Scheffé's Method

Scheffé's Method:

Suppose

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where

$$\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}).$$

Let

$$\boldsymbol{c}_1'\boldsymbol{\beta}, \ldots, \boldsymbol{c}_q'\boldsymbol{\beta}$$

be q estimable functions, where

$$\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q$$

are linearly independent.

Let

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{c}_1' \\ \vdots \\ \boldsymbol{c}_q' \end{bmatrix}$$

and define

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_q \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1'\boldsymbol{\beta} \\ \vdots \\ \boldsymbol{c}_q'\boldsymbol{\beta} \end{bmatrix} = \boldsymbol{C}\boldsymbol{\beta}.$$

Let $\boldsymbol{W} = \boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'$ with diagonal elements denoted

$$w_{11}, \ldots, w_{qq}.$$

Then

$$\text{Var}(\boldsymbol{C}\hat{\boldsymbol{\beta}}) = \sigma^2 \boldsymbol{W}$$

and

$$\text{Var}(\hat{\theta}_j) = \text{Var}(\boldsymbol{c}_j'\hat{\boldsymbol{\beta}}) = \sigma^2 w_{jj} \quad j = 1, \ldots, q.$$

For any $q \times 1$ vector $\boldsymbol{u}$ and any $k \in \mathbb{R}$, let $L(\boldsymbol{u}, k)$ denote the interval

$$[\boldsymbol{u}'\hat{\boldsymbol{\theta}} - k\sqrt{\hat{\sigma}^2\boldsymbol{u}'\boldsymbol{W}\boldsymbol{u}}, \boldsymbol{u}'\hat{\boldsymbol{\theta}} + k\sqrt{\hat{\sigma}^2\boldsymbol{u}'\boldsymbol{W}\boldsymbol{u}}].$$

We want to find $k \ni$

$$\mathbb{P}(\boldsymbol{u}'\boldsymbol{\theta} \in L(\boldsymbol{u}, k) \quad \forall\, \boldsymbol{u} \in \mathbb{R}^q) = 1 - \alpha.$$

Thus, we seek simultaneously coverage probability $1 - \alpha$ for an infinite set of intervals.

Show that

$$\boldsymbol{u}'\boldsymbol{\theta} \in L(\boldsymbol{u}, k) \quad \forall\, \boldsymbol{u} \in \mathbb{R}^q$$

$$\Longleftrightarrow$$

$$\frac{(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{C}\boldsymbol{\beta})'(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}')^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{C}\boldsymbol{\beta})}{q\hat{\sigma}^2} \leq \frac{k^2}{q}.$$

Thus, we have

$$\mathbb{P}(\boldsymbol{u}'\boldsymbol{\theta} \in L(\boldsymbol{u}, k) \quad \forall\, \boldsymbol{u} \in \mathbb{R}^q)$$
$$= \mathbb{P}\left[\frac{(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{C}\boldsymbol{\beta})'(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}')^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{C}\boldsymbol{\beta})}{q\hat{\sigma}^2} \leq \frac{k^2}{q}\right].$$

What shall we choose for k to make this probability equal to $1 - \alpha$?

Example:

Suppose an experiment was conducted using a completely randomized design with 10 subjects in each of 4 treatment groups.

The treatment groups were defined by the combinations of levels from 2 factors: diet (1 or 2) and exercise program (1 or 2).

The model

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

was fit the response data

$$\begin{aligned} y_{ijk} \quad = \quad & \text{measure of overall health} \\ & \text{for diet } i\text{, exercise program } j\text{,} \\ & \text{subject } k \ (i = 1, 2;\ j = 1, 2;\ k = 1, \ldots, 10). \end{aligned}$$

The parameter μ_{ij} represents the mean response for diet i, exercise program j $(i = 1, 2;\ j = 1, 2)$.

The ε_{ijk} terms are assumed to be iid $N(0, \sigma^2)$.

A summary of the data is as follows:

$$\bar{y}_{11\cdot} = 9 \quad \bar{y}_{12\cdot} = 7$$
$$\bar{y}_{21\cdot} = 8 \quad \bar{y}_{22\cdot} = 3$$
$$\hat{\sigma}^2 = 5.$$

Suppose we want to construct a set of confidence intervals using a method that gives simultaneous coverage probability at least 95%.

Suppose the confidence intervals will be used to address the following questions:

1. Diet main effect?
2. Exercise program main effect?
3. Diet-by-exercise program interaction?
4. Difference between diet 1 and diet 2 under exercise program 1?
5. Difference between exercise program 1 and 2 under diet 1?
6. Diet 1, exercise program 1 vs. mean of other treatments?

What estimable function of

$$\boldsymbol{\beta} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix}$$

is of interest in each of the questions 1 through 6, respectively?

Compute an estimate and standard error for each estimable function of interest.

Determine appropriate intervals to address questions 1 through 6.

Each interval is of the form

$$\boldsymbol{c}_j'\hat{\boldsymbol{\beta}} \pm k se_j.$$

How shall we choose k?

If we use the Bonferroni approach, then

$$k = t_{40-4,\frac{0.05}{(2)(6)}} \approx 2.79.$$

This approach would be legitimate if we were interested in these 6 intervals, and only these 6 intervals, prior to observing the data.

Alternatively, we can consider Scheffé intervals, $k = \sqrt{qF_{q,40-4,0.05}}$.

What is the value of q in our situation?