DOE 3

Completely Randomized Design

1 Model and parameter estimation

1.1 Point estimation

Suppose that there are r different treatments. We are interested in testing the equality of the r population means of the treatments. Let μ_i be the population mean of ith treatment. Then the null and alternative hypotheses are

$$\boxed{\begin{aligned} &H_0 : \mu_1 = \mu_2 = \cdots = \mu_r \\ &H_1 : \mu_i \neq \mu_j \quad \text{for at least one pair } (i, j) \end{aligned}} \tag{1}$$

where $i \neq j$. For the hypothesis testing above, suppose that we collected r independent normal samples as below:

$$\begin{aligned} &\text{Sample } 1: \; Y_{11}, Y_{12}, \ldots, Y_{1n_1} \overset{iid}{\sim} N(\mu_1, \sigma^2) \\ &\text{Sample } 2: \; Y_{21}, Y_{22}, \ldots, Y_{2n_2} \overset{iid}{\sim} N(\mu_2, \sigma^2) \\ &\qquad\vdots \qquad\qquad\qquad\qquad\quad \vdots \\ &\text{Sample } r: \; Y_{r1}, Y_{r2}, \ldots, Y_{rn_r} \overset{iid}{\sim} N(\mu_r, \sigma^2) \end{aligned}$$

Thus, it is reasonable to assume the following model

$$Y_{ij} = \mu_i + \epsilon_{ij}, \tag{2}$$

where $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, $i = 1, 2, \ldots, r$, and $j = 1, 2, \ldots, n_i$. This model is called the *one-way analysis of variance* (also known as the *one-way classification* or the *single-factor analysis* *of variance*).

There is an alternative version of $Y_{ij} = \mu_i + \epsilon_{ij}$ in (2). Let $\mu_i = \mu + \tau_i$. That is, the ith treatment mean is equal to the *overall mean* (baseline) plus the ith treatment mean (effect) so that we have

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij},$$

where we *assume* the following constraint

$$\sum_{i=1}^{r} n_i \tau_i = 0. \tag{3}$$

Note that the null hypothesis in (1) is equivalent to

$$\boxed{\begin{aligned} &H_0 : \tau_1 = \tau_2 = \cdots = \tau_r = 0 \\ &H_1 : \tau_i \neq 0 \quad \text{for at least one } i, \end{aligned}} \tag{4}$$

which is more simple than (1). A possible question can be *what if the constraint in (3) is not used?* In **Remark** 2, we will consider the other case.

One can estimate the parameters using the maximum likelihood estimation. The likelihood function is given by

$$L = \prod_{i=1}^{r} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(- \frac{\epsilon_{ij}^2}{2\sigma^2} \Big) \tag{5}$$

and its log-likelihood is

$$\begin{aligned} \ell &= C - N \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{r} \sum_{j=1}^{n_i} \epsilon_{ij}^2 \\ &= C - N \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \mu_i)^2, \end{aligned} \tag{6}$$

where $N = \sum_{i=1}^{r} n_i$. Thus, maximizing the likelihood in (5) or the log-likelihood (6) is equivalent to minimizing Q_2 below

$$Q_2 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} \epsilon_{ij}^2 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \mu_i)^2. \tag{7}$$

Differentiating Q_2 with respect to μ_i and setting it zero, we have

$$\frac{\partial Q_2}{\partial \mu_i} = 2 \sum_{j=1}^{n_i} (Y_{ij} - \mu_i) \cdot (-1) = (-2) \cdot \Big(\sum_{j=1}^{n_i} Y_{ij} - n_i \mu_i \Big) = 0.$$

Solving the above for μ_i, we have

$$\hat{\mu}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n_i} Y_{i\bullet} = \overline{Y}_{i\bullet},$$

where $Y_{i\bullet} = \sum_{j=1}^{n_i} Y_{ij}$.

We can also reparametrized μ_i with μ and τ_i. Since $\mu_i = \mu + \tau_i$ and $\sum_{i=1}^{r} n_i\tau_i = 0$ due to (3), we have $\sum_{i=1}^{r} n_i\mu_i = \sum_{i=1}^{r} n_i\mu + \sum_{i=1}^{r} n_i\tau_i = N\mu$. Thus, we have $\mu = \sum_{i=1}^{r} n_i\mu_i/N$,

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{r} n_i\hat{\mu}_i = \frac{1}{N}\sum_{i=1}^{r} n_i\overline{Y}_{i\bullet} = \frac{1}{N}\sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij} = \overline{Y}_{\bullet\bullet} \tag{8}$$

and

$$\hat{\tau}_i = \hat{\mu}_i - \hat{\mu} = \overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet}, \tag{9}$$

where $\overline{Y}_{\bullet\bullet} = Y_{\bullet\bullet}/N$ and $Y_{\bullet\bullet} = \sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij}$.

It should be noted that the distribution of $\hat{\mu}$ is $\hat{\mu} \sim N(\mu, \sigma^2/N)$. For convenience, we denote $\hat{Y}_{ij} = \hat{\mu}_i = \hat{\mu} + \hat{\tau}_i$. Then we have

$$\hat{Y}_{ij} = \overline{Y}_{i\bullet}.$$

We also denote the residuals by $\hat{\epsilon}_{ij} = Y_{ij} - \hat{\mu}_i$. Then we have

$$\hat{\epsilon}_{ij} = Y_{ij} - \overline{Y}_{i\bullet}.$$

Remark 1. It should be noted that if the balanced design is used (that is, $n_1 = n_2 = \cdots = n_r = n$), then the constraint in (3) becomes

$$\sum_{i=1}^{r} \tau_i = 0. \tag{10}$$

△

Remark 2. When we reparametrize μ_i with $\mu_i = \mu + \tau_i$, we used the constraint $\sum_{i=1}^{r} n_i\tau_i = 0$ in (3). Then a natural question arises: *What if the constraint $\sum_{i=1}^{r}\tau_i = 0$ is used?* If this constraint is used, we have $\sum_{i=1}^{r}\mu_i = \sum_{i=1}^{r}\mu + \sum_{i=1}^{r}\tau_i = r\mu$. Then we can estimate μ and τ_i as below.

$$\tilde{\mu} = \frac{1}{r}\sum_{i=1}^{r}\hat{\mu}_i = \frac{1}{r}\sum_{i=1}^{r}\overline{Y}_{i\bullet} = \overline{\overline{Y}}_{\bullet\bullet}$$

and

$$\tilde{\tau}_i = \hat{\mu}_i - \tilde{\mu} = \overline{Y}_{i\bullet} - \overline{\overline{Y}}_{\bullet\bullet}.$$

Notice that $\overline{\overline{Y}}_{\bullet\bullet}$ and $\overline{Y}_{\bullet\bullet}$ are not equal in general since

$$\overline{\overline{Y}}_{\bullet\bullet} = \frac{1}{r}\sum_{i=1}^{r}\frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} \quad \text{and} \quad \overline{Y}_{\bullet\bullet} = \frac{1}{\sum_{i=1}^{r} n_i}\sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij}.$$

Thus, $\tilde{\mu}$ and $\tilde{\tau}_i$ are different from $\hat{\mu}$ in (8) and $\hat{\tau}_i$ (9), respectively. However, it is easily seen that $\overline{\overline{Y}}_{\bullet\bullet} = \overline{Y}_{\bullet\bullet}$ when the design is balanced.

Note that the distribution of $\overline{Y}_{i\bullet}$ is $\overline{Y}_{i\bullet} \overset{iid}{\sim} N(\mu+\tau_i, \sigma^2/n_i)$ due to (2). Then, under the constraint $\sum_{i=1}^{r}\tau_i = 0$, we have $\tilde{\mu} \sim N(\mu, \frac{\sigma^2}{r^2}\sum_{i=1}^{r} n_i^{-1})$ while $\hat{\mu} \sim N(\mu, \sigma^2/N)$. Thus, both $\hat{\mu}$ and $\tilde{\mu}$ are unbiased, but $\text{Var}(\hat{\mu}) \leq \text{Var}(\tilde{\mu})$ where the equality holds when the balanced design is used. △

1.2 Interval estimation

Recall that $Y_{ij} = \mu_i + \epsilon_{ij}$, where $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, $i = 1, 2, \ldots, r$, and $j = 1, 2, \ldots, n_i$. Then we have $Y_{i\bullet} \overset{iid}{\sim} N(n_i\mu_i, n_i\sigma^2)$ and thus $\overline{Y}_{i\bullet} \overset{iid}{\sim} N(\mu_i, \sigma^2/n_i)$. Standardizing $\overline{Y}_{i\bullet}$, we have

$$\frac{\overline{Y}_{i\bullet} - \mu_i}{\sqrt{\sigma^2/n_i}} \sim N(0,1).$$

One can show that

$$\frac{1}{\sigma^2}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\bullet})^2 = \frac{1}{\sigma^2}\text{SSE} \sim \chi^2(N-r), \tag{11}$$

which will be detailed in Section 2.1.

Definition 1. The *covariance* of the random variables, U and V is defined by

$$\text{Cov}(U,V) = E[(U-\mu)(V-\nu)],$$

where $\mu = E(U)$ and $\nu = E(V)$.

Lemma 1. *If $U_1, U_2, \ldots, U_n$ and $V_1, V_2, \ldots, V_m$ are random variables and $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_m$ are constants, then we have*

$$\text{Cov}\left(\sum_{i=1}^{n} a_i U_i, \sum_{j=1}^{m} b_j V_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \text{Cov}(U_i, V_j).$$

Proof. See Section 7.4 of Ross (2014). □

Remark 3. Using Definition 1 and Lemma 1, it is easily seen that

$$\begin{aligned}
\mathrm{Cov}(U_1, a_1) &= 0 \\
\mathrm{Cov}(U_1, U_1) &= \mathrm{Var}(U_1) \\
\mathrm{Cov}(U_1, U_2) &= \mathrm{Cov}(U_2, U_1) \\
\mathrm{Cov}(a_1U_1, a_2U_2) &= a_1a_2\mathrm{Cov}(U_1, U_2) \\
\mathrm{Cov}(U_1 + a_1, U_2 + a_2) &= \mathrm{Cov}(U_1, U_2) \\
\mathrm{Cov}(a_1U_1 + a_2U_2, b_1V_1 + b_2V_2) &= \\
&\quad a_1b_1\mathrm{Cov}(U_1, V_1) + a_1b_2\mathrm{Cov}(U_1, V_2) + a_2b_1\mathrm{Cov}(U_2, V_1) + a_2b_2\mathrm{Cov}(U_2, V_2).
\end{aligned}$$

△

Lemma 2. *Under the assumption that $Y_{ij} = \mu_i + \epsilon_{ij}$ with $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, we have the following.*

(a) $\mathrm{Cov}(\overline{Y}_{i\bullet}, Y_{ij} - \overline{Y}_{i\bullet}) = 0$.

(b) $\overline{Y}_{i\bullet}$ *and* $Y_{ij} - \overline{Y}_{i\bullet}$ *are independent.*

(c) $\overline{Y}_{i\bullet}$ *and* $\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\bullet})^2$ *are independent.*

Proof. (a) Refer to Example 4e in Section 7.4 of Ross (2014).
(b) The proof of this part is based on (a) in this lemma. See Section 7.8.2 of Ross (2014). Also see Theorem 4.5-1 of Hogg et al. (2015).
(c) Note that this is very similar to Exercise 11.6 (b) of Casella and Berger (2002). For this proof, refer to Theorem 4.6.12 and Lemma 5.3.3 of Casella and Berger (2002) along with (b) in this lemma. □

We can easily show that $\mathrm{Cov}(\overline{Y}_{i\bullet}, Y_{ij} - \overline{Y}_{i\bullet}) = 0$ because

$$\begin{aligned}
\mathrm{Cov}(\overline{Y}_{i\bullet}, Y_{ij} - \overline{Y}_{i\bullet}) &= \mathrm{Cov}(\overline{Y}_{i\bullet}, Y_{ij}) - \mathrm{Cov}(\overline{Y}_{i\bullet}, \overline{Y}_{i\bullet}) \\
&= \mathrm{Cov}\Big(\frac{1}{n_i}\sum_{j'=1}^{n_i} Y_{ij'}, Y_{ij}\Big) - \mathrm{Var}(\overline{Y}_{i\bullet}) \\
&= \frac{1}{n_i}\mathrm{Cov}(Y_{ij}, Y_{ij}) - \frac{\sigma^2}{n_i} = \frac{1}{n_i}\mathrm{Var}(Y_{ij}) - \frac{\sigma^2}{n_i} = 0.
\end{aligned}$$

It should be noted that $\mathrm{Cov}(U, V) = 0$ does not guarantee the independence of U and V in general. However, as seen in Lemma 2, the condition $\mathrm{Cov}(U, V) = 0$ implies that independence of U and V especially for this one-way ANOVA model.

It is immediate from Lemma 2 (c) that $(\overline{Y}_{i\bullet}-\mu_i)/\sqrt{\sigma^2/n_i}$ and $\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i\bullet})^2/\sigma^2$ are also independent. Then we can Studentize $\overline{Y}_{i\bullet}$ as below

$$\frac{\frac{\overline{Y}_{i\bullet}-\mu_i}{\sqrt{\sigma^2/n_i}}}{\sqrt{\frac{1}{\sigma^2}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i\bullet})^2/(N-r)}} \sim t(\mathrm{df}=N-r). \tag{12}$$

Note that we denote $\mathrm{SSE}=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i\bullet})^2$ and $\mathrm{MSE}=\mathrm{SSE}/(N-r)$ as will be given in (15) and (20), respectively. Using this MSE, we can rewrite (12) as

$$\frac{\overline{Y}_{i\bullet}-\mu_i}{\sqrt{\mathrm{MSE}/n_i}} \sim t(\mathrm{df}=N-r).$$

The endpoints for the interval estimation of μ_i with $100(1-\alpha)\%$ confidence level are given by

$$\boxed{\overline{Y}_{i\bullet} \ \pm\ t(1-\tfrac{\alpha}{2};N-r)\sqrt{\frac{\mathrm{MSE}}{n_i}},}$$

where $t(\gamma;\nu)$ is the *lower* γth quantile of the t distribution with ν degrees of freedom. This is also called the $100(1-\alpha)\%$ confidence interval of μ_i.

