One-way ANOVA — random effects model.

Example: milk

\[ Y_{ij} = \mu + A_i + E_{ij} \]

\[ \sim N(0, \sigma^2_A) \]

main effect of the
randomly selected
i-th level of Factor A
(batch)

\[ \sigma^2_A \]

Also: \( A_i \) and \( E_{ij} \) are mutually
independent.

- Variance components:

\[ \text{Var}(Y_{ij}) = \text{Var}(A_i) + \text{Var}(E_{ij}) \]

\[ = \sigma^2_A + \sigma^2 \]

- Inference:

\[ H_0: \sigma^2_A = 0, \quad H_1: \sigma^2_A > 0 \]

- Expected mean squares (EMS)

\[ \text{EMS}_A = \sigma^2 + \nu \sigma^2_A, \quad \text{EMS}_E = \sigma^2 \]

(repetition)
Estimation of variance components

\[ \text{EMS}_A = \sigma^2 + r \sigma_A^2 \]

\[ \text{EMS}_E = \sigma^2 \]

- Equate the EMS's in (11.3) and (11.4) to their observed values in the ANOVA table.

- Solve for \( \sigma_A^2 \) and \( \sigma^2 \).

\[ \hat{\sigma}_A^2 \]

\[ \hat{\sigma}^2 \]

Example 11.1.3 (Milk)

(a) Test whether different batches of milk have different percentage protein content.

(b) Estimate the variance components for the random effect-model.
(a) $H_0: \sigma^2 = 0$
$H_1: \sigma^2 > 0$

$$F = \frac{MS_A}{MS_E} = \frac{0.0637}{0.0163} = 3.91$$

Null distribution is $F_{(3, 8)}$
and $F_{(3, 8)}(0.95) = 4.066 > 3.91$

Hence, Accept $H_0$. No evidence for a difference between batches.

\[ \text{Var}(E_{ij}) \]

(b) $\text{EMS}_E = \sigma^2$, \hspace{1cm} $\hat{\sigma}^2 = MS_E = 0.0163$

$\text{EMS}_A = \sigma^2 + r\hat{\sigma}^2$

$$\Rightarrow \hat{\sigma}^2 + 3\hat{\sigma}^2_A = 0.0637$$

$$\hat{\sigma}^2_A = \frac{0.0637 - 0.0163}{3} = 0.0158$$

Variability among
the effects of the population
of levels of factor A (batch)

\[ \text{Var}(A_{ij}) \]
Minitab can be used to analyse the random-effect model.

\[ c_1 = \text{protein} \]
\[ c_2 = \text{batch} \]

**Analysis of Variance (Balanced Designs)**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Type</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch</td>
<td>random</td>
<td>4</td>
<td>1 2 3 4</td>
</tr>
</tbody>
</table>

**Analysis of Variance for protein**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch</td>
<td>3</td>
<td>0.19103</td>
<td>0.06368</td>
<td>3.91</td>
<td>0.064</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>0.13013</td>
<td>0.01627</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>0.32117</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The above output gives the EMS's as well as estimates of the variance components.

- You ought to know how to calculate the latter yourself.
- Note the subcommand random used in the above.
- Compare the above output with the following for the fixed effect model.
MTB > anova c1 = c2;
SUBC> ems.

Analysis of Variance (Balanced Designs)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Type</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch</td>
<td>fixed</td>
<td>4</td>
<td>1  2   3  4</td>
</tr>
</tbody>
</table>

Analysis of Variance for protein

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch</td>
<td>3</td>
<td>0.19103</td>
<td>0.06368</td>
<td>3.91</td>
<td>0.054</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>0.13013</td>
<td>0.01627</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>0.32117</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source  Variance Error Expected Mean Square
component term (using unrestricted model)

1 batch 2 (2) + Q[1] Function of $\alpha$, $\gamma$, ..., fixed effects.
Inference on $\mu$

Recall that $\mu$ is the mean response of a randomly chosen item from the population. In the Milk example, it is the mean protein content of a randomly chosen carton from a randomly chosen batch. For a balanced design,

$$\hat{\mu} = \bar{y} = \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{r} y_{ij}$$

where $n = ra = \text{sample size}$. It can be shown that

$$\text{var}(\bar{y}) = \frac{\sigma^2 + r \sigma_A^2}{n} \quad (= \frac{\text{EMS}_A}{n})$$

Hence,

$$\text{s.e.}(\bar{y}) = \sqrt{\frac{\text{MS}_A}{n}}$$
Example 11.1.4 (Milk)

Obtain a 95% confidence interval for the mean milk content of all batches, 

\[ \hat{\mu} = \bar{y} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{b} y_{ij} = 3.348 \]

\[ \text{se} (\bar{y}) = \sqrt{\frac{HSA}{n}} = \sqrt{\frac{0.0637}{12}} \]

\[ \therefore \text{95\% C.I. for } \mu \text{ is: } \]

\[ \bar{y} \pm (\text{table value}) \times \text{se} (\bar{y}) \]

\[ \bar{y} \pm 1.975 \times \frac{0.0637}{\sqrt{12}} \]

\[ = 3.348 \pm 3.182 \sqrt{\frac{0.0637}{12}} \]

\[ = (3.116, 3.5801) \]
11.2 Two-way

We consider a completely randomized two-factor factorial experiment.

