3. Special cases of parallel groups study

§3.1. A general discussion

Three special cases are to be discussed in this notes:

- Several treatments are investigated against a standard treatment or a control;
- Two factors (two drugs, two medical means, etc., which can be combined for the treatment of patients) are investigated to see whether there is any interaction;
- Different treatments correspond to different doses of quantitative factor.

In all these special cases, only pre-specified contrasts are of the investigator's concern. A different approach other than Scheffe and Tukey's methods is in order.

The F test for overall treatment significance is irrelevant in these special cases. The pre-specified specific contrasts should be investigated regardless whether or not the overall F test is significant.

There are cases where the overall F test is non-significant, but still there are some significant pre-specified specific contrasts which are non-artifacts. (Why?)

§3.2. Several treatments versus a control

Suppose there are p groups each of whose means is to be compared with the same control mean. The parallel groups design is applicable for the assignment of patients to the different groups. The data is summarized as follows:

Group	Sample size	Mean	Var	L
0	n_0	$\bar{X}_0$	s_0^2	—
1	n_1	$\bar{X}_1$	s_1^2	L_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
p	n_p	$\bar{X}_p$	s_p^2	L_p

- **Test statistics and Dunnett's multiple comparison procedure**

  The test statistics:

$$L_i = \frac{\bar{X}_i - \bar{X}_0}{s}\sqrt{\frac{n_0 n_i}{n_0 + n_i}}, \quad i = 1, \ldots, p,$$

where

$$s^2 = \frac{\sum_{i=0}^{p}(n_i - 1)s_i^2}{n. - p - 1}, \quad n. = \sum_{i=0}^{p} n_i.$$

Note: Under the assumption f normality, $\bar{X}_i - \bar{X}_0$ follows a normal distribution with mean $\mu_i - \mu_0$ and variance $\frac{n_0+n_i}{n_0 n_i}\sigma^2$;

s^2 is an unbiased estimate of σ^2;

$(n. - p - 1)s^2$ follows a χ^2 distribution with df $n. - p - 1$;

s^2 is independent of $\bar{X}_i, i = 0, 1, \ldots, p$.

L_i follows a t distribution with df $n. - p - 1$.

The Dunnett's multiple comparison procedure:

The procedure is especially developed for the comparison of several treatments versus a control. It compares the L_i's with the Dunnett's critical values given in Table A. 6 of Fleiss.

The critical values for one-sided tests and two-sided tests are given separately. For one-sided tests, $|L_i|$'s are compared with one-sided critical values, and for two-sided test, $|L_i|$'s are compared with two-sided critical values.

The tabled critical values are for the case of equal group sizes, i.e., $n_0 = n_1 = \ldots n_p$. For unequal group sizes, the critical values are adjusted by a factor m defined below:

$$m = 1 + 0.07(1 - \frac{n_i}{n_0}).$$

Example: A total of 60 cockrels were assigned at random to receive either no treatment or one of three drugs in diets. These 60 experimental birds war sacrificed at the end of the experiment and the fat content of the breast muscle was measured. The data is summarized below:

Group	Sample size	Mean	Standard deviation	L
0	15	2.580	0.258	—
1	15	2.461	0.409	-0.92
2	15	2.232	0.381	-2.70
3	15	2.573	0.348	-0.05

One-sided critical value $d_{3,56,0.05} = 2.106$.

The critical value is obtained by interpolation between $d_{3,40,0.05}$ and $d_{3,60,0.05}$ as

$$\begin{aligned} d_{3,56,0.05} &= 2.13 + \frac{2.10 - 2.13}{60 - 40}(56 - 40) \\ &= 2.106. \end{aligned}$$

The treatment 2 is significantly different from the control (reducing the fat content significantly).

- **Optimal allocation of group sizes**

The variance of $\bar{X}_i - \bar{X}_0$ is proportional to $\frac{n_0+n_i}{n_0 n_i}$. A reduction of this variance has the effect of increasing the power of the comparison. Let $n_1 = \cdots = n_p$. The optimal group sizes are

$$n_0 = \frac{n.}{1+\sqrt{p}}, \quad n_i = \frac{n.}{p+\sqrt{p}}.$$

- **Randomization scheme with optimal group sizes**

  The randomly permuted blocks scheme can be carried out in batches of size $(p + \sqrt{p})r$ for randomly chosen $r \leq 4$. Since $\sqrt{p}$ is not necessarily an integer, some adjustment is needed to minimize the gap between theoretical optimal size and the actual allocated size.