We can also obtain the $100(1-\alpha)\%$ confidence interval of $\mu_\ell-\mu_m$ as follows. We have

$$\overline{Y}_{\ell\bullet}-\overline{Y}_{m\bullet} \sim N\left(\mu_\ell-\mu_m, \sigma^2\Big(\frac{1}{n_\ell}+\frac{1}{n_m}\Big)\right)$$

and

$$\frac{(\overline{Y}_{\ell\bullet}-\overline{Y}_{m\bullet})-(\mu_\ell-\mu_m)}{\sqrt{\sigma^2\Big(\frac{1}{n_\ell}+\frac{1}{n_m}\Big)}} \sim N(0,1).$$

Using (11), we have

$$\frac{(\overline{Y}_{\ell\bullet}-\overline{Y}_{m\bullet})-(\mu_\ell-\mu_m)}{\sqrt{\mathrm{MSE}\Big(\frac{1}{n_\ell}+\frac{1}{n_m}\Big)}} \sim t(\mathrm{df}=N-r). \tag{13}$$

Thus, the endpoints for the $100(1-\alpha)\%$ confidence interval of $\mu_\ell-\mu_m$ are given by

$$\boxed{\overline{Y}_{\ell\bullet}-\overline{Y}_{m\bullet} \ \pm\ t(1-\tfrac{\alpha}{2};N-r)\sqrt{\mathrm{MSE}\left(\frac{1}{n_\ell}+\frac{1}{n_m}\right)}.} \tag{14}$$

2 Analysis of variance (ANOVA)

2.1 Decomposition of the total sum of squares

Since $Y_{ij} - \overline{Y}_{\bullet\bullet} = (Y_{ij} - \overline{Y}_{i\bullet}) + (\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet}) = (\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet}) + (Y_{ij} - \overline{Y}_{i\bullet})$, we have

$$\begin{aligned}
\underbrace{\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{\bullet\bullet})^2}_{\mathrm{SS_{Total}}} &= \underbrace{\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2}_{\mathrm{SS_{Treatment}}} + \underbrace{\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\bullet})^2}_{\mathrm{SS_{Error}}} \\
&= \sum_{i=1}^{r} n_i(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2 + \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\bullet})^2 \qquad (15) \\
\mathrm{SSTo} &= \mathrm{SStr} + \mathrm{SSE}.
\end{aligned}$$

Lemma 3. *Let Z_i be* iid *standard normal random variables for $i = 1, 2, \ldots, \nu$ and $\sum_{i=1}^{\nu} Z_i^2 = V_1 + V_2 + \cdots + V_s$, where V_j has ν_j degrees of freedom for $j = 1, 2, \ldots, s$ and $\nu_j > 0$. Then V_j are independent with chi-squared random variables each with ν_j degrees of freedom* if and only if $\nu = \nu_1 + \nu_2 + \cdots + \nu_s$.

Proof. See Cochran (1934) and Chapter 15 of Kendall and Stuart (1979). □

It is immediate from (2) that we have $Y_{ij} \overset{iid}{\sim} N(\mu_i, \sigma^2)$ for each i. Thus, we have $Y_{i\bullet} \overset{iid}{\sim} N(n_i\mu_i, n_i\sigma^2)$ and $Y_{\bullet\bullet} \sim N(\sum_{i=1}^{r} n_i\mu_i, N\sigma^2)$ which is equivalent to $Y_{\bullet\bullet} \sim N(N\mu, N\sigma^2)$ since $\sum_{i=1}^{r} n_i\mu_i = \sum_{i=1}^{r} n_i(\mu + \tau_i) = N\mu$ due to the constraint $\sum_{i=1}^{r} n_i\tau_i = 0$ from (3). Thus, $\overline{Y}_{i\bullet} \overset{iid}{\sim} N(\mu_i, \sigma^2/n_i)$ and $\overline{Y}_{\bullet\bullet} \sim N(\mu, \sigma^2/N)$. It should be emphasized that $\overline{Y}_{\bullet\bullet} \sim N(\mu, \sigma^2/N)$ due to the constraint $\sum_{i=1}^{r} n_i\tau_i = 0$, regardless of the null or alternative hypotheses, H_0 or H_1, in (1).

We have $(Y_{ij} - \mu)/\sigma \sim N(0, 1)$ under H_0, but $\sqrt{N}(\overline{Y}_{\bullet\bullet} - \mu)/\sigma \sim N(0, 1)$ regardless of H_0 or H_1. We have

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \mu)^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left\{(Y_{ij} - \overline{Y}_{\bullet\bullet}) + (\overline{Y}_{\bullet\bullet} - \mu)\right\}^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{\bullet\bullet})^2 + N(\overline{Y}_{\bullet\bullet} - \mu)^2$$

so that

$$\underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \mu)^2}_{\chi^2(N)\text{ under }H_0} = \frac{1}{\sigma^2}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{\bullet\bullet})^2 + \underbrace{\frac{1}{\sigma^2}N(\overline{Y}_{\bullet\bullet} - \mu)^2}_{\chi^2(1)}.$$

Using Lemma 3, we have

$$\frac{1}{\sigma^2}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{\bullet\bullet})^2 = \frac{1}{\sigma^2}\cdot \text{SSTo} \ \sim\ \chi^2(N-1) \text{ under } H_0. \tag{16}$$

We have $\sqrt{n_i}(\overline{Y}_{i\bullet}-\mu)/\sigma \sim N(0,1)$ under H_0, but $\sqrt{N}(\overline{Y}_{\bullet\bullet}-\mu)/\sigma \sim N(0,1)$ regardless of H_0 or H_1. We have

$$\sum_{i=1}^{r} n_i(\overline{Y}_{i\bullet}-\mu)^2 = \sum_{i=1}^{r} n_i\Big\{(\overline{Y}_{i\bullet}-\overline{Y}_{\bullet\bullet})^2 + (\overline{Y}_{\bullet\bullet}-\mu)^2\Big\} = \sum_{i=1}^{r} n_i(\overline{Y}_{i\bullet}-\overline{Y}_{\bullet\bullet})^2 + N(\overline{Y}_{\bullet\bullet}-\mu)^2.$$

so that

$$\underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{r} n_i(\overline{Y}_{i\bullet}-\mu)^2}_{\chi^2(r)\text{ under } H_0} = \frac{1}{\sigma^2}\sum_{i=1}^{r} n_i(\overline{Y}_{i\bullet}-\overline{Y}_{\bullet\bullet})^2 + \underbrace{\frac{1}{\sigma^2}N(\overline{Y}_{\bullet\bullet}-\mu)^2}_{\chi^2(1)}.$$

Then it is immediate from Lemma 3 that we have

$$\frac{1}{\sigma^2}\sum_{i=1}^{r} n_i(\overline{Y}_{i\bullet}-\overline{Y}_{\bullet\bullet})^2 = \frac{1}{\sigma^2}\cdot \text{SStr} \ \sim\ \chi^2(r-1) \text{ under } H_0. \tag{17}$$

Again, we have SSTo = SStr + SSE from (15), that is,

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{r} n_i(\overline{Y}_{i\bullet}-\overline{Y}_{\bullet\bullet})^2 + \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i\bullet})^2,$$

which results in

$$\underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{\bullet\bullet})^2}_{\chi^2(N-1)\text{ under } H_0} = \underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{r} n_i(\overline{Y}_{i\bullet}-\overline{Y}_{\bullet\bullet})^2}_{\chi^2(r-1)\text{ under } H_0} + \frac{1}{\sigma^2}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i\bullet})^2$$

due to (16) and (17). Thus, we have

$$\frac{1}{\sigma^2}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i\bullet})^2 \sim \chi^2(N-r) \text{ under } H_0. \tag{18}$$

We have shown that the statistic in (18) has a chi-squared distribution under H_0, but we can show that it has a chi-squared distribution regardless of H_0 and H_1 as below. We have

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\mu_i)^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i\bullet})^2 + \sum_{i=1}^{r} n_i(\overline{Y}_{i\bullet}-\mu_i)^2.$$

Again, recall that we have $Y_{ij} \sim N(\mu_i, \sigma^2)$ and $\overline{Y}_{i\bullet} \sim N(\mu_i, \sigma^2/n_i)$ regardless of H_0 and H_1. Note that the distributions of the statistics in (16) and (17) were based on the $(Y_{ij}-\mu)/\sigma \sim N(0,1)$ under H_0, but Y_{ij} is not restricted to H_0 here. We have

$$\underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\mu_i)^2}_{\chi^2(N)} = \frac{1}{\sigma^2}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i\bullet})^2 + \underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{r} n_i(\overline{Y}_{i\bullet}-\mu_i)^2}_{\chi^2(r)}.$$

The above works under either H_0 and H_1 as aforementioned. Thus, we have

$$\frac{1}{\sigma^2}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i\bullet})^2 = \frac{1}{\sigma^2}\cdot \text{SSE} \ \sim\ \chi^2(N-r) \text{ under either } H_0 \text{ or } H_1. \tag{19}$$

In summary, we have

$$\boxed{\underbrace{\frac{1}{\sigma^2}\cdot\text{SSTo}}_{\chi^2(N-1)\text{ under }H_0} \quad=\quad \underbrace{\frac{1}{\sigma^2}\cdot\text{SStr}}_{\chi^2(r-1)\text{ under }H_0} \quad+\quad \underbrace{\frac{1}{\sigma^2}\cdot\text{SSE}}_{\chi^2(N-r)}.}$$

2.2 Expected mean square (EMS)

It is well known that the expected value of a chi-squared random variable is its degrees of freedom. Thus, using the above results, we have

$$E\Big(\frac{\text{SSTo}}{N-1}\Big) = E\Big(\frac{\text{SStr}}{r-1}\Big) = E\Big(\frac{\text{SSE}}{N-r}\Big) = \sigma^2 \text{ under } H_0.$$

Thus, SSTo$/(N-1)$, SStr$/(r-1)$ and SSE$/(N-r)$ are all unbiased for σ^2 under H_0. Then, what are the expected values of these quantities in general? First, for convenience, we denote

$$\text{MStr} = \frac{\text{SStr}}{r-1} \qquad \text{and} \qquad \text{MSE} = \frac{\text{SSE}}{N-r}, \tag{20}$$

which are called mean squares. The expected values of these mean squares are called EMS.

We can rewrite SSTo as

$$\text{SSTo} = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}^2 - 2\overline{Y}_{\bullet\bullet}Y_{ij} + \overline{Y}_{\bullet\bullet}^2) = \sum_{i=1}^{r}\sum_{j=1}^{n_i}Y_{ij}^2 - N\overline{Y}_{\bullet\bullet}^2.$$

Since $Y_{ij}\sim N(\mu_i,\sigma^2)$ and $\overline{Y}_{\bullet\bullet}\sim N(\mu,\sigma^2/N)$, we have $E(Y_{ij}^2) = \text{Var}(Y_{ij}) + \mu_i^2 = \sigma^2+\mu_i^2$ and $E(\overline{Y}_{\bullet\bullet}^2) = \text{Var}(\overline{Y}_{\bullet\bullet}) + \mu^2 = \sigma^2/N + \mu^2$. Using these, we have

$$E(\text{SSTo}) = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(\sigma^2+\mu_i^2) - (\sigma^2+N\mu^2) = (N-1)\sigma^2 + \sum_{i=1}^{r} n_i\mu_i^2 - N\mu^2. \tag{21}$$

We also have

$$\sum_{i=1}^{r} n_i\mu_i^2 = \sum_{i=1}^{r} n_i(\mu+\tau_i)^2 = \sum_{i=1}^{r} n_i(\mu^2+\tau_i^2+2\mu\tau_i) = N\mu^2 + \sum_{i=1}^{r} n_i\tau_i^2 + 2\mu\sum_{i=1}^{r} n_i\tau_i,$$

where the term, $\sum_{i=1}^{r} n_i\tau_i$, is zero by (3) so that we have

$$\sum_{i=1}^{r} n_i\mu_i^2 = N\mu^2 + \sum_{i=1}^{r} n_i\tau_i^2. \tag{22}$$

Substituting (22) into (21), we have

$$E(\text{SSTo}) = (N-1)\sigma^2 + \sum_{i=1}^{r} n_i\tau_i^2 \quad \text{and} \quad E\Big(\frac{\text{SSTo}}{N-1}\Big) = \sigma^2 + \frac{\sum_{i=1}^{r} n_i\tau_i^2}{N-1}.$$

As expected, $E\big(\text{SSTo}/(N-1)\big)$ becomes σ^2 under H_0.

We rewrite SStr as

$$\text{SStr} = \sum_{i=1}^{r} n_i(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{r} n_i\overline{Y}_{i\bullet}^2 - N\overline{Y}_{\bullet\bullet}^2.$$

Since $\overline{Y}_{i\bullet} \sim N(\mu_i, \sigma^2/n_i)$, we have $E(\overline{Y}_{i\bullet}^2) = \text{Var}(\overline{Y}_{i\bullet}) + \mu_i^2 = \sigma^2/n_i + \mu_i^2$. Using this along with (22) and $E(\overline{Y}_{\bullet\bullet}^2) = \sigma^2/N + \mu^2$, we have

$$E(\text{SStr}) = \sum_{i=1}^{r}(\sigma^2 + n_i\mu_i^2) - (\sigma^2 + N\mu^2) = (r-1)\sigma^2 + \sum_{i=1}^{r} n_i\mu_i^2 - N\mu^2 = (r-1)\sigma^2 + \sum_{i=1}^{r} n_i\tau_i^2,$$

which results in

$$\boxed{E(\text{MStr}) = \sigma^2 + \frac{\sum_{i=1}^{r} n_i\tau_i^2}{r-1}.}$$

This quantity also becomes σ^2 under H_0.

Next, we rewrite SSE as

$$\text{SSE} = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\bullet})^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^{r} n_i\overline{Y}_{i\bullet}^2$$

Using $Y_{ij} \sim N(\mu_i, \sigma^2)$ and $\overline{Y}_{i\bullet} \sim N(\mu_i, \sigma^2/n_i)$, we have

$$E(\text{SSE}) = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(\sigma^2 + \mu_i^2) - \sum_{i=1}^{r} n_i(\sigma^2/n_i + \mu_i^2) = (N-r)\sigma^2,$$

which results in

$$\boxed{E(\text{MSE}) = \sigma^2.}$$

Note that the above works under H_0 and H_1 as well.

ANOVA Decomposition

Source	SS	df	MS	F	EMS
Treatment	SStr	$r-1$	$\text{MStr} = \text{SStr}/(r-1)$	$F = \frac{\text{MStr}}{\text{MSE}}$	$\sigma^2 + \frac{\sum_{i=1}^{r} n_i\tau_i^2}{r-1}$
Error	SSE	$N-r$	$\text{MSE} = \text{SSE}/(N-r)$		σ^2
Total	SSTo	$N-1$			

In summary, we recall the decomposition of the total sum of squares in (15):

$$\boxed{\underbrace{\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{\bullet\bullet})^2}_{\text{SSTo}} = \underbrace{\sum_{i=1}^{r} n_i(\overline{Y}_{i\bullet}-\overline{Y}_{\bullet\bullet})^2}_{\text{SStr}} + \underbrace{\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i\bullet})^2}_{\text{SSE}}} \tag{23}$$

3 One-way ANOVA as regression

3.1 Cell-means coding

A very important application of regression analysis involves a list of predictors which can include categorical variables for treatments. The categorical variables are also called *indicator* or *dummy* variables or qualitative variables. We re-analyze the above one-way ANOVA model using the linear regression model. For notational brevity, we omit ij index and we then have

$$Y = \beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_r Z_r + \epsilon \tag{24}$$

where $\epsilon \overset{iid}{\sim} N(0,\sigma^2)$. We define Z_i as follows:

$$Z_i = \begin{cases} 1 & : \text{ for the } i\text{th treatment} \\ 0 & : \text{ otherwise} \end{cases},$$

where $i = 1, 2, \ldots, r$. That is, the dummy variables Z_i are coded as below. This scheme is called *cell-means coding*

Cell-means coding

Treatment	Z_1	Z_2	Z_3	$\cdots$	Z_r
$i=1$	1	0	0	$\cdots$	0
$i=2$	0	1	0	$\cdots$	0
$i=3$	0	0	1	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
$i=r$	0	0	0	$\cdots$	1

It is easy to see that the model in (24) implies that

$$\begin{aligned}
\mu_1 &= E(Y_{1j}) = \beta_1 \\
\mu_2 &= E(Y_{2j}) = \beta_2 \\
\vdots &= \quad \vdots \quad = \vdots \\
\mu_r &= E(Y_{rj}) = \beta_r
\end{aligned}$$

where $Y_{ij} = \mu_i + \epsilon_{ij}$ and $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$ again. This clearly shows that the linear regression can handle the one-way ANOVA model using the cell-means coding.

