Chapter 2

General Linear Hypothesis and Analysis of Variance

Regression model for the general linear hypothesis

Let $Y_1, Y_2, ..., Y_n$ be a sequence of n independent random variables associated with responses. Then we can write it as

$$E(Y_i) = \sum_{j=1}^{p} \beta_j x_{ij}, \; i = 1, 2, ..., n, \; j = 1, 2, ..., p$$
$$Var(Y_i) = \sigma^2.$$

This is the linear model in the expectation form where $\beta_1, \beta_2, ..., \beta_p$ are the unknown parameters and x_{ij}'s are the known values of independent covariates $X_1, X_2, ..., X_p$.

Alternatively, the linear model can be expressed as

$$Y_i = \sum_{j=1}^{p} \beta_j x_{ij}, + \varepsilon_i, \; i = 1, 2, ..., n; \; j = 1, 2, ..., p$$

where ε_i's are identically and independently distributed random error component with mean 0 and variance σ^2, i.e., $E(\varepsilon_i) = 0$ $Var(\varepsilon_i) = \sigma^2$ and $Cov(\varepsilon_i, \varepsilon_j) = 0 (i \neq j).$

In matrix notations, the linear model can be expressed as

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

where

- $Y = (Y_1, Y_2, ..., Y_n)'$ is a $n \times 1$ vector of observations on the response variable,
- the matrix $X = \begin{pmatrix} X_{11} & X_{12} \ldots X_{1p} \\ X_{21} & X_{22} \ldots X_{2p} \\ \vdots & \vdots \; \ddots \; \vdots \\ X_{n1} & X_{n2} \ldots X_{np} \end{pmatrix}$ is a $n \times p$ matrix of n observations on p independent covariates $X_1, X_2, ..., X_p$,
- $\beta = (\beta_1, \beta_2, ..., \beta_p)'$ is a $p \times 1$ vector of unknown regression parameters (or regression coefficients) $\beta_1, \beta_2, ..., \beta_p$ associated with $X_1, X_2, ..., X_p$, respectively and
- $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)'$ is a $n \times 1$ vector of random errors or disturbances.
- We assume that $E(\varepsilon) = 0$, the covariance matrix $V(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I_p$, $rank(X) = p$.

In the context of analysis of variance and design of experiments,

- the matrix X is termed as the **design matrix**;
- unknown $\beta_1, \beta_2, ..., \beta_p$ are termed as **effects**,
- the covariates $X_1, X_2, ..., X_p,$ are **counter variables** or **indicator variables** where x_{ij} counts the number of times the effect β_j occurs in the i^{th} observation x_i.
- x_{ij} mostly takes the values 1 or 0 but not always.
- The value $x_{ij} = 1$ indicates the presence of effect β_j in x_i and $x_{ij} = 0$ indicates the absence of effect β_j in x_i.

Note that in the **linear regression model**, the covariates are usually continuous variables.

When some of the covariates are counter variables, and rest are continuous variables, then the model is called a **mixed model** and is used in the analysis of covariance.

Relationship between the regression model and analysis of variance model

The same linear model is used in the linear regression analysis as well as in the analysis of variance. So it is important to understand the role of a linear model in the context of linear regression analysis and analysis of variance.

Consider the multiple linear model

$$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + ... + X_p\beta_p + \varepsilon .$$

In the case of analysis of variance model,

- the one-way classification considers only one covariate,
- two-way classification model considers two covariates,
- three-way classification model considers three covariates and so on.

If β, γ and δ denote the effects associated with the covariates X, Z and W which are the counter variables, then in

One-way model: $Y = \alpha + X\beta + \varepsilon$

Two-way model: $Y = \alpha + X\beta + Z\gamma + \varepsilon$

Three-way model: $Y = \alpha + X\beta + Z\gamma + W\delta + \varepsilon$ and so on.

Consider an example of agricultural yield. The study variable Y denotes the yield which depends on various covariates $X_1, X_2,..., X_p$. In the case of regression analysis, the covariates $X_1, X_2,..., X_p$ are the different variables like temperature, the quantity of fertilizer, amount of irrigation etc.

Now consider the case of one-way model and try to understand its interpretation in terms of the multiple regression model. The covariate X is now measured at different levels, e.g., if X is the quantity of fertilizer then suppose there are p possible values, say 1 Kg., 2 Kg.,..., p Kg. then $X_1, X_2,..., X_p$ denotes these p values in the following way.

The linear model now can be expressed as

$$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon$$

by defining

$$X_1 = \begin{cases} 1 \text{ if effect of 1 Kg. fertilizer is present} \\ 0 \text{ if effect of 1 Kg. fertilizer is absent} \end{cases}$$

$$X_2 = \begin{cases} 1 \text{ if effect of 2 Kg. fertilizer is present} \\ 0 \text{ if effect of 2 Kg. fertilizer is absent} \end{cases}$$

$$\vdots$$

$$X_p = \begin{cases} 1 \text{ if effect of } p \text{ Kg. fertilizer is present} \\ 0 \text{ if effect of } p \text{ Kg. fertilizer is absent.} \end{cases}$$

If the effect of 1 Kg. of fertilizer is present, then other effects will obviously be absent and the linear model is expressible as

$$\begin{aligned} Y &= \beta_0 + \beta_1 (X_1 = 1) + \beta_2 (X_2 = 0) + ... + \beta_p (X_p = 0) + \varepsilon \\ &= \beta_0 + \beta_1 + \varepsilon \end{aligned}$$

If the effect of 2 Kg. of fertilizer is present then

$$\begin{aligned} Y &= \beta_0 + \beta_1 (X_1 = 0) + \beta_2 (X_2 = 1) + ... + \beta_p (X_p = 0) + \varepsilon \\ &= \beta_0 + \beta_2 + \varepsilon \end{aligned}$$

If the effect of p Kg. of fertilizer is present then

$$\begin{aligned} Y &= \beta_0 + \beta_1 (X_1 = 0) + \beta_2 (X_2 = 0) + ... + \beta_p (X_p = 1) + \varepsilon \\ &= \beta_0 + \beta_p + \varepsilon \end{aligned}$$

and so on.

If the experiment with 1 Kg. of fertilizer is repeated n_1 number of times then n_1 observation on response variables are recorded which can be represented as

$$
\begin{aligned}
Y_{11} &= \beta_0 + \beta_1.1 + \beta_2.0 + ... + \beta_p.0 + \varepsilon_{11} \\
Y_{12} &= \beta_0 + \beta_1.1 + \beta_2.0 + ... + \beta_p.0 + \varepsilon_{12} \\
&\vdots \\
Y_{1n_1} &= \beta_0 + \beta_1.1 + \beta_2.0 + ... + \beta_p.0 + \varepsilon_{1n1}
\end{aligned}
$$

If $X_2 = 1$ is repeated n_2 times, then on the same lines n_2 number of times then n_2 observation on response variables are recorded which can be represented as

$$
\begin{aligned}
Y_{21} &= \beta_0 + \beta_1.0 + \beta_2.1 + ... + \beta_p.0 + \varepsilon_{21} \\
Y_{22} &= \beta_0 + \beta_1.0 + \beta_2.1 + ... + \beta_p.0 + \varepsilon_{22} \\
&\vdots \\
Y_{2n_2} &= \beta_0 + \beta_1.0 + \beta_2.1 + ... + \beta_p.0 + \varepsilon_{2n_2}
\end{aligned}
$$

The experiment is continued and if $X_p = 1$ is repeated n_p times, then on the same lines

$$
\begin{aligned}
Y_{p1} &= \beta_0 + \beta_1.0 + \beta_2.0 + ... + \beta_p.1 + \varepsilon_{P1} \\
Y_{p2} &= \beta_0 + \beta_1.0 + \beta_2.0 + ... + \beta_p.1 + \varepsilon_{P2} \\
&\vdots \\
Y_{pn_p} &= \beta_0 + \beta_1.0 + \beta_2.0 + ... + \beta_p.1 + \varepsilon_{pn_p}
\end{aligned}
$$

All these $n_1, n_2, .., n_p$ observations can be represented as

$$
\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{p1} \\ y_{p2} \\ \vdots \\ y_{pn_p} \end{pmatrix}
=
\begin{pmatrix}
1 & 1 & 0 & 0 \cdots 0 & 0 \\
1 & 1 & 0 & 0 \cdots 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \ddots \vdots & \vdots \\
1 & 1 & 0 & 0 \cdots 0 & 0 \\
1 & 0 & 1 & 0 \cdots 0 & 0 \\
1 & 0 & 1 & 0 \cdots 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \ddots \vdots & \vdots \\
1 & 0 & 1 & 0 \cdots 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \ddots \vdots & \vdots \\
1 & 0 & 0 & 0 \cdots 0 & 1 \\
1 & 0 & 0 & 0 \cdots 0 & 1 \\
\vdots & \vdots & \vdots & \vdots \ddots \vdots & \vdots \\
1 & 0 & 0 & 0 \cdots 0 & 1
\end{pmatrix}
\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}
+
\begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n_2} \\ \vdots \\ \varepsilon_{p1} \\ \varepsilon_{p2} \\ \vdots \\ \varepsilon_{pn_p} \end{pmatrix}
$$

or

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

In the two-way analysis of variance model, there are two covariates and the linear model is expressible as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \gamma_1 Z_1 + \gamma_2 Z_2 + ... + \gamma_q Z_q + \varepsilon$$

where $X_1, X_2, ..., X_p$ denotes, e.g., the p levels of the quantity of fertilizer, say 1 Kg., 2 Kg.,...,p Kg. and $Z_1, Z_2, ..., Z_q$ denotes, e.g., the q levels of level of irrigation, say 10 Cms., 20 Cms.,…$10q$ Cms. etc. The levels $X_1, X_2, ..., X_p$ $Z_1, Z_2, ..., Z_q$ are the counter variable indicating the presence or absence of the effect as in the earlier case. If the effect of X_1 and Z_1 are present, i.e., 1 Kg of fertilizer and 10 Cms. of irrigation is used then the linear model is written as

$$\begin{aligned} Y &= \beta_0 + \beta_1.1 + \beta_2.0 + ... + \beta_p.0 + \gamma_1.1 + \gamma_2.0 + ... + \gamma_p.0 + \varepsilon \\ &= \beta_0 + \beta_1 + \gamma_1 + \varepsilon. \end{aligned}$$

If $X_2 = 1$ and $Z_2 = 1$ is used, then the model is

$$Y = \beta_0 + \beta_2 + \gamma_2 + \varepsilon.$$

The design matrix can be written accordingly as in the one-way analysis of variance case.
In the three-way analysis of variance model

$$Y = \alpha + \beta_1 X_1 + ... + \beta_p X_p + \gamma_1 Z_1 + ... + \gamma_q Z_q + \delta_1 W_1 + ... + \delta_r W_r + \varepsilon$$

The regression parameters β's can be fixed or random.

- If all β's are unknown constants, they are called as **parameters** of the model and the model is called as a **fixed effect model** or **model I**. The objective, in this case, is to make inferences about the parameters and the error variance σ^2 .
- If for some j, $x_{ij} = 1$ for all $i = 1, 2, ..., n$ then β_j is termed an **additive constant**. In this case, β_j occurs with every observation and so it is also called a **general mean effect**.
- If all β 's are observable random variables except the additive constant, then the linear model is termed as **random effect model**, **model II** or **variance components model**. The objective, in this case, is to make inferences about the variances of β's, i.e., $\sigma^2_{\beta_1}, \sigma^2_{\beta_2}, ..., \sigma^2_{\beta_p}$ and error variance σ^2 and/or certain functions of them..
- If some parameters are fixed and some are random variables, then the model is called a **mixed effect model** or **model III**. In the mixed effect model, at least one β_j is constant and at least one β_j is a random variable. The objective is to make inference about the fixed effect parameters, variance of random effects and error variance σ^2.

Analysis of variance

Analysis of variance is a body of statistical methods of analyzing the measurements assumed to be structured as

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

where x_{ij} are integers, generally 0 or 1 indicating usually the absence or presence of effects β_j; and ε_i's are assumed to be identically and independently distributed with mean 0 and variance σ^2. It may be noted that the ε_i's can be assumed additionally to follow a normal distribution $N(0, \sigma^2)$. It is needed for the maximum likelihood estimation of parameters from the beginning of the analysis, but in the least-squares estimation, it is needed only when conducting the tests of hypothesis and the confidence interval estimation of parameters. The least-squares method does not require any knowledge of distribution like normal up to the stage of estimation of parameters.

We need some basic concepts to develop tools.

