Analysis of Variance and Design of Experiments

General Linear Hypothesis and Analysis of Variance

:::

Lecture 9

CCD and Multiple Comparison Tests

Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Slides can be downloaded from http://home.iitk.ac.in/~shalab/sp

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

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

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

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.

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

Thus under H_0

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

where

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

$$Q_2 = \sum_{i=1}^{p} S_i^2$$

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

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

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

The likelihood ratio test rejects H_0

whenever $\dfrac{q_1}{q_2} > F_{1-\alpha}(p-1, n-p).$

If $F > F_{1-\alpha}(p-1, n-p)$, **then** $H_0 : \beta_1 = \beta_2 = ... = \beta_p$ **is rejected.**

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 the same.**

This can be done by pairwise testing of $\beta' s.$

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

Case of rejection of H_0

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**

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

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 <u>critical difference.</u>**

Case of rejection of H_0

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*.

Case of rejection of H_0

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

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

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.**

Multiple comparison test:

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.

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 $\bar{y}_{1o}, \bar{y}_{2o}, \ldots, \bar{y}_{po}$.

These means are ranked and arranged in an ascending order as $\bar{y}_1^*, \bar{y}_2^*, \ldots, \bar{y}_p^*$ where $\bar{y}_1^* = \underset{i}{Min}\, \bar{y}_{io}$ and $\bar{y}_p^* = \underset{i}{Max}\, \bar{y}_{io}, \; i = 1, 2, \ldots, p.$

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

Studentized range test:

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.**

Studentized range test:

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

- **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.**

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 $\beta_1, \beta_2, \ldots \beta_p$ are denoted as $\beta_1^*, \beta_2^*, \ldots, \beta_p^*$ corresponding to $\bar{y}_1^*, \bar{y}_2^*, \ldots, \bar{y}_p^*$ respectively in the context of Student-Newman-Keuls test.

For example, the largest mean $\bar{y}_p^*$ may be $\bar{y}_3$ and so $\beta_p^* = \beta_3$.

Student-Newman-Keuls test:

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

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

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}.

Student-Newman-Keuls test:

- **If either range R_1 or R_2 is smaller than W_{p-1}, then 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}.**

