C. One-way ANOVA Models

1 Completely Randomized Experiments

An experiment is to be designed and conducted to compare \( k \) treatments with respective sample sizes \( n_i, i = 1, \cdots, k \) and total sample size of \( N = \sum_{i=1}^{k} n_i \).

1.1 Selection of Experimental Units

From a homogeneous population, select randomly \( N \) experimental units. If the population is known to be heterogeneous, the consideration of block factor(s) or concomitant variable(s) (i.e., covariates) that define/reflect such heterogeneity should be in place (see discussion in later topics). To these randomly selected experimental units, the \( k \) treatments will be applied.

1.2 Allocation of Experimental Units to Treatments

Randomly allocate units to the \( k \) treatments with the respective sample sizes. Note that the total number of possible allocations is

\[
\binom{N}{n_1, \cdots, n_k} = \frac{N!}{n_1! \times \cdots \times n_k!}
\]

1.3 Scheduling the Experimentation

The order in which the treatments are applied in executing the experiment should be randomized. The following is an example of a schedule to carry out such experiment with \( k = 3, n_1 = 5, n_2 = 4, n_3 = 7 \), and \( N = 16 \) (treatments are A, B, and C).

| run order: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| treatment: | C | B | A | C | B | C | C | A | A | A | B | B | C | C | A | C |

1.4 Measuring The Responses

Quantitative responses (i.e., continuous measurements) are generally preferable unless a response is nominal in nature. For example, in an experiment involving the improvement of opening door of cars, a continuous force measurement for opening a door is better than an ordinal (easy, normal, hard to open) judgement. However, an ordinal judgement is preferable to delaying the experiment until a good measurement system is developed.

If there are multiple responses, the order in which the responses are measured should be randomized.
1.5 Rationale for Randomization

Randomization provides protection against unknown factors that may have effects on responses. It is used as a device of concealment so that the unwanted influence of subjective judgement in treatment allocation (e.g., a physician’s assignment of medical treatments to patients) is minimized. Moreover, randomization ensures the validity of the estimate of experimental error and provides a basis for statistical inference (to be discussed in a later section).

1.6 Replication

Note that each treatment is repeated a number of times to provides enough power for the inference (the estimation precision and power of tests will be discussed in later sections). This is based on the principle of replication. The sample size $n_i$ for treatment $i$ is called the number of replications or simply replicates.

1.7 Tabulating Experimental Result

Assume single response.

Arrange outcomes from each treatment in rows (can also do in columns). Also record actual run order (for example, can put run order enclosed in parentheses in superscripts). Run order is important in assessing model validity. For the example above

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Replication</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$y_{13}^{(3)}$, $y_{12}^{(8)}$, $y_{13}^{(9)}$, $y_{14}^{(10)}$, $y_{15}^{(15)}$</td>
</tr>
<tr>
<td>B</td>
<td>$y_{21}^{(2)}$, $y_{22}^{(5)}$, $y_{23}^{(11)}$, $y_{24}^{(12)}$</td>
</tr>
<tr>
<td>C</td>
<td>$y_{31}^{(1)}$, $y_{32}^{(4)}$, $y_{33}^{(6)}$, $y_{34}^{(7)}$, $y_{35}^{(13)}$, $y_{36}^{(14)}$, $y_{37}^{(16)}$</td>
</tr>
</tbody>
</table>

2 Models and Assumptions

| One-way (fixed-effects) ANOVA model

\[
Y_{ij} = \mu_i + \varepsilon_{ij} \\
= \mu + \alpha_i + \varepsilon_{ij}, \ i = 1, \cdots, k, \ j = 1, \cdots, n_i, \\
\text{with} \ \sum_{i=1}^{k} \alpha_i = 0 \text{ where } \varepsilon_{ij} \text{'s are i.i.d. } N(0, \sigma^2).
\]

Note that $Y_{ij}$’s are independent and for $i = 1, \cdots, k, \ Y_{ij} \sim N(\mu_i, \sigma^2), \ j = 1, \cdots, n_i$.

2.1 Notation and Terminology

1. If $n = n_1 = \cdots = n_k$ then the design/model is called balanced. Otherwise, it’s unbalanced.
2. Total sample size \( N = \sum_{i=1}^{k} n_i \). For balanced model, \( N = nk \).