Least squares estimate of β:

Let $y_1, y_2, ..., y_n$ be a sample of observations on $Y_1, Y_2, ..., Y_n$. The least-squares estimate of β is the values $\hat{\beta}$ of β for which the sum of squares due to errors, i.e.,

$$\begin{aligned} S^2 &= \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta) \\ &= y'y - 2X'y + \beta' X'X\beta \end{aligned}$$

is minimum where $y = (y_1, y_2, ..., y_n)'$. Differentiating S^2 with respect to β and substituting it to be zero, the normal equations are obtained as

$$\frac{dS^2}{d\beta} = 2X'X\beta - 2X'y = 0$$

or $X'X\beta = X'y$.

If X has full rank p, then $(X'X)$ has a unique inverse and the unique least squares estimate of β is

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

which is the best linear unbiased estimator of β in the sense of having minimum variance in the class of linear and unbiased estimator. If the rank of X is not full, then generalized inverse is used for finding the inverse of $(X'X)$.

If $L'\beta$ is a linear parametric function where $L = (\ell_1, \ell_2, ..., \ell_p)'$ is a non-null vector, then the least-squares estimate of $L'\beta$ is $L'\hat{\beta}$.

A question arises that what are the conditions under which a linear parametric function $L'\beta$ admits a unique least-squares estimate in the general case.

The concept of estimable function is needed to find such conditions.

Estimable functions:

A linear function $\lambda'\beta$ of the parameters with known λ is said to be an estimable parametric function (or estimable) if there exists a linear function $L'Y$ of Y such that

$E(L'Y) = \lambda'\beta$ for all $\beta \in R^p$.

Note that not all parametric functions are estimable.

Following results will be useful in understanding further topics.

Theorem 1: A linear parametric function $L'\beta$ admits a unique least squares estimate if and only if $L'\beta$ is estimable.

Theorem 2 (Gauss Markoff theorem):

If the linear parametric function $L'\beta$ is estimable then the linear estimator $L'\hat{\beta}$ where $\hat{\beta}$ is a solution of

$$X'X\hat{\beta} = X'Y$$

is the best linear unbiased estimator of $L'\beta$ in the sense of having minimum variance in the class of all linear and unbiased estimators of $L'\beta$.

Theorem 3: If the linear parametric function $\phi_1 = l_1'\beta, \phi_2 = l_2'\beta, ..., \phi_k = l_k'\beta$ are estimable, then any linear combination of $\phi_1, \phi_2, ..., \phi_k$ is also estimable.

Theorem 4: All linear parametric functions in β are estimable if and only if X has full rank.

If X is not of full rank, then some linear parametric functions do not admit the unbiased linear estimators and nothing can be inferred about them. The linear parametric functions which are not estimable are said to be **confounded**. A possible solution to this problem is to add linear restrictions on β so as to reduce the linear model to full rank.

Theorem 5: Let $L_1'\beta$ and $L_2'\beta$ be two estimable parametric functions and let $L_1'\hat{\beta}$ and $L_2'\hat{\beta}$ be their least squares estimators. Then

$$Var(L_1'\hat{\beta}) = \sigma^2 \; L_1'(X'X)^{-1}L_1$$
$$Cov(L_1'\hat{\beta}, L_2'\hat{\beta}) = \sigma^2 \; L_1'(X'X)^{-1}L_2$$

assuming that X is a full rank matrix. If not, the generalized inverse of $X'X$ can be used in place of the unique inverse.

Estimator of σ^2 based on least squares estimation:

Consider an estimator of σ^2 as,

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-p}(y - X\hat{\beta})'(y - X\hat{\beta}) \\
&= \frac{1}{n-p}[y - X(X'X)^{-1}X'y]'[y - X(X'X)^{-1}X'y] \\
&= \frac{1}{n-p}y'[I - X(X'X)^{-1}X'][I - X(X'X)^{-1}X']y \\
&= \frac{1}{n-p}y'[I - X(X'X)^{-1}X']y
\end{aligned}$$

where the hat matrix $[I - X(X'X)^{-1}X']$ is an idempotent matrix with its trace as

$$\begin{aligned}
tr[I - X(X'X)^{-1}] &= trI - trX(X'X)^{-1}X' \\
&= n - tr(X'X)^{-1}X'X \text{ (using the result } tr(AB) = tr(BA)) \\
&= n - tr\, I_p \\
&= n - p.
\end{aligned}$$

Note that using $E(y'Ay) = tr(A\Sigma)$, with $Cov(y) = \Sigma$, we have

$$\begin{aligned}
E(\hat{\sigma}^2) &= \frac{\sigma^2}{n-p}tr[I - X(X'X)^{-1}X'] \\
&= \sigma^2
\end{aligned}$$

and so $\hat{\sigma}^2$ is an unbiased estimator of σ^2.

Maximum Likelihood Estimation

The least-squares method does not use any distribution of the random variables in case of the estimation of parameters. We need the distributional assumption in case of least squares only while constructing the tests for hypothesis and the confidence intervals. For maximum likelihood estimation, we need the distributional assumption from the beginning.

Suppose $y_1, y_2, ..., y_n$ are independently and identically distributed following a normal distribution with mean $E(y_i) = \sum_{j=1}^{p} \beta_j x_{ij}$ and variance $Var(y_i) = \sigma^2$ *(i* =1, 2,…, *n).* Then the likelihood function of $y_1, y_2, ..., y_n$ is

$$L(y \big| \beta, \sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \right]$$

where $y = (y_1, y_2, ..., y_n)'$. Then

$$L = \ln L(y \big| \beta, \sigma^2) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta).$$

Differentiating the log-likelihood with respect to β and σ^2, we have

$$\frac{\partial L}{\partial \beta} = 0 \Rightarrow X'X\tilde{\beta} = X'y$$

$$\frac{\partial L}{\partial \sigma^2} = 0 \Rightarrow \tilde{\sigma}^2 = \frac{1}{n} (y - X\tilde{\beta})'(y - X\tilde{\beta})$$

Assuming the full rank of X, the normal equations are solved and the maximum likelihood estimators are obtained as

$$\tilde{\beta} = (X'X)^{-1} X'y$$

$$\tilde{\sigma}^2 = \frac{1}{n} (y - X\tilde{\beta})'(y - X\tilde{\beta})$$

$$= \frac{1}{n} y' \left[I - X(X'X)^{-1} X' \right] y.$$

The second-order differentiation conditions can be checked and they are satisfied for $\hat{\beta}$ and $\tilde{\sigma}^2$ to be the maximum likelihood estimators.

Note that in the maximum likelihood estimator $\tilde{\beta}$ is the same as the least-squares estimator $\tilde{\beta}$ and

- $\tilde{\beta}$ is an unbiased estimator of β ,i.e., $E(\tilde{\beta}) = \beta$ unlike the least-squares estimator but
- $\tilde{\sigma}^2$ is not an unbiased estimator of σ^2 ,i.e., $E(\tilde{\sigma}^2) = \frac{n-p}{n}\sigma^2 \neq \sigma^2$ like the least-squares estimator.

Now we use the following theorems for developing the test of hypothesis.

Theorem 6: Let $Y = (Y_1, Y_2, ..., Y_n)'$ follow a multivariate normal distribution $N(\mu, \Sigma)$ with mean vector μ and positive definite covariance matrix Σ. Then $Y'AY$ follows a noncentral chi-square distribution with p degrees of freedom and noncentrality parameter $\mu' A\mu$, i.e., $\chi^2(p, \mu' A\mu)$ if and only if ΣA is an idempotent matrix of rank p.

Theorem 7: Let $Y = (Y_1, Y_2, ..., Y_n)'$ follows a multivariate normal distribution $N(\mu, \Sigma)$ with mean vector μ and positive definite covariance matrix Σ. Let $Y'A_1Y$ follows $\chi^2(p_1, \mu' A_1\mu)$ and $Y'A_2Y$ follows $\chi^2(p_2, \mu' A_2\mu)$. Then $Y'A_1Y$ and $Y'A_2Y$ are independently distributed if $A_1\Sigma A_2 = 0$.

Theorem 8: Let $Y = (Y_1, Y_2, ..., Y_n)'$ follows a multivariate normal distribution $N(\mu, \sigma^2 I)$, then the maximum likelihood (or least squares) estimator $L'\hat{\beta}$ of estimable linear parametric function is independently distributed of $\hat{\sigma}^2$; $L\hat{\beta}$ follows $N\left[L'\beta, L'(X'X)^{-1}L\right]$ and $\frac{n\hat{\sigma}^2}{\sigma^2}$ follows $\chi^2(n-p)$ where $rank(X) = p$.

Proof: Consider $\hat{\beta} = (X'X)^{-1}X'Y$, then

$$\begin{aligned}
E(L\hat{\beta}) &= L(X'X)^{-1}X'E(Y) \\
&= L(X'X)^{-1}X'X\beta \\
&= L\beta \\
Var(L\hat{\beta}) &= L'Var(\hat{\beta})L \\
&= L'E(\hat{\beta}-\beta)'(\hat{\beta}-\beta)L \\
&= \sigma^2 L'(X'X)^{-1}L.
\end{aligned}$$

Since $\hat{\beta}$ is a linear function of y and $L'\hat{\beta}$ is a linear function of $\hat{\beta}$, so $L'\hat{\beta}$ follows a normal distribution $N\left[L\beta, \sigma^2 L'(X'X)^{-1}L\right]$. Let $A = I - X(X'X)^{-1}X'$ and $B = L(X'X)^{-1}X'$, then

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

and $n\hat{\sigma}^2 = (Y - X\beta)'\left[I - X(X'X)^{-1}X'\right](Y - X\beta) = Y'AY.$

So, using Theorem 6 with *rank*(*A*) = *n* - *p*, $\frac{n\hat{\sigma}^2}{\sigma^2}$ follows a $\chi^2(n-p)$. Also

$$\begin{aligned} BA &= L'(X'X)^{-1}X' - L'(X'X)^{-1}X'X(X'X)^{-1}X' \\ &= 0. \end{aligned}$$

So using Theorem 7, $Y'AY$ and BY are independently distributed.

Tests of Hypothesis in the Linear Regression Model

First, we discuss the development of the tests of hypothesis concerning the parameters of a linear regression model. These tests of the hypothesis will be used later in the development of tests based on the analysis of variance.

Analysis of Variance

The technique in the analysis of variance involves the breaking down of total variation into orthogonal components. Each orthogonal factor represents the variation due to a particular factor contributing in the total variation.

Model

Let $Y_1, Y_2, ..., Y_n$ be independently distributed following a normal distribution with mean $E(Y_i) = \sum_{j=1}^{p} \beta_j x_{ij}$ and variance σ^2. Denoting $Y = (Y_1, Y_2, ..., Y_n)'$ a $n \times 1$ column vector, such assumption can be expressed in the form of a linear regression model

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

where X is a $n \times p$ matrix, β is a $p \times 1$ vector and ε is a $n \times 1$ vector of disturbances with

$E(\varepsilon) = 0$

$Cov(\varepsilon) = \sigma^2 I$ and ε follows a normal distribution.

This implies that

$$E(Y) = X\beta$$

$$E(Y - X\beta)(Y - X\beta)' = \sigma^2 I.$$

Now we consider four different types of tests of hypothesis.

In the first two cases, we develop the likelihood ratio test for the null hypothesis related to the analysis of variance. Note that, later we will derive the same test on the basis of least squares principle also. An important idea behind the development of this test is to demonstrate that the test used in the analysis of variance can be derived using the least-squares principle as well as the likelihood ratio test.

Case 1: Consider the null hypothesis for testing $H_0 : \beta = \beta^0$ where $\beta = (\beta_1, \beta_2, ..., \beta_p)', \beta^0 = (\beta_1^0, \beta_2^0, ..., \beta_p^0)'$ is specified and σ^2 is unknown. This null hypothesis is equivalent to $H_0 : \beta_1 = \beta_1^0, \beta_2 = \beta_2^0, ..., \beta_p = \beta_p^0.$

Assume that all β_i 's are estimable, i.e., $rank(X) = p$ (full column rank). We now develop the likelihood ratio test.