§3.3. Investigation of interaction

When more than one means can be used simultaneously for the treatment of patients, there is a concern whether there is any interaction effects between the different means, good or bad. A factorial design can investigate interactions. Besides, it has the advantage of investigating more than one means in a single experiment.

- 2×2 **factorial design**

A 2×2 factorial design is a special case of factorial designs. It involves two factors, each of two levels. E.g., the first factor is whether or not drug A is administered, the second factor is whether or not drug B is administered. The treatments are the four combinations of the levels of the two factors.

An example: In an experiment, male mice were randomly assigned to receive one of four treatments (non operation, castration only, adrenalectomy only, both castration and adrenalectomy). Each animal was sacrificed at a fixed time after the operations and its thymus gland was weighed. Sample sizes in the four groups are equal at the time of operation, but mortality in the immediate postoperative period produced unequal sample sizes at the time of sacrifice. The data is summarized below:

Group	Castration	Adrenalectomy	n_i	$\bar{X}_i$	s_i
1	no	no	13	26.2	4.51
2	yes	no	14	37.4	4.85
3	no	yes	10	32.1	7.56
4	yes	yes	8	43.0	6.55

- **Interaction and its test**

  Effects of one factor within a level of the other factor:

  Effect of Castration without adrenalectomy:

  $$E_{C|\bar{A}} = \mu_2 - \mu_1.$$

  Effect of Castration with adrenalectomy:

  $$E_{C|A} = \mu_4 - \mu_3.$$

  Effect of adrenalectomy without Castration:

  $$E_{A|\bar{C}} = \mu_3 - \mu_1.$$

  Effect of adrenalectomy with Castration:

  $$E_{A|C} = \mu_4 - \mu_2.$$

Interaction effect:

$$\begin{aligned} E_{AC} &= E_{C|A} - E_{C|\bar{A}} = E_{A|C} - E_{A|\bar{C}} \\ &= \mu_4 - \mu_3 - \mu_2 + \mu_1. \end{aligned}$$

Test statistic for interaction:

Let μ_i be estimated by $\bar{X}_i$. Then

$$\hat{E}_{AC} = \bar{X}_4 - \bar{X}_3 - \bar{X}_2 + \bar{X}_1.$$

Test statistic:

$$L_{AC} = \frac{\hat{E}_{AC}}{\sqrt{\widehat{\mathrm{Var}}(\hat{E}_{AC})}},$$

where

$$\widehat{\mathrm{Var}}(\hat{E}_{AC}) = s^2 \sum_{i=1}^{4} \frac{1}{n_i}.$$

Under the hypothesis of no interaction, L_{AC} has a t-distribution with df $n. - 4$. The interaction effect is claimed significant at level α

if

$$|L_{AC}| \geq t_{n.-4,\alpha/2}.$$

Back to the example:

$$\begin{aligned}
\hat{E}_{C|\bar{A}} &= \bar{X}_2 - \bar{X}_1 = 11.2. \\
\hat{E}_{C|A} &= \bar{X}_4 - \bar{X}_3 = 10.9. \\
\hat{E}_{AC} &= \hat{E}_{C|A} - \hat{E}_{C|\bar{A}} = -0.3. \\
s^2 &= \frac{12 \times 4.51^2 + \cdots + 7 \times 6.55^2}{12 + \cdots + 7} \\
&= 33.2823. \\
L_{AC} &= \frac{-0.3}{\sqrt{33.2823 \times (\frac{1}{13} + \cdots + \frac{1}{8})}} \\
&= -0.08.
\end{aligned}$$

$|L_{AC}| = 0.08 < t_{41,0.025} = 2.019541$. The interaction effect is not significant.

- **Inference on main effects**

The main effect of a factor is defined as its averaged within-level effects. For example, the main effect of castration is

$$E_C = \frac{E_{C|A} + E_{C|\bar{A}}}{2}.$$

Note:

1. The concept of "main effect" is defined when more than one factors are involved.
2. The "main" effect of a factor is, in general, not the "pure" effect of that factor when it is administered alone. The "main" effect is equal to the "pure" effect only when the interaction effect is zero.
3. Analysis of "main" effect when interaction effect is non-zero is meaningless. It only makes sense when the interaction is found non-significant.

Estimation of main effects when interaction is zero:

Intuitively, main effect can be estimated by the simple average of the estimated within-level effects. Due to different sample sizes, a more efficient estimate is given by the weighted average of the two estimated within-level effects, with weight being inversely proportional to their variances.