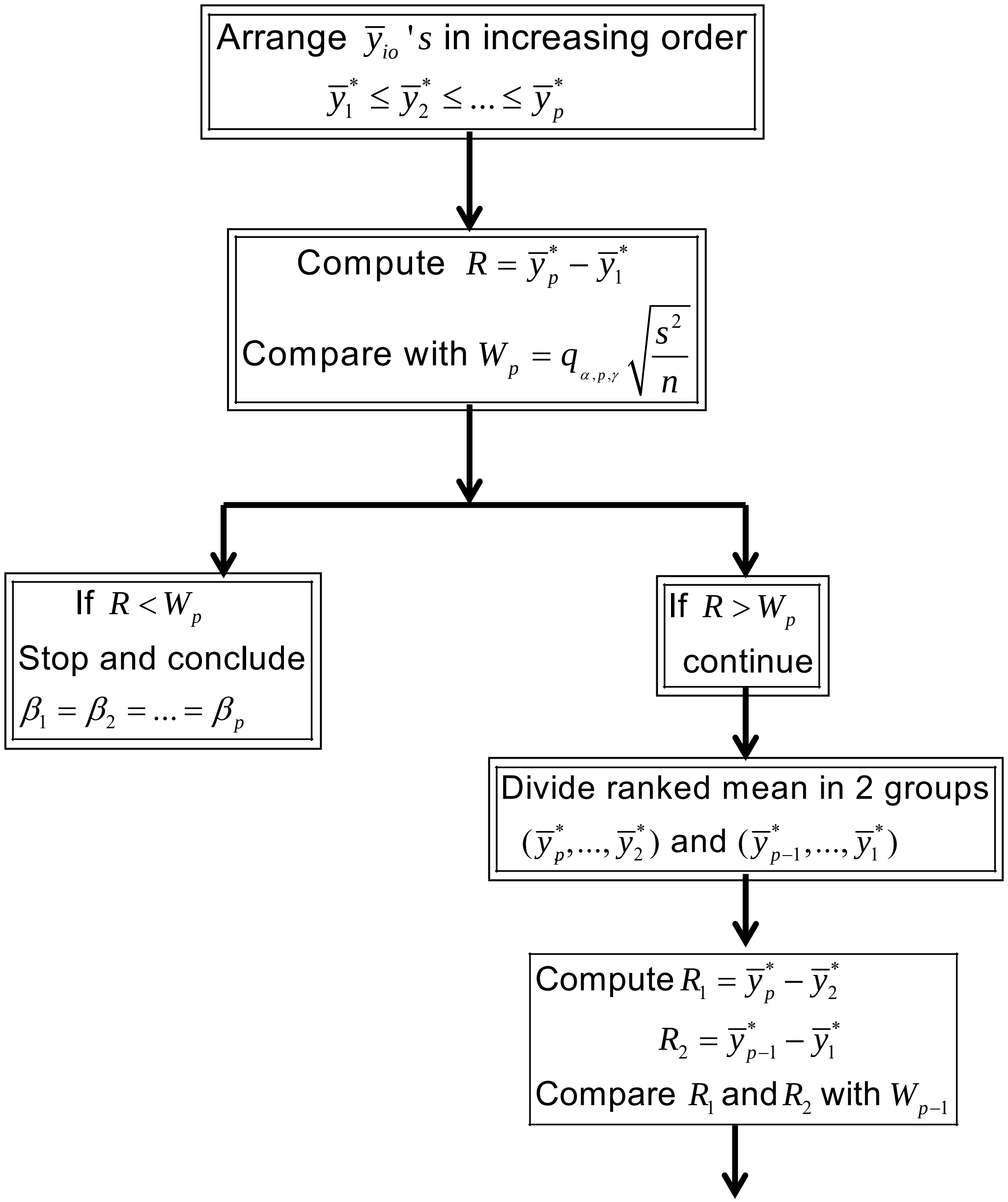
Continue with this procedure until a group of *i* means (or β_i 's) is found whose range does not exceed W_i.

Student-Newman-Keuls test:

