Chapter 11
Random-effects and mixed-effects models 139

11.1. One-way 139
11.2. Two-way 142
3 randomly selected cartons from each batch

4 randomly selected batches

(5) Response v.: percentage protein content (numerical) in a carton.
   Expl v.: batch (categorical) four levels.

(6) Do different batches of milk have different mean % protein content?
11.1 One-way

Example 11.1.1 (Milk)
To assess the batch-to-batch variation in the protein content of 250-ml cartons of pasteurized buttermilk produced by a dairy farm, four batches of cartons were randomly selected from the production line. From each batch, the protein contents were determined for three randomly selected cartons. The resulting data (percentage protein content) are as follows:

<table>
<thead>
<tr>
<th>Batch $i$</th>
<th>percentage protein content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.42 3.41 3.57</td>
</tr>
<tr>
<td>2</td>
<td>3.05 3.14 3.23</td>
</tr>
<tr>
<td>3</td>
<td>3.23 3.48 3.37</td>
</tr>
<tr>
<td>4</td>
<td>3.46 3.59 3.23</td>
</tr>
</tbody>
</table>

(a) What are the response variable and explanatory variable here? What type are they?

(b) What are the questions of interest?
What is the difference between this example and the Melon example?

**Melon example:**

- The factor is Variety and it has four level A, B, C and D.
- We want to know if they give different mean yield.
- We are interested in these four particular varieties.
- We say the levels of the factor Variety fixed.
- The model used for analysis is

\[ Y_{ij} = \mu + \alpha_i + E_{ij} \quad \text{(11.1)} \]

where

- \( Y_{ij} \) = j-th observed response for the factor at i-th level,
- \( \mu \) = overall mean,
- \( \alpha_i \) = main effect of i-th level of factor A,
- \( E_{ij} \) = random error.

**Fixed effects model:** Data are classified into groups on the basis of the levels of the experimental factor (Variety A, B, C, D).
Example 5.1.1 (Melon)

Six plots of each of four variety of melons gave the following yields:

A 25 17 26 16 22 16
B 40 35 32 37 43 37
C 18 23 26 15 11 24
D 28 29 33 32 30 28

(a) What are the experimental units?
(b) What are the response variable and explanatory variable here? What type are they?
(c) What are the questions of interest?

\[ Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \]

levels of factor (variety) are fixed

\[ \bar{y}_4 = \hat{\mu} \]

\[ i = 4, \ j = 1, \ldots, 6 \]
Production from 1 dairy farm.

Levels of factor (batch) are random.

\[ y_{ij} = \mu + \alpha_i + E_{ij} \]

Main effect of the randomly selected \( i \)'th level of factor A.

The responses observed are random samples selected in two stages:

1. Random sample, size \( a \), selected from a population of levels of exp. factor \( \alpha \) = 4
2. Random sample of \( \bar{X} \) responses is observed for the \( i \)'th level selected in stage 1

Randomly select

Batch 1
Batch 2
Batch 3
Batch 4
Random-effects model:

- interested in the effects of many levels of an experimental factor
- measuring responses at every level may be difficult, impossible, or prohibitively expensive.

- regard the set of all levels (treatments) under consideration as a statistical population in its own right.

- draw conclusions about this population on the basis of the observed responses to a random sample of levels selected from this population.
Examples of fixed and random effects:

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Random effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car manufacturers, e.g. Ford, Toyota</td>
<td>Engines sampled from a production line</td>
</tr>
<tr>
<td>Varieties of wheat $1, 2, 3$</td>
<td>Rows of plants in a crop</td>
</tr>
<tr>
<td>Chemicals for controlling insect pests</td>
<td>Batches of insects collected from a site</td>
</tr>
<tr>
<td>Methods of teaching</td>
<td>Students performing a test</td>
</tr>
<tr>
<td>Types of exercise</td>
<td>Basketballers jumping as high as possible</td>
</tr>
<tr>
<td>Strains of a fungus (?)</td>
<td>Strains of a fungus (?)</td>
</tr>
</tbody>
</table>
Milk example:

- The factor is Batch and it has four levels 1, 2, 3 and 4.
- We want to know if different batches have different mean percentage milk protein. (There lies the similarity between the two examples.)
- There is a population of different batches and we are more interested in whether the batches in the population have different mean percentage milk protein, rather than the four selected batches per se.
- We consider these four batches as a random sample from the population of batches and use it to gain information about the population.
- We say the levels of the factor Batch are random: they are drawn randomly from a population of levels.
- The model used for analysis is

\[ Y_{ij} = \mu + A_i + E_{ij} \]  \hspace{1cm} (11.2)

where \(A_i\) denote the main effect of the randomly selected \(i\)-th level of factor A.
The model is called a fixed-effect model.