Since

$$\mathrm{Var}(\hat{E}_{C|\bar{A}}) = \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}),$$
$$\mathrm{Var}(\hat{E}_{C|A}) = \sigma^2(\frac{1}{n_3} + \frac{1}{n_4}).$$

the weights are determined as

$$w_{C|\bar{A}} = \frac{n_1 n_2}{n_1 + n_2}, \quad w_{C|A} = \frac{n_3 n_4}{n_3 + n_4}.$$

The estimate of the main effect E_C is then given by

$$\hat{E}_C = \frac{w_{C|\bar{A}}\hat{E}_{C|\bar{A}} + w_{C|A}\hat{E}_{C|A}}{w_{C|\bar{A}} + w_{C|A}}.$$

$$\widehat{\mathrm{Var}}(\hat{E}_C) = \frac{s^2}{w_{C|\bar{A}} + w_{C|A}}.$$

Test statistic:

$$L_C = \frac{\hat{E}_C}{\sqrt{\widehat{\mathrm{Var}}(\hat{E}_C)}}.$$

$L_C \sim t_{n.-4}$ under the null hypothesis of no main effect of C.

Back to the example:

$$w_{C|\bar{A}} = \frac{13 \times 14}{13 + 14} = 6.74,\ w_{C|A} = \frac{10 \times 8}{10 + 8} = 4.44,$$
$$\hat{E}_C = \frac{6.74 \times 11.2 + 4.44 \times 10.9}{6.74 + 4.44} = 11.08,$$
$$\widehat{\mathrm{Var}}(\hat{E}_C) = \frac{33.2823}{11.18} = 2.9769,$$
$$L_C = \frac{11.08}{\sqrt{2.9769}} = 6.40 > t_{41,0.025} = 2.019541.$$

- **The invalidity of the analysis of variance table with unequal sample sizes**

  When sample sizes are unequal, the ANOVA table is not valid for the analysis. The SS of interaction and main effects do not necessarily provide valid measurement for the effect they intend to measure. For example, the SS of interaction for a 2×2 factorial design is given by

$$\mathrm{SS}_I = \sum_{i=1}^{2} \sum_{j=1}^{2} n_{ij} (\bar{X}_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}_{\cdot\cdot})^2.$$

  where, in terms of the notation in the previous summary table,

$$\begin{aligned}
&\bar{X}_{11} = \bar{X}_1, \bar{X}_{21} = \bar{X}_2,\\
&\bar{X}_{12} = \bar{X}_3, \bar{X}_{22} = \bar{X}_4,\\
&\bar{X}_{1.} = \frac{n_1\bar{X}_1 + n_3\bar{X}_3}{n_1 + n_3}, \ \cdots,\\
&\bar{X}_{..} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + n_3\bar{X}_3 + n_4\bar{X}_4}{n_1 + n_2 + n_3 + n_4}.
\end{aligned}$$

The quantity in the parenthesis of SS_I is, for instance,

$$\begin{aligned}
& \bar{X}_{11} - \bar{X}_{1.} - \bar{X}_{.1} + \bar{X}_{..}\\
= \ & \bar{X}_1 - \frac{n_1\bar{X}_1 + n_3\bar{X}_3}{n_1 + n_3} -\\
& \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2} +\\
& \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + n_3\bar{X}_3 + n_4\bar{X}_4}{n_1 + n_2 + n_3 + n_4}.\\
= \ & (1 - \frac{n_1}{n_1 + n_3} - \frac{n_1}{n_1 + n_2} + \frac{n_1}{n_1 + n_2 + n_3 + n_4})\bar{X}_1 -\\
& (\frac{n_2}{n_1 + n_2} - \frac{n_2}{n_1 + n_2 + n_3 + n_4})\bar{X}_2 -\\
& (\frac{n_3}{n_1 + n_3} - \frac{n_3}{n_1 + n_2 + n_3 + n_4})\bar{X}_2 +\\
& \frac{n_3}{n_1 + n_2 + n_3 + n_4}\bar{X}_4.
\end{aligned}$$

When $n_1 = \cdots = n_4 = n$, it reduces to

$$\frac{1}{4}(\bar{X}_1 - \bar{X}_2 - \bar{X}_3 + \bar{X}_4),$$

which is an unbiased estimate of $-4E_{AC}$ and

$$\mathrm{SS}_I = \frac{n}{4}\hat{E}^2_{AC}, \ \ \frac{\mathrm{SS}_I}{s^2} = L^2_{AC}.$$

But, in general, SS_I does not provide any meaningful measure on the interaction effect.

