Analysis of Variance for Some Standard Experimental Designs

What is an experiment? It is a planned inquiry to obtain new facts or to confirm or deny results of previous experiments.

Three Categories of Experiments

1. Preliminary – The investigator tries out a large number of treatments in order to obtain leads for future work.

2. Critical – The investigator compares responses to different treatments using a sufficient number of observations of the responses to give reasonable assurance of detecting meaningful differences in treatment effects.

3. Demonstrational – Intended to compare a new treatment(s) with a standard.

The critical experiment is most usual.

Every experiment is set up to provide answers to one or more questions. With this in mind, the experimenter decides what treatment comparisons provide relevant information.

The experimental design is the set of rules used to draw the sample from the target population(s) of experimental units.

Experimental Units are the material to which treatments are applied, one treatment to a unit.
The treatment is the procedure whose effect is to be measured and compared with the effects of other treatments (of course “effect” means the treatment’s influence on the way in which the unit responds.)

A characteristic of all experimental units is variation. The term experimental error is used to mean the variation among observations of experimental units which are treated alike. We often use $\sigma^2$ in our models to represent experimental error.

It is easy to talk about “experimental units treated alike” at a conceptual level. As a practical matter the story is different. If you deal with animals, for example, placing them in different pens, or using different kinds of halters, ... etc., may contribute to variation in the way they respond to the same treatment, over and above their inherent variation.

We want to avoid, as much as possible, inflating experimental error simply by the way we avoidably mishandle experimental units. Certainly this is a very practical consideration, and is situation dependent.

In the most ideal situation experimental error would be zero ($\sigma^2 = 0$). Then the response to treatment would be solely due to the effect of the treatment. We would then need to apply each treatment to only one unit.
The ideal situation doesn’t occur in any practical cases. Thus $\sigma^2_\epsilon > 0$ and we must estimate it by computing $s^2_\epsilon \equiv s^2_W$, the MSE, because it is needed for testing significance and setting confidence intervals.

Computing $s^2_\epsilon$ is only possible when at least one treatment is applied more than one experimental unit in the experiment. Mathematics aside, the need for replication (the term used to say treatments are applied more than once) is obvious when one notes that it will be impossible to assess the differences in response to units treated differently unless we see how they respond when treated alike.

The investigator is certainly well advised to do the best that he or she can to control (reduce) experimental error and to use other devices in experimental design to reduce variation in mean response.

For example, trying to insure uniformity of experimental material and the way it is handled helps control $\sigma^2_\epsilon$.

Recall that the variance of a treatment mean observation, $V(\bar{y}_i) = \sigma^2_\epsilon/n_i$ is reduced as the number $n_i$ of replications of that treatment increases, and estimated differences are more precisely estimated using more replications.

These and other considerations are a part of the design and conduct of experiments.
The simplest design is one that we have already discussed, it is called the *completely randomized design* (CRD). Let us review CRD for a moment using material you have seen before.

In the past we have talked about populations without worrying about how they might have materialized. Here we are simply saying that we can imagine some populations as arising thru experimentation.

The simplest experimental design is named the *completely randomized design*. Here we have (say) $t$ treatments and (say) $nt$ experimental units. These experimental units are divided into $t$ sets of $n$ and the sets are assigned at random, each to a treatment.

We can, equivalently, characterize this design as one that acquires a simple random sample of size $n$ from each of $t$ different treated populations $T_1, T_2, \ldots, T_t$. Each of these populations has mean $\mu_i$ and variance $\sigma^2_\varepsilon, t = 1, 2, \ldots, t$.