3. Population grand mean \( \mu = \sum_{i=1}^{k} n_i \mu_i / N \).

4. Population \( i \)th (treatment) mean \( \mu_i \).

5. Population \( i \)th (treatment) effect \( \alpha_i = \mu_i - \mu \).

6. The two model representations defined by Equations 1 and 2 are called, respectively, means model and factor effects model.

### 2.2 Graphical Representation of One-Way Model

Consider \( k = 3 \), \( n_1 = 5 \), \( n_2 = 4 \), and \( n_3 = 7 \). The following gives a graphical representation of such one-way fixed-effects model with samples:

![One-Way ANOVA Model with Samples](image)

3 Estimations

#### 3.1 Estimators

1. Grand average (Sample grand mean) \( \hat{\mu} = \overline{Y}_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij} / N = \sum_{i=1}^{k} n_i \overline{Y}_{i.} / N \).

2. \( i \)th treatment average (Sample \( i \)th treatment mean) \( \hat{\mu}_i = \overline{Y}_{i.} = \sum_{j=1}^{n_i} Y_{ij} / n_i \).
3. (Sample) $i$th treatment effect $\hat{\alpha}_i = Y_{i.} - \bar{Y}..$.

4. $i$th sample variance $S_i^2 = \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2 / (n_i - 1)$.

5. (Between) treatments Sum of Squares $SS_B = \sum_{i=1}^{k} n_i (\bar{Y}_{i.} - \bar{Y}..)^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \hat{\alpha}_i^2$.

6. Within treatments Sum of Squares (Error Sum of Squares) $SS_E = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2$.

7. Total Sum of Squares $SS_T = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}..)^2$.

8. $SS_T = SS_B + SS_E$.

9. (Between) treatments Mean Squares $MS_B = SS_B / (k - 1)$.

10. Within treatments Mean Squares (Error Mean Squares) $MS_E = SS_E / (N - k)$ ($= \sum_{i=1}^{k} (n_i - 1) S_i^2 / (N - k)$).

### 3.2 Distributions of Estimators

1. **Theorem** For the one-way fixed-effects model, the following distributional results hold.

   (a) Total Sum of Squares.
   
   \[
   \frac{SS_T}{\sigma^2} \sim \chi^2(N - 1, \lambda = \sum_{i=1}^{k} \frac{n_i \alpha_i^2}{\sigma^2}).
   \]

   (b) Between treatments Sum of Squares.
   
   \[
   \frac{SS_B}{\sigma^2} \sim \chi^2(k - 1, \lambda = \sum_{i=1}^{k} \frac{n_i \alpha_i^2}{\sigma^2}).
   \]

   (c) Within treatments Sum of Squares.
   
   \[
   \frac{SS_E}{\sigma^2} \sim \chi^2(N - k)
   \]

   (d) Independence. $SS_B$ and $SS_E$ are independent.

   (e) F ratio.
   
   \[
   \frac{MS_B}{MS_E} \sim F(k - 1, N - k; \lambda = \frac{\sum_{i=1}^{k} n_i \alpha_i^2}{\sigma^2}).
   \]

   (f) Hypothesis test. $MS_B/MS_E > F_{\alpha;k-1,N-k}$ defines a level-$\alpha$ LR (Likelihood Ratio) test for $H_0 : \mu_1 = \cdots = \mu_k$ versus $H_1 : \mu_i \neq \mu_j$ for some $i \neq j$.
Expected Mean Squares.

\[ E(\text{MS}_E) = \sigma^2 \quad (\text{i.e., \ MS}_E \ \text{is unbiased}); \]
\[ E(\text{MS}_B) = \sigma^2 + \frac{\sum_{i=1}^{k} n_i \alpha_i^2}{k-1}. \]

**Proof**

First note that

- \( Y_{ij} \)'s are independent \( N(\mu_i, \sigma^2); \)
- \( \overline{Y}_i \)'s are independent \( N(\mu_i, \sigma^2/n_i); \)
- \( \overline{Y}_.. \sim N(\mu,\sigma^2/N). \)

\[ \text{(a)} \quad \frac{SS_T}{\sigma^2} \sim \chi^2(N-1, \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\mu_i - \mu)^2/\sigma^2). \]

Now, \( \sum_{j=1}^{n_i} (\mu_i - \mu)^2 = n_i \alpha_i^2 \) and claim.

\[ \text{(b)} \quad \text{Observe that} \]

\[ w_i = n_1, \ w = N = \sum_{i=1}^{k} n_i, \ \overline{Y}_.. = \overline{Y}_w, \ \text{and } SS_B = SS_w. \]