Suppose that we stack the responses (Y_i) from the r samples in such a way that the first sample comes first, the second sample comes next, and the last rth sample comes last. Then the response vector of length $N = \sum_{i=1}^{r} n_i$, the data matrix (or design matrix) and the vector of the error terms are given by

$$\underset{N\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{n_1} \\ Y_{n_1+1} \\ \vdots \\ Y_{n_1+n_2} \\ \vdots \\ \vdots \\ Y_{N-n_r+1} \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \\ \vdots \\ \vdots \\ Y_{r1} \\ \vdots \\ Y_{rn_r} \end{bmatrix}, \underset{N\times r}{\mathbf{Z}} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \underset{N\times 1}{\boldsymbol{\epsilon}} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_{n_1} \\ \epsilon_{n_1+1} \\ \vdots \\ \epsilon_{n_1+n_2} \\ \vdots \\ \vdots \\ \epsilon_{N-n_r+1} \\ \vdots \\ \epsilon_N \end{bmatrix} = \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1n_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2n_2} \\ \vdots \\ \vdots \\ \epsilon_{r1} \\ \vdots \\ \epsilon_{rn_r} \end{bmatrix}.$$

Using the response vector and the data matrix, we can rewrite the linear regression model as

$$\boxed{\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon}.}$$

To estimate the regression parameters, we use the least squares method which is equivalent to the maximum likelihood method under the normal distribution assumption. It is easily seen that the sum of squared errors is given by

$$Q_2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i} \epsilon_{ij}^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i} (Y_{ij} - \beta_i)^2,$$

which is essentially the same as the (7). Differentiating Q_2 with respect to β_i and solving the estimating equations, we have

$$\hat{\beta}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n_i} Y_{i\bullet} = \overline{Y_{i\bullet}},$$

where $Y_{i\bullet} = \sum_{j=1}^{n_i} Y_{ij}$ again.

Remark 4. It is easy to show that

$$\mathbf{Z}'\mathbf{Z} = \text{diag}(n_1, n_2, \ldots, n_r) = \begin{bmatrix} n_1 & 0 & \ldots & 0 \\ 0 & n_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & n_r \end{bmatrix}.$$

Thus, we have $(\mathbf{Z}'\mathbf{Z})^{-1} = \text{diag}(1/n_1, 1/n_2, \ldots, 1/n_r)$. Let $\mathbf{1}_n$ be a n-dimensional vector with all the elements being ones. Then $\mathbf{J}_n = \mathbf{1}_n\mathbf{1}_n'$ is a $n \times n$ square matrix of ones, which is often called *all-ones matrix.* Then the hat matrix for the cell-means coding is given by

$$\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' = \text{diag}\left(\frac{1}{n_1}\mathbf{J}_{n_1}, \frac{1}{n_2}\mathbf{J}_{n_2}, \ldots, \frac{1}{n_r}\mathbf{J}_{n_r}\right).$$

△

Remark 5. The cell-means coding has nice features. However, it is rarely used in practice because of the following reasons.

1. The data matrix $\mathbf{Z}$ does not include $\mathbf{1} = (1, 1, \ldots, 1)'$ in the column which implies that the intercept (grand mean) is not used. Thus, a usual way of decomposing the total sum of squares (SSTo) does not work.

2. Based on the regression model in (24), one can test the hypothesis that a cell mean is zero in such a way that

   $$H_0 : \beta_i = 0,$$

   whose usual testing statistic is given by

   $$T = \frac{\hat{\beta}_i - 0}{\sqrt{\text{MSE} \cdot (\mathbf{Z}'\mathbf{Z})^{-1}_{ii}}} = \frac{\hat{\beta}_i - 0}{\sqrt{\text{MSE}/n_i}}.$$

   Unlike the conventional regression applications, it does not often make sense to perform the hypothesis above in real-world applications of the design of experiments. Usually, it is more interesting to test whether $\tau_i = 0$ instead of $\beta_i = 0$.

For these reasons, the cell-means coding is not widely-used in practice. However, the use of the cell-means coding along with the regression method present an idea of how to incorporate the regression method into the design of experiments. △

3.2 Effect coding

The effect is another popular coding scheme to incorporate the regression method into the design of experiments. We re-analyze the above one-way ANOVA model using the linear regression model *with the intercept*. For notational brevity, we omit ij index and we then have

$$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_{r-1} Z_{r-1} + \epsilon \tag{25}$$

where $\epsilon \overset{iid}{\sim} N(0, \sigma^2)$. For the ith treatment where $i = 1, 2, \ldots, r-1$, we set up $Z_i = 1$ and all other dummy variables are zero, but we set $Z_i = -n_i/n_r$ for the rth treatment. Then we have

$$Z_i = \begin{cases} 1 & : \text{ for the } i\text{th treatment} \\ -n_i/n_r & : \text{ for the } r\text{th treatment} \\ 0 & : \text{ otherwise} \end{cases},$$

where $i = 1, 2, \ldots, r-1$. This scheme is called *effect coding*. Note that, including the intercept β_0, there are r regression coefficients all told. For convenience, let $Z_0 = 1$ which is considered as the predictor with β_0. That is, the dummy variables Z_i are coded as below.

Effect coding

Treatment	Z_0	Z_1	Z_2	$\cdots$	Z_{r-1}
$i=1$	1	1	0	$\cdots$	0
$i=2$	1	0	1	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
$i=r-1$	1	0	0	$\cdots$	1
$i=r$	1	$-n_1/n_r$	$-n_2/n_r$	$\cdots$	$-n_{r-1}/n_r$

It is easy to see that the model in (25) implies that

$$\begin{aligned} \mu_1 = E(Y_{1j}) &= \beta_0 + \beta_1 \\ \mu_2 = E(Y_{2j}) &= \beta_0 + \beta_2 \\ \vdots = \quad \vdots \quad &= \quad \vdots \\ \mu_{r-1} = E(Y_{r-1,j}) &= \beta_0 + \beta_{r-1} \\ \mu_r = E(Y_{rj}) &= \beta_0 - \frac{n_1}{n_r}\beta_1 - \frac{n_2}{n_r}\beta_2 - \cdots - \frac{n_{r-1}}{n_r}\beta_{r-1} \end{aligned}$$

where $Y_{ij} = \mu_i + \epsilon_{ij}$ and $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$ again.

Again, suppose that we stack the responses (Y_i) from the r samples in such a way that the first sample comes first, the second sample comes next, and the last rth sample comes last. Then the response vector of length $N = \sum_{i=1}^{r} n_i$ and the data or design matrix are given by

$$\underset{N\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{n_1} \\ Y_{n_1+1} \\ \vdots \\ Y_{n_1+n_2} \\ \vdots \\ \vdots \\ Y_{N-n_r+1} \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \\ \vdots \\ \vdots \\ Y_{r1} \\ \vdots \\ Y_{rn_r} \end{bmatrix}, \underset{N\times r}{\mathbf{Z}} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \\ \vdots & \vdots & & \vdots & \\ 1 & -n_1/n_r & -n_2/n_r & \cdots & -n_{r-1}/n_r \\ \vdots & \vdots & & \vdots & \\ 1 & -n_1/n_r & -n_2/n_r & \cdots & -n_{r-1}/n_r \end{bmatrix}.$$

Using the response vector and the data matrix, we can rewrite the linear regression model as

$$\boxed{\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon}.}$$

Analogous with the cell-means coding case, we can estimate the regression parameters using the least squares method. Then the sum of squared errors is given by

$$\begin{aligned} Q_2 &= \sum_{i=1}^{r}\sum_{j=1}^{n_i} \epsilon_{ij}^2 \\ &= \sum_{i=1}^{r-1}\sum_{j=1}^{n_i} \epsilon_{ij}^2 + \sum_{j=1}^{n_i} \epsilon_{rj}^2 \\ &= \sum_{i=1}^{r-1}\sum_{j=1}^{n_i} (Y_{ij} - \beta_0 - \beta_i)^2 + \sum_{j=1}^{n_i} \Big(Y_{rj} - \beta_0 + \frac{n_1}{n_r}\beta_1 + \frac{n_2}{n_r}\beta_2 + \cdots + \frac{n_{r-1}}{n_r}\beta_{r-1}\Big)^2. \end{aligned}$$

Let $\beta_r = -(n_1\beta_1 + n_2\beta_2 + \cdots + n_{r-1}\beta_{r-1})/n_r$. Then the minimization of the above Q_2 is equivalent to minimizing it

$$Q_2^* = \sum_{i=1}^{r}\sum_{j=1}^{n_i} (Y_{ij} - \beta_0 - \beta_i)^2 \quad \text{subject to} \quad \sum_{i=1}^{r} n_i\beta_i = 0.$$

The auxiliary function with Lagrange multiplier λ is

$$\Psi = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \beta_0 - \beta_i)^2 - \lambda \sum_{i=1}^{r} n_i\beta_i.$$

Differentiating Ψ with respect to β_0 and setting it to zero, we have

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \beta_0 - \beta_i) = 0,$$

which results in

$$Y_{\bullet\bullet} - N\beta_0 - \sum_{i=1}^{r} n_i\beta_i = Y_{\bullet\bullet} - N\beta_0 = 0.$$

Thus, we have

$$\hat{\beta}_0 = \overline{Y}_{\bullet\bullet}. \tag{26}$$

Next, differentiating Ψ with respect to β_i and setting it to zero, we have

$$\sum_{j=1}^{n_i}(Y_{ij} - \beta_0 - \beta_i) = 0,$$

which results in

$$Y_{i\bullet} - n_i\beta_0 - n_i\beta_i = 0.$$

Solving the above for β_i and substituting $\hat{\beta}_0 = \overline{Y}_{\bullet\bullet}$ into the above, we have

$$\hat{\beta}_i = \overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet}. \tag{27}$$

Comparing (26) and (27) with (8) and (9), respectively, it is easily seen that

$$\hat{\beta}_0 = \hat{\mu} \quad \text{and} \quad \hat{\beta}_i = \hat{\tau}_i.$$

Remark 6. Unlike the the cell-means coding case, it is quite complex to obtain $\mathbf{Z}'\mathbf{Z}$ which is certainly different from the cell-means coding case. However, the hat matrix for the effect coding is the same as that for the cell-means coding which is again given by

$$\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' = \operatorname{diag}\left(\frac{1}{n_1}\mathbf{J}_{n_1}, \frac{1}{n_2}\mathbf{J}_{n_2}, \ldots, \frac{1}{n_r}\mathbf{J}_{n_r}\right),$$

where $\mathbf{J}_n = \mathbf{1}_n\mathbf{1}_n'$ and $\mathbf{1}_n$ is a n-dimensional vector with all the elements being ones. △

4 Two sample t-test as DOE

We briefly review the two-sample t-test statistic. Suppose that $Y_{1j} \overset{iid}{\sim} N(\mu_1, \sigma^2)$ for $j = 1, 2, \ldots, n_1$ and $Y_{2j} \overset{iid}{\sim} N(\mu_2, \sigma^2)$ for $j = 1, 2, \ldots, n_2$. The two-sample t-test statistic under $H_0 : \mu_1 = \mu_2$ is given by

$$\boxed{T = \frac{\overline{Y}_{1\bullet} - \overline{Y}_{2\bullet}}{\sqrt{S_p^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t(\mathrm{df} = n_1 + n_2 - 2),}$$

where

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2},\ S_1^2 = \frac{1}{n_1-1}\sum_{j=1}^{n_1}(Y_{1j} - \overline{Y}_{1\bullet})^2,\ S_2^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(Y_{2j} - \overline{Y}_{2\bullet})^2.$$

Then S_p^2 can be rewritten as

$$S_p^2 = \frac{\sum_{j=1}^{n_1}(Y_{1j} - \overline{Y}_{1\bullet})^2 + \sum_{j=1}^{n_2}(Y_{2j} - \overline{Y}_{2\bullet})^2}{n_1+n_2-2} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\bullet})^2}{N-r} = \frac{\mathrm{SSE}}{N-r} = \mathrm{MSE},$$

where $N = n_1 + n_2$ and $r = 2$. We also have $T^2 \sim F(r-1, N-r)$ and

$$T^2 = \frac{(\overline{Y}_{1\bullet} - \overline{Y}_{2\bullet})^2}{\mathrm{MSE} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \frac{(\overline{Y}_{1\bullet} - \overline{Y}_{2\bullet})^2}{\frac{\mathrm{SSE}}{N-r} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \frac{\frac{n_1 n_2}{N} \cdot (\overline{Y}_{1\bullet} - \overline{Y}_{2\bullet})^2}{\mathrm{SSE}/(N-r)} \sim F(r-1, N-r). \tag{28}$$

We can also rewrite SStr for $r = 2$ as follows. Substituting $\overline{Y}_{\bullet\bullet} = (n_1\overline{Y}_{1\bullet} + n_2\overline{Y}_{2\bullet})/N$ into

$$\mathrm{SStr} = \sum_{i=1}^{r} n_i(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2 = n_1(\overline{Y}_{1\bullet} - \overline{Y}_{\bullet\bullet})^2 + n_2(\overline{Y}_{2\bullet} - \overline{Y}_{\bullet\bullet})^2,$$

we have

$$\begin{aligned}
\mathrm{SStr} &= n_1\left(\overline{Y}_{1\bullet} - \frac{n_1\overline{Y}_{1\bullet} + n_2\overline{Y}_{2\bullet}}{N}\right)^2 + n_2\left(\overline{Y}_{2\bullet} - \frac{n_1\overline{Y}_{1\bullet} + n_2\overline{Y}_{2\bullet}}{N}\right)^2 \\
&= n_1\left[\frac{(n_1+n_2)\overline{Y}_{1\bullet}}{N} - \frac{n_1\overline{Y}_{1\bullet} + n_2\overline{Y}_{2\bullet}}{N}\right]^2 + n_2\left[\frac{(n_1+n_2)\overline{Y}_{2\bullet}}{N} - \frac{n_1\overline{Y}_{1\bullet} + n_2\overline{Y}_{2\bullet}}{N}\right]^2 \\
&= n_1\left[\frac{n_2(\overline{Y}_{1\bullet} - \overline{Y}_{2\bullet})}{N}\right]^2 + n_2\left[\frac{n_1(\overline{Y}_{2\bullet} - \overline{Y}_{1\bullet})}{N}\right]^2 \\
&= \left(\frac{n_1 n_2^2}{N^2} + \frac{n_2 n_1^2}{N^2}\right)(\overline{Y}_{1\bullet} - \overline{Y}_{2\bullet})^2 = \frac{n_1 n_2}{N}(\overline{Y}_{1\bullet} - \overline{Y}_{2\bullet})^2.
\end{aligned}$$

Comparing the above with (28), we have

$$T^2 = \frac{\mathrm{SStr}/(r-1)}{\mathrm{SSE}/(N-r)} \sim F(r-1, N-r),$$

where $r = 2$ again. That is,

$$\boxed{F = \frac{\text{MStr}}{\text{MSE}} \sim F(r-1, N-r).}$$

5 Examples

Example 1. As a real-data example, we consider the two-sample t-test problem in Example 8.2-2 of Hogg et al. (2015).