- Let A and B be the two factors with \( a \) and \( b \) levels respectively.
- There are \( a \times b \) treatment combinations.
- The factors A and B may be fixed or random.
- Their effects may be additive or interactive.
- Suppose the design is balanced with \( r \) replications for each treatment.

\[
\text{n} = a \times b \times r
\]

\[r \times e \times m\]
Two-way fixed-effect models:
Both A and B are fixed.

**Additive:** \[ Y_{ijk} = \mu + \alpha_i + \beta_j + E_{ijk}, \]
\[ i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, r. \]

where \( Y_{ijk} \) = \( k \)-th observed response for \( i \)-th level of factor A and \( j \)-th level of factor B

\[ \mu = \text{overall mean}, \]

\[ \alpha_i = \text{fixed effect of } i \text{-th level of factor A}, \]

\[ \beta_j = \text{fixed effect of } j \text{-th level of factor B}, \]

\[ E_{ijk} = \text{random error}, \]

- Constraints: \( \sum \alpha_i = 0 \) and \( \sum \beta_j = 0 \).
- Assumption: \( E_{ijk} \) are independent \( \text{N}(0, \sigma^2) \).
- All others are fixed parameters.

**Interactive:** \[ Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + E_{ijk}, \]
\[ i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, r. \]

where, in addition,

- we have \((\alpha\beta)_{ij}\) = fixed interaction effect.
- Constraints: \( \Sigma_i (\alpha\beta)_{ij} = 0 \) and \( \Sigma_j (\alpha\beta)_{ij} = 0 \). \( (\Sigma_i \alpha_i = 0, \Sigma_j \beta_j = 0). \)
- Assumption: \( E_{ijk} \) are independent \( \text{N}(0, \sigma^2) \).

We used these two models in Chapter 6.
11.2. TWO-WAY

**Two-way random-effect models:**

Both A and B are random.

\[
\begin{align*}
\text{Additive:} & \quad Y_{ijk} = \mu + A_i + B_j + E_{ijk}, \\
i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, n.
\end{align*}
\]

where \(A_i\) = random effect of \(i\)-th level of factor A

\(B_j\) = random effect of \(j\)-th level of factor B.

The assumptions are

1. \(E_{ijk}\) are independent \(N(0, \sigma^2)\);

2. \(A_i\) are independent \(N(0, \sigma^2_A)\);

3. \(B_j\) are independent \(N(0, \sigma^2_B)\);

4. \(E_{ijk}, A_i\) and \(B_j\) are mutually independent.

For this model,

\[\text{var}(Y_{ijk}) = \sigma^2_A + \sigma^2_B + \sigma^2.\]

The RHS shows the variance components of this model.
\[ \hat{\text{Interactive:}} \quad Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + E_{ijk}, \quad i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, n. \]

where, in addition, we have

\((AB)_{ij} = \text{random interaction effect of } i\text{-th level of factor } A \text{ and } j\text{-th level factor } B.\)

The additional assumptions are

5. \((AB)_{ij} \text{ are independent } N(0, \sigma^2_{AB});\)

6. \((AB)_{ij} \text{ are independent of } E_{ij}, A_i \text{ and } B_j.\)

For this model,

\[ \text{var}(Y_{ijk}) = \sigma_A^2 + \sigma_B^2 + \sigma^2_{AB} + \sigma^2. \]

There are therefore four variance components to this model.
Example 11.2.3 (DNA) 

Suppose an experiment to examine the effect of different analysts and subjects in chemical analyses for the DNA content of plaque. Three female subjects (aged 18–20 years) were chosen for the study. Each subject was allowed to maintain her usual diet, supplemented with 30 mg (15 tablets) of sucrose per day. No toothbrushing or mouthwashing was allowed during the study. At the end of the week, plaque was scraped from the entire dentition of each subject and divided into three samples. Each of three analysts chosen at random was then given an unmarked sample of plaque from each of the subjects and asked to perform an analysis for the DNA content (in micrograms).

(a) What are the factors being investigated? Are they random or fixed?