- \( Y_{ij} = \mu + \alpha_i + E_{ij} \)

\( \alpha_i \) is considered as a constant (fixed), albeit unknown and to be estimated.

\( Y_{ij} = \mu + A_i + E_{ij} \)

\( A_i \) is considered as a random variable.

The model is called a random-effect model.
One-way random effect model

The assumptions for the model

\[ \tilde{Y}_{ij} = \mu + (\hat{A}_i) + (\hat{E}_{ij}), \quad i = 1, \ldots, a; \quad j = 1, \ldots, n_i. \]

are

1. \( E_{ij} \) are independent \( N(0, \sigma^2) \);
2. \( A_i \) are independent \( N(0, \sigma_A^2) \);
3. \( E_{ij} \) and \( A_i \) are mutually independent.

(Compare these with the assumptions for the fixed-effect model which consist of 1. only.)

\[ y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad \text{where} \]

\( \epsilon_{ij} \) are independent random draws from \( N(0, \sigma^2) \).
Variance components

Under the above assumptions,
\[
\text{var}(Y_i) = \text{var}(A_i) + \text{var}(E_{ij}) = \sigma_A^2 + \sigma^2.
\]

- The variability of \( Y \) is the sum of two components, \( \sigma_A^2 \) and \( \sigma^2 \).

- \( \sigma_A^2 \)
  - Gives the variability among the effects of the population of levels of factor A.
  - In the Milk example, gives the variability between the batches in the population of batches.

- \( \sigma^2 \)
  - Gives the variability among the replicated measurements for the same treatment.
  - In the Milk example, it gives the variability within a batch.

- \( \sigma_A^2 \) and \( \sigma^2 \) are called the variance components of \( \text{var}(Y) \).
Null hypothesis

In a one-way random-effect model, the null hypothesis is

\[ H_0 : \sigma_A^2 = 0, \]

which corresponds to saying that there is no difference among
the effects of different levels of A. i.e. between batches.

(Compare this with the null hypothesis of the fixed-effect model \( H_0 : \alpha_1 = \cdots = \alpha_a = 0. \))

The hypothesis can be tested by an ANOVA table which is
exactly the same as for the fixed-effect model.

If we repeat an exp. with

- **fixed effects model:** the same set of treatments is used in each repetition.
  
  Parameters: \( \mu, \alpha_1, \alpha_2, \ldots, \alpha_k \) and \( \sigma^2 \).

- **random effects model:** a different set of treatments can, and very likely will, enter in different repetitions.
  
  Parameters: \( \mu, \sigma_A^2, \sigma^2 \).
Example 11.1.2 (Milk)

Use Minitab to construct an ANOVA table for the data.

Solution:

MTB > oneway c1 levels in c2

Analysis of Variance on 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>0.1910</td>
<td>0.0637</td>
<td>3.91</td>
<td>0.054</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>0.1301</td>
<td>0.0163</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>0.3212</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
MS_A = \frac{\text{SS}_{\text{Batch}}}{\text{df}_{\text{Batch}}}
\]

\[
MS_E = \frac{\text{SS}_{\text{Error}}}{\text{df}_{\text{Error}}}
\]

\[
\text{SS}_{\text{Batch}} = \sum_i n_i (\bar{y}_i - \bar{y})^2
\]

\[
\text{SS}_{\text{Total}} = \sum_i \sum_j (y_{ij} - \bar{y})^2
\]

\[
\text{SS}_{\text{Error}} = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2
\]

\[
\text{SS}_{\text{Total}} = \text{SS}_{\text{Batch}} + \text{SS}_{\text{Error}}
\]
Expected mean squares (EMS)

\[
\text{[mean of the MS]} \quad \left[ \text{Expected variance} \right]
\]

- There are a few mean squares (MS) in any ANOVA table.
- They are random quantities, as they vary from sample to sample.
- Their expected values (over sampling variation) are called the expected mean squares (EMS).

\[
\text{Remember} \quad MS \quad \text{is an estimate of variance}
\]

\[
\frac{SS}{df}
\]
Suppose

- the design is balanced,
- factor A has \( a \) levels and \( \gamma \) replications at each level of A.

Then,

\[
E + A
\]

\[
EMS_A = \sigma^2 + \frac{\gamma}{a} \sigma_A^2 \quad (11.3)
\]

\[
EMS_E = \sigma^2 \quad (11.4)
\]

If \( H_0 : \sigma_A^2 = 0 \) is true, then

\[
EMS_A = EMS_E.
\]

If \( \sigma_A^2 > 0 \), then

\[
EMS_A > EMS_E.
\]

Thus, the \( F \)-statistic

\[
F = \frac{MS_A}{MS_E}
\]

can be used to test \( H_0 \) and it has an \( F \)-distribution with the usual \( df \)'s.