> *A machine with filler heads packages a product. This machine has 12 fillers on the left side and another 12 fillers on the right side. Let Y_{1j} and Y_{2j} be the fill weights in grams when a machine fills a package by the left and right heads, respectively. We assume that Y_{1j} and Y_{2j} are normally distributed with a common variance. The data sets are given by*
>
> $Y_{1j} = \{1071, 1076, 1070, 1083, 1082, 1067, 1078, 1080, 1075, 1084, 1075, 1080\}$ *and*
>
> $Y_{2j} = \{1074, 1069, 1075, 1067, 1068, 1079, 1082, 1064, 1070, 1073, 1072, 1075\}$.

Ⓡ: Reading Data

```
> Y1 = c(1071,1076,1070,1083,1082,1067,1078,1080,1075,1084,1075,1080)
> Y2 = c(1074,1069,1075,1067,1068,1079,1082,1064,1070,1073,1072,1075)
> boxplot(Y1,Y2, names=c("X","Y") )
```

Ⓡ: Two sample t-test

```
> t.test(Y1,Y2, alternative="two.sided", var.equal=TRUE)
        Two Sample t-test
data:  Y1 and Y2
t = 2.053, df = 22, p-value = 0.05215
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.04488773  8.87822107
sample estimates:
mean of x mean of y
 1076.750  1072.333
```

Ⓡ: ANOVA with aov() function with factor

```
> group = factor( c(rep("Y1",12), rep("Y2",12)) )
> Y = c(Y1,Y2)
> plot(Y~group)  # Boxplot

> myaov = aov(Y ~ group)
> summary(myaov)
            Df Sum Sq Mean Sq F value Pr(>F)
group        1  117.0  117.04   4.215 0.0522 .
Residuals   22  610.9   27.77
---

> # ---------------------------------------------
> # Parameter Estimation
> # ---------------------------------------------
> mui = tapply(Y, group, mean)
> tau = mui - mean(Y)
> cbind(mui, tau)
```

```
         mui        tau
Y1 1076.750   2.208333
Y2 1072.333  -2.208333
```

Ⓡ: ANOVA with lm() function with factor

```
> group = factor( c(rep("Y1",12), rep("Y2",12)) )
> mylm = lm (Y ~ group)
> anova(mylm)     # same as the summary(myaov) above
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value  Pr(>F)
group      1 117.04 117.042  4.2148 0.05215 .
Residuals 22 610.92  27.769
---

> summary(mylm)  # ANOVA is OK, but parameter estimates are not
Call:
lm(formula = Y ~ group)
Residuals:
    Min      1Q  Median      3Q     Max
-9.7500 -3.5833  0.1667  3.2500  9.6667

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1076.750      1.521 707.825   <2e-16 ***
groupY2       -4.417      2.151  -2.053   0.0522 .
---

Residual standard error: 5.27 on 22 degrees of freedom
Multiple R-squared:  0.1608,    Adjusted R-squared:  0.1226
F-statistic: 4.215 on 1 and 22 DF,  p-value: 0.05215
```

Ⓡ: Regression: cell-means coding

```
> z1 =  c(rep(1,12), rep(0,12))
> z2 =  c(rep(0,12), rep(1,12))
>
> # under H0
> mylm0 = lm( Y ~ 1 )
> anova(mylm0)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 23 727.96   31.65
>
> # under H1
> mylm1 = lm( Y ~ 0 + z1 + z2 )
> mylm1
Call:
lm(formula = Y ~ 0 + z1 + z2)
Coefficients:
  z1    z2
1077  1072

> anova(mylm1)
Analysis of Variance Table
Response: Y
          Df   Sum Sq  Mean Sq F value    Pr(>F)
z1         1 13912687 13912687  501016 < 2.2e-16 ***
z2         1 13798785 13798785  496914 < 2.2e-16 ***
Residuals 22      611       28
---

> # ---------------------------------------------
> # Parameter Estimation
> # ---------------------------------------------
> mu0 = mean(Y)
> mui = coef(mylm1)
> tau = mui - mu0
> cbind(mui, tau)
```

```
          mui        tau
z1 1076.750   2.208333
z2 1072.333  -2.208333
>
> # ---------------------------------------------------
> # Difference between H0 and H1  (Note: anova() function)
> # ---------------------------------------------------
> anova(mylm0, mylm1)
Analysis of Variance Table
Model 1: Y ~ 1
Model 2: Y ~ 0 + z1 + z2
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     23 727.96
2     22 610.92  1    117.04 4.2148 0.05215 .
---
```

Ⓡ: Regression: Effect coding

```
> z1 =  c(rep(1,12), rep(-1,12))
>
> # under H0
> mylm0 = lm( Y ~ 1 )
> anova(mylm0)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 23 727.96   31.65

> # under H1 (note: with intercept)
> mylm1 = lm( Y ~ 1 + z1 )
> mylm1
Call:
lm(formula = Y ~ 1 + z1)
Coefficients:
(Intercept)           z1
   1074.542        2.208

> anova(mylm1)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value  Pr(>F)
z1         1 117.04 117.042  4.2148 0.05215 .
Residuals 22 610.92  27.769
---

> # ---------------------------------------------------
> # Parameter Estimation
> # ---------------------------------------------------
> mu0 = coef(mylm1)[1]
> tau = c( coef(mylm1)[2], -coef(mylm1)[2] )
> mui = c(mu0+tau)
> cbind(mui, tau)
          mui        tau
z1 1076.750   2.208333
z1 1072.333  -2.208333

> # ---------------------------------------------------
> # Difference between H0 and H1
> # ---------------------------------------------------
> anova(mylm0, mylm1)
Analysis of Variance Table
Model 1: Y ~ 1
Model 2: Y ~ 1 + z1
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     23 727.96
2     22 610.92  1    117.04 4.2148 0.05215 .
---
```

Example 2. As a real-data example, we consider Example 3.6 of the textbook by Kim (2014).

> *To determine the effect of iris color on critical flicker frequency (CFF), iris colours and CFFs for 19 persons are recorded. The subjects are divided into three groups on the basis of iris color (blue, brown and green). Let Y_{1j}, Y_{2j} and Y_{3j} be the CFFs of blue, brown and green iris colors, respectively. We assume that Y_{1j}, Y_{2j} and Y_{3j} are normally distributed with a common variance. As will be shown below, the analysis indicates that iris color is a statistically significant factor for CFF. The original data set is provided in Smith and Misiak (1973).*

Ⓡ: Reading Data

```
> Y1 = c(25.7, 27.2, 29.9, 28.5, 29.4, 28.3)           # Blue
> Y2 = c(26.8, 27.9, 23.7, 25, 26.3, 24.8, 25.7, 24.5) # Brown
> Y3 = c(26.4, 24.2, 28.0, 26.9, 29.1)                 # Green
> Y = c(Y1, Y2, Y3)
> n1 = length(Y1); n2 = length(Y2); n3 = length(Y3)    # no. of subjects
```

Ⓡ: ANOVA with aov() function with factor

```
> # Check the input
> color = factor( rep(c("Blue","Brown","Green"), c(n1,n2,n3)) )
> levels(color)
[1] "Blue"   "Brown"  "Green"
> cbind(Y,color)
          Y color
 [1,]  25.7     1
 [2,]  27.2     1
      .........
      .........
# Plots
# stripchart(Y ~ color, vertical=T)
# boxplot(Y ~ color, ylab="Flicker")

> myaov = aov(Y ~ color)
> summary(myaov)
            Df  Sum Sq  Mean Sq  F value  Pr(>F)
color        2  23.00   11.499   4.802  0.0232 *
Residuals   16  38.31    2.394
---

> # mu.hat and tau.hat
> mui = tapply(Y, color, mean)
> tau = mui - mean(Y)
> cbind(mui, tau)
           mui        tau
Blue   28.16667  1.4140351
Brown  25.58750 -1.1651316
Green  26.92000  0.1673684

# Diagnostic
# par ( mfrow=c(2,2) )
# plot(myaov)
```

Ⓡ: ANOVA with lm() function with factor

```
> color = factor( rep(c("Blue","Brown","Green"), c(n1,n2,n3)) )
> mylm = lm (Y ~ color)
> anova(mylm)     # same as the summary(myaov) above
Analysis of Variance Table
```

```
Response: Y
          Df Sum Sq Mean Sq F value  Pr(>F)
color      2 22.997 11.4986  4.8023 0.02325 *
Residuals 16 38.310  2.3944
---

> summary(mylm)  # ANOVA is OK, but parameter estimates are not
Call:
lm(formula = Y ~ color)
Residuals:
    Min      1Q  Median      3Q     Max
-2.7200 -0.8771  0.1125  1.1462  2.3125

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  28.1667     0.6317  44.588  < 2e-16 ***
colorBrown   -2.5792     0.8357  -3.086  0.00708 **
colorGreen   -1.2467     0.9370  -1.331  0.20200
---

Residual standard error: 1.547 on 16 degrees of freedom
Multiple R-squared:  0.3751,    Adjusted R-squared:  0.297
F-statistic: 4.802 on 2 and 16 DF,  p-value: 0.02325
```

Ⓡ: Regression: cell-means coding

```
> z1 = rep( c(1,0,0), c(n1,n2,n3) )
> z2 = rep( c(0,1,0), c(n1,n2,n3) )
> z3 = rep( c(0,0,1), c(n1,n2,n3) )

> # under H0
> mylm0 = lm( Y ~ 1 )
> anova(mylm0)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 18 61.307   3.406

> # under H1 (Note: no-intercept model)
> mylm1 = lm( Y ~ 0 + z1 + z2 + z3 )
> anova(mylm1)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
z1         1 4760.2  4760.2  1988.1 < 2.2e-16 ***
z2         1 5237.8  5237.8  2187.5 < 2.2e-16 ***
z3         1 3623.4  3623.4  1513.3 < 2.2e-16 ***
Residuals 16   38.3     2.4
---

> # mu.hat and tau.hat
> mui = coef(mylm1)
> tau = mui - mean(Y)
> cbind(mui, tau)
        mui        tau
z1 28.16667  1.4140351
z2 25.58750 -1.1651316
z3 26.92000  0.1673684

> # Difference between H0 and H1  (Note: anova() function)
> anova(mylm0, mylm1)
Analysis of Variance Table
Model 1: Y ~ 1
Model 2: Y ~ 0 + z1 + z2 + z3
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     18 61.307
2     16 38.310  2    22.997 4.8023 0.02325 *
---
```

Ⓡ: Regression: effect coding

```
> # Wrong Effect coding (factor-effect model)
> z1 = rep( c(1,0,-1), c(n1,n2,n3) ) # This is only for balanced sample
> z2 = rep( c(0,1,-1), c(n1,n2,n3) ) # This is only for balanced sample

> # Correct Effect coding
> z1 = rep( c(1,0,-n1/n3), c(n1,n2,n3) )
> z2 = rep( c(0,1,-n2/n3), c(n1,n2,n3) )

> # under H0
> mylm0 = lm( Y ~ 1 )
> anova(mylm0)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 18 61.307   3.406

> # under H1 (note: with intercept)
> mylm1 = lm( Y ~ 1 + z1 + z2)
> mylm1   # beta0 is a grand mean.
Call:
lm(formula = Y ~ 1 + z1 + z2)
Coefficients:
(Intercept)           z1           z2
     26.753        1.414       -1.165

> anova(mylm1)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value  Pr(>F)
z1         1  4.239  4.2387  1.7703 0.20200
z2         1 18.759 18.7586  7.8344 0.01287 *
Residuals 16 38.310  2.3944
---

> mu0 = coef(mylm1)[1]
> c(mu0, mean(Y))
(Intercept)
   26.75263    26.75263
> tau = c( coef(mylm1)[-1], -sum(c(n1,n2)*coef(mylm1)[-1])/n3 )
> mui = c(mu0+tau)
> cbind(mui, tau)
        mui        tau
z1 28.16667  1.4140351
z2 25.58750 -1.1651316
   26.92000  0.1673684

> # Difference between H0 and H1
> # Note: ANOVA decomposition is the same as the cell-means coding
> anova(mylm0, mylm1)
Analysis of Variance Table
Model 1: Y ~ 1
Model 2: Y ~ 1 + z1 + z2
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     18 61.307
2     16 38.310  2    22.997 4.8023 0.02325 *
```

∥

Example 3. Revisit Example 2. In Section 1.2, we considered the confidence interval of μ_i. One may obtain the confidence interval using `t.test()` function in R language. But, this is *not* recommended because it does not use all the samples to estimate σ^2.

Again, the interval estimation of μ_i with $100(1-\alpha)\%$ coverage is given by

$$\overline{Y}_{i\bullet} \ \pm \ t(1-\tfrac{\alpha}{2}; N-r)\sqrt{\frac{\text{MSE}}{n_i}},$$

Note that the degrees of freedom is $N-r$ when all the samples are used. On the other hand, the interval estimation of μ_i using the ith sample only is give by

$$\overline{Y}_{i\bullet} \ \pm \ t(1-\tfrac{\alpha}{2}; n_i-1)\sqrt{\frac{S_i^2}{n_i}},$$

As an illustration, we obtain the 95% confidence interval of μ_1 as below.