- **Power for testing interaction effect**

Let Δ_0 be the size of effect considered physically meaningful and the error variance σ^2 is known. To test the main effect of this size at level α in the absence of interaction, the power is

$$1 - \beta_M = Pr(Z > z_{\alpha/2} - \frac{\Delta_0\sqrt{n}}{\sigma}).$$

To test an interaction of the same size, the power is

$$1 - \beta_I = Pr(Z > z_{\alpha/2} - \frac{\Delta_0\sqrt{n}}{2\sigma}).$$

For a fixed power $1-\beta$, the sample size needed for the test on main effect is

$$n_M = [\frac{\sigma}{\Delta_0}(z_{\alpha/2} - z_{1-\beta})]^2.$$

But the sample size needed for the test on interaction is

$$n_I = 4[\frac{\sigma}{\Delta_0}(z_{\alpha/2} - z_{1-\beta})]^2 = 4n_M.$$

§3.4. Bonferroni criterion for multiple comparisons

In multiple comparison, if the contrasts to be tested are pre-specified and the number of these contrasts is small, the Scheffe's approach to control overall error rate is conservative.

Let K be the number of contrasts to be tested. Suppose α is the overall error rate to be controlled. To control the overall error rate, the Bonferroni criterion controls the error rate for each individual contrasts at a lower level than α, see, for the contrast k, the rate is controlled at $\alpha_k (< \alpha)$, subject to that the individual error rates sum up to α, i.e.,

$$\sum_{k=1}^{K} \alpha_k = \alpha.$$

The allocation of the α_k's can be determined by a consideration of the seriousness of the type I error associated with each of the contrasts. In general, α_k can be taken equal for all k, i.e.,

$$\alpha_k = \alpha/k.$$

The rationale of Bonferroni criterion:

Let A_k be the event that contrast k is erroneously claimed significant. The probability that at least one such event occurs, which is the overall error rate to be controlled, is given by

$$Pr(\cup_{k=1}^{K} A_k) \leq \sum_{k=1}^{K} Pr(A_k) = \sum_{k=1}^{K} \alpha_k.$$

§3.5. Dose-response studies

- **Purpose of dose-response studies**

  In dose-response studies, the treatments are different levels of a quantitative variable, i.e., the doses of a drug.

  The purpose of the studies is not to see whether there is any difference among the responses to the different doses, but rather to find a relationship between the response and the dose.

Let X_i denote the response to dose Z_i. The following regression relationship is to be investigated:

$$E(X_i) = \beta_0 + \beta_1 Z_i + \beta_2 Z_i^2.$$

By convention, Z_i is taken as $\ln_2(\text{dose}_i/\text{dose}_1)$ where dose_1 is the smallest dose, since dose is usually increased by doubling the previous dose.

In a dose-response study, parallel groups design is applied to different dose levels. The observations are assumed to have the following relationship:

$$X_{ij} = \beta_0 + \beta_1 Z_i + \beta_2 Z_i^2 + \epsilon_{ij},$$
$$i = 1, \ldots, g; \; j = 1, \ldots, n_i.$$

Example: Rickets was experimentally induced in a sample of rats by feeding them a diet deficient in vitamin D. The diet was ten supplemented by one of three doses of vitamin D and the degree to which the right knee was healed was measured. The doses are 3.5, 7 and 14. The $\ln_2$ transformed doses are 0, 1 and 2. The data is summarized in the following table:

Group	Dose	n_i	$\bar{X}_i$	s_i
1	3.5 (0)	10	1.80	1.09
2	7.0 (1)	8	3.31	1.63
3	14.0 (2)	13	3.42	1.02

There is a key difference between the dose-response regression model above and a general regression model. That is, in the dose-response study, at each value of the covariate, there are multiple independent observations, while in a general regression model, there is usually only one observation at each value of the covariate.

The task of analysis:

1. To evaluate whether the regression relationship is adequate for describing the relationship between response and dose.
2. To test whether it should be a quadratic model or and linear model.