The $(p+1)\times 1$ dimensional parametric space Ω is a collection of points such that

$$\Omega = \left\{(\beta, \sigma^2); -\infty < \beta_i < \infty, \sigma^2 > 0, i = 1, 2, ... p\right\}.$$

Under H_0, all β's 's are known and equal, say β^0 all are known and the Ω reduces to one-dimensional space given by

$$\omega = \left\{(\beta^0, \sigma^2); \sigma^2 > 0\right\}.$$

The likelihood function of $y_1, y_2, ..., y_n$ is

$$L(y|\beta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta^0)'(y - X\beta^0)\right]$$

The likelihood function is maximum over Ω when β and σ^2 are substituted with their maximum likelihood estimators, i.e.,

$$\hat{\beta} = (X'X)^{-1}X'y$$
$$\hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta}).$$

Substituting $\hat{\beta}$ and $\hat{\sigma}^2$ in $L(y|\beta, \sigma^2)$ gives

$$\underset{\Omega}{Max}\ L(y|\beta, \sigma^2) = \left(\frac{1}{2\pi\hat{\sigma}^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\hat{\sigma}^2}(y - X\hat{\beta})'(y - X\hat{\beta})\right)$$
$$= \left(\frac{n}{2\pi(y - X\hat{\beta})'(y - X\hat{\beta})}\right)^{\frac{n}{2}} \exp\left(-\frac{n}{2}\right).$$

Under H_0, the maximum likelihood estimator of σ^2 is $\hat{\sigma}^2 = \frac{1}{n}(y - X\beta^0)'(y - X\beta^0).$

The maximum value of the likelihood function under H_0 is

$$\underset{\omega}{Max}\ L(y|\beta,\sigma^2) = \left(\frac{1}{2\pi\hat{\sigma}^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\hat{\sigma}^2}(y - X\beta^0)'(y - X\beta^0)\right)$$

$$= \left(\frac{n}{2\pi(y - X\beta^0)'(y - X\beta^0)}\right)^{\frac{n}{2}} \exp\left(-\frac{n}{2}\right)$$

The likelihood ratio test statistic is

$$\lambda = \frac{\underset{\omega}{Max}\, L(y|\beta,\sigma^2)}{\underset{\Omega}{Max}\, L(y|\beta,\sigma^2)}$$

$$= \left[\frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{(y - X\beta^0)'(y - X\beta^0)}\right]^{\frac{n}{2}}$$

$$= \left[\frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{\left[(y - X\hat{\beta}) + (X\hat{\beta} - X\beta^0)\right]'\left[(y - X\hat{\beta}) + (X\hat{\beta} - X\beta^0)\right]}\right]^{\frac{n}{2}}$$

$$= \left[1 + \frac{(\hat{\beta} - \beta^0)'X'X(\hat{\beta} - \beta^0)}{(y - X\hat{\beta})'(y - X\hat{\beta})}\right]^{-\frac{n}{2}}$$

$$= \left[1 + \frac{q_1}{q_2}\right]^{-\frac{n}{2}}$$

where $q_2 = (y - X\hat{\beta})'(y - X\hat{\beta})$

and $q_1 = (\hat{\beta} - \beta^0)'X'X(\hat{\beta} - \beta^0).$

The expression of q_1 and q_2 can be further simplified as follows.

Consider

$$q_1 = (\hat{\beta} - \beta^0)'X'X(\hat{\beta} - \beta^0)$$

$$= \left[(X'X)^{-1}X'y - \beta^0\right]' X'X \left[(X'X)^{-1}X'y - \beta^0\right]$$

$$= \left[(X'X)^{-1}X'(y - X\beta^0)\right]' X'X \left[(X'X)^{-1}X'(y - X\beta^0)\right]$$

$$= (y - X\beta^0)'X(X'X)^{-1}X'X(X'X)^{-1}X'(y - X\beta^0)$$

$$= (y - X\beta^0)'X(X'X)^{-1}X'(y - X\beta^0)$$

$$
\begin{aligned}
q_2 &= (y - X\hat{\beta})'(y - X\hat{\beta}) \\
&= \left[y - X(X'X)^{-1}X'y\right]'\left[y - X(X'X)^{-1}X'y\right] \\
&= y'\left[I - X(X'X)^{-1}X'\right]y \\
&= [(y - X\beta^0) + X\beta^0]'[I - X(X'X)^{-1}X'][(y - X\beta^0)' + X\beta^0] \\
&= (y - X\beta^0)'[I - X(X'X)^{-1}X'](y - X\beta^0)
\end{aligned}
$$

Other two terms become zero using

$$[I - X(X'X)^{-1}X']X = 0$$

In order to find out the decision rule for H_0 based on λ, first, we need to find if λ is a monotonic increasing or decreasing function of $\frac{q_1}{q_2}$. So we proceed as follows:

Let $g = \frac{q_1}{q_2}$, so that

$$
\begin{aligned}
\lambda &= \left(1 + \frac{q_1}{q_2}\right)^{-\frac{n}{2}} \\
&= (1+g)^{-\frac{n}{2}}
\end{aligned}
$$

then

$$\frac{d\lambda}{dg} = -\frac{n}{2}\frac{1}{(1+g)^{\frac{n}{2}+1}}$$

So as g increases, λ decreases.

Thus λ is a monotonic decreasing function of $\frac{q_1}{q_2}$.

The decision rule is to reject H_0 if $\lambda \le \lambda_0$ where λ_0 is a constant to be determined on the basis of size of the test α. Let us simplify this in our context.

$$\lambda \le \lambda_0$$

or $$\left(1 + \frac{q_1}{q_2}\right)^{-\frac{n}{2}} \le \lambda_o$$

or $$\frac{1}{(1+g)^{\frac{n}{2}}} \le \lambda_o$$

or $\quad (1+g) \geq \lambda_0^{-\frac{2}{n}}$

or $\quad g \geq \lambda_0^{-\frac{2}{n}} - 1$

or $\quad g \geq C$

where C is a constant to be determined by the size α condition of the test.

So reject H_0 whenever $\dfrac{q_1}{q_2} \geq C.$

Note that the statistic $\dfrac{q_1}{q_2}$ can also be obtained by the least-squares method as follows. The least-squares methodology will also be discussed in further lectures.

$$q_1 = (\hat{\beta} - \beta^0)' X'X (\hat{\beta} - \beta^0)$$

$$q_1 \quad = \quad \underset{\omega}{Min}(y - X\beta)'(y - X\beta) \quad - \quad \underset{\Omega}{Min}(y - X\beta)'(y - X\beta)$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

sum of squares due to deviation from H_o OR sum of squares due to β	sum of squares due to H_o OR Total sum of squares	sum of squares due to error

It will be seen later that the test statistic will be based on the ratio $\dfrac{q_1}{q_2}$. In order to find an appropriate distribution of $\dfrac{q_1}{q_2}$, we use the following theorem:

Theorem 9: Let

$$Z = Y - X\beta^0$$

$$Q_1 = Z'X(X'X)^{-1}X'Z$$

$$Q_2 = Z'[I - X(X'X)^{-1}X']Z.$$

Then $\dfrac{Q_1}{\sigma^2}$ and $\dfrac{Q_2}{\sigma^2}$ are independently distributed. Further, when H_0 is true, then $\dfrac{Q_1}{\sigma^2} \sim \chi^2(p)$ and $\dfrac{Q_2}{\sigma^2} \sim \chi^2(n-p)$ where $\chi^2(m)$ denotes the χ^2 distribution with 'm' degrees of freedom.

Proof: Under H_0,

$$E(Z) = X\beta^0 - X\beta^0 = 0$$
$$Var(Z) = Var(Y) = \sigma^2 I.$$

Further Z is a linear function of Y and Y follows a normal distribution. So

$$Z \sim N(0, \sigma^2 I)$$

The matrices $X(X'X)^{-1}X'$ and $[I - X(X'X)^{-1}X']$ are idempotent matrices. So

$$tr[X(X'X)^{-1}X'] = tr[(X'X)^{-1}X'X] = tr(I_p) = p$$
$$tr[I - X(X'X)^{-1}X'] = tr\, I_n - tr[X(X'X)^{-1}X'] = n - p$$

So using theorem 6, we can write that under H_0

$$\frac{Q_1}{\sigma^2} \sim \chi^2(p) \quad \text{and} \quad \frac{Q_2}{\sigma^2} \sim \chi^2(n-p)$$

where the degrees of freedom p and $(n-p)$ are obtained by the trace of $\left[X(X'X)^{-1}X'\right]$ and trace of $\left[I - X(X'X)^{-1}X'\right]$, respectively.

Since

$$\left[I - X(X'X)^{-1}X'\right]\left[X(X'X)^{-1}X'\right] = 0,$$

So using theorem 7, the quadratic forms Q_1 and Q_2 are independent under H_0.

Hence the theorem is proved.

Since Q_1 and Q_2 are independently distributed, so under H_0

$\dfrac{Q_1/p}{Q_2/(n-p)}$ follows a central F distribution, i.e.

$$\left(\frac{n-p}{p}\right)\frac{Q_1}{Q_2} \sim F(p, n-p).$$

Hence the constant C in the likelihood ratio test statistic λ is given by

$$C = F_{1-\alpha}(p, n-p)$$

where $F_{1-\alpha}(n_1, n_2)$ denotes the upper $100\alpha\%$ points of F-distribution with n_1 and n_2 degrees of freedom.

The computations of this test of hypothesis can be represented in the form of an analysis of variance table.

ANOVA for testing $H_0 : \beta = \beta^0$

Source of variation	Degrees of freedom	Sum of squares	Mean squares	F-value
Due to β	p	q_1	$\frac{q_1}{p}$	$\left(\frac{n-p}{p}\right)\frac{q_1}{q_2}$
Error	$n-p$	q_2	$\frac{q_2}{(n-p)}$	
Total	n	$(y - X\beta^0)'(y - X\beta^0)$		

Case 2: Test of a subset of parameters $H_0 : \beta_k = \beta_k^0, k = 1, 2, .., r < p$ when $\beta_{r+1}, \beta_{r+2}, ..., \beta_p$ and σ^2 are unknown.

In case 1 , the test of hypothesis was developed when all β 's are considered in the sense that we test for each $\beta_i = \beta_i^0,\ i = 1, 2, ..., p.$ Now consider another situation, in which the interest is to test only a subset of $\beta_1, \beta_2, ..., \beta_p$, i.e., not all but only a few parameters. This type of test of hypothesis can be used, e.g., in the following situation. Suppose five levels of voltage are applied to check the rotations per minute (rpm) of a fan at 160 volts, 180 volts, 200 volts, 220 volts and 240 volts. It can be realized in practice that when the voltage is low, then the rpm at 160, 180 and 200 volts can be observed easily. At 220 and 240 volts, the fan rotates at the full speed and there is not much difference in the rotations per minute at these voltages. So the interest of the experimenter lies in testing the hypothesis related to only first three effects, viz., β_1, for 160 volts, β_2 for 180 volts and β_3 for 200 volts. The null hypothesis in this case can be written as:

$H_0 : \beta_1 = \beta_1^0, \beta_2^0 = \beta_2^0, \beta_3 = \beta_3^0$

when β_4, β_5 and σ^2 are unknown.

Note that under case 1, the null hypothesis will be

$H_0 : \beta_1 = \beta_1^0, \beta_2^0 = \beta_2^0, \beta_3 = \beta_3^0, \beta_4 = \beta_4^0, \beta_5 = \beta_5^0$.

Let $\beta_1, \beta_2, ..., \beta_p$ be the p parameters.

We can divide them into two parts such that out of $\beta_1, \beta_2, ..., \beta_r, \beta_{r+1}, ..., \beta_p$ and we are interested in testing a hypothesis of a subset of it.

Suppose, We want to test the null hypothesis

$H_0 : \beta_k = \beta_k^0, k = 1, 2, .., r < p$ when $\beta_{r+1}, \beta_{r+2}, ..., \beta_p$ and σ^2 are unknown.

The alternative hypothesis under consideration is $H_1 : \beta_k \neq \beta_k^0$ for at least one $k = 1, 2, .., r < p$.

In order to develop a test for such a hypothesis, the linear model

$Y = X\beta + \varepsilon$

under the usual assumptions can be rewritten as follows:

Partition $X = (X_1 \ X_2)$, $\beta = \begin{pmatrix} \beta_{(1)} \\ \beta_{(2)} \end{pmatrix}$

where $\beta_{(1)} = (\beta_1, \beta_2, ..., \beta_r)'$, $\beta_{(2)} = (\beta_{r+1}, \beta_{r+2}, ..., \beta_p)'$

with the order as $X_1 : n \times r$, $X_2 : n \times (p-r)$, $\beta_{(1)} : r \times 1$ and $\beta_{(2)} : (p-r) \times 1$.

The model can be rewritten as

$$\begin{aligned} Y &= X\beta + \varepsilon \\ &= (X_1 \ X_2)\begin{pmatrix} \beta_{(1)} \\ \beta_{(2)} \end{pmatrix} + \varepsilon \\ &= X_1\beta_{(1)} + X_2\beta_{(2)} + \varepsilon \end{aligned}$$

The null hypothesis of interest is now

$H_0 : \beta_{(1)} = \beta_{(1)}^0 = (\beta_1^0, \beta_2^0, ..., \beta_r^0)$ where $\beta_{(2)}$ and σ^2 are unknown.