Ⓡ: Reading Data

```
> Y1 = c(25.7, 27.2, 29.9, 28.5, 29.4, 28.3)          # Blue
> Y2 = c(26.8, 27.9, 23.7, 25, 26.3, 24.8, 25.7, 24.5) # Brown
> Y3 = c(26.4, 24.2, 28.0, 26.9, 29.1)                 # Green
> Y = c(Y1, Y2, Y3)
> n1 = length(Y1); n2 = length(Y2); n3 = length(Y3)    # no. of subjects
```

Ⓡ: CI from `t.test()`

```
> # Bad CI
> t.test(Y1)
        One Sample t-test
data:  Y1
t = 45.154, df = 5, p-value = 1.006e-07
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 26.56317 29.77016
sample estimates:
mean of x
 28.16667
```

Ⓡ: CI using S_1 only.

```
> # Same as CI from t.test()
> a = 0.05
> D = qt(1-a/2,df=n1-1) * sqrt(var(Y1)/n1)
> c(mean(Y1)-D, mean(Y1)+D)
[1] 26.56317 29.77016
```

Ⓡ: CI using MSE (better).

```
> # Better CI
> a = 0.05; df = n1+n2+n3-3
> SSE = sum((Y1-mean(Y1))^2) + sum((Y2-mean(Y2))^2) + sum((Y3-mean(Y3))^2)
> MSE = SSE / df
> D = qt(1-a/2,df=df) * sqrt(MSE/n1)
> c(mean(Y1)-D, mean(Y1)+D)
[1] 26.82749 29.50584
```

Again, the $100(1-\alpha)\%$ confidence interval of $\mu_\ell - \mu_m$ is given by

$$\overline{Y}_{\ell\bullet} - \overline{Y}_{m\bullet} \ \pm \ t(1-\tfrac{\alpha}{2}; N-r)\sqrt{\text{MSE}\left(\frac{1}{n_\ell}+\frac{1}{n_m}\right)}.$$

On the other hand, based on the two-sample t-test statistic, the confidence interval of $\mu_\ell - \mu_m$ is given by

$$\overline{Y}_{\ell\bullet} - \overline{Y}_{m\bullet} \ \pm\ t(1-\tfrac{\alpha}{2}; n_\ell + n_m - 2)\sqrt{S_p^2\left(\frac{1}{n_\ell} + \frac{1}{n_m}\right)},$$

which, however, is not recommended because it S_p^2 does not incorporate all the samples provided. As an illustration, we obtain the 95% confidence interval of $\mu_1 - \mu_2$ as below.

Ⓡ: CI from `t.test()`

```
> # Bad CI
> t.test(Y1,Y2, var.equal=TRUE)

        Two Sample t-test
data:  Y1 and Y2
t = 3.3272, df = 12, p-value = 0.006029
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.8902227 4.2681106
sample estimates:
mean of x mean of y
 28.16667  25.58750
```

Ⓡ: CI using S_p^2 (with two samples).

```
> # Same as CI from t.test()
> a = 0.05
> Sp2 = ((n1-1)*var(Y1)+(n2-1)*var(Y2))/(n1+n2-2)
> D = qt(1-a/2,df=n1+n2-2) * sqrt(Sp2*(1/n1+1/n2))
> c(mean(Y1)-mean(Y2)-D, mean(Y1)-mean(Y2)+D)
[1] 0.8902227 4.2681106
```

Ⓡ: CI using MSE (better).

```
> a = 0.05; df = n1+n2+n3-3
> SSE = sum((Y1-mean(Y1))^2) + sum((Y2-mean(Y2))^2) + sum((Y3-mean(Y3))^2)
> MSE = SSE / df
> D = qt(1-a/2,df=df) * sqrt(MSE*(1/n1+1/n2))
> c(mean(Y1)-mean(Y2)-D, mean(Y1)-mean(Y2)+D)
[1] 0.8076044 4.3507289
```

‖

6 Comparisons among treatment means

6.1 Contrasts

When we compare the treatment means, the idea of a contrast is widely used because we can compare the treatment means using them. As an illustration, suppose that one would like to test the hypothesis

$$H_0 : \mu_1 = \mu_2 \quad \text{versus} \quad H_1 : \mu_1 \neq \mu_2,$$

which is equivalent to

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 \neq 0.$$

Thus, this hypothesis testing can be carried out by using a linear combination of the parameters. This linear combination is called a contrast.

Definition 2. We denote $\boldsymbol{\theta} = (\mu_1, \mu_2, \ldots, \mu_r)$ be a set of parameters (or statistics) and $\mathbf{a} = (a_1, a_2, \ldots, a_r)$ be a collection of known constants with its sum begin zero, that is,

$$\sum_{i=1}^{r} a_i = 0.$$

Then $\mathbf{a} = (a_1, a_2, \ldots, a_r)$ is called *contrast constants* and its linear combination below is called a *contrast*

$$\Gamma = \sum_{i=1}^{r} a_i \mu_i.$$

For example, with $\mathbf{a} = (1, -1, 0, \ldots, 0)$, we have

$$\Gamma = \sum_{i=1}^{r} a_i \mu_i = \mu_1 - \mu_2.$$

Thus, the hypothesis testing $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ can be rewritten as a contrast as below

$$H_0 : \Gamma = 0 \quad \text{versus} \quad H_1 : \Gamma \neq 0.$$

Definition 3. Let $\overline{Y}_{i\bullet}$ be the estimator of the i treatment μ_i based on n_i observations and $\mathbf{a} = (a_1, a_2, \ldots, a_r)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_r)$ be contrast constants satisfying

$$\sum_{i=1}^{r} \frac{a_i b_i}{n_i} = 0.$$

Then two contrasts below are called *orthogonal*

$$\Gamma_{\mathbf{a}} = \sum_{i=1}^{r} a_i \mu_i \quad \text{and} \quad \Gamma_{\mathbf{b}} = \sum_{i=1}^{r} b_i \mu_i.$$

Remark 7. Some textbooks state that $\Gamma_{\mathbf{a}}$ and $\Gamma_{\mathbf{b}}$ are *orthogonal* especially when $\sum_{i=1}^{r} a_i b_i = 0$ and *uncorrelated* when $\sum_{i=1}^{r} a_i b_i / n_i = 0$. Of course, if the balanced design is used (that is, $n_1 = n_2 = \cdots = n_r = n$), then the uncorrelation condition becomes $\sum_{i=1}^{r} a_i b_i = 0$. △

6.2 Inferences regarding contrasts

We assumed that $Y_{ij} = \mu_i + \epsilon_{ij}$, where $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, $i = 1, 2, \ldots, r$, and $j = 1, 2, \ldots, n_i$ so that

$$\overline{Y}_{i\bullet} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \sim N\left(\mu_i, \frac{\sigma^2}{n_i}\right). \tag{29}$$

Thus, for any contrast constants $\mathbf{a} = (a_1, a_2, \ldots, a_r)$, $\sum_{i=1}^r a_i \overline{Y}_{i\bullet}$ is also normally distributed with

$$E\left(\sum_{i=1}^r a_i \overline{Y}_{i\bullet}\right) = \sum_{i=1}^r a_i \mu_i \quad \text{and} \quad \text{Var}\left(\sum_{i=1}^r a_i \overline{Y}_{i\bullet}\right) = \sigma^2 \sum_{i=1}^r \frac{a_i^2}{n_i}.$$

Standardizing $\sum_{i=1}^r a_i \overline{Y}_{i\bullet}$, we have

$$\frac{\sum_{i=1}^r a_i \overline{Y}_{i\bullet} - \sum_{i=1}^r a_i \mu_i}{\sqrt{\sigma^2 \sum_{i=1}^r a_i^2 / n_i}} \sim N(0, 1). \tag{30}$$

Note that (30) defines a pivot but it includes a nuisance parameter σ^2.

Lemma 4. *Under the assumption that $Y_{ij} = \mu_i + \epsilon_{ij}$ with $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, $\sum_{i=1}^r a_i \overline{Y}_{i\bullet} - \sum_{i=1}^r a_i \mu_i$ and SSE in (19) are independent.*

Proof. The proof is very similar to that of Lemma 2 (c). □

Recall that $\text{SSE}/\sigma^2 \sim \chi^2(N - r)$ from (19). Thus, using Lemma 4, we can Studentize (30) as

$$\frac{\frac{\sum_{i=1}^r a_i \overline{Y}_{i\bullet} - \sum_{i=1}^r a_i \mu_i}{\sqrt{\sigma^2 \sum_{i=1}^r a_i^2 / n_i}}}{\sqrt{\frac{1}{\sigma^2} \cdot \text{SSE}/(N - r)}} = \frac{\sum_{i=1}^r a_i \overline{Y}_{i\bullet} - \sum_{i=1}^r a_i \mu_i}{\sqrt{\text{MSE} \sum_{i=1}^r a_i^2 / n_i}} \sim t(\text{df} = N - r). \tag{31}$$

Note that MSE can be viewed as an extended version of the pooled sample variance because

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + \cdots + (n_r - 1)S_r^2}{(n_1 - 1) + (n_2 - 1) + \cdots + (n_r - 1)} = \frac{\text{SSE}}{N - r},$$

where $S_i^2 = \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i\bullet})^2 / (n_i - 1)$ is the sample variance of the ith sample. One can perform the hypothesis testing of

$$H_0 : \sum_{i=1}^r a_i \mu_i = 0 \text{ versus } H_1 : \sum_{i=1}^r a_i \mu_i \neq 0$$

by rejecting H_0 if

$$\left| \frac{\sum_{i=1}^r a_i \overline{Y}_{i\bullet}}{\sqrt{\text{MSE} \sum_{i=1}^r a_i^2/n_i}} \right| > t(1-\tfrac{\alpha}{2}; N-r)$$

at the significance level of α.

Both (30) and (31) define a pivot, but the former includes a nuisance parameter. Thus, we invert (31) for the contrast $\sum_{i=1}^r a_i \mu_i$ to obtain an interval estimator whose endpoints with $100(1-\alpha)\%$ confidence level are given by

$$\sum_{i=1}^r a_i \overline{Y}_{i\bullet} \ \pm \ t(1-\tfrac{\alpha}{2}; N-r) \cdot \sqrt{\text{MSE} \sum_{i=1}^r \frac{a_i^2}{n_i}}. \tag{32}$$

It should be noted that the endpoints in (14) are easily obtained from (32) with $a_\ell = 1$, $a_m = -1$ and $a_i = 0$ for $i \neq \ell, m$.

Lemma 5. *Under the assumption that $Y_{ij} = \mu_i + \epsilon_{ij}$ with $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, we have*

$$\text{Cov}\left(\sum_{i=1}^r a_i \overline{Y}_{i\bullet}, \sum_{j=1}^r b_j \overline{Y}_{j\bullet} \right) = \sigma^2 \sum_{i=1}^r \frac{a_i b_i}{n_i}.$$

Proof. It is immediate from Lemma 1 that

$$\text{Cov}\left(\sum_{i=1}^r a_i \overline{Y}_{i\bullet}, \sum_{j=1}^r b_j \overline{Y}_{j\bullet} \right) = \sum_{i=1}^r \sum_{j=1}^r a_i b_j \text{Cov}\left(\overline{Y}_{i\bullet}, \overline{Y}_{j\bullet} \right)$$

Since $\overline{Y}_{i\bullet}$ and $\overline{Y}_{j\bullet}$ are independent for $i \neq j$, the above becomes

$$\sum_{i=1}^r a_i b_i \text{Cov}\left(\overline{Y}_{i\bullet}, \overline{Y}_{i\bullet} \right) = \sum_{i=1}^r a_i b_i \text{Var}(\overline{Y}_{i\bullet}) = \sum_{i=1}^r a_i b_i \frac{\sigma^2}{n_i},$$

which completes the proof. □

Hence, the contrasts are uncorrelated if they are orthogonal by Definition 3.

Definition 4. For a contrast $\Gamma = \sum_{i=1}^r a_i \mu_i$ with the estimate $\hat{\Gamma} = \sum_{i=1}^r a_i \overline{Y}_{i\bullet}$, the term

$$\frac{\left(\sum_{i=1}^r a_i \overline{Y}_{i\bullet} \right)^2}{\sum_{i=1}^r a_i^2/n_i}$$

is called the *contrast sum of squares* due to the contrast $\hat{\Gamma}$.

Note that it is immediate from (30) that under $H_0 : \sum_{i=1}^r a_i \mu_i = 0$

$$\frac{1}{\sigma^2} \frac{\left(\sum_{i=1}^r a_i \overline{Y}_{i\bullet}\right)^2}{\sum_{i=1}^r a_i^2/n_i} \sim \chi^2(1).$$

In general, if the sum of squares has a chi-squared distribution with ν degrees of freedom, we can decompose it into ν independent chi-squared random variables with all one degree of freedom using orthogonal (uncorrelated) contrasts. Thus, the treatment sum of squares in (15) can be decomposed into $r-1$ contrast sums of squares with uncorrelated contrasts.

If there are r treatments, we can find $r-1$ orthogonal (uncorrelated) sets of contrast constants. Denote these sets by $\mathbf{a}^{(\ell)} = (a_1^{(\ell)}, a_2^{(\ell)}, \ldots, a_r^{(\ell)})$ for $\ell = 1, 2, \ldots, r-1$ which satisfy Definition 3, that is,

$$\sum_{i=1}^r \frac{a_i^{(\ell)} a_i^{(\ell')}}{n_i} = 0,$$

for all $\ell \neq \ell'$. Note that $\left[\mathbf{1}_r, \mathbf{a}^{(1)'}, \mathbf{a}^{(2)'}, \ldots, \mathbf{a}^{(r-1)'}\right]$ constitute a $r \times r$ square matrix of full rank r, where $\mathbf{1}_r$ is a r-dimensional vector with all the elements being ones and we also have $\mathbf{1}_r' \mathbf{a}^{(\ell)} = 0$ for $\ell = 1, 2, \ldots, r-1$. Then we can decompose SStr in (15) into contrast sums of squares as follows

$$\begin{aligned}\sum_{i=1}^r n_i(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2 &= \frac{\left(\sum_{i=1}^r a_i^{(1)} \overline{Y}_{i\bullet}\right)^2}{\sum_{i=1}^r \left(a_i^{(1)}\right)^2/n_i} + \frac{\left(\sum_{i=1}^r a_i^{(2)} \overline{Y}_{i\bullet}\right)^2}{\sum_{i=1}^r \left(a_i^{(2)}\right)^2/n_i} + \cdots + \frac{\left(\sum_{i=1}^r a_i^{(r-1)} \overline{Y}_{i\bullet}\right)^2}{\sum_{i=1}^r \left(a_i^{(r-1)}\right)^2/n_i} \\ &= \sum_{\ell=1}^{r-1} \frac{\left(\sum_{i=1}^r a_i^{(\ell)} \overline{Y}_{i\bullet}\right)^2}{\sum_{i=1}^r \left(a_i^{(\ell)}\right)^2/n_i}. \qquad (33)\end{aligned}$$

Example 4. As a real-data example, we consider Example 3.4 of the textbook by Kim (2014).