- **Estimation of the regression model**

The estimation of the regression coefficients can be made by least square (LS) method based either on the raw individual data or based on the summarized data. Note:

$$\begin{aligned}
& \sum_{i=1}^{g}\sum_{j=1}^{n_i}[X_{ij}-\beta_0-\beta_1 Z_i-\beta_2 Z_i^2]^2 \\
= & \sum_{i=1}^{g}\sum_{j=1}^{n_i}[X_{ij}-\bar{X}_i]^2+\sum_{i=1}^{g} n_i[\bar{X}_i-\beta_0-\beta_1 Z_i-\beta_2 Z_i^2]^2.
\end{aligned}$$

Let

$$\boldsymbol{Z} = \begin{bmatrix} \sqrt{n_1} & \sqrt{n_1}Z_1 & \sqrt{n_1}Z_1^2 \\ \sqrt{n_2} & \sqrt{n_2}Z_2 & \sqrt{n_2}Z_2^2 \\ \vdots & \vdots & \vdots \\ \sqrt{n_g} & \sqrt{n_g}Z_g & \sqrt{n_g}Z_g^2 \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_g \end{bmatrix}.$$

The LS estimate of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{x}.$$

- **Checking the adequacy of the regression model**.

In general regression analysis, the adequacy of the model is checked by residual analysis. The parallel groups design for the dose-response study allows a particular F statistic to be used for this purpose. The statistic is given by

$$F = \frac{\sum_{i=1}^{g} n_i(\bar{X}_i - \hat{X}_i)^2}{(g-3)s^2},$$

where s^2 is the usual pooled estimate of the error variance σ^2, and $\hat{X}_i = \hat{\beta}_0 + \hat{\beta}_1 Z_i + \hat{\beta}_2 Z_i^2$.

If the regression model is adequate, the F statistic above follows a F-distribution with df $g-3$ and $\nu = n. - g$. The hypothesis of adequacy is rejected if $F > F_{g-3,\nu,\alpha/3}$.

The significance level is adjusted by Bonferroni criterion, taking into account three tests are to be conducted in the whole analysis.

If the F test is significant, other models or variable transformations should be explored to seek for an appropriate model.

If $g = 3$, then $F = 0$. The adequacy checking step is skipped.

- **Determining the order of the regression function**

  If the F-test is non-significant, further tests need to be carried out to determine whether a quadratic or a linear model is appropriate.

  By the LS theory, when the model is adequate, the LS estimate $\hat{\boldsymbol{\beta}}$ follows a multivariate normal distribution with mean $\boldsymbol{\beta}$ and variance-covariance matrix $\sigma^2 W^{-1}$, where $W = Z'Z$.

  The test statistic for the significance of a component of $\boldsymbol{\beta}$ is given by

$$L_j = \frac{\hat{\beta}_j}{s\sqrt{W^{[j+1,j+1]}}}, \ j = 1, 2,$$

  where $W^{[j+1,j+1]}$ is the $(j+1)$th diagonal element of W^{-1}.

If $g = 3$, the L_j is compared with $t_{\nu,\alpha/4}$, otherwise, it is compared with $t_{\nu,\alpha/6}$.

If $\beta_2 = 0$ is not rejected, the simple linear regression model should be re-fitted to get new estimates of β_0 and β_1.

- **Confidence band of the regression function**

A confidence band for the regression curve $f(Z) = \beta_0 + \beta_1 Z + \beta_2 Z^2$ is given by

$$\hat{f}(Z) \pm \sqrt{3F_{g-3,\nu,\alpha}}\mathrm{se}(\hat{f}(Z)),$$

where

$$\mathrm{se}(\hat{f}(Z)) = s\sqrt{\boldsymbol{z}'W^{-1}\boldsymbol{z}}, \boldsymbol{z} = (1, Z, Z^2)'.$$

- **Back to the example:**

In the example,

$$W = \begin{bmatrix} 31 & 34 & 60 \\ 34 & 60 & 112 \\ 60 & 112 & 216 \end{bmatrix},$$

$$W^{-1} = \begin{bmatrix} .1000 & -.1500 & .0500 \\ -.1500 & .7442 & -.3442 \\ .0500 & -.3442 & .1692 \end{bmatrix},$$

$$\hat{\boldsymbol{\beta}}' = (1.800, 2.213, -0.703),$$

$$s^2 = 1.4920.$$

The test statistic for β_2 is

$$L_2 = \frac{-.703}{\sqrt{1.4920 \times 0.1692}} = -1.41.$$

$|L_2| < t_{\nu,\alpha/4} = t_{28,0.0125} = 2.36$, for $\alpha = 0.05$.

A simple linear regression model is re-fitted. The fitted regression function is given by

$$\hat{X} = 2.007 + 0.786Z.$$