It follows from item 2 of §A.4.1 that

\[ \frac{SS_B}{\sigma^2} \sim \chi^2(k-1, \sum_{i=1}^{k} n_i (\mu_i - \mu)^2/\sigma^2). \]

Now, \( \sum_{i=1}^{k} n_i (\mu_i - \mu)^2/\sigma^2 = \sum_{i=1}^{k} n_i \alpha_i^2/\sigma^2 \) and claim.

\[ \text{(c)} \quad \text{Observe that} \]

\[ \frac{(n_i - 1)S_i^2}{\sigma^2} \sim \chi^2(n_i - 1), \ i = 1, \cdots, k \ \text{and are independent (why?).} \]

Hence,

\[ \frac{SS_E}{\sigma^2} = \sum_{i=1}^{k} \frac{(n_i - 1)S_i^2}{\sigma^2} \sim \chi^2(N - k = \sum_{i=1}^{k} (n_i - 1)). \]
Note that $\bar{Y}_i$ is independent of $S_i^2$ for $i = 1, \ldots, k$. Thus, $SS_B = g(\bar{Y}_1, \ldots, \bar{Y}_k)$ and $SS_E = h(S_1^2, \ldots, S_k^2)$ are independent. Note: from (d), having shown (a) and (c), one can show (b).

(f) Denote

$$\mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{i_{n_i}} \end{bmatrix} \sim N_{n_i}(\mu_{1n_i}, \sigma^2 I_{n_i})$$

where $\mathbf{1}_{n_i}$ is an $n_i \times 1$ vector of 1’s.

The observation vector

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_k \end{bmatrix} \sim N_k(\mu, \sigma^2 I_N)$$

where $\mathbf{\mu} = \begin{bmatrix} \mu_1 \mathbf{1}_{1n_1} \\ \vdots \\ \mu_k \mathbf{1}_{kn_k} \end{bmatrix}$.

The likelihood function $L = L(\mu_1, \ldots, \mu_k, \sigma^2; \mathbf{Y})$ is given by

$$L = (2\pi)^{-\frac{N}{2}} |\sigma^2 I_N|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{Y} - \mathbf{\mu})' (\sigma^2 I_N)^{-1} (\mathbf{Y} - \mathbf{\mu}) \right\}$$

$$= c(\sigma^2) \exp \left\{ -\frac{1}{2\sigma^2} ||\mathbf{Y} - \mathbf{\mu}||^2 \right\}, \text{ where}$$

$$c(\sigma^2) = (2\pi)^{-\frac{N}{2}} |\sigma^2 I_N|^{-\frac{1}{2}} = (2\pi \sigma^2)^{-\frac{N}{2}},$$

$$||\mathbf{Y} - \mathbf{\mu}||^2 = (\mathbf{Y} - \mathbf{\mu})' (\mathbf{Y} - \mathbf{\mu}) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_i)^2.$$

Use the log-likelihood, $\ln L$, one can show that

$$\hat{\mu}_i = \bar{Y}_i, \ i = 1, \ldots, k, \ \text{and} \ \hat{\sigma}^2 = MS_E \ \text{are the MLEs (why?)}. $$

Therefore,

$$\hat{L} = c(\hat{\sigma}^2) \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_i)^2 \right\}$$

Under $H_0$, applying similar arguments, one can show that

$$\hat{L}_0 = c(\hat{\sigma}^2) \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu})^2 \right\}$$
where \( \hat{\mu} = \overline{Y} \) is the MLE of \( \mu \). The LR test statistic is \( \hat{L}_0 / \hat{L} \) and we reject \( H_0 \) if \( \hat{L}_0 / \hat{L} < c_1 \) or if \( -\frac{2}{k-1} \ln(\hat{L}_0 / \hat{L}) > c_2 \) where \( c_2 = -\frac{2}{k-1} \ln c_1 \). Now,

\[
-\frac{2}{k-1} \ln(\hat{L}_0 / \hat{L}) = \frac{MS_B}{MS_E} \sim F(k-1, N-k) \text{ under } H_0.
\]

\[\text{[g]}\]

From (c), \( E(\frac{SS_E}{\sigma^2}) = N - k \). Hence, \( EMS_E = \sigma^2 \).

From (b), \( E(\frac{SS_B}{\sigma^2}) = (k-1) + \lambda \).