Ⓡ: Reading Data

```
> Y1 = c(2.4, 2.7, 3.1, 3.1)
> Y2 = c(0.7, 1.6, 1.7, 1.8)
> Y3 = c(2.4, 3.1, 5.4, 6.1)
> Y4 = c(0.3, 0.3, 2.4, 2.7)
> Y5 = c(0.5, 0.9, 1.4, 2.0)
> Y = c(Y1, Y2, Y3, Y4, Y5)
> n = length(Y1)
> treat = factor( rep(LETTERS[1:5], rep(n,5)) )
```

Ⓡ: Calculating SStr

```
# Calculating SStr with aov() function
> myaov = aov(Y~treat)
> summary(myaov)
```

```
            Df Sum Sq Mean Sq F value  Pr(>F)
treat        4  27.01   6.752   5.966 0.00444 **
Residuals   15  16.98   1.132
---
> SStr = summary(myaov)[[1]][["Sum Sq"]][1]
> SStr
[1] 27.007

# Calculating SStr manually
> Yi.bar = c(mean(Y1), mean(Y2), mean(Y3), mean(Y4), mean(Y5))
> Ybarbar = mean(Y)
> SStr = sum( n*(Yi.bar-Ybarbar)^2 )
> SStr
[1] 27.007
```

Ⓡ: Decomposition of SStr with Contrasts

```
> # ----------------------------------------------
> # Orthogonal contrasts (manual contrasts)
> # ----------------------------------------------
> a1 = c(1, -1/4, -1/4, -1/4, -1/4)
> a2 = c(0,    1, -1/3, -1/3, -1/3)
> a3 = c(0,    0,    1, -1/2, -1/2)
> a4 = c(0,    0,    0,    1,   -1)
> contrasts(treat) = cbind(a1,a2,a3,a4)
> contrasts(treat)
     a1         a2   a3 a4
A  1.00  0.0000000  0.0  0
B -0.25  1.0000000  0.0  0
C -0.25 -0.3333333  1.0  0
D -0.25 -0.3333333 -0.5  1
E -0.25 -0.3333333 -0.5 -1

> # Check the validity of contrasts (sum=0)
> cnt = contrasts(treat)
> apply(cnt, 2, sum)
          a1           a2           a3           a4
0.000000e+00 5.551115e-17 0.000000e+00 0.000000e+00

> # Check the orthogonality
> crossprod(cnt)
              a1            a2  a3 a4
a1  1.250000e+00 -2.775558e-17 0.0  0
a2 -2.775558e-17  1.333333e+00 0.0  0
a3  0.000000e+00  0.000000e+00 1.5  0
a4  0.000000e+00  0.000000e+00 0.0  2

> # Check the decomposition of SStr
> myaov = aov( Y ~ treat )
> summary.lm(myaov)
Call:
aov(formula = Y ~ treat)
Residuals:
    Min      1Q  Median      3Q     Max
-1.8500 -0.7125  0.1750  0.4625  1.8500
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.2300     0.2379   9.375 1.16e-07 ***
treata1       0.5950     0.4757   1.251 0.230213
treata2      -0.6313     0.4606  -1.370 0.190729
treata3       1.9583     0.4343   4.509 0.000416 ***
treata4       0.1125     0.3761   0.299 0.768957
---
Residual standard error: 1.064 on 15 degrees of freedom
Multiple R-squared:  0.614,    Adjusted R-squared:  0.5111
F-statistic: 5.966 on 4 and 15 DF,  p-value: 0.004442

> MSE = summary(myaov)[[1]][["Mean Sq"]][2]
> MSE
[1] 1.131667
> t.stat = summary.lm(myaov)[[4]][-1,3]
> SStr.for.contrast = t.stat^2 * MSE
> SStr.for.contrast
  treata1   treata2   treata3   treata4
 1.770125  2.125208 23.010417  0.101250
> sum(SStr.for.contrast)  # Note: SStr = 27.007
[1] 27.007
```

We have used the following orthogonal contrast constants: $\mathbf{a}^{(1)} = (1, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4})$, $\mathbf{a}^{(2)} = (0, 1, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3})$, $\mathbf{a}^{(3)} = (0, 0, 1, -\frac{1}{2}, -\frac{1}{2})$, and $\mathbf{a}^{(4)} = (0, 0, 0, 1, -1)$. As shown in the R program above, the treatment sum of squares was decomposed as follows

$$27.007 \text{ (SStr)} = 1.770125 + 2.125208 + 23.010417 + 0.101250.$$

The R program reports only t-test statistic

$$T_\ell = \frac{\sum_{i=1}^{r} a_i^{(\ell)} \overline{Y}_{i\bullet}}{\sqrt{\text{MSE} \sum_{i=1}^{r} \left(a_i^{(\ell)}\right)^2 / n_i}} \tag{34}$$

which is from (31) under $H_0 : \sum_{i=1}^{r} a_i^{(\ell)} \mu_i = 0$. Along with the above orthogonal contrast constants, we can test the below

$$\begin{aligned}
&T_1 \text{ for testing } H_0^{(1)} : \ \mu_1 = (\mu_2 + \mu_3 + \mu_4 + \mu_5)/4, \\
&T_2 \text{ for testing } H_0^{(2)} : \ \mu_2 = (\mu_3 + \mu_4 + \mu_5)/3, \\
&T_3 \text{ for testing } H_0^{(3)} : \ \mu_3 = (\mu_4 + \mu_5)/2, \\
&T_4 \text{ for testing } H_0^{(4)} : \ \mu_4 = \mu_5.
\end{aligned}$$

Ⓡ: Decomposition of SStr with Contrasts (using t-test statistics)

```
> # Check the t-test statistics
> T1 = sum(a1*Yi.bar) / sqrt(MSE* sum(a1^2/n) )
> T2 = sum(a2*Yi.bar) / sqrt(MSE* sum(a2^2/n) )
> T3 = sum(a3*Yi.bar) / sqrt(MSE* sum(a3^2/n) )
> T4 = sum(a4*Yi.bar) / sqrt(MSE* sum(a4^2/n) )
> c(T1,T2,T3,T4)
[1]  1.250670 -1.370382  4.509236  0.299115
> c(T1,T2,T3,T4)^2 * MSE
[1]  1.770125  2.125208 23.010417  0.101250
> sum( c(T1,T2,T3,T4)^2 * MSE )
[1] 27.007
```

We calculated the t-test statistics manually as above, but the `summary.lm()` function reported them as shown earlier. Since it is easily seen from (34) that

$$\frac{\left(\sum_{i=1}^{r} a_i \overline{Y}_{i\bullet}\right)^2}{\sum_{i=1}^{r} a_i^2 / n_i} = T^2 \cdot \text{MSE},$$

we can calculate the contrast sum of squares using the t-test statistic and decompose SStr as below.

$$\frac{\left(\sum_{i=1}^{r} a_i^{(1)}\overline{Y}_{i\bullet}\right)^2}{\sum_{i=1}^{r}\left(a_i^{(1)}\right)^2/n_i} = T_1^2 \cdot \text{MSE} = 1.250670^2 \times 1.131667 = 1.770125$$

$$\frac{\left(\sum_{i=1}^{r} a_i^{(2)}\overline{Y}_{i\bullet}\right)^2}{\sum_{i=1}^{r}\left(a_i^{(2)}\right)^2/n_i} = T_2^2 \cdot \text{MSE} = (-1.370382)^2 \times 1.131667 = 2.125208$$

$$\frac{\left(\sum_{i=1}^{r} a_i^{(3)}\overline{Y}_{i\bullet}\right)^2}{\sum_{i=1}^{r}\left(a_i^{(3)}\right)^2/n_i} = T_3^2 \cdot \text{MSE} = 4.509236^2 \times 1.131667 = 23.010417$$

$$\frac{\left(\sum_{i=1}^{r} a_i^{(4)}\overline{Y}_{i\bullet}\right)^2}{\sum_{i=1}^{r}\left(a_i^{(4)}\right)^2/n_i} = T_4^2 \cdot \text{MSE} = 0.299115^2 \times 1.131667 = 0.101250$$

‖

Contrasts are not uniquely determined. Thus, they should be determined considering the statistical hypothesis of interest. There are several well-known contrasts such as Helmert contrasts (`contr.helmert()`), polynomial contrasts (`contr.poly()`), sum-to-zero contrasts (`contr.sum()`), etc. The sum-to-zero contrasts are similar to effect coding in regression so they are also called effect-coding contrasts. The polynomial contrasts are widely used for testing polynomial patterns in treatment means with more than two treatments (e.g., linear, quadratic, cubic, quartic, etc.). The Helmert and polynomial contrasts are orthogonal contrasts while sum-to-zero contrasts are not. Note that R language also provides the other functions for contrasts such as `contr.treatment()` and `contr.SAS()`, but these are not perpendicular to the intercept vector $\mathbf{1}_r$ which implies that $\mathbf{1}_r'\mathbf{a} = \sum_{i=1}^{r} a_i \neq 0$. Thus, `contr.treatment()` and `contr.SAS()` can not be considered as contrasts in this course!

Example 5. Revisit Example 4. In this example, we analyze the data again with the Helmert, polynomial and sum-to-zero contrasts.

Ⓡ: Reading Data

```
> Y1 = c(2.4, 2.7, 3.1, 3.1)
> Y2 = c(0.7, 1.6, 1.7, 1.8)
> Y3 = c(2.4, 3.1, 5.4, 6.1)
> Y4 = c(0.3, 0.3, 2.4, 2.7)
> Y5 = c(0.5, 0.9, 1.4, 2.0)
> Y = c(Y1, Y2, Y3, Y4, Y5)
> n = length(Y1)
> treat = factor( rep(LETTERS[1:5], rep(n,5)) )
```

Ⓡ: (1) contrasts using `contr.helmert()` function

```
> contrasts(treat) = contr.helmert(5)
> contrasts(treat)
  [,1] [,2] [,3] [,4]
A   -1   -1   -1   -1
```

```
B    1   -1   -1   -1
C    0    2   -1   -1
D    0    0    3   -1
E    0    0    0    4
>
> # Check the validity of contrasts (sum=0)
> cnt = contrasts(treat)
> apply(cnt, 2, sum)
[1] 0 0 0 0


> # Check the orthogonality
> crossprod(cnt)
     [,1] [,2] [,3] [,4]
[1,]    2    0    0    0
[2,]    0    6    0    0
[3,]    0    0   12    0
[4,]    0    0    0   20

> # Check the decomposition of SStr
> myaov = aov( Y ~ treat )
> summary.lm(myaov)
Call:
aov(formula = Y ~ treat)
Residuals:
    Min      1Q  Median      3Q     Max
-1.8500 -0.7125  0.1750  0.4625  1.8500
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.2300     0.2379   9.375 1.16e-07 ***
treat1       -0.6875     0.3761  -1.828  0.08752 .
treat2        0.7042     0.2171   3.243  0.00546 **
treat3       -0.3542     0.1535  -2.307  0.03577 *
treat4       -0.2575     0.1189  -2.165  0.04692 *
---
Residual standard error: 1.064 on 15 degrees of freedom
Multiple R-squared:  0.614,     Adjusted R-squared:  0.5111
F-statistic: 5.966 on 4 and 15 DF,  p-value: 0.004442

> MSE = summary(myaov)[[1]][["Mean Sq"]][2]
> t.stat = summary.lm(myaov)[[4]][-1,3]
> SStr.for.contrast = t.stat^2 * MSE
> SStr.for.contrast
   treat1    treat2    treat3    treat4
 3.781250 11.900417  6.020833  5.304500
> sum(SStr.for.contrast)  ## Note: SStr = 27.007
[1] 27.007
```

Ⓡ: (2) contrasts using contr.poly() function

```
> contrasts(treat) = contr.poly(5)
> contrasts(treat)
          .L         .Q            .C         ^4
A -0.6324555  0.5345225 -3.162278e-01  0.1195229
B -0.3162278 -0.2672612  6.324555e-01 -0.4780914
C  0.0000000 -0.5345225 -4.095972e-16  0.7171372
D  0.3162278 -0.2672612 -6.324555e-01 -0.4780914
E  0.6324555  0.5345225  3.162278e-01  0.1195229

> # Check the validity of contrasts (sum=0)
> cnt = contrasts(treat)
> apply(cnt, 2, sum)
          .L           .Q           .C           ^4
0.000000e+00 1.110223e-16 9.001589e-17 6.938894e-17

> # Check the orthogonality
> crossprod(cnt)
              .L            .Q           .C            ^4
.L  1.000000e+00 -1.110223e-16 5.551115e-17 -2.081668e-16
.Q -1.110223e-16  1.000000e+00 8.326673e-17 -1.665335e-16
.C  5.551115e-17  8.326673e-17 1.000000e+00  1.387779e-17
```

```
^4 -2.081668e-16 -1.665335e-16 1.387779e-17  1.000000e+00

> # Check the decomposition of SStr
> myaov = aov( Y ~ treat )
> summary.lm(myaov)
Call:
aov(formula = Y ~ treat)
Residuals:
    Min      1Q  Median      3Q     Max
-1.8500 -0.7125  0.1750  0.4625  1.8500
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.2300     0.2379   9.375 1.16e-07 ***
treat.L      -1.0356     0.5319  -1.947  0.07050 .
treat.Q      -0.8886     0.5319  -1.671  0.11551
treat.C      -0.4981     0.5319  -0.936  0.36391
treat^4       2.1544     0.5319   4.050  0.00105 **
---
Residual standard error: 1.064 on 15 degrees of freedom
Multiple R-squared:  0.614,     Adjusted R-squared:  0.5111
F-statistic: 5.966 on 4 and 15 DF,  p-value: 0.004442

> MSE = summary(myaov)[[1]][["Mean Sq"]][2]
> t.stat = summary.lm(myaov)[[4]][-1,3]
> SStr.for.contrast = t.stat^2 * MSE
> SStr.for.contrast
 treat.L  treat.Q  treat.C  treat^4
 4.29025  3.15875  0.99225 18.56575
> sum(SStr.for.contrast)  ## Note: SStr = 27.007
[1] 27.007
```

Ⓡ: (3) contrasts using contr.sum() function

```
> contrasts(treat) = contr.sum(5)
> contrasts(treat)
  [,1] [,2] [,3] [,4]
A    1    0    0    0
B    0    1    0    0
C    0    0    1    0
D    0    0    0    1
E   -1   -1   -1   -1
>
> # Check the validity of contrasts (sum=0)
> cnt = contrasts(treat)
> apply(cnt, 2, sum)
[1] 0 0 0 0


> # Check the orthogonality (not orthogonal as shown below)
> crossprod(cnt)
     [,1] [,2] [,3] [,4]
[1,]    2    1    1    1
[2,]    1    2    1    1
[3,]    1    1    2    1
[4,]    1    1    1    2

> # Check the decomposition of SStr
> myaov = aov( Y ~ treat )
> summary.lm(myaov)
Call:
aov(formula = Y ~ treat)
Residuals:
    Min      1Q  Median      3Q     Max
-1.8500 -0.7125  0.1750  0.4625  1.8500
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.2300     0.2379   9.375 1.16e-07 ***
treat1        0.5950     0.4757   1.251 0.230213
treat2       -0.7800     0.4757  -1.640 0.121901
treat3        2.0200     0.4757   4.246 0.000704 ***
treat4       -0.8050     0.4757  -1.692 0.111293
```

```
---
Residual standard error: 1.064 on 15 degrees of freedom
Multiple R-squared:  0.614,     Adjusted R-squared:  0.5111
F-statistic: 5.966 on 4 and 15 DF,  p-value: 0.004442

> MSE = summary(myaov)[[1]][["Mean Sq"]][2]
> t.stat = summary.lm(myaov)[[4]][-1,3]
> SStr.for.contrast = t.stat^2 * MSE
> SStr.for.contrast
   treat1    treat2    treat3    treat4
 1.770125  3.042000 20.402000  3.240125
> sum(SStr.for.contrast)  ## Note: SStr = 27.007
[1] 28.45425
```

∥

6.3 Contrasts – different version

When samples are unbalanced, great care should be taken. Contrast constants and their orthogonality are defined in Definitions 2 and 3. However, some textbooks use a different version as below.