Factor A: Female subjects — random
(3 levels)

Factor B: Analysts — random
(3 levels)
17.3 Extensions of Random-Effects Models

17.3 Suppose that the pharmaceutical company of Exercise 17.1 wishes to estimate the expected potency for a measurement made on a vat selected at random from a month's production of a liquid medication.
   a. Using the sample data of Exercise 17.1, form a point estimate of the average potency for a measurement made on a randomly selected vat.
   b. Place a 95% confidence interval on the average potency for a measurement made on a randomly selected vat.

Extensions of Random-Effects Models

The ideas presented for a random-effects model in a one-factor experiment can be extended to any of the block designs and factorial experiments covered in Chapter 15. Although we will not have time to cover all such situations, we will consider first a randomized block design in which the block effects and the treatment effects are random.

Consider an experiment to examine the effects of different analysts and subjects in chemical analyses for the DNA content of plaque. Three female subjects (ages 18–20 years) were chosen for the study. Each subject was allowed to maintain her usual diet, supplemented with 30 mg (15 tablets) of sucrose per day. No toothbrushing or mouthwashing was allowed during the study. At the end of the week, plaque was scraped from the entire dentition of each subject and divided into three samples. Each of three analysts chosen at random was then given an unmarked sample of plaque from each of the subjects and asked to perform an analysis for the DNA content (in micrograms). The two-factor experiment of sample data could then be organized as shown in Table 17.6.

<table>
<thead>
<tr>
<th>Table 17.6</th>
<th>DNA concentrations for samples of plaque</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Analyst 1</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>1</td>
<td>13.2</td>
</tr>
<tr>
<td>2</td>
<td>12.5</td>
</tr>
<tr>
<td>3</td>
<td>13.0</td>
</tr>
<tr>
<td>Means</td>
<td>12.9</td>
</tr>
</tbody>
</table>

This experimental design is recognized as a randomized block design, with subjects representing blocks and analysts being the treatments. The experimental units are samples of plaque scraped from the dentition of subjects. If we assume that the three subjects represent a random sample from a large population of possible subjects, and, similarly, that the three analysts represent a random sample from a large population of possible analysts, we can write the following random-effects model relating DNA concentration to the two factors "analysts" and "subjects":

\[ y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \]

We assume the following:

1. \( \mu \) is an overall unknown concentration mean.
2. \( \alpha_i \) is a random effect due to the \( i \)th analyst. \( \alpha_i \) is normally distributed, with mean 0 and variance \( \sigma^2_\alpha \).
Two-factor (3 x 3 factorial) random-effects additive model.

\[ Y_{ijk} = \mu + A_i + B_j + E_{ijk} \]

\[ A_i \text{ indep. } N(0, \sigma_A^2) \]
\[ B_j \text{ indep. } N(0, \sigma_B^2) \]
\[ E_{ijk} \text{ indep. } N(0, \sigma_e^2) \]

(b) The following partial Minitab output is obtained on the data. What is the model that is fitted?

MTB > anova c1 = c2 c3;
SUBC> random c2 c3;
SUBC> ems.
Analysis of Variance (Balanced Designs)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Type</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyst</td>
<td>random</td>
<td>3</td>
<td>1 2 3</td>
</tr>
<tr>
<td>Subject</td>
<td>random</td>
<td>3</td>
<td>1 2 3</td>
</tr>
</tbody>
</table>

Analysis of Variance for DNA

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyst</td>
<td>2</td>
<td>0.8822</td>
<td>0.4411</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>2</td>
<td>33.2366</td>
<td>16.6178</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>4</td>
<td>0.0911</td>
<td>0.0228</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td>34.2089</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c) Evaluate the variance components for the model.

(d) Is there a significant difference of results in the population of analysts?
(C) \[ \hat{\sigma}^2 = 0.0228 \]

Subject (A): \[ \hat{\sigma}_A^2 = \frac{16.678 - 0.0228}{3} \]
\[ = 5.5317 \]

Analyst (B): \[ \hat{\sigma}_B^2 = \frac{0.4411 - 0.0228}{3} \]
\[ = 0.1394 \]

(b) \( H_0 \): no difference between analysts \( (\sigma_B^2 = 0) \)
\( H_1 \): \( \sigma_B^2 > 0 \)

\[ F\text{-statistic} = \frac{MS_B}{MS_E} = \frac{0.4411}{0.0228} = 19.35 \]

Under \( H_0 \), \( F \approx F_{(2,4)} \)
and \( F_{(2,4)} (0.99) = 18.0 \)

\[ \therefore \text{We reject } H_0 \text{ at the 0.01 level.} \]

There is evidence of significant differences between analysts.