2)$,as

$$Y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + ... + x_{ip}\beta_p + \varepsilon_i, \; i = 1, 2, ..., n.$$

These equations can be written using the matrix notations as

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

where Y is a $n \times 1$ vector of observation, X is a $n \times p$ matrix of n observations on each of $X_1, X_2, ..., X_p$ variables, β is a $p \times 1$ vector of parameters and ε is a $n \times 1$ vector of random error components with

$\varepsilon \sim N(0, \sigma^2 I_n)$. Here Y is called **study or dependent. variable**, $X_1, X_2, ..., X_p$ are called **explanatory or independent variables** and $\beta_1, \beta_2, ..., \beta_p$ are called as **regression coefficients**.

Alternatively since $Y \sim N(X\beta, \sigma^2 I)$ so the linear model can also be expressed in the expectation form as a normal random variable Y with

$$E(y) = X\beta$$
$$Var(y) = \sigma^2 I.$$

Note that β and σ^2 are unknown but X is known.

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.

Best linear unbiased estimates (BLUE)

The unbiased minimum variance linear estimate $\ell' Y$ of an estimable function $\lambda'\beta$ is called the best linear unbiased estimate of $\lambda'\beta$.

- Suppose $\ell_1' Y$ and $\ell_2' Y$ are the BLUE of $\lambda_1'\beta$ and $\lambda_2'\beta$ respectively. Then $(a_1\ell_1 + a_2\ell_2)'Y$ is the BLUE of $(a_1\lambda_1 + a_2\lambda_2)'\beta$.
- If $\lambda'\beta$ is estimable, its best estimate is $\lambda'\hat{\beta}$ where $\hat{\beta}$ is any solution of the equations $X'X\beta = X'Y$.

Least squares estimation

The least-squares estimate of β is $Y = X\beta + \varepsilon$ is the value of β which minimizes the error sum of squares $\varepsilon'\varepsilon$.

Let

$$\begin{aligned} S = \varepsilon'\varepsilon &= (Y - X\beta)'(Y - X\beta) \\ &= Y'Y - 2\beta' X'Y + \beta' X'X\beta. \end{aligned}$$

Minimizing S with respect to β involves

$$\frac{\partial S}{\partial \beta} = 0$$

$$\Rightarrow X'X\beta = X'Y$$

which is termed as **normal equation**. This normal equation has a unique solution given by

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

assuming $rank(X) = p$. Note that $\frac{\partial^2 S}{\partial\beta\partial\beta'} = X'X$ is a positive definite matrix. So $\hat{\beta} = (X'X)^{-1}X'Y$ is the value of β which minimizes $\varepsilon'\varepsilon$ and is termed as **ordinary least squares estimator** of β.

- In this case, $\beta_1, \beta_2, ..., \beta_p$ are estimable and consequently, all the linear parametric function are estimable.
- $E(\hat{\beta}) = (X'X)^{-1}X'E(Y) = (X'X)^{-1}X'X\beta = \beta$
- $Var(\hat{\beta}) = (X'X)^{-1}X'Var(Y)X(X'X)^{-1} = \sigma^2(X'X)^{-1}$
- If $\lambda'\hat{\beta}$ and $\mu'\hat{\beta}$ are the estimates of $\lambda'\beta$ and $\mu'\beta$ respectively, then
  - ❖ $Var(\lambda'\hat{\beta}) = \lambda' Var(\hat{\beta})\lambda = \sigma^2[\lambda'(X'X)^{-1}\lambda]$
  - ❖ $Cov(\lambda'\hat{\beta}, \mu'\hat{\beta}) = \sigma^2[\mu'(X'X)^{-1}\lambda]$.
- $Y - X\hat{\beta}$ is called the **residual vector**
- $E(Y - X\hat{\beta}) = 0.$

Linear model with correlated observations:

In the linear model

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

with $E(\varepsilon) = 0, Var(\varepsilon) = \Sigma$ and ε is normally distributed, we find

$$E(Y) = X\beta, Var(Y) = \Sigma$$

Assuming Σ to be positive definite, so we can write

$$\Sigma = P'P$$

where P is a nonsingular matrix. Premultiplying $Y = X\beta + \varepsilon$ by P, we get

$$PY = PX\beta + P\varepsilon$$

or $\quad Y^* = X^*\beta + \varepsilon^*$

where $Y^* = PY$, $X^* = PX$ and $\varepsilon^* = P\varepsilon$.

Note that in this model $E(\varepsilon^* 0 = 0$ and $Var(\varepsilon^*) = \sigma^2 I$.

Distribution of $\ell'Y$:

In the linear model $Y = X\beta + \varepsilon, \; \varepsilon \sim N(0, \sigma^2 I)$ consider a linear function $\ell'Y$ which is normally distributed with

$$E(\ell'Y) = \ell' X\beta,$$
$$Var(\ell'Y) = \sigma^2(\ell'\ell).$$

Then

$$\frac{\ell'Y}{\sigma\sqrt{\ell'\ell}} \sim N\left(\frac{\ell' X\beta}{\sigma\sqrt{\ell'\ell}}, 1\right).$$

Further, $\dfrac{(\ell'Y)^2}{\sigma^2\ell'\ell}$ has a noncentral Chi-square distribution with one degree of freedom and noncentrality parameter $\dfrac{(\ell' X\beta)^2}{\sigma^2\ell'\ell}$.

Degrees of freedom:

A linear function $\ell'Y$ of the observations $(\ell \neq 0)$ is said to carry one degrees of freedom. A set of linear functions $L'Y$ where L is r x n matrix, is said to have M degrees of freedom if there exist M linearly independent functions in the set and no more. Alternatively, the degrees of freedom carried by the set $L'Y$ equals $rank(L)$. When the set $L'Y$ are the estimates of $\Lambda'\beta$, the degrees of freedom of the set $L'Y$ will also be called the degrees of freedom for the estimates of $\Lambda'\beta$.

Sum of squares:

If $\ell'Y$ is a linear function of observations, then the projection of Y on ℓ is the vector $\dfrac{Y'\ell}{\ell'\ell}.\ell$. The square of this projection is called the sum of squares (SS) due to $\ell' y$ is given by $\dfrac{(\ell'Y)^2}{\ell'\ell}$. Since $\ell'Y$ has one degree of freedom, so the SS due $\ell'Y$ to has one degree of freedom.

The sum of squares and the degrees of freedom arising out of the mutually orthogonal sets of functions can be added together to give the sum of squares and degrees of freedom for the set of all the function together and vice versa.

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.

- $X'A_1 X$ is distribution 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 n_2 degrees of freedom and noncentrality parameter $\mu' A_2\mu$.

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

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 $Q_i = X'A_iX$ with $\text{rank}(A_i) = N_i,\ i = 1,2,...,k.$ 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.