Definition 5. We denote $\boldsymbol{\theta} = (\mu_1, \mu_2, \ldots, \mu_r)$ be a set of parameters (or statistics) and $\mathbf{c} = (c_1, c_2, \ldots, c_r)$ be a collection of known constants. We assume that μ_i is estimated with a sample of size n_i. Then $\mathbf{c} = (c_1, c_2, \ldots, c_r)$ is called *contrast constants* and its linear combination below is called a *contrast*

$$\Gamma = \sum_{i=1}^{r} n_i c_i \mu_i,$$

where $\sum_{i=1}^{r} n_i c_i = 0$.

Remark 8. We can notice the following.

1. The condition $\sum_{i=1}^{r} n_i c_i = 0$ is similar to the condition in (3).
2. The textbook uses $\sum_{i=1}^{r} c_i = 0$ (Definition 2) condition in 정의 3.2 on Page 68.
3. It is easily seen that a_i in Definition 2 and c_i in Definition 5 has the relation $a_i = n_i c_i$.

△

Definition 6. Let $\overline{Y}_{i\bullet}$ be the estimator of the i treatment μ_i based on n_i observations and $\mathbf{c} = (c_1, c_2, \ldots, c_r)$ and $\mathbf{d} = (d_1, d_2, \ldots, d_r)$ be contrast constants satisfying

$$\sum_{i=1}^{r} n_i c_i d_i = 0.$$

Then two contrasts below are called *orthogonal*

$$\Gamma_{\mathbf{a}} = \sum_{i=1}^{r} n_i c_i \mu_i \quad \text{and} \quad \Gamma_{\mathbf{b}} = \sum_{i=1}^{r} n_i d_i \mu_i.$$

Remark 9. It should be noted that the textbook uses $\sum_{i=1}^{r} n_i c_i d_i = 0$ (Definition 6) in 정의 3.3 on Page 70. △

Lemma 6. *Under the assumption that* $Y_{ij} = \mu_i + \epsilon_{ij}$ *with* $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, *we have*

$$\mathrm{Cov}\left(\sum_{i=1}^{r} n_i c_i \overline{Y}_{i\bullet}, \sum_{j=1}^{r} n_i d_j \overline{Y}_{j\bullet}\right) = \sigma^2 \sum_{i=1}^{r} n_i c_i d_i.$$

Proof. The proof is very similar to that of Lemma 5. □

Analogous to (29), we have $n_i \overline{Y}_{i\bullet} \sim N\left(n_i \mu_i, n_i \sigma^2\right)$. Thus, for any contrast constants $\mathbf{c} = (c_1, c_2, \ldots, c_r)$, $\sum_{i=1}^{r} n_i c_i \overline{Y}_{i\bullet}$ is also normally distributed with

$$E\left(\sum_{i=1}^{r} n_i c_i \overline{Y}_{i\bullet}\right) = \sum_{i=1}^{r} n_i c_i \mu_i \quad \text{and} \quad \mathrm{Var}\left(\sum_{i=1}^{r} n_i c_i \overline{Y}_{i\bullet}\right) = \sigma^2 \sum_{i=1}^{r} n_i c_i^2.$$

Standardizing $\sum_{i=1}^{r} n_i c_i \overline{Y}_{i\bullet}$, we have

$$\frac{\sum_{i=1}^{r} n_i c_i \overline{Y}_{i\bullet} - \sum_{i=1}^{r} n_i c_i \mu_i}{\sqrt{\sigma^2 \sum_{i=1}^{r} n_i c_i^2}} \sim N(0, 1). \tag{35}$$

Note that (35) defines a pivot but it includes a nuisance parameter σ^2. As we Studentized (30), we can also do (35) so that we have

$$\frac{\sum_{i=1}^{r} n_i c_i \overline{Y}_{i\bullet} - \sum_{i=1}^{r} n_i c_i \mu_i}{\sqrt{\mathrm{MSE} \sum_{i=1}^{r} n_i c_i^2}} \sim t(\mathrm{df} = N - r).$$

As we decompose the treatment sum of squares (SStr) into $r - 1$ independent random variables with all one degree of freedom with orthogonal (uncorrelated) contrasts in Definition 2, we can also do this decomposition with orthogonal (uncorrelated) contrasts in

Definition 5. Denote these sets by $\mathbf{c}^{(\ell)} = (c_1^{(\ell)}, c_2^{(\ell)}, \dots, c_r^{(\ell)})$ for $\ell = 1, 2, \dots, r-1$ which satisfy Definition 6, that is,

$$\sum_{i=1}^{r} n_i c_i^{(\ell)} c_i^{(\ell')} = 0,$$

for all $\ell \neq \ell'$. Analogous to (33), we can decompose SStr in (15) into contrast sums of squares as follows

$$\sum_{i=1}^{r} n_i (\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2 = \sum_{\ell=1}^{r-1} \frac{\left(\sum_{i=1}^{r} n_i c_i^{(\ell)} \overline{Y}_{i\bullet}\right)^2}{\sum_{i=1}^{r} n_i \left(c_i^{(\ell)}\right)^2}.$$

Example 6. Revisit Example 2. In this example, we calculated SStr $= 22.997 \approx 23$.

Ⓡ: Reading Data

```
> Y1 = c(25.7, 27.2, 29.9, 28.5, 29.4, 28.3)           # Blue
> Y2 = c(26.8, 27.9, 23.7, 25, 26.3, 24.8, 25.7, 24.5) # Brown
> Y3 = c(26.4, 24.2, 28.0, 26.9, 29.1)                 # Green
> Y = c(Y1, Y2, Y3)
> n1 = length(Y1); n2 = length(Y2); n3 = length(Y3)
> ni = c(n1,n2,n3)
```

Ⓡ: SStr, MSE, etc.

```
> # ni and Yi.bar
> Y1bar = mean(Y1); Y2bar = mean(Y2); Y3bar = mean(Y3)
> Yi.bar = c(Y1bar, Y2bar, Y3bar)
> rbind(ni, Yi.bar)
           [,1]    [,2]  [,3]
ni      6.00000  8.0000  5.00
Yi.bar 28.16667 25.5875 26.92
>
> # SStr
> SStr = sum( ni*(Yi.bar-mean(Y))^2 )
> SStr
[1] 22.99729
>
> # SSE
> df = n1+n2+n3-3
> SSE = sum((Y1-Y1bar)^2) + sum((Y2-Y2bar)^2) + sum((Y3-Y3bar)^2)
> MSE = SSE / df
> MSE
[1] 2.39438
```

Ⓡ: (1) Contrasts a_i (correct)

```
> a1 = c(-0.6, 1.1, -0.5);  a2 = c(-1, 0, 1)
> sum(a1)            # zero
[1] 1.110223e-16
> sum(a2)            # zero
[1] 0
> sum( a1*a2 / ni ) # zero
[1] -1.387779e-17
> sum( ni*a1*a2 )    # not zero (does not satisfy the textbook definition)
[1] 1.1
>
> SSa1 = sum( a1*Yi.bar )^2 / sum( a1^2/ni )
> SSa2 = sum( a2*Yi.bar )^2 / sum( a2^2/ni )
> c(SSa1,SSa2, SSa1+SSa2) # Correct decomposition (SStr=22.997)
[1] 18.758618  4.238667 22.997285
>
```

```
> # Double check with t-test statistics
> T1 = sum(a1*Yi.bar) / sqrt( sum(MSE*a1^2*(1/ni)) )
> T2 = sum(a2*Yi.bar) / sqrt( sum(MSE*a2^2*(1/ni)) )
> c(T1^2*MSE, T2^2*MSE, T1^2*MSE+T2^2*MSE)    ## SStr=22.997
[1] 18.758618  4.238667 22.997285
> # Correct decomposition: 18.758618 + 4.238667 = 22.997285 (SStr)
```

Ⓡ: (2) Contrasts c_i from the textbook (wrong)

```
> c1 = c(-0.5, 1.1, -0.6);  c2 = c(-1, 0, 1)
> sum(c1)            # zero (satisfies Definition 3.2 in the textbook)
[1] 1.110223e-16
> sum(c2)            # zero (satisfies Definition 3.2 in the textbook)
[1] 0
> sum( ni*c1*c2 )    # zero (satisfies Definition 3.3 in the textbook)
[1] 0
>
> SSc1 = sum( c1*Yi.bar )^2 / sum( c1^2/ni )
> SSc2 = sum( c2*Yi.bar )^2 / sum( c2^2/ni )
> c(SSc1,SSc2, SSc1+SSc2) # Wrong decomposition (SStr=22.997)
[1] 16.474121  4.238667 20.712787
>
> # Double check with t-test statistics
> T1 = sum(c1*Yi.bar) / sqrt( sum(MSE*c1^2*(1/ni)) )
> T2 = sum(c2*Yi.bar) / sqrt( sum(MSE*c2^2*(1/ni)) )
> c(T1^2*MSE, T2^2*MSE, T1^2*MSE+T2^2*MSE)
[1] 16.474121  4.238667 20.712787
> # Wrong decomposition: 16.474121 + 4.238667 = 20.712787  (not 22.997)
```

Ⓡ: (3) Contrasts c_i from the handout (correct)

```
> a1 = c(-0.6, 1.1, -0.5);  a2 = c(-1.0, 0.0,  1.0)
> c1 = a1/ni;               c2 = a2/ni
> sum(ni*c1)          # zero (satisfies the definition in the handout)
[1] 1.110223e-16
> sum(ni*c2)          # zero (satisfies the definition in the handout)
[1] 0
> sum(ni*c1*c2)       # zero (satisfies the definition in the handout)
[1] -1.387779e-17
>
> SSc1 = sum( c1*ni*Yi.bar )^2 / sum( ni*c1^2 )
> SSc2 = sum( c2*ni*Yi.bar )^2 / sum( ni*c2^2 )
> c(SSc1, SSc2, SSc1+SSc2)
[1] 18.758618  4.238667 22.997285
> # Correct decomposition: 18.758618 + 4.238667 = 22.997285 (SStr)
```

‖

Example 7. We solve Example 6 again using the `aov()` function. As shown in the R programs below, the orthogonal contrasts (a_i) in Definition 3 fail to decompose the SStr. Only the orthogonal contrasts (c_i) satisfying Definitions 5 and 6 work properly when the `aov()` function is used. Note that both a_i or c_i decomposed the SStr successfully in Example 6. However, the contrasts c_i based on $\boxed{\text{정의 3.2}}$ of the textbook fail to decompose the SStr in both Examples 6 and 7.

Ⓡ: Reading Data

```
> Y1 = c(25.7, 27.2, 29.9, 28.5, 29.4, 28.3)           # Blue
> Y2 = c(26.8, 27.9, 23.7, 25, 26.3, 24.8, 25.7, 24.5) # Brown
> Y3 = c(26.4, 24.2, 28.0, 26.9, 29.1)                 # Green
> Y = c(Y1, Y2, Y3)
> n1 = length(Y1); n2 = length(Y2); n3 = length(Y3)
> ni = c(n1,n2,n3)
> color = factor( rep(c("Blue","Brown","Green"), ni) )
```

Ⓡ: (1) Contrasts a_i (wrong)

```
> # Contrasts
> a1 = c(-0.6, 1.1, -0.5);  a2 = c(-1, 0, 1)
> sum(a1)             # zero
[1] 1.110223e-16
> sum(a2)             # zero
[1] 0
> sum( a1*a2 / ni ) # zero
[1] -1.387779e-17

> contrasts(color) = cbind(a1,a2)
> myaov = aov( Y ~ color )
> myaov
Call:
   aov(formula = Y ~ color)
Terms:
                   color Residuals
Sum of Squares  22.99729  38.31008
Deg. of Freedom        2        16

Residual standard error: 1.547378
Estimated effects may be unbalanced

> summary(myaov)
            Df Sum Sq Mean Sq F value Pr(>F)
color        2  23.00  11.499   4.802 0.0232 *
Residuals   16  38.31   2.394
---

> # t-test statistics
> summary.lm(myaov)
Call:
aov(formula = Y ~ color)
Residuals:
    Min      1Q  Median      3Q     Max
-2.7200 -0.8771  0.1125  1.1462  2.3125

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  26.8914     0.3617  74.354   <2e-16 ***
colora1      -1.1854     0.4365  -2.715   0.0153 *
colora2      -0.5641     0.4703  -1.199   0.2478
---
Residual standard error: 1.547 on 16 degrees of freedom
Multiple R-squared:  0.3751,   Adjusted R-squared:  0.297
F-statistic: 4.802 on 2 and 16 DF,  p-value: 0.02325

> # Decomposition
> MSE = summary(myaov)[[1]][["Mean Sq"]][2]
> t.stat = summary.lm(myaov)[[4]][-1,3]
> SStr.for.contrast = t.stat^2 * MSE
> SStr.for.contrast
  colora1   colora2
17.655157  3.444492
> sum(SStr.for.contrast)
[1] 21.09965
> # Wrong decomposition: 17.655157  3.444492 = 21.09965 (not 22.997)
```

Ⓡ: (2) Contrasts c_i from the textbook (wrong)

```
> # Contrasts
> c1 = c(-0.5, 1.1, -0.6);  c2 = c(-1, 0, 1)
> sum(c1)             # zero (satisfies Definition 3.2 in the textbook)
[1] 1.110223e-16
> sum(c2)             # zero (satisfies Definition 3.2 in the textbook)
[1] 0
```

```
> sum( ni*c1*c2 )   # zero (satisfies Definition 3.3 in the textbook)
[1] 0

> contrasts(color) = cbind(c1,c2)
> myaov = aov( Y ~ color )
> myaov
Call:
   aov(formula = Y ~ color)
Terms:
                   color Residuals
Sum of Squares  22.99729  38.31008
Deg. of Freedom        2        16

Residual standard error: 1.547378
Estimated effects may be unbalanced
> summary(myaov)
            Df Sum Sq Mean Sq F value Pr(>F)
color        2  23.00  11.499   4.802 0.0232 *
Residuals   16  38.31   2.394
---

> # t-test statistics
> summary.lm(myaov)
Call:
aov(formula = Y ~ color)
Residuals:
    Min      1Q  Median      3Q     Max
-2.7200 -0.8771  0.1125  1.1462  2.3125
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  26.8914     0.3617  74.354   <2e-16 ***
colorc1      -1.1854     0.4365  -2.715   0.0153 *
colorc2      -0.6826     0.4677  -1.459   0.1638
---
Residual standard error: 1.547 on 16 degrees of freedom
Multiple R-squared:  0.3751,    Adjusted R-squared:  0.297
F-statistic: 4.802 on 2 and 16 DF,  p-value: 0.02325

> # Decomposition
> MSE = summary(myaov)[[1]][["Mean Sq"]][2]
> t.stat = summary.lm(myaov)[[4]][-1,3]
> SStr.for.contrast = t.stat^2 * MSE
> SStr.for.contrast
  colorc1   colorc2
17.655157  5.100057
> sum(SStr.for.contrast)
[1] 22.75521
> # Wrong decomposition: 17.655157 + 5.100057 = 22.75521 (not 22.997)
```