The model of this design which we will use is:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, 2, \ldots, t; j = 1, 2, \ldots, n$$

$$E(\epsilon_{ij}) = 0, \quad V(\epsilon_{ij}) = \sigma^2_\varepsilon, \quad \text{with independent } \epsilon_{ij} \text{'s}.$$  

Here $\alpha_i$ is the effect on experimental unit due to treatment $i$. Thus the treated population means are $\mu_i = \mu + \alpha_i$ in this notation.
CRD is a simple design, easily implemented. The design itself does nothing to help reduce experimental error. Replications \( (n_i) \) of treatments \( T_i \) can be equal or unequal, giving flexibility to the design. Unless the experimental units are very homogeneous which is unlikely to occur with many experimental subjects, the MSE \( s_{\epsilon}^2 \) will usually be quite large. This reduces our ability to reliably detect small to moderate differences in treatment effects \( [(\alpha_i - \alpha_j) \text{ or } (\mu_i - \mu_j)] \) depending on notation used in the model.

When experimental units tend not to be homogeneous (usually the case) we can often employ a different design than CRD which helps control \( \sigma_{\epsilon}^2 \).

One such design is called the \textit{tt} randomized complete block design (RCBD). This design is used when there is an identified variable that seems to strongly contribute to experimental error. By grouping experimental units in an appropriate manner, the contribution to experimental error by this variable can be reduced or eliminated. Groups of units that have similar in values for this variable (usually measured but estimated), comprise what are called \textit{blocks}. The simplest randomized block design is one we looked at in Chapter 6 – namely the paired design.

The model for an RCBD is, in your textbook’s notation:

\[
y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \quad i = 1, 2, \ldots, t \\
\quad \quad \quad \quad j = 1, 2, \ldots b
\]
where $\mu$ is the parent population mean, $\alpha_i$ is the effect on $y$ due to the $i$-th treatment, $\beta_j$ is the effect due to block $j$ and

$$\epsilon_{ij} \sim \text{NID}(0, \sigma^2_\epsilon).$$

The expected value of $y_{ij}$ is

$$E(y_{ij}) = \mu + \alpha_i + \beta_j.$$  

For a treatment mean $\bar{y}_i$, we have

$$\bar{y}_i = \frac{\sum_{j=1}^{b} (\mu - \alpha_i + \beta_j + \epsilon_{ij})}{b} = \mu + \alpha_i + \bar{\beta} + \bar{\epsilon}_i.$$  

The difference $\bar{y}_i - \bar{y}_j$ has the form

$$\bar{y}_i - \bar{y}_j = (\alpha_i - \alpha_j) + (\epsilon_i - \epsilon_j).$$

Note that the effect of blocks is not present in this mean difference. We already used this in the 2-treatment case by using differences between pairs ( $d_j$’s), i.e., the difference within a pair.

The AOV for the randomized block design is

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Betw. Treats</td>
<td>$t - 1$</td>
<td>$b \sum_{i=1}^{t} (\bar{y}_i - \bar{y})^2$</td>
<td>$SS_T/(t - 1)$</td>
</tr>
<tr>
<td>Betw. Blocks</td>
<td>$b - 1$</td>
<td>$t \sum_{j=1}^{b} (\bar{y}_j - \bar{y})^2$</td>
<td>$SS_B/(b - 1)$</td>
</tr>
<tr>
<td>Error</td>
<td>$(b - 1)(t - 1)$</td>
<td>$\sum_i \sum_j (y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y})^2$</td>
<td>$SSE/(b - 1)(t - 1)$</td>
</tr>
<tr>
<td>Total</td>
<td>$bt - 1$</td>
<td>$\sum_i \sum_j (y_{ij} - \bar{y})^2$</td>
<td></td>
</tr>
</tbody>
</table>
The expected value of the mean square for error is the experimental error

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

and for the mean square for treatments

$$E(\text{MST}) = \sigma^2_e + b \sum_{i=1}^{t} (\alpha_i - \bar{\alpha})^2$$

To test $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_t$ we thus can use the $F$-test statistic $F = \text{MST}/\text{MSE}$ with df$_1 = t - 1$ and df$_2 = (b - 1)(t - 1)$, and reject $H_0$ for observed $F$ values larger than an $F_{\alpha}$ in the upper tail of the distribution. The assumption of normality of the $\epsilon_{ij}$’s gives us the $F$ distribution of the ratio of these mean squares when $H_0$ is true.