The complete parametric space is

$$\Omega = \left\{ (\beta, \sigma^2); -\infty < \beta_i < \infty, \sigma^2 > 0, i = 1, 2, ..., p \right\}$$

and sample space under H_0 is

$$\omega = \left\{ (\beta_{(1)}^0, \beta_{(2)}, \sigma^2); -\infty < \beta_i < \infty, \sigma^2 > 0, \ i = r+1, r+2, ..., p \right\}.$$

The likelihood function is

$$L(y \mid \beta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta) \right].$$

The maximum value of likelihood function under Ω is obtained by substituting the maximum likelihood estimates of β and σ^2, i.e.,

$$\hat{\beta} = (X'X)^{-1}X'y$$
$$\hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta})$$

as

$$\underset{\Omega}{Max}\, L(y\big|\beta,\sigma^2) = \left(\frac{1}{2\pi\hat{\sigma}^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\hat{\sigma}^2}(y - X\hat{\beta})'(y - X\hat{\beta})\right]$$
$$= \left(\frac{n}{2\pi(y - X\hat{\beta})'(y - X\hat{\beta})}\right)^{\frac{n}{2}} \exp\left(-\frac{n}{2}\right).$$

Now we find the maximum value of likelihood function under H_0. The model under H_0 becomes $Y = X_1\beta_{(1)}^0 + X_2\beta_2 + \varepsilon$. The likelihood function under H_0 is

$$L(y\big|\beta,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}(y - X_1\beta_{(1)}^0 - X_2\beta_{(2)})'(y - X_1\beta_{(1)}^0 - X_2\beta_{(2)})\right]$$
$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}(y^* - X_2\beta_{(2)})'(y^* - X_2\beta_{(2)})\right]$$

where $y^* = y - X_1\beta_{(1)}^{(0)}$. Note that $\beta_{(2)}$ and σ^2 are the unknown parameters. This likelihood function looks like as if it is written for $y^* \sim N(X_2\beta_{(2)}, \sigma^2)$.

This helps is writing the maximum likelihood estimators of $\beta_{(2)}$ and σ^2 directly as

$$\hat{\beta}_{(2)} = (X_2'X_2)^{-1}X_2'y^*$$
$$\hat{\sigma}^2 = \frac{1}{n}(y^* - X_2\hat{\beta}_{(2)})'(y^* - X_2\hat{\beta}_{(2)}).$$

Note that $X_2'X_2$ is a principal minor of $X'X$. Since $X'X$ is a positive definite matrix, so $X_2'X_2$ is also positive definite. Thus $(X_2'X_2)^{-1}$ exists and is unique.

Thus the maximum value of likelihood function under H_0 is obtained as

$$\underset{\omega}{Max}\, L(y^*\big|\hat{\beta},\sigma^2) = \left(\frac{1}{2\pi\hat{\sigma}^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\hat{\sigma}^2}(y^* - X_2\hat{\beta}_{(2)})'(y^* - X_2\hat{\beta}_{(2)})\right]$$
$$= \left(\frac{n}{2\pi(y^* - X_2\hat{\beta}_{(2)})'(y^* - X_2\hat{\beta}_{(2)})}\right)^{\frac{n}{2}} \exp\left(-\frac{n}{2}\right)$$

The likelihood ratio test statistic for $H_0 : \beta_{(1)} = \beta_{(1)}^0$ is

$$\begin{aligned}
\lambda &= \frac{\max\limits_{\omega} L(y\,|\,\beta,\sigma^2)}{\max\limits_{\Omega} L(y\,|\,\beta,\sigma^2)} \\
&= \left[\frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{(y^* - X_2\hat{\beta}_{(2)})'(y^* - X_2\hat{\beta}_{(2)})} \right]^{\frac{n}{2}} \\
&= \left[\frac{(y^* - X_2\hat{\beta}_{(2)})'(y^* - X_2\hat{\beta}_{(2)}) - (y - X\hat{\beta})'(y - X\hat{\beta}) + (y - X\hat{\beta})'(y - X\hat{\beta})}{(y - X\hat{\beta})'(y - X\hat{\beta})} \right]^{-\frac{n}{2}} \\
&= \left[1 + \frac{(y^* - X_2\hat{\beta}_{(2)})'(y^* - X_2\hat{\beta}_{(2)}) - (y - X\hat{\beta})'(y - X\hat{\beta})}{(y - X\hat{\beta})'(y - X\hat{\beta})} \right]^{-\frac{n}{2}} \\
&= \left[1 + \frac{q_1}{q_2} \right]^{-\frac{n}{2}}
\end{aligned}$$

where $q_1 = (y^* - X_2\hat{\beta}_{(2)})'(y^* - X_2\hat{\beta}_{(2)}) - (y - X\hat{\beta})'(y - X\hat{\beta})$ and $q_2 = (y - X\hat{\beta})'(y - X\hat{\beta})$.

Now we simplify q_1 and q_2.

Consider

$$\begin{aligned}
&(y^* - X_2\hat{\beta}_{(2)})'(y^* - X_2\hat{\beta}_{(2)})' = \\
&= (y^* - X_2(X_2'X_2)^{-1}X_2'y^*)'(y^* - X_2(X_2'X_2)^{-1}X_2'y^*) \\
&= y^{*\prime}\left[I - X_2(X_2'X_2)^{-1}X_2' \right] y^* \\
&= \left[(y - X_1\beta_{(1)}^0 - X_2\beta_{(2)}) + X_2\beta_{(2)} \right]' \left[I - X_2(X_2'X_2)^{-1}X_2' \right] \left[(y - X_1\beta_{(1)}^0 - X_2\beta_{(2)}) + X_2\beta_{(2)} \right] \\
&= (y - X_1\beta_{(1)}^0 - X_2\beta_{(2)})' \left[I - X_2(X_2'X_2)^{-1}X_2' \right] (y - X_1\beta_{(1)}^0 - X_2\beta_{(2)}).
\end{aligned}$$

The other terms become zero using the result $X_2'\left[I - X_2(X_2'X_2)^{-1}X_2' \right] = 0$.

Note that under H_0, $X_1\beta_{(1)}^0 + X_2\beta_{(2)}$ can be expressed as $(X_1 \quad X_2)(\beta_{(1)}^0 \quad \beta_{(2)})'$, Consider

$$
\begin{aligned}
&(y - X\hat{\beta})'(y - X\hat{\beta})' = \\
&= (y - X(X'X)^{-1}X'y)'(y - X(X'X)^{-1}X'y) \\
&= y'\left[I - X(X'X)^{-1}X'\right]y \\
&= \left[(y - X_1\beta_{(1)}^0 - X_2\hat{\beta}_{(2)}) + X_1\beta_{(1)}^0 + X_2\beta_{(2)})\right]' \\
&\quad \times\left[I - X(X'X)^{-1}X'\right]\left[(y - X_1\beta_{(1)}^0 - X_2\hat{\beta}_{(2)}) + X_1\beta_{(1)}^0 + X_2\beta_{(2)})\right] \\
&\quad -(y - X_1\beta_{(1)}^0 - X_2\beta_{(2)})\left[I - X(X'X)^{-1}X'\right](y - X_1\beta_{(1)}^0 - X_2\beta_{(2)}) \\
&= (y - X_1\beta_{(1)}^0 - X_2\beta_{(2)})'\left[I - X(X'X)^{-1}X'\right](y - X_1\beta_{(1)}^0 - X_2\beta_{(2)})
\end{aligned}
$$

and another term becomes zero using the result $X'\left[I - X(X'X)^{-1}X'\right] = 0.$

Note that under H_0, the term $X_1\beta_{(1)}^0 + X_2\beta_{(2)}$ can be expressed as $(X_1 \quad X_2)(\beta_{(1)}^0 \quad \beta_{(2)})'$. Thus

$$
\begin{aligned}
q_1 &= (y^* - X_2\hat{\beta}_{(2)})'(y^* - X_2\hat{\beta}_{(2)})' - (y - X\hat{\beta})'(y - X\hat{\beta})' \\
&= y^{*\prime}\left[I - X_2(X_2'X_2)^{-1}X_2'\right]y^* - y'\left[I - X(X'X)^{-1}X'\right]y \\
&= (y - X_1\beta_{(1)}^0 - X_2\beta_{(2)})'\left[I - X_2(X_2'X_2)^{-1}X_2'\right](y - X_1\beta_{(1)}^0 - X_2\beta_{(2)}) \\
&\quad -(y - X_1\beta_{(1)}^0 - X_2\beta_{(2)})'\left[I - X(X'X)^{-1}X'\right](y - X_1\beta_{(1)}^0 - X_2\beta_{(2)}) \\
&= (y - X_1\beta_{(1)}^0 - X_2\beta_{(2)})'\left[X(X'X)^{-1}X' - X_2(X_2'X_2)^{-1}X_2'\right](y - X_1\beta_{(1)}^0 - X_2\beta_{(2)})
\end{aligned}
$$

$$
\begin{aligned}
q_2 &= (y - X\hat{\beta})'(y - X\hat{\beta}) \\
&= y'\left[I - X(X'X)'X'\right]y \\
&= \left[(y - X_1\beta_{(1)}^0 - X_2\beta_{(2)}) + (X_1\beta_{(1)}^0 + X_2\beta_{(2)})\right]'\left[I - X(X'X)'X'\right] \\
&\quad \times\left[(y - X_1\beta_{(1)}^0 - X_2\beta_{(2)}) + (X_1\beta_{(1)}^0 + X_2\beta_{(2)})\right] \\
&= (y - X_1\beta_{(1)}^0 - X_2\beta_{(2)})'\left[I - X(X'X)^{-1}X'\right](y - X_1\beta_{(1)}^0 - X_2\beta_{(2)}).
\end{aligned}
$$

Other terms become zero. Note that in simplifying the terms q_1 and q_2, we tried to write them in the quadratic form with the same variable $(y - X_1\beta_{(1)}^0 - X_2\beta_{(2)})$.

Using the same argument as in case 1, we can say that since λ is a monotonic decreasing function of $\frac{q_1}{q_2}$, so the likelihood ratio test rejects H_0 whenever

$$\frac{q_1}{q_2} > C$$

where C is a constant to be determined by the size α of the test.

The likelihood ratio test statistic can also be obtained through least-squares method as follows:

$(q_1 + q_2)$: Minimum value of $(y - X\beta)'(y - X\beta)$ when $H_0 : \beta_{(1)} = \beta_{(1)}^0$ holds true.

: Sum of squares due to H_0

q_2 : Sum of squares due to error.

q_1 : Sum of squares due to the deviation from H_0 or sum of squares due to $\beta_{(1)}$ adjusted for $\beta_{(2)}$.

If $\beta_{(1)}^0 = 0$ then

$$
\begin{aligned}
q_1 &= (y - X_2\hat{\beta}_{(2)})'(y - X_2\hat{\beta}_{(2)}) - (y - X\hat{\beta})'(y - X\hat{\beta}) \\
&= (y'y - \hat{\beta}_{(2)}' X_2' y) - (y'y - \hat{\beta}' X' y) \\
&= \hat{\beta}' X' y \qquad - \qquad \hat{\beta}_{(2)}' X_2' y.
\end{aligned}
$$

$\downarrow$ $\downarrow$

Reduction sum of squares or sum of squares due to β	sum of squares due to $\beta_{(2)}$ ignoring $\beta_{(1)}$

Now we have the following theorem based on Theorems 6 and 7.

Theorem 10: Let $Z = Y - X_1\beta_{(1)}^0 - X_2\beta_{(2)}$; $Q_1 = Z'AZ$ and $Q_2 = Z'BZ$ where $A = X(X'X)^{-1}X' - X_2(X_2'X_2)^{-1}X_2'$ and $B = I - X(X'X)^{-1}X'$.

Then $\frac{Q_1}{\sigma^2}$ and $\frac{Q_2}{\sigma^2}$ are independently distributed. Further $\frac{Q_1}{\sigma^2} \sim \chi^2(r)$ and $\frac{Q_2}{\sigma^2} \sim \chi^2(n-p)$.

Thus under H_0, $\frac{Q_1/r}{Q_2/(n-p)} = \frac{n-p}{r}\frac{Q_1}{Q_2}$ follows an F-distribution $F(r, n-p)$.

Hence the constant C in λ is

$$C = F_{1-\alpha}(r, n-p),$$

where $F_{1-\alpha}(r, n-p)$ denotes the upper $100\alpha\%$ points on F-distribution with r and $(n-p)$ degrees of freedom.

The analysis of variance table for this null hypothesis is as follows:

ANOVA for testing $H_0 : \beta_{(1)} = \beta_{(1)}^0$

Source of Variation	Degrees of Freedom	Sum of squares	Mean squares	F-value
Due to $\beta_{(1)}$	r	q_1	$\frac{q_1}{r}$	$\left(\frac{n-p}{r}\right)\frac{q_1}{q_2}$
Error	$n-p$	q_2	$\frac{q_2}{(n-p)}$	
Total	$n-(p-r)$	q_1+q_2		

Case 3: Test of $H_0 : L'\beta = \delta$

Let us consider the test of hypothesis related to a linear parametric function. Assuming that the linear parameter function $L'\beta$ is estimable where $L = (\ell_1, \ell_2, ..., \ell_p)'$ is a $p \times 1$ vector of known constants and $\beta = (\beta_1, \beta_2, ..., \beta_p)'$. The null hypothesis of interest is

$$H_0 : L'\beta = \delta.$$

where δ is some specified constant.