Consequently, \( EMS_B = \sigma^2 (1 + \frac{\lambda}{k-1}) = \sigma^2 + \sum_{i=1}^{k} \frac{n_i \alpha_i^2}{k-1} \).

Remark: The power of the level-\( \alpha \) F test of equal means is calculated by

\[
P(\frac{MS_B}{MS_E} > F_{\alpha,k-1,N-k}| H_1) = P(F(k-1, N-k; \lambda = \sum_{i=1}^{k} \frac{n_i \alpha_i^2}{\sigma^2} > F_{\alpha,k-1,N-k})
\]

2. F test as the generalization of two sample (pooled) t test For \( k = 2 \),

\[
t = \frac{\overline{Y}_1 - \overline{Y}_2}{S_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(\overline{Y}_1 - \overline{Y}_2) / \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}{S_{\text{pooled}}}.
\]

Observe that

\[
\overline{Y} = \frac{n_1 \overline{Y}_1 + n_2 \overline{Y}_2}{n_1 + n_2}.
\]

Now,

\[
(\text{denominator})^2 = S_{\text{pooled}}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = MS_E.
\]

Express

\[
n_1(\overline{Y}_1 - \overline{Y} \cdot)^2 = n_1 \left[ \overline{Y}_1 - \frac{n_1 \overline{Y}_1 + n_2 \overline{Y}_2}{n_1 + n_2} \right]^2 = \frac{n_1 n_2^2}{(n_1 + n_2)^2} (\overline{Y}_1 - \overline{Y}_2)^2.
\]

By symmetry,

\[
n_2(\overline{Y}_2 - \overline{Y} \cdot)^2 = \frac{n_2^2 n_2}{(n_1 + n_2)^2} (\overline{Y}_1 - \overline{Y}_2)^2.
\]
Hence,
\[(\text{numerator})^2 = \frac{n_1n_2}{n_1 + n_2} (\bar{Y}_1 - \bar{Y}_2)^2\]
\[= \frac{n_1n_2}{(n_1 + n_2)^2} (\bar{Y}_1 - \bar{Y}_2)^2 + \frac{n_1^2n_2}{(n_1 + n_2)^2} (\bar{Y}_1 - \bar{Y}_2)^2\]
\[= n_1(\bar{Y}_1 - \bar{Y}_.)^2 + n_2(\bar{Y}_2 - \bar{Y}_.)^2\]
\[= SS_B = MS_B \text{ (since } 2 - 1 = 1)\]

It follows then
\[t^2 = \frac{MS_B}{MS_E}\]

3. Geometric view of the decomposition of deviations in the model

Consider the decomposition: for \(i = 1, \ldots, k\), for \(j = 1, \ldots, n_i\),

**Model** \(Y_{ij} - \mu = \alpha_i + \varepsilon_{ij}\)

**Data** \(Y_{ij} - \bar{Y}_- = (\bar{Y}_i - \bar{Y}_-) + (Y_{ij} - \bar{Y}_i)\)

deviation between within
from treatments treatments
grand deviations deviations
average (fit, \(\hat{Y}_{ij}\)) (residual, \(\hat{\varepsilon}_{ij}\))

Write the decomposition in vector form

\[D = T + R\]

with

\[D = \begin{bmatrix} D_1 \\ \vdots \\ D_k \end{bmatrix}, \quad T = \begin{bmatrix} T_1 \\ \vdots \\ T_k \end{bmatrix}, \quad R = \begin{bmatrix} R_1 \\ \vdots \\ R_k \end{bmatrix}, \quad \text{where}\]

\[D_i = \begin{bmatrix} Y_{i1} - \bar{Y}_- \\ \vdots \\ Y_{in_i} - \bar{Y}_- \end{bmatrix}_{n_i \times 1}, \quad T_i = \begin{bmatrix} \hat{Y}_{i1} \\ \vdots \\ \hat{Y}_{in_i} \end{bmatrix}_{n_i \times 1}, \quad R_i = \begin{bmatrix} \hat{\varepsilon}_{i1} \\ \vdots \\ \hat{\varepsilon}_{in_i} \end{bmatrix}_{n_i \times 1}\]

The inner product of \(T\) and \(R\):

\[T'R = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \hat{Y}_{ij} \hat{\varepsilon}_{ij} = 0 \text{ (why?)}\]

and hence \(T\) is orthogonal to \(R\). By Pythagorean Theorem,

\[\|D\|_2^2 = \|T\|_2^2 + \|R\|_2^2.\]