Ⓡ: (3) Contrasts c_i from the handout (correct)

```
> # Try the contrasts below
> a1 = c(-0.6, 1.1, -0.5);  a2 = c(-1.0, 0.0,  1.0)
> c1 = a1/ni;               c2 = a2/ni
> sum(ni*c1)         # zero (satisfies the definition in the handout)
[1] 1.110223e-16
> sum(ni*c2)         # zero (satisfies the definition in the handout)
[1] 0
> sum(ni*c1*c2)      # zero (satisfies the definition in the handout)
[1] -1.387779e-17

> contrasts(color) = cbind(c1,c2)
> myaov = aov( Y ~ color )
> myaov
Call:
   aov(formula = Y ~ color)
Terms:
                   color Residuals
Sum of Squares  22.99729  38.31008
Deg. of Freedom        2        16

```

```
Residual standard error: 1.547378
Estimated effects are balanced
> summary(myaov)
            Df Sum Sq Mean Sq F value Pr(>F)
color        2  23.00  11.499   4.802 0.0232 *
Residuals   16  38.31   2.394
---

> # t-test statistics
> summary.lm(myaov)
Call:
aov(formula = Y ~ color)
Residuals:
    Min      1Q  Median      3Q     Max
-2.7200 -0.8771  0.1125  1.1462  2.3125

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   26.753      0.355  75.361   <2e-16 ***
colorc1       -8.474      3.027  -2.799   0.0129 *
colorc2       -3.400      2.555  -1.331   0.2020
---
Residual standard error: 1.547 on 16 degrees of freedom
Multiple R-squared:  0.3751,    Adjusted R-squared:  0.297
F-statistic: 4.802 on 2 and 16 DF,  p-value: 0.02325

> # Decomposition
> MSE = summary(myaov)[[1]][["Mean Sq"]][2]
> t.stat = summary.lm(myaov)[[4]][-1,3]
> SStr.for.contrast = t.stat^2 * MSE
> SStr.for.contrast
  colorc1   colorc2
18.758618  4.238667
> sum(SStr.for.contrast)
[1] 22.99729
> # Correct decomposition: 18.758618 + 4.238667 = 22.997285 (SStr)
```

‖

6.4 Comparing pairs of treatment means

Using contrasts, we compare any differences in treatment means. However, in many practice, it is very difficult to determine which contrasts are to be used. One appealing idea proposed by R. A. Fisher is comparing only *pairs* of treatment means which can be done with the contrasts $\Gamma = \mu_\ell - \mu_m$ for $\ell \neq m$. The Fisher method compares all the pairs of treatment means with the significance level α for each individual pair. The underlying theory on this method was already investigated in Section 1.2. Under $H_0 : \mu_\ell = \mu_m$, the t-test statistic in (13) becomes

$$\frac{\overline{Y}_{\ell\bullet} - \overline{Y}_{m\bullet}}{\sqrt{\mathrm{MSE}\left(\frac{1}{n_\ell} + \frac{1}{n_m}\right)}} \sim t(\mathrm{df} = N - r).$$

Also, we obtained the endpoints for the $100(1-\alpha)\%$ confidence interval of $\mu_\ell - \mu_m$ in (14). Notice that the deviation from the center of the interval is given by

$$t(1-\tfrac{\alpha}{2}; N-r)\sqrt{\mathrm{MSE}\left(\frac{1}{n_\ell}+\frac{1}{n_m}\right)}.$$

which is called the least significance difference (LSD) and we denote it by $\mathrm{LSD}_{\ell m}$. To perform this method, we simply compare the observed difference between each pair of means to the corresponding LSD. That is, if $|\overline{Y}_{\ell\bullet} - \overline{Y}_{m\bullet}| > \mathrm{LSD}_{\ell m}$, then we can conclude that the treatment means μ_ℓ and μ_m are different.

These days, with the advent of powerful and accessible computers, it is quite easy to calculate p-value. Thus, we also carry out this method by comparing the p-values of the t-test statistic for each pair of means. In general, p-values are preferred because they provide better information than test statistics. Let $T_{\ell m} = (\overline{Y}_{\ell\bullet} - \overline{Y}_{m\bullet})/\sqrt{\mathrm{MSE}(1/n_\ell + 1/n_m)}$. Then the corresponding p-value is calculated as

$$2 \cdot \Big\{1 - F_{N-r}\big(|T_{\ell m}|\big)\Big\},$$

where $F_\nu(\,\cdot\,)$ is the cumulative distribution function of the t-distribution with ν degrees of freedom. Note that $1 - F_\nu(\,\cdot\,)$ is easily calculated with `pt(q, df=`ν`, lower.tail=FALSE)` function in R language.

Remark 10. There are several other methods for this pairwise comparison. A natural question is: *Which method is recommended?* There is no clear answer for this question. Carmer and Swanson (1973) have carried out extensive Monte Carlo simulations whose results show that the Fisher's least significant difference (LSD) method is very effective. But, the Fisher's LSD method can not control the overall error rate (say, at the selected level α) which is also known as *family* or *experiment-wise* error rate. Thus, the overall error rate is needed, refer to other methods such as Scheffé (1953) and Tukey (1953). △

Example 8. We solve Example 6 again to use the Fisher LSD procedure.

Ⓡ: Reading Data

```
> Y1 = c(25.7, 27.2, 29.9, 28.5, 29.4, 28.3)           # Blue
> Y2 = c(26.8, 27.9, 23.7, 25, 26.3, 24.8, 25.7, 24.5) # Brown
> Y3 = c(26.4, 24.2, 28.0, 26.9, 29.1)                 # Green
> Y = c(Y1, Y2, Y3)
> n1 = length(Y1); n2 = length(Y2); n3 = length(Y3)
> df = n1+n2+n3-3
```

Ⓡ: Using t-test statistics

```
> # Don't use: t.test(Y1,Y2, var.equal=TRUE)
> # MSE
> SSE = sum((Y1-mean(Y1))^2) + sum((Y2-mean(Y2))^2) + sum((Y3-mean(Y3))^2)
> MSE = SSE / df
> MSE
[1] 2.39438
>
> T12 = (mean(Y1)-mean(Y2))/sqrt(MSE*(1/n1+1/n2))
> p12 = 2*pt(abs(T12), df=df, lower.tail=FALSE)
> c(T12, p12) # Blue vs. Brown
[1] 3.086309326 0.007079982
>
> T13 = (mean(Y1)-mean(Y3))/sqrt(MSE*(1/n1+1/n3))
> p13 = 2*pt(abs(T13), df=df, lower.tail=FALSE)
> c(T13, p13) # Blue vs. Green
[1] 1.3305098 0.2020033
>
> T23 = (mean(Y2)-mean(Y3))/sqrt(MSE*(1/n2+1/n3))
> p23 = 2*pt(abs(T23), df=df, lower.tail=FALSE)
> c(T23, p23) # Brown vs. Green
[1] -1.5105287  0.1504046
>
> # Compare
> # t.test(Y1,Y2, var.equal=TRUE)$p.value  vs. p12
> # t.test(Y1,Y3, var.equal=TRUE)$p.value  vs. p13
> # t.test(Y2,Y3, var.equal=TRUE)$p.value  vs. p23
```

Ⓡ: Using pairwise.t.test() function

```
> color = factor( rep(c("Blue","Brown","Green"), c(n1,n2,n3)) )
> pairwise.t.test(Y,color, p.adjust="none", pool.sd=TRUE)

        Pairwise comparisons using t tests with pooled SD

data:  Y and color

      Blue   Brown
Brown 0.0071 -
Green 0.2020 0.1504

P value adjustment method: none
```

Example 9. We consider Example 4 again to use the Fisher LSD procedure.

Ⓡ: Reading Data

```
> Y1 = c(2.4, 2.7, 3.1, 3.1)
> Y2 = c(0.7, 1.6, 1.7, 1.8)
> Y3 = c(2.4, 3.1, 5.4, 6.1)
> Y4 = c(0.3, 0.3, 2.4, 2.7)
> Y5 = c(0.5, 0.9, 1.4, 2.0)
> Y = c(Y1, Y2, Y3, Y4, Y5)
> n1=length(Y1);n2=length(Y2);n3=length(Y3);n4=length(Y4);n5=length(Y5)
> df = n1+n2+n3+n4+n5-5
```

Ⓡ: Using t-test statistics

```
> SSE = sum((Y1-mean(Y1))^2) + sum((Y2-mean(Y2))^2) +
    sum((Y3-mean(Y3))^2) + sum((Y4-mean(Y4))^2) + sum((Y5-mean(Y5))^2)
> MSE = SSE / df
> MSE
[1] 1.131667
```

```
> # 1 vs. {2,3,4,5}
> T12 = (mean(Y1)-mean(Y2))/sqrt(MSE*(1/n1+1/n2))
> p12 = 2*pt(abs(T12), df=df, lower.tail=FALSE)
> c(T12, p12)
[1] 1.82792526 0.08751812

> T13 = (mean(Y1)-mean(Y3))/sqrt(MSE*(1/n1+1/n3))
> p13 = 2*pt(abs(T13), df=df, lower.tail=FALSE)
> c(T13, p13)
[1] -1.89439527  0.07761809

> T14 = (mean(Y1)-mean(Y4))/sqrt(MSE*(1/n1+1/n4))
> p14 = 2*pt(abs(T14), df=df, lower.tail=FALSE)
> c(T14, p14)
[1] 1.86116026 0.08243548

> T15 = (mean(Y1)-mean(Y5))/sqrt(MSE*(1/n1+1/n5))
> p15 = 2*pt(abs(T15), df=df, lower.tail=FALSE)
> c(T15, p15)
[1] 2.16027531 0.04734312

> # 2 vs. {3,4,5}
> T23 = (mean(Y2)-mean(Y3))/sqrt(MSE*(1/n2+1/n3))
> p23 = 2*pt(abs(T23), df=df, lower.tail=FALSE)
> c(T23, p23)
[1] -3.722320527  0.002043516
>
> T24 = (mean(Y2)-mean(Y4))/sqrt(MSE*(1/n2+1/n4))
> p24 = 2*pt(abs(T24), df=df, lower.tail=FALSE)
> c(T24, p24)
[1] 0.0332350 0.9739254
>
> T25 = (mean(Y2)-mean(Y5))/sqrt(MSE*(1/n2+1/n5))
> p25 = 2*pt(abs(T25), df=df, lower.tail=FALSE)
> c(T25, p25)
[1] 0.332350 0.744224

> # 3 vs. {4,5}
> T34 = (mean(Y3)-mean(Y4))/sqrt(MSE*(1/n3+1/n4))
> p34 = 2*pt(abs(T34), df=df, lower.tail=FALSE)
> c(T34, p34)
[1] 3.755555531 0.001909124
>
> T35 = (mean(Y3)-mean(Y5))/sqrt(MSE*(1/n3+1/n5))
> p35 = 2*pt(abs(T35), df=df, lower.tail=FALSE)
> c(T35, p35)
[1] 4.054670574 0.001037412

> # 4 vs. 5
> T45 = (mean(Y4)-mean(Y5))/sqrt(MSE*(1/n4+1/n5))
> p45 = 2*pt(abs(T45), df=df, lower.tail=FALSE)
> c(T45, p45)
[1] 0.2991150 0.7689566

> cbind( c(p12,p13,p14,p15), c(NA,p23,p24,p25), c(NA,NA,p34,p35),
    c(NA,NA,NA,p45))
           [,1]        [,2]        [,3]      [,4]
[1,] 0.08751812          NA          NA        NA
[2,] 0.07761809 0.002043516          NA        NA
[3,] 0.08243548 0.973925402 0.001909124        NA
[4,] 0.04734312 0.744223954 0.001037412 0.7689566
```

Ⓡ: Using pairwise.t.test() function

```
> treat = factor( rep(1:5, c(n1,n2,n3,n4,n5)) )
> pairwise.t.test(Y, treat, p.adjust="none", pool.sd=TRUE)

        Pairwise comparisons using t tests with pooled SD
data:  Y and treat
  1      2      3      4
2 0.0875 -      -      -
```

```
3 0.0776 0.0020 -      -
4 0.0824 0.9739 0.0019 -
5 0.0473 0.7442 0.0010 0.7690
```

Where there are many groups, it is more convenient to sort the factors based on their corresponding treatment means from the smallest to the largest. Then, it is easier to cluster treatments.

Ⓡ: Using pairwise.t.test() function after soring the group means

```
> treat = factor( rep(1:5, c(n1,n2,n3,n4,n5)) )

> by(Y, treat, mean)  # sample means for each treatment
treat: 1
[1] 2.825
--------------------------------------------------------
treat: 2
[1] 1.45
--------------------------------------------------------
treat: 3
[1] 4.25
--------------------------------------------------------
treat: 4
[1] 1.425
--------------------------------------------------------
treat: 5
[1] 1.2
> gr.order = order(by(Y, treat, mean)) # the order of the means
> gr.order
[1] 5 4 2 1 3

> treat2 = factor(treat, levels=gr.order)
> levels(treat2)
[1] "5" "4" "2" "1" "3"
> pairwise.t.test(Y, treat2, p.adjust="none", pool.sd=TRUE)

        Pairwise comparisons using t tests with pooled SD
data:  Y and treat2
  5      4      2      1
4 0.7690 -      -      -
2 0.7442 0.9739 -      -
1 0.0473 0.0824 0.0875 -
3 0.0010 0.0019 0.0020 0.0776
```

‖

7 Note on heteroscedasticity case

For testing the effect of the treatments, we assumed the model $Y_{ij} = \mu_i + \epsilon_{ij}$ where $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$. This model assumes that the variances of all the samples are equal, which is called *homoscedasticity*. If this homoscedasticity is not satisfied so that $\epsilon_{ij} \sim N(0, \sigma_i^2)$, we have to use other methods. In the statistics literature, several methods are recommended. However, there is no clear winner for this heteroscedasticity case. Some well-known methods are (i) Box-Cox transform, (ii) Welch-type ANOVA, (iii) weighted linear regression model, (iv) bootstrap, etc.

References

Carmer, S. G. and Swanson, M. R. (1973). An evaluation of ten pairwise multiple comparison procedures by monte carlo methods. *Journal of the American Statistical Association*, 68(341):66–74.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury, Pacific Grove, CA, second edition.

Cochran, W. G. (1934). The distribution of quadratic forms in a normal system with applications to the analysis of variance. *Proceedings of the Cambridge Philosophical Society*, 30:178–191.

Hogg, R. V., Tanis, E. A., and Zimmerman, D. L. (2015). *Probability and Statistical Inference*. Pearson, 9th edition.

Kendall, M. and Stuart, A. (1979). *The Advanced Theory of Statistics*, volume 2. Charles Griffin, fourth edition.

Ross, S. M. (2014). *A First Course in Probability*. Prentice Hall, 9th edition.

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, 40:87–104.

Smith, J. M. and Misiak, H. (1973). The effect of iris color on critical flicker frequency. *The Journal of General Psychology*, 89:91–95.

Tukey, J. W. (1953). The problem of multiple comparisons. Unpublished notes, Princeton University.