Consider the set up of linear model $Y = X\beta + \varepsilon$ where $Y = (Y_1, Y_2, ..., Y_n)'$ follows $N(X\beta, \sigma^2 I)$. The maximum likelihood estimators of β and σ^2 are

$$\hat{\beta} = (X'X)^{-1}X'y \text{ and } \hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta})$$

respectively. The maximum likelihood estimate of estimable $L'\beta$ is $L'\hat{\beta}$, with

$$E(L'\hat{\beta}) = L'\beta$$
$$Cov(L'\hat{\beta}) = \sigma^2 L'(X'X)^{-1}L$$
$$L'\hat{\beta} \sim N\left[L'\beta, \sigma^2 L'(X'X)^{-1}L\right]$$

and

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p)$$

assuming X to be the full column rank matrix. Further, $L'\hat{\beta}$ and $\frac{n\hat{\sigma}^2}{\sigma^2}$ are also independently distributed.

Under $H_0 : L'\beta = \delta$, the statistic

$$t = \frac{\sqrt{(n-p)}(L'\hat{\beta} - \delta)}{\sqrt{n\hat{\sigma}^2 L'(X'X)^{-1}L}}$$

follows a *t*-distribution with $(n-p)$ degrees of freedom. So the test for $H_0 : L'\beta = \delta$ against $H_1 : L'\beta \neq \delta$ rejects H_0 whenever

$$|t| \geq t_{1-\frac{\alpha}{2}}(n-p)$$

where $t_{1-\alpha}(n_1)$ denotes the upper 100α% points on *t*-distribution with n_1 degrees of freedom.

Case 4: Test of $H_0 : \phi_1 = \delta_1, \phi_2 = \delta_2, ..., \phi_k = \delta_k$

Now we develop the test of hypothesis related to more than one linear parametric functions. Let the i^{th} estimable linear parametric function is

$\phi_i = L_i'\beta$ and there are *k* such functions with L_i and β both being $p \times 1$ vectors as in the Case 3. Our interest is to test the hypothesis

$$H_0 : \phi_1 = \delta_1, \phi_2 = \delta_2, ..., \phi_k = \delta_k$$

where $\delta_1, \delta_2, ..., \delta_k$ are the known constants.

Let $\phi = (\phi_1, \phi_2, ..., \phi_k)'$ and $\delta = (\delta_1, \delta_2, .., \delta_k)'$.

Then H_0 is expressible as $H_0 : \phi = L'\beta = \delta$

where L' is a $k \times p$ matrix of constants associated with $L_1, L_2, ..., L_k$. The maximum likelihood estimator of ϕ_i is : $\hat{\phi}_i = L_i'\hat{\beta}$ where $\hat{\beta} = (X'X)^{-1}X'y$.

Then $\hat{\phi} = (\hat{\phi}_1, \hat{\phi}_2, ..., \hat{\phi}_k) = L\hat{\beta}$.

Also $E(\hat{\phi}) = \phi$

$$Cov(\hat{\phi}) = \sigma^2 V$$

where $V = ((L_i'(X'X)^{-1}L_j))$ where $(L_i'(X'X)^{-1}L_j)$ is the $(i,j)^{th}$ element of V. Thus

$$\frac{(\hat{\phi} - \phi)'V^{-1}(\hat{\phi} - \phi)}{\sigma^2}$$

follows a χ^2-distribution with *k* degrees of freedom and

$\frac{n\hat{\sigma}^2}{\sigma^2}$ follows a χ^2-distribution with $(n-p)$ degrees of freedom where

$\hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta})$ is the maximum likelihood estimator of σ^2.

Further $\frac{(\hat{\phi}-\phi)'V^{-1}(\hat{\phi}-\phi)}{\sigma^2}$ and $\frac{n\hat{\sigma}^2}{\sigma^2}$ are also independently distributed.

Thus under $H_0 : \phi = \delta$

$$\frac{\left(\frac{\frac{(\hat{\phi}-\delta)'V^{-1}(\hat{\phi}-\delta)}{\sigma^2}}{k}\right)}{\left(\frac{\frac{n\hat{\sigma}^2}{\sigma^2}}{(n-p)}\right)}$$

or $\left(\frac{n-p}{k}\right)\frac{(\hat{\phi}-\delta)'V^{-1}(\hat{\phi}-\delta)}{n\hat{\sigma}^2}$

follows F- distribution with k and $(n-p)$ degrees of freedom. So the hypothesis $H_0 : \phi = \delta$ is rejected against H_1 : At least one $\phi_i \neq \delta_i$ for $i = 1, 2, ..., k$ whenever $F \geq F_{1-\alpha}(k, n-p)$ where $F_{1-\alpha}(k, n-p)$ denotes the $100\alpha\%$ points on F-distribution with k and $(n-p)$ degrees of freedom.

One-way classification with fixed effect linear models of full rank:

The objective in the one-way classification is to test the hypothesis about the equality of means on the basis of several samples which have been drawn from univariate normal populations with different means but the same variances.

Let there be p univariate normal populations and samples of different sizes are drawn from each of the population. Let $y_{ij}\,(j = 1, 2, ..., n_i)$ be a random sample from the i^{th} normal population with mean β_i and variance $\sigma^2, i = 1, 2, ..., p$, i.e.,

$Y_{ij} \sim N(\beta_i, \sigma^2),\ j = 1, 2, ..., n_i;\ i = 1, 2, ..., p.$

The random samples from different populations are assumed to be independent of each other.

These observations follow the set up of linear model

$Y = X\beta + \varepsilon$

where

$$Y = (Y_{11}, Y_{12}, ..., Y_{1n_1}, Y_{21}, ..., Y_{2n_2}, ..., Y_{p1}, Y_{p2}, ..., Y_{pn_p})'$$
$$y = (y_{11}, y_{12}, ..., y_{1n_1}, y_{21}, ..., y_{2n_2}, ..., y_{p1}, y_{p2}, ..., y_{pn_p})'$$
$$\beta = (\beta_1, \beta_2, ..., \beta_p)'$$
$$\varepsilon = (\varepsilon_{11}, \varepsilon_{12}, ..., \varepsilon_{1n_1}, \varepsilon_{21}, ..., \varepsilon_{2n_2}, ..., \varepsilon_{p1}, \varepsilon_{p2}, ..., \varepsilon_{pn_p})'$$

$$X = \begin{pmatrix} \left.\begin{matrix} 1\ 0 \ldots 0 \\ \vdots\ \ddots \vdots \\ 1\ 0\ \ 0 \end{matrix}\right\} n_1 \text{ values} \\ \left.\begin{matrix} 0\ 1 \ldots 0 \\ \vdots\ \ddots \vdots \\ 0\ 1 \ldots 0 \end{matrix}\right\} n_2 \text{ values} \\ \vdots\ \vdots\ \vdots \\ \left.\begin{matrix} 0\ 0 \ldots 1 \\ \vdots\ \ddots \vdots \\ 0\ 0 \ldots 1 \end{matrix}\right\} n_p \text{ values} \end{pmatrix}$$

$$x_{ij} = \begin{cases} 1 & \text{if } \beta_i \text{ occurs in the } j^{th} \text{ observation } x_j \\ & \text{or if effect } \beta_i \text{ is present in } x_j \\ 0 & \text{if effect } \beta_i \text{ is absent in } x_j \end{cases}$$

$$n = \sum_{i=1}^{p} n_i.$$

So X is a matrix of order $n \times p$, β is fixed and

- first n_1 rows of ε are $\varepsilon_1' = (1, 0, 0, ..., 0)$,
- next n_2 rows of ε are $\varepsilon_2' = (0, 1, 0, ..., 0)$
- and similarly, the last n_p rows of ε are $\varepsilon_p' = (0, 0, ..., 0, 1)$.

Obviously, $rank(X) = p$, $E(Y) = X\beta$ and $Cov(Y) = \sigma^2 I.$

This completes the representation of a **fixed effect linear model of full rank**.

The null hypothesis of interest is $H_0 : \beta_1 = \beta_2 = ... = \beta_p = \beta$ (say)

and H_1 : At least one $\beta_i \neq \beta_j (i \neq j)$

where β and σ^2 are unknown.

We would develop here the likelihood ratio test. It may be noted that the same test can also be derived through the least-squares method. This will be demonstrated in the next module. This way the readers will understand both the methods.

We already have developed the likelihood ratio for the hypothesis $H_0 : \beta_1 = \beta_2 = ... = \beta_p$ in case 1.

The whole parametric space Ω is a $(p+1)$ dimensional space $\Omega = \left\{(\beta, \sigma^2) : -\infty < \beta < \infty, \sigma^2 > 0, i = 1, 2, ..., p\right\}$.

Note that there are $(p+1)$ parameters are $\beta_1, \beta_2, ..., \beta_p$ and σ^2 .

Under H_0, Ω reduces to two dimensional space ω as

$\omega = \left\{(\beta, \sigma^2); -\infty < \beta < \infty,\ \sigma^2 > 0\right\}$..

The likelihood function under Ω is

$$L(y\,|\,\beta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij} - \beta_i)^2\right]$$

$$L = \ln\ L(y\,|\,\beta, \sigma^2) = -\frac{n}{2}\ \ln\ (2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij} - \beta_i)^2$$

$$\frac{\partial L}{\partial \beta_i} = 0 \Rightarrow \hat{\beta}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} = \overline{y}_{io}$$

$$\frac{\partial L}{\partial \sigma^2} = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij} - \overline{y}_{io})^2.$$

The dot sign (o) in $\overline{y}_{io}$ indicates that the average has been taken over the second subscript *j*. The Hessian matrix of second-order partial derivation of $\ln L$ with respect to β_i and σ^2 is negative definite at $\beta = \overline{y}_{io}$ and $\sigma^2 = \hat{\sigma}^2$ which ensures that the likelihood function is maximized at these values.

Thus the maximum value of $L(y\,|\,\beta, \sigma^2)$ over Ω is

$$\underset{\Omega}{Max}\ L(y\,|\,\beta, \sigma^2) = \left(\frac{1}{2\pi\hat{\sigma}^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij} - \hat{\beta}_i)^2\right]$$

$$= \left[\frac{n}{2\pi\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij} - \overline{y}_{io})^2}\right]^{n/2} \exp\left(-\frac{n}{2}\right).$$

The likelihood function under ω is

$$L(y\left|\beta,\sigma^2\right.)=\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\beta)^2\right]$$

$$\ln L(y\left|\beta,\sigma^2\right.)=-\frac{n}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\beta)^2$$

The normal equations and the least-squares are obtained as follows:

$$\frac{\partial\ln L(y\left|\beta,\sigma^2\right.)}{\partial\beta}=0\Rightarrow\hat{\beta}=\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i}y_{ij}=\overline{y}_{oo}$$

$$\frac{\partial\ln L(y\left|\beta,\sigma^2\right.)}{\partial\sigma^2}=0\Rightarrow\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_{oo})^2.$$

The maximum value of the likelihood function over ω under H_0 is

$$\underset{\omega}{Max}\ L(y\left|\beta,\sigma^2\right.)=\left(\frac{1}{2\pi\hat{\sigma}^2}\right)^{\frac{n}{2}}\exp\left[-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\hat{\beta})^2\right]$$

$$=\left[\frac{n}{2\pi\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_{oo})^2}\right]^{n/2}\exp\left(-\frac{n}{2}\right).$$

The likelihood ratio test statistic is

$$\lambda=\frac{\underset{\omega}{Max}\,L(y\left|\beta,\sigma^2\right.)}{\underset{\Omega}{Max}\,L(y\left|\beta,\sigma^2\right.)}$$

$$=\left[\frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_{io})^2}{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_{oo})^2}\right]^{n/2}$$

We have that

$$\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_{oo})^2=\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left[(y_{ij}-\overline{y}_{io})+(\overline{y}_{io}-\overline{y}_{oo})\right]^2$$

$$=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_{io})^2+\sum_{i=1}^{p}n_i(\overline{y}_{io}-\overline{y}_{oo})^2$$

Thus

$$\lambda = \left[\frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 + \sum_{I=1}^{p} n_i(\bar{y}_{io} - \bar{y}_{oo})^2}{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{io})^2} \right]^{-\frac{n}{2}}$$

$$= \left[1 + \frac{q_1}{q_2} \right]^{-\frac{n}{2}}$$

where

$$q_1 = \sum_{i=1}^{p} n_i(\bar{y}_{io} - \bar{y}_{oo})^2 \text{ , and } \quad q_2 = \sum_{i=1}^{p}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{io})^2 .$$

Note that if the least-squares principal is used, then

q_1 : sum of squares due to deviations from H_0 or the between population sum of squares,

q_2 : sum of squares due to error or the within-population sum of squares,

q_1+q_2 *:* sum of squares due to H_0 or the total sum of squares.