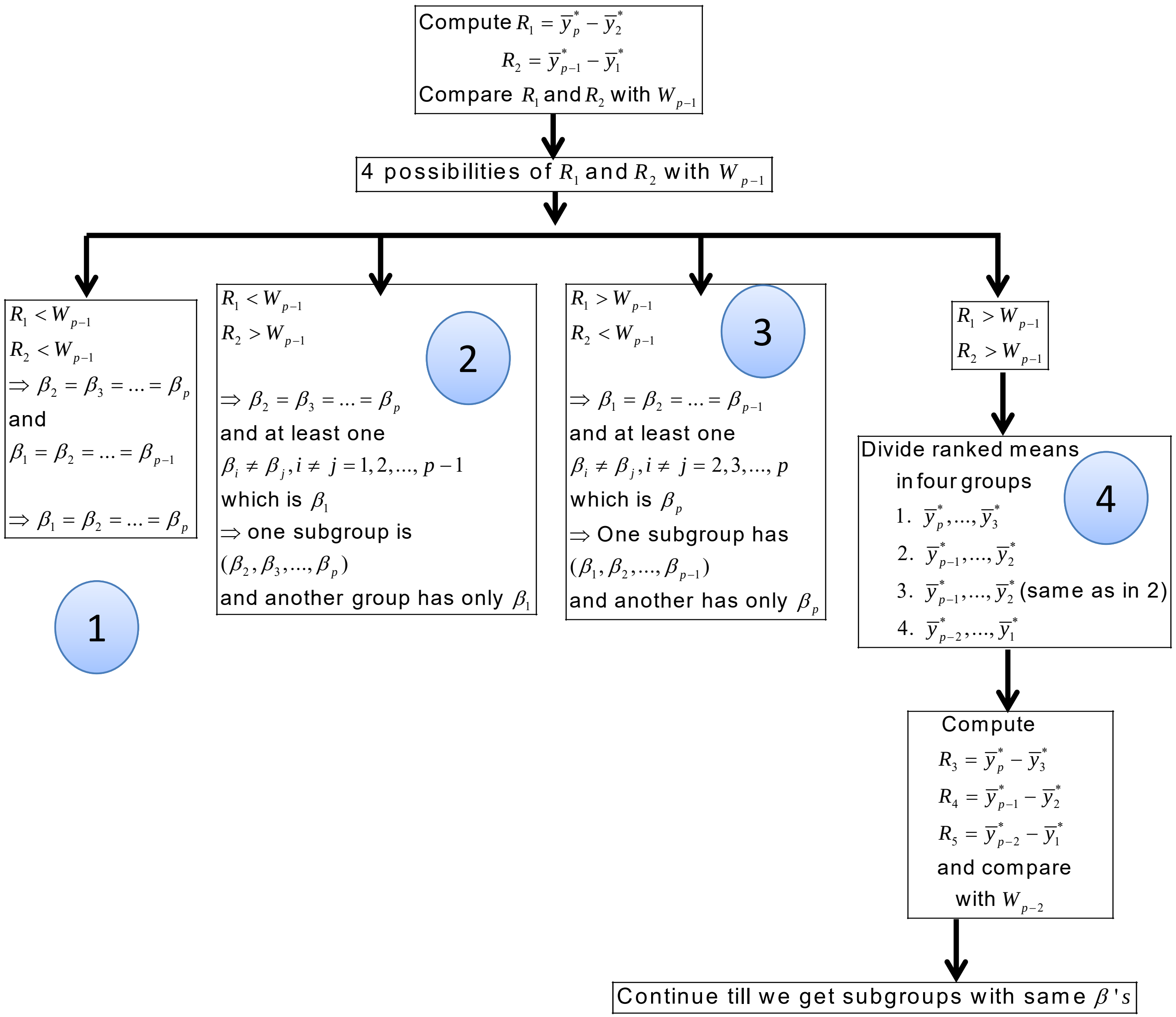
By this method, the difference between any two means under test is significant when the range of the observed means of each and every subgroup containing the two means under test is significant according to the studentized critical range.

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

Student-Newman-Keuls test:



Student-Newman-Keuls test:



Student-Newman-Keuls test:

1

$R_1 < W_{p-1}$

$R_2 < W_{p-1}$

$\Rightarrow \beta_2 = \beta_3 = \ldots = \beta_p$

and

$\beta_1 = \beta_2 = \ldots = \beta_{p-1}$

$\Rightarrow \beta_1 = \beta_2 = \ldots = \beta_p$

2

$R_1 < W_{p-1}$

$R_2 > W_{p-1}$

$\Rightarrow \beta_2 = \beta_3 = \ldots = \beta_p$

and at least one

$\beta_i \neq \beta_j, i \neq j = 1, 2, \ldots, p-1$

which is β_1

$\Rightarrow$ one subgroup is

$(\beta_2, \beta_3, \ldots, \beta_p)$

and another group has only β_1

Student-Newman-Keuls test:

$R_1 > W_{p-1}$
$R_2 < W_{p-1}$

3

$\Rightarrow \beta_1 = \beta_2 = \ldots = \beta_{p-1}$
and at least one
$\beta_i \neq \beta_j, i \neq j = 2, 3, \ldots, p$
which is β_p
$\Rightarrow$ One subgroup has
$(\beta_1, \beta_2, \ldots, \beta_{p-1})$
and another has only β_p

$R_1 > W_{p-1}$
$R_2 > W_{p-1}$

Divide ranked means in four groups

4

1. $\bar{y}^*_p, \ldots, \bar{y}^*_3$
2. $\bar{y}^*_{p-1}, \ldots, \bar{y}^*_2$
3. $\bar{y}^*_{p-1}, \ldots, \bar{y}^*_2$ (same as in 2)
4. $\bar{y}^*_{p-2}, \ldots, \bar{y}^*_1$