Using theorems 6 and 7, let

$$Q_1 = \sum_{i=1}^{p} n_i(\bar{Y}_{io} - \bar{Y}_{oo})^2, \quad Q_2 = \sum_{i=1}^{p} S_i^2$$

where

$$S_i^2 = \sum_{i=1}^{n_i} (\bar{Y}_{ij} - \bar{Y}_{io})^2$$

$$\bar{Y}_{oo} = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i} Y_{ij}$$

$$\bar{Y}_{io} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$$

then under H_0

$$\frac{Q_1}{\sigma^2} \sim \chi^2(p-1)$$

$$\frac{Q_2}{\sigma^2} \sim \chi^2(n-p)$$

and $\frac{Q_1}{\sigma^2}$ and $\frac{Q_2}{\sigma^2}$ are independently distributed.

Thus under H_0

$$\frac{\left(\dfrac{\frac{Q_1}{\sigma^2}}{p-1}\right)}{\left(\dfrac{\frac{Q_2}{\sigma^2}}{n-p}\right)} \sim F(p-1, n-p).$$

The likelihood ratio test reject H_0 whenever

$$\frac{q_1}{q_2} > C$$

where the constant $C = F_{1-\alpha}(p-1, n-p).$

The analysis of variance table for the one-way classification in fixed effect model is

Source of Variation	Degrees of freedom	Sum of squares	Mean sum of squares	F
Between Population	$p-1$	q_1	$\frac{q_1}{p-1}$	$\left(\frac{n-p}{p-1}\right).\frac{q_1}{q_2}$
Within Population	$n-p$	q_2	$\frac{q_2}{n-p}$	
Total	$n-1$	q_1+q_2		

Note that

$$E\left[\frac{Q_2}{n-p}\right] = \sigma^2$$

$$E\left[\frac{Q_1}{p-1}\right] = \sigma^2 + \frac{\sum_{i=1}^{p}(\beta_i - \overline{\beta})^2}{p-1}; \overline{\beta} = \frac{1}{p}\sum_{i=1}^{p}\beta_i$$

Case of rejection of H_0

If $F > F_{1-\alpha}(p-1, n-p)$, then $H_0: \beta_1 = \beta_2 = ... = \beta_p$ is rejected. This means that at least one β_i is different from others which is responsible for the rejection. So the objective is to investigate and find out such β_i and divide the population into groups such that the means of populations within the groups are same. This can be done by pairwise testing of β's.

Test $H_0 : \beta_i = \beta_k \; (i \neq k)$ against $H_1 : \beta_i \neq \beta_k$.

This can be tested using following *t*-statistic

$$t = \frac{\bar{Y}_{io} - \bar{Y}_{ko}}{\sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_k} \right)}}$$

which follows the *t* distribution with $(n - p)$ degrees of freedom under H_0 and $s^2 = \frac{q_2}{n-p}$. Thus

the decision rule is to reject H_0 at the level α if the observed difference

$$|(\bar{y}_{io} - \bar{y}_{ko})| > t_{1-\frac{\alpha}{2},\, n-p} \sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_k} \right)}$$

The quantity $t_{1-\frac{\alpha}{2},\, n-p} \sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_k} \right)}$ is called the **critical difference**.

Thus following steps are followed :

1. Compute all possible critical differences arising out of all possible pair $(\beta_i, \beta_k),\; i \neq k = 1, 2, ..., p$.
2. Compare them with their observed differences
3. Divide the *p* populations into different groups such that the populations in the same group have the same means.

The computation are simplified if $n_i = n$ for all *i*. In such a case, the **common critical difference (CCD)** is

$$CCD = t_{1-\frac{\alpha}{2},\, n-p} \sqrt{\frac{2s^2}{n}}$$

and the observed difference $(\bar{y}_{io} - \bar{y}_{ko}), i \neq k$ are compared with CCD.

If $|\bar{y}_{io} - \bar{y}_{ko}| > CCD$

then the corresponding effects/means $\bar{y}_{io}$ and $\bar{y}_{ko}$ are coming from populations with different means.

Note: In general we say that if there are three effects $\beta_1, \beta_2, \beta_3$ then

if $H_{01} : \beta_1 = \beta_2 (\equiv$ denote as event $A)$ is accepted

and if $H_{02} : \beta_2 = \beta_3 (\equiv$ denote as event $B)$ is accepted

then $H_{03} : \beta_1 = \beta_2 (\equiv$ denote as event $C)$ will be accepted.

The question arises here that in what sense do we conclude such a statement about the acceptance of H_{03}. The reason is as follows:

Since event $A \bigcap B \subset C$, so

$$P(A \bigcap B) \leq P(C)$$

In this sense if the probability of an event is higher than the intersection of the events, i.e., the probability that H_{03} is accepted is higher than the probability of acceptance of H_{01} and H_{02} both, so we conclude, in general, that the acceptance of H_{01} and H_{02} implies the acceptance of H_{03}.

Multiple comparison test:

One interest in the analysis of variance is to decide whether population means are equal or not. If the hypothesis of equal means is rejected then one would like to divide the populations into subgroups such that all populations with the same means come to the same subgroup. This can be achieved by the multiple comparison tests.

A multiple comparison test procedure conducts the test of hypothesis for all the pairs of effects and compares them at a significance level α i.e., it works on per comparison basis.

This is based mainly on the *t*-statistic. If we want to ensure that the significance level α simultaneously for all group comparison of interest, the approximate multiple test procedure is one that controls the error rate per experiment basis.

There are various available multiple comparison tests. We will discuss some of them in the context of one-way classification. In two-way or higher classification, they can be used on similar lines.

1. Studentized range test:

It is assumed in the Studentized range test that the *p* samples, each of size *n*, have been drawn from *p* normal populations. Let their sample means be $\bar{y}_{1o}, \bar{y}_{2o}, ..., \bar{y}_{po}$ These means are ranked and arranged in ascending order as $\bar{y}_1^*, \bar{y}_2^*, ..., \bar{y}_p^*$ where $\bar{y}_1^* = \underset{i}{Min}\, \bar{y}_{io}$ and $\bar{y}_p^* = \underset{i}{Max}\, \bar{y}_{io}$, $i = 1, 2, ..., p$.

Find the range as $R = \bar{y}_p^* - \bar{y}_1^*$.

The Studentized range is defined as

$$q_{p,\, n-p} = \frac{R\sqrt{n}}{s}$$

where $q_{\alpha,p,\gamma}$ is the upper $100\alpha\%$ point of Studentized range when $\gamma = n - p$. The tables for $q_{\alpha,p,\gamma}$ are available.

The testing procedure involves the comparison of $q_{p,\gamma}$ with $q_{\alpha,p,\gamma}$ in the usual way as-

- if $q_{p,n-p} < q_{\alpha,p,n-p}$ then conclude that $\beta_1 = \beta_2 = ... = \beta_p$.
- if $q_{p,n-p} > q_{\alpha,p,n-p}$ then all $\beta's$ in the group are not the same.

2. Student-Newman-Keuls test:

The Student-Newman-Keuls test is similar to Studentized range test in the sense that the range is compared with $100\alpha\%$ points on critical Studentized range W_P given by

$$W_p = q_{\alpha,p,\gamma}\sqrt{\frac{s^2}{n}}.$$

The observed range $R = \bar{y}_p^* - \bar{y}_1^*$ is now compared with W_p. Let the effects $\beta_1, \beta_2, ...\beta_p$ are denoted as $\beta_1^*, \beta_2^*, ..., \beta_p^*$ corresponding to $\bar{y}_1^*, \bar{y}_2^*, ..., \bar{y}_p^*$ respectively in the context of Student-Newman-Keuls test. For example, the largest mean $\bar{y}_p^*$ maybe $\bar{y}_3$ and so $\beta_p^* = \beta_3$.

- If $R < W_p$ then stop the process of comparison and conclude that $\beta_1^* = \beta_2^* = ... = \beta_p^*$.
- If $R > W_p$ then

(i) divide the ranked means $\bar{y}_1^*, \bar{y}_2^*, ..., \bar{y}_p^*$ into two subgroups containing - $(\bar{y}_p^*, \bar{y}_{p-1}^*, ..., \bar{y}_2^*)$ and $(\bar{y}_{p-1}^*, \bar{y}_{p-2}^*, ..., \bar{y}_1^*)$

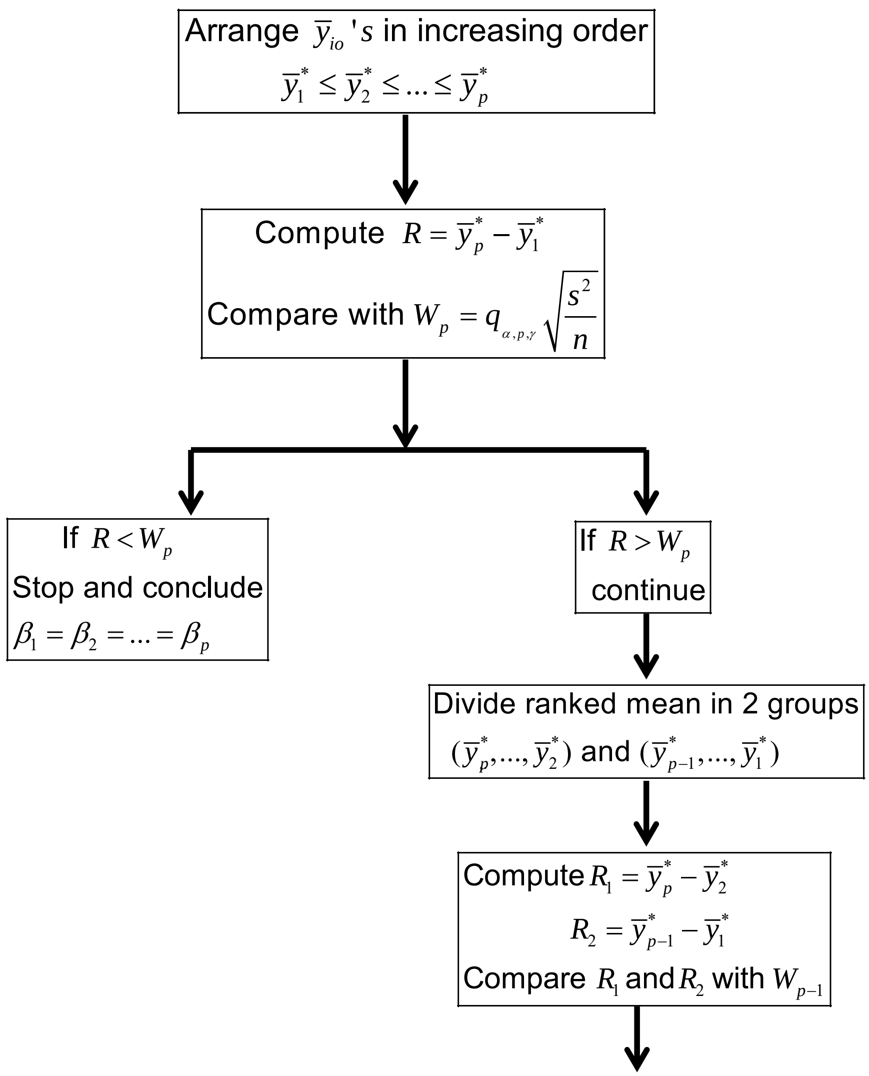
(ii) Compute the ranges $R_1 = \bar{y}_p^* - \bar{y}_2^*$ and $R_2 = \bar{y}_{p-1}^* - \bar{y}_1^*$. Then compare the ranges R_1 and R_2 with W_{p-1}.

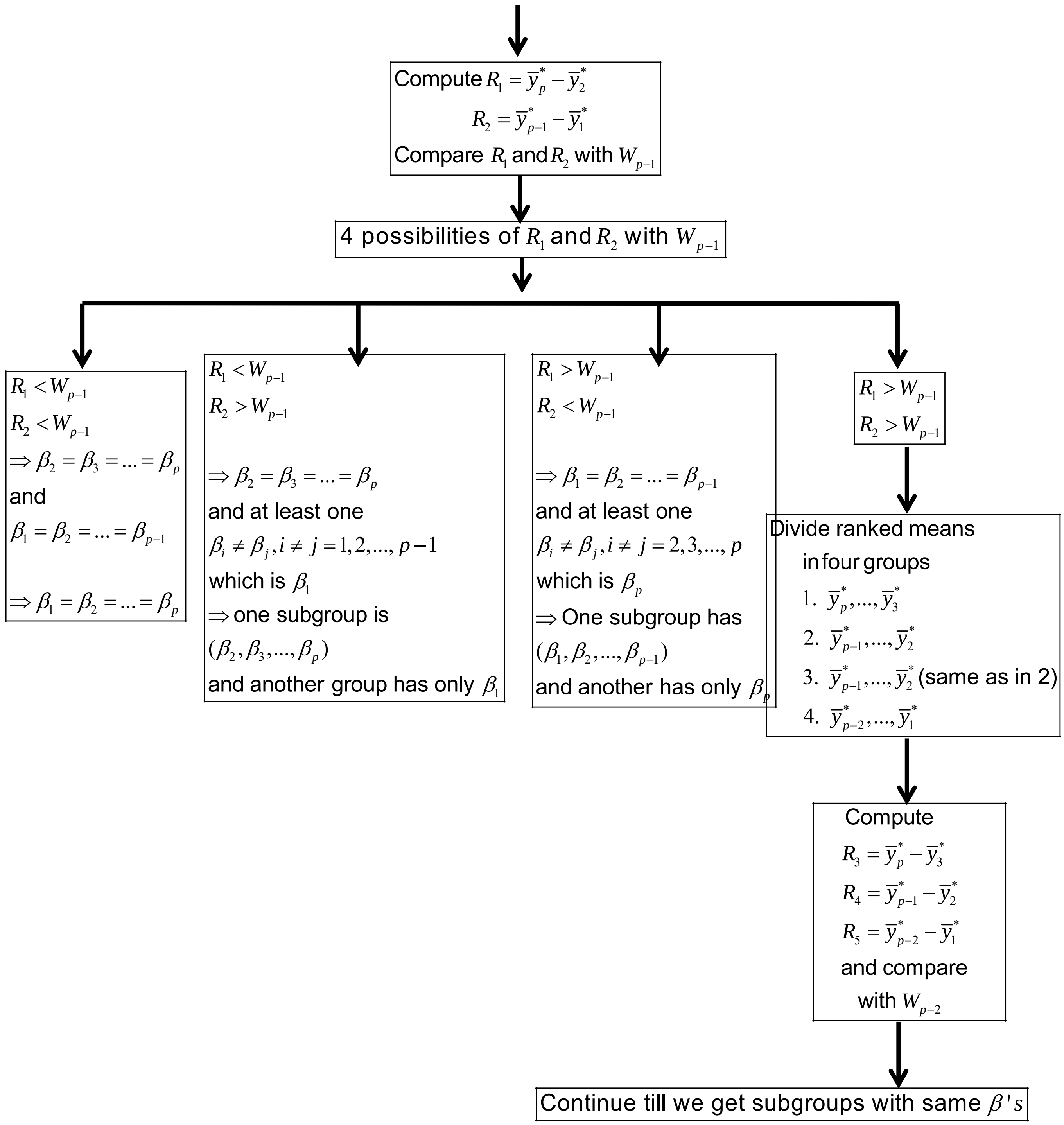
- If either range (R_1 or R_2) is smaller than W_{p-1}, then the means (or β_i's) in each of the groups are equal.
- If R_1 and/or R_2 are greater then W_{p-1}, then the $(p-1)$ means (or β_i's) in the group concerned are divided into two groups of $(p-2)$ means (or β_i's) each and compare the range of the two groups with W_{p-2}.

Continue with this procedure until a group of *i* means (or β_i's) is found whose range does not exceed W_i.

By this method, the difference between any two means under test is significant when the range of the observed means of each and every subgroup containing the two means under test is significant according to the studentized critical range.

This procedure can be easily understood by the following flow chart.




Compute $R_1 = \bar{y}_p^* - \bar{y}_2^*$
$R_2 = \bar{y}_{p-1}^* - \bar{y}_1^*$
Compare R_1 and R_2 with W_{p-1}
4 possibilities of R_1 and R_2 with W_{p-1}
$R_1 < W_{p-1}$
$R_2 < W_{p-1}$
$\Rightarrow \beta_2 = \beta_3 = ... = \beta_p$
and
$\beta_1 = \beta_2 = ... = \beta_{p-1}$
$\Rightarrow \beta_1 = \beta_2 = ... = \beta_p$
$R_1 < W_{p-1}$
$R_2 > W_{p-1}$
$\Rightarrow \beta_2 = \beta_3 = ... = \beta_p$
and at least one
$\beta_i \neq \beta_j, i \neq j = 1, 2, ..., p-1$
which is β_1
$\Rightarrow$ one subgroup is
$(\beta_2, \beta_3, ..., \beta_p)$
and another group has only β_1
$R_1 > W_{p-1}$
$R_2 < W_{p-1}$
$\Rightarrow \beta_1 = \beta_2 = ... = \beta_{p-1}$
and at least one
$\beta_i \neq \beta_j, i \neq j = 2, 3, ..., p$
which is β_p
$\Rightarrow$ One subgroup has
$(\beta_1, \beta_2, ..., \beta_{p-1})$
and another has only β_p
$R_1 > W_{p-1}$
$R_2 > W_{p-1}$
Divide ranked means
in four groups
1. $\bar{y}_p^*, ..., \bar{y}_3^*$
2. $\bar{y}_{p-1}^*, ..., \bar{y}_2^*$
3. $\bar{y}_{p-1}^*, ..., \bar{y}_2^*$ (same as in 2)
4. $\bar{y}_{p-2}^*, ..., \bar{y}_1^*$
Compute
$R_3 = \bar{y}_p^* - \bar{y}_3^*$
$R_4 = \bar{y}_{p-1}^* - \bar{y}_2^*$
$R_5 = \bar{y}_{p-2}^* - \bar{y}_1^*$
and compare
with W_{p-2}
Continue till we get subgroups with same β's

3. Duncan's multiple comparison test:

The test procedure in Duncan's multiple comparison test is the same as in the Student-Newman-Keuls test except the observed ranges are compared with Duncan's 100α% critical range

$$D_p = q^*_{\alpha_p, p, \gamma}\sqrt{\frac{s^2}{n}}$$

where $\alpha_p = 1-(1-\alpha)^{p-1}$, $q^*_{\alpha_p, p,\gamma}$ denotes the upper $100\alpha\%$ points of the Studentized range based on Duncan's range.

Tables for Duncan's range are available.

Duncan felt that this test is better than the Student-Newman-Keuls test for comparing the differences between any two ranked means. Duncan regarded that the Student-Newman-Keuls method is too stringent in the sense that the true differences between the means will tend to be missed too often. Duncan notes that in testing the equality of a subset $k, (2 \le k \le p)$ means through null hypothesis, we are in fact testing whether $(p-1)$ orthogonal contrasts between the $\beta's$ differ from zero or not. If these contrasts were tested in separate independent experiments, each at level α, the probability of incorrectly rejecting the null hypothesis would be $\left[1-(1-\alpha)^{p-1}\right]$. So Duncan proposed to use $\left[1-(1-\alpha)^{p-1}\right]$ in place of α in the Student-Newman-Keuls test.

[Reference: Contributions to order statistics, Wiley 1962, Chapter 9 (Multiple decision and multiple comparisons, H.A. David, pages 147-148)].

Case of unequal sample sizes:

When sample means are not based on the same number of observations, the procedures based on Studentized range, Student-Newman-Keuls test and Duncan's test are not applicable. Kramer proposed that in Duncan's method, if a set of p means is to be tested for equality, then replace

$$q^*_{\alpha^*_p, p, \gamma}\frac{s}{\sqrt{n}} \text{ by } q^*_{\alpha^*_p, p, \gamma}\, s\sqrt{\frac{1}{2}\left(\frac{1}{n_U}+\frac{1}{n_L}\right)}$$

where n_U and n_L are the number of observations corresponding to the largest and smallest means in the data. This procedure is only an approximate procedure but will tend to be conservative since means based on a small number of observations will tend to be overrepresented in the extreme groups of means.

Another option is to replace n by the harmonic mean of $n_1, n_2, ..., n_p$, i.e., $\dfrac{p}{\sum_{i=1}^{p}\left(\dfrac{1}{n_i}\right)}$.

4. The "Least Significant Difference" (LSD):

In the usual testing of $H_0 : \beta_i = \beta_k$ against $H_1 : \beta_i \neq \beta_k$, the *t*-statistic

$$t = \frac{\bar{y}_{io} - \bar{y}_{ko}}{\sqrt{\widehat{Var}(\bar{y}_{io} - \bar{y}_{ko})}}$$

is used which follows a *t*-distribution , say with degrees of freedom $'df'$. Thus H_0 is rejected whenever

$$|t| > t_{df,\, 1-\frac{\alpha}{2}}$$

and it is concluded that β_1 and β_2 are significantly different. The inequality $|t| > t_{df,\, 1-\frac{\alpha}{2}}$ can be equivalently written as

$$|\bar{y}_{io} - \bar{y}_{ko}| > t_{df,\, 1-\frac{\alpha}{2}} \sqrt{\widehat{Var}(\bar{y}_{io} - \bar{y}_{ko})}\,.$$

If every pair of sample for which

$$|\bar{y}_{io} - \bar{y}_{ko}| \text{ exceeds } t_{df,\, 1-\frac{\alpha}{2}} \sqrt{\widehat{Var}(\bar{y}_{io} - \bar{y}_{ko})}$$

then this will indicate that the difference between β_i and β_k is significantly different. So according to this, the quantity $t_{df,\, 1-\frac{\alpha}{2}} \sqrt{\widehat{Var}(\bar{y}_{io} - \bar{y}_{ko})}$ would be the least difference of $\bar{y}_{io}$ and $\bar{y}_{ko}$ for which it will be declared that the difference between β_i and β_k is significant. Based on this idea, we use the pooled variance of the two samples $Var\left(\bar{y}_{io} - \bar{y}_{ko}\right)$ as s^2 and the Least Significant Difference (LSD) is defined as

$$LSD = t_{df,\, 1-\frac{\alpha}{2}} \sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_k} \right)}\,.$$

If $n_1 = n_2 = n,$ then

$$LSD = t_{df,\, 1-\frac{\alpha}{2}} \sqrt{\frac{2s^2}{n}}\,.$$

Now all $\frac{p(p-1)}{2}$ pairs of $\bar{y}_{io}$ and $\bar{y}_{ko}$, $(i \neq k = 1, 2, ..., p)$ are compared with *LSD*. Use of *LSD* criterion may not lead to good results if it is used for comparisons suggested by the data (largest/smallest sample mean) or if all pairwise comparisons are done without correction of the test level. If *LSD* is used for all the pairwise comparisons then these tests are not independent. Such correction for test levels was incorporated in Duncan's test.

5. Tukey's "Honestly significant Difference" (*HSD*)

In this procedure, the Studentized rank values $q_{\alpha,n,\gamma}$ are used in place of *t*-quantiles and the standard error of the difference of pooled mean is used in place of standard error of mean in common critical difference for testing $H_0: \beta_i = \beta_k$ against $H_0: \beta_i \neq \beta_k$ and Tukey's Honestly Significant Difference is computed as

$$HSD = q_{1-\frac{\alpha}{2},p,\gamma}\sqrt{\frac{MS_{error}}{n}}$$

assuming all samples are of the same size *n*. All $\frac{p(p-1)}{2}$ pairs $|\bar{y}_{io} - \bar{y}_{ko}|$ are compared with *HSD*.

If $|\bar{y}_{io} - \bar{y}_{ko}| > HSD$ then β_i and β_k are significantly different.

We notice that all the multiple comparison test procedure discussed up to now are based on the testing of hypothesis. There is one-to-one relationship between the testing of hypothesis and the confidence interval estimation. So the confidence interval can also be used for such comparisons. Since $H_0: \beta_i = \beta_k$ is same as $H_0: \beta_i - \beta_k = 0$ so first we establish the relationship and then describe the Tukey's and Scheffe's procedures for multiple comparison test which are based on the confidence interval. We need the following concepts.

Contrast:

A linear parametric function $L = l'\beta = \sum_{i=1}^{p} \ell_i \beta_i$ where $\beta = (\beta_1, \beta_2, ..., \beta_P)$ and $\ell = (\ell_1, \ell_2, ..., \ell_p)$ are the $p \times 1$ vectors of parameters and constants respectively is said to be a contrast when $\sum_{i=1}^{p} \ell_i = 0.$

For example, $\beta_1 - \beta_2 = 0,\ \beta_1 + \beta_2 - \beta_3 - \beta_1 = 0,\ \beta_1 + 2\beta_2 - 3\beta_3 = 0$ etc. are contrast whereas $\beta_1 + \beta_2 = 0, \beta_1 + \beta_2 + \beta_3 + \beta_4 = 0,\ \beta_1 - 2\beta_2 - 3\beta_3 = 0$ etc. are not contrasts.

Orthogonal contrast:

If $L_1 = \ell'\beta = \sum_{i=1}^{p} \ell_i \beta_i$ and $L_2 = m'\beta = \sum_{i=1}^{p} m_i \beta_i$ are contrasts such that $\ell' m = 0$ or $\sum_{i=1}^{p} \ell_i m_i = 0$ then L_1 and L_2 are called orthogonal contrasts.

For example, $L_1 = \beta_1 + \beta_2 - \beta_3 - \beta_4$ and $L_2 = \beta_1 - \beta_2 + \beta_3 - \beta_4$ are contrasts. They are also the orthogonal contrasts.

The condition $\sum_{i=1}^{p} \ell_i m_i = 0$ ensures that L_1 and L_2 are independent in the sense that

$$Cov(L_1, L_2) = \sigma^2 \sum_{i=1}^{p} \ell_i m_i = 0.$$

Mutually orthogonal contrasts:

If there are more than two contrasts then they are said to be mutually orthogonal, if they are pair-wise orthogonal.

It may be noted that the number of mutually orthogonal contrasts is the number of degrees of freedom.

Coming back to the multiple comparison test, if the null hypothesis of equality of all effects is rejected then it is reasonable to look for the contrasts which are responsible for the rejection. In terms of contrasts, it is desirable to have a procedure

(i) that permits the selection of the contrasts after the data is available.

(ii) with which a known level of significance is associated.

Such procedures are Tukey's and Scheffe's procedures. Before discussing these procedures, let us consider the following example which illustrates the relationship between the testing of hypothesis and confidence intervals.

Example: Consider the test of hypothesis for

$H_0 : \beta_i = \beta_j \ (i \neq j = 1, 2, ..., p)$
or $H_0 : \beta_i - \beta_j = 0$
or $H_0 :$ contrast $= 0$
or $H_0 : L = 0.$

The test statistic for $H_0 : \beta_i = \beta_j$ is

$$t = \frac{(\hat{\beta}_i - \hat{\beta}_j) - (\beta_i - \beta_j)}{\sqrt{\widehat{Var}(\bar{y}_{io} - \bar{y}_{ko})}} = \frac{\hat{L} - L}{\sqrt{\widehat{Var}(\hat{L})}}$$

where $\hat{\beta}$ denotes the maximum likelihood (or least-squares) estimator of β and t follows a t-distribution with df degrees of freedom. This statistic, in fact, can be extended to any linear contrast, say e.g., $L = \beta_1 + \beta_2 - \beta_3 - \beta_4$, $\hat{L} = \hat{\beta}_1 + \hat{\beta}_2 - \hat{\beta}_3 - \hat{\beta}_4$.

The decision rule is

reject $H_0 : L = 0$ against $H_1 : L \neq 0$.

If $\left|\hat{L}\right| > t_{df}\sqrt{\widehat{Var}(\hat{L})}$.

The $100(1-\alpha)\%$ confidence interval of L is obtained as

$$P\left[-t_{df} \leq \frac{\hat{L}-L}{\sqrt{\widehat{Var}(\hat{L})}} \leq t_{df}\right] = 1-\alpha$$

or $P\left[\hat{L} - t_{df}\sqrt{\widehat{Var}(\hat{L})} \leq L \leq \hat{L} + t_{df}\sqrt{\widehat{Var}(\hat{L})}\right] = 1-\alpha$

so that the $100(1-\alpha)\%$ confidence interval of L is

$$\left[\hat{L} - t_{df}\sqrt{\widehat{Var}(\hat{L})},\ \hat{L} + t_{df}\sqrt{\widehat{Var}(\hat{L})}\right]$$

and

$$\hat{L} - t_{df}\sqrt{\widehat{Var}(\hat{L})} \leq L \leq \hat{L} + t_{df}\sqrt{\widehat{Var}(\hat{L})}$$

If this interval includes $L = 0$ between lower and upper confidence limits, then $H_0 : L = 0$ is accepted. Our objective is to know if the confidence interval contains zero or not.

Suppose for some given data the confidence intervals for $\beta_1 - \beta_2$ and $\beta_1 - \beta_3$ are obtained as

$-3 \leq \beta_1 - \beta_2 \leq 2$ and $2 \leq \beta_1 - \beta_3 \leq 4.$

Thus we find that the interval of $\beta_1 - \beta_2$ includes zero which implies that $H_0 : \beta_1 - \beta_2 = 0$ is accepted. Thus $\beta_1 = \beta_2$. On the other hand interval of $\beta_1 - \beta_3$ does not include zero and so $H_0 : \beta_1 - \beta_3 = 0$ is not accepted. Thus $\beta_1 \neq \beta_3$.

If the interval of $\beta_1 - \beta_3$ is $-1 \leq \beta_1 - \beta_3 \leq 1$ then $H_0 : \beta_1 = \beta_3$ is accepted. If both $H_0 : \beta_1 = \beta_2$ and $H_0 : \beta_1 = \beta_3$ are accepted then we can conclude that $\beta_1 = \beta_2 = \beta_3$.

Tukey's procedure for multiple comparisons (*T*-method)

The *T*-method uses the distribution of the studentized range statistic . (The *S*-method (discussed next) utilizes the *F*-distribution). The *T*-method can be used to make the simultaneous confidence statements about contrasts $(\beta_i - \beta_j)$ among a set of parameters $\{\beta_1, \beta_2, ..., \beta_p\}$ and an estimate s^2 of error variance if certain restrictions are satisfied.

These restrictions have to be viewed according to the given conditions. For example, one of the restrictions is that all $\hat{\beta}_i 's$ have equal variances. In the setup of one-way classification, $\hat{\beta}_i$ has it's mean $\bar{Y}_i$ and its variance is $\dfrac{\sigma^2}{n_i}$. This reduces to a simple condition that all $n_i 's$ are same, i.e., $n_i = n$ for all *i*. so that all the variances are the same.

Another assumption is to assume that $\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_p$ are statistically independent and the only contrasts considered are the $\dfrac{p(p-1)}{2}$ differences $\{\beta_i - \beta_j, i \neq j = 1, 2, ..., p\}$.

We make the following assumptions:

(i) The $\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_p$ are statistically independent

(ii) $\hat{\beta}_i \sim N(\beta_i, a^2\sigma^2), i = 1, 2, ..., p, \ a > 0$ is a known constant.

(iii) s^2 is an independent estimate of σ^2 with γ degrees of freedom (here $\gamma = n - p$), i.e.,

$$\frac{\gamma s^2}{\sigma^2} \sim \chi^2(\gamma)$$

and

(iv) s^2 is statistically independent of $\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_p$.

The statement of *T*-method is as follows:

Under the assumptions (i)-(iv), the probability is $(1-\alpha)$ that the values of contrasts $L = \sum_{i=1}^{p} C_i \beta_i \quad (\sum_{i=1}^{p} C_i = 0)$ simultaneously satisfy

$$\hat{L} - Ts\left(\frac{1}{2}\sum_{i=1}^{p}|C_i|\right) \leq L \leq \hat{L} + Ts\left(\frac{1}{2}\sum_{i=1}^{p}|C_i|\right)$$

where $\hat{L} = \sum_{i=0}^{p} C_i \hat{\beta}_i$, $\hat{\beta}_i$ is the maximum likelihood (or least squares) estimate of β_i, $T = aq_{\alpha,p,\gamma}$, with $q_{\alpha,p,\gamma}$ being the upper α point of the distribution of studentized range.

Note that if L is a contrast like $\beta_i - \beta_j (i \neq j)$ then $\frac{1}{2}\sum_{i=1}^{p}|C_i| = 1$ and the variance is σ^2 so that $a = 1$ and the interval simplifies to

$$(\hat{\beta}_i - \hat{\beta}_j) - Ts \leq \beta_i - \beta_j \leq (\hat{\beta}_i - \hat{\beta}_j) + Ts$$

where $T = q_{\alpha,p,\gamma}$. Thus the maximum likelihood (or least squares) estimate $\hat{L} = \hat{\beta}_i - \hat{\beta}_j$ of $L = \beta_i - \beta_j$ is said to be significantly different from zero according to T-criterion if the interval $(\hat{\beta}_i - \hat{\beta}_j - Ts,\ \hat{\beta}_i - \hat{\beta}_j + Ts)$ does not cover $\beta_i - \beta_j = 0$, i.e.,

if $\left|\hat{\beta}_i - \hat{\beta}_j\right| > Ts$

or more general if $\left|\hat{L}\right| > Ts\left(\frac{1}{2}\sum_{i=1}^{p}|C_i|\right)$.

The steps involved in the testing now involve the following steps:

- Compute $\hat{L}$ or $\left(\hat{\beta}_i - \hat{\beta}_j\right)$.
- Compute all possible pairwise differences.
- Compare all the differences with

$$q_{\alpha,p,\gamma}.\frac{s}{\sqrt{n}}\left(\frac{1}{2}\sum_{i=1}^{p}|C_i|\right)$$

- If $\left|\hat{L}\right|$ or $(\hat{\beta}_i - \hat{\beta}_j) > Ts\left(\frac{1}{2}\sum_{i=1}^{p}|C_i|\right)$

then $\hat{\beta}_i$ and $\hat{\beta}_j$ are significantly different where $T = \frac{q_{\alpha,p,\gamma}}{\sqrt{n}}$.

Tables for T are available.

When sample sizes are not equal, then **Tukey-Kramer Procedure** suggests to compare $\left|\hat{L}\right|$ with

$$q_{\alpha,p,\gamma}\ s\sqrt{\frac{1}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}\ \left(\frac{1}{2}\sum_{i=1}^{p}|C_i|\right)$$

or

$$T\sqrt{\frac{1}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}\ \left(\frac{1}{2}\sum_{i=1}^{p}|C_i|\right).$$

The Scheffe's method (*S*-method) of multiple comparisons

S-method generally gives shorter confidence intervals then *T*-method. It can be used in a number of situations where *T*-method is not applicable, e.g., when the sample sizes are not equal.

A set L of estimable functions $\{\psi\}$ is called a p-dimensional space of estimable functions if there exists p linearly independent estimable functions $(\psi_1, \psi_2, ..., \psi_p)$ such that every ψ in L is of the form $\psi = \sum_{i=1}^{p} C_i y_i$ where $C_1, C_2, ..., C_p$ are known constants. In other words, L is the set of all linear combinations of $\psi_1, \psi_2, ..., \psi_p$.

Under the assumption that the parametric space Ω is $Y \sim N(X\beta, \sigma^2 I)$ with $rank(X) = p$, $\beta = (\beta_1, ..., \beta_p)$, X is $n \times p$ matrix, consider a p-dimensional space L of estimable functions generated by a set of p linearly independent estimable functions $\{\psi_1, \psi_2, ..., \psi_p\}$.

For any $\psi \in L$, Let $\hat{\psi} = \sum_{i=1}^{n} C_i y_i$ be its least squares (or maximum likelihood) estimator,

$$\begin{aligned} Var(\hat{\psi}) &= \sigma^2 \sum_{i=1}^{n} C_i^2 \\ &= \sigma_{\psi}^2 \ (say) \end{aligned}$$

and $\hat{\sigma}_{\psi}^2 = s^2 \sum_{i=1}^{n} C_i^2$

where s^2 is the mean square due to error with $(n-p)$ degrees of freedom.

The statement of *S*-method is as follows:

Under the parametric space Ω, the probability is $(1-\alpha)$ that simultaneously for all $\psi \in L$, $\hat{\psi} - S\hat{\sigma}_{\hat{\psi}} \le \psi \le \hat{\psi} + S\hat{\sigma}_{\hat{\psi}}$ where the constant $S = \sqrt{pF_{1-\alpha}(p, n-p)}$.

Method: For a given space L of estimable functions and confidence coefficient $(1-\alpha)$, the least square (or maximum likelihood) estimate $\hat{\psi}$ of $\psi \in L$ will be said to be significantly different from zero according to *S*-criterion if the confidence interval

$$(\hat{\psi} - S\hat{\sigma}_{\hat{\psi}} \le \psi \le \hat{\psi} + S\hat{\sigma}_{\hat{\psi}})$$

does not cover $\psi = 0$, i.e., if $|\hat{\psi}| > S\hat{\sigma}_{\hat{\psi}}$.

The *S*-method is less sensitive to the violation of assumptions of normality and homogeneity of variances.

Comparison of Tukey's and Scheffe's methods:

1. Tukey's method can be used only with equal sample size for all factor level but S-method is applicable whether the sample sizes are equal or not.
2. Although, Tukey's method is applicable for any general contrast, the procedure is more powerful when comparing simple pairwise differences and not making more complex comparisons.
3. It only pairwise comparisons are of interest, and all factor levels have equal sample sizes, Tukey's method gives shorter confidence interval and thus is more powerful.
4. In the case of comparisons involving general contrasts, Scheffe's method tends to give narrower confidence interval and provides a more powerful test.
5. Scheffe's method is less sensitive to the violations of assumptions of normal distribution and homogeneity of variances.