Design and Analysis of Experiments

After one brings a process to physical stability and quantifies what it is capable of doing, it’s reasonable to consider fundamental changes to its configuration/how it is run. Intelligent/efficient data collection and analysis aimed at finding fundamental process improvements is the subject of the final set of modules of this course. The topic is the "design and analysis of experiments," with the goal of eventually addressing complex situations where there are many "process knobs" (factors), each with multiple settings (levels) and thus many many potential ways that things could be done, and the object is to find good combinations of levels of important factors.
The Big Problem

The figure below illustrates the problem addressed in these last four modules. The noisy process output $y$ is affected by variables $x_1, x_2, x_3$ and potentially other variables (both recognized and unrecognized). The question is how to set up the "control panel" (the settings of the "knobs" or values of some variables $x_1, x_2, x_3$) to collect data and efficiently learn how to optimize the process to get desired values of $y$.

Figure: A Process With Many Inputs x or Factors Affecting a Response y
We begin with a most basic experimental scenario, where one has data consisting of observed responses, \( y \), for some number, \( r \), different processes conditions. We’ll write

\[
y_{ij} = \text{the } j\text{th response in sample } i \text{ (made under the } i\text{th set of process conditions)}
\]

where sample sizes are \( n_1, n_2, \ldots, n_r \).
A classic data set from Devore’s *Probability and Statistics for Engineering and the Sciences* concerns the current required to achieve a target brightness on a type of television tube. All combinations of 2 types of glass and 3 types of phosphor created $r = 6$ types of tubes. Tests on 3 tubes of each type produced the measured current requirements $y_{ij}$ (in $\mu$A) recorded in the following table with corresponding summary statistics.
**Samples from r Different Experimental Conditions**

Example 20-1 continued

<table>
<thead>
<tr>
<th>Type 1 Tubes</th>
<th>Type 2 Tubes</th>
<th>Type 3 Tubes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Glass 1, Phosphor 1)</td>
<td>(Glass 1, Phosphor 2)</td>
<td>(Glass 1, Phosphor 3)</td>
</tr>
<tr>
<td>( y_{11} = 280 )</td>
<td>( y_{21} = 300 )</td>
<td>( y_{31} = 270 )</td>
</tr>
<tr>
<td>( y_{12} = 290 )</td>
<td>( y_{22} = 310 )</td>
<td>( y_{32} = 285 )</td>
</tr>
<tr>
<td>( y_{13} = 285 )</td>
<td>( y_{23} = 295 )</td>
<td>( y_{33} = 290 )</td>
</tr>
<tr>
<td>( \bar{y}_1 = 285 )</td>
<td>( \bar{y}_2 = 301.67 )</td>
<td>( \bar{y}_3 = 281.67 )</td>
</tr>
<tr>
<td>( s^2_1 = 25 )</td>
<td>( s^2_2 = 58.33 )</td>
<td>( s^2_3 = 108.33 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type 4 Tubes</th>
<th>Type 5 Tubes</th>
<th>Type 6 Tubes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Glass 2, Phosphor 1)</td>
<td>(Glass 2, Phosphor 2)</td>
<td>(Glass 2, Phosphor 3)</td>
</tr>
<tr>
<td>( y_{41} = 230 )</td>
<td>( y_{51} = 260 )</td>
<td>( y_{61} = 220 )</td>
</tr>
<tr>
<td>( y_{42} = 235 )</td>
<td>( y_{52} = 240 )</td>
<td>( y_{62} = 225 )</td>
</tr>
<tr>
<td>( y_{43} = 240 )</td>
<td>( y_{53} = 235 )</td>
<td>( y_{63} = 230 )</td>
</tr>
<tr>
<td>( \bar{y}_4 = 235 )</td>
<td>( \bar{y}_5 = 245 )</td>
<td>( \bar{y}_6 = 225 )</td>
</tr>
<tr>
<td>( s^2_4 = 25 )</td>
<td>( s^2_5 = 175 )</td>
<td>( s^2_6 = 25 )</td>
</tr>
</tbody>
</table>
It is often useful to model observations from $r$ samples of respective sizes $n_1, n_2, \ldots, n_r$ as independent random samples from normal distributions with possibly different means $\mu_1, \mu_2, \ldots, \mu_r$ but a common standard deviation $\sigma$. The figure below illustrates these distributional assumptions.

**Figure:** Distributions of Responses Under $r$ Different Sets of Process Conditions
This basic "one-way normal model" is sometimes expressed in symbolic form as

\[ y_{ij} = \mu_i + \epsilon_{ij} \]

where the \( \epsilon_{ij} \) are independent normal random variables with mean 0 and standard deviation \( \sigma \).

Section 6.1 of *SQAME* discusses ways (based on examination of residuals much as in the regression analysis of Stat 231) for investigating the reasonableness of the one-way normal model in a particular application. In this module and the ones that follow, we will take for granted that such work has been taken care of, and consider what then can be done in the way of statistical inference and planning.
Where the one-way normal model is appropriate, it makes sense to pool together the $r$ sample standard deviations $s_1, s_2, \ldots, s_r$ to make a single pooled estimate of the common group standard deviation, $\sigma$. The way we will do this is to use

$$s_{\text{pooled}}^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2 + \cdots + (n_r - 1) s_r^2}{(n_1 - 1) + (n_2 - 1) + \cdots + (n_r - 1)}$$

as an estimate of $\sigma^2$, where $n = n_1 + n_2 + \cdots + n_r$ is the total number of observations in the study. This estimate of $\sigma^2$ is a weighted average of the $r$ sample variances. Corresponding to it is the estimate of $\sigma$ (the standard deviation of responses for any fixed one of the conditions $1, 2, \ldots, r$)

$$s_{\text{pooled}} = \sqrt{s_{\text{pooled}}^2}$$
The pooled sample standard deviation can be used to make confidence limits for \( \sigma \). These are

\[
\begin{align*}
    s_{\text{pooled}} \sqrt{\frac{n - r}{\chi^2_{\text{upper}}}} & \quad \text{and} \quad s_{\text{pooled}} \sqrt{\frac{n - r}{\chi^2_{\text{lower}}}}
\end{align*}
\]

where the appropriate degrees of freedom are \( \nu = n - r \).
The $r = 6$ sample standard deviations for the different tube types in the glass-phosphor study are pooled to make

$$s_{\text{pooled}} = \sqrt{\frac{2 \times 25 + 2 \times 58.33 + 2 \times 108.33 + 2 \times 25 + 2 \times 175 + 2 \times 25}{18 - 6}}$$

$$= 8.3 \mu A$$

This intends to measure the variation in current required to produce the standard brightness in tubes of any single type. Appropriate degrees of freedom for this estimate are $\nu = 18 - 6 = 12$ and 95% confidence limits are

$$8.3 \sqrt{\frac{12}{23.337}} \quad \text{and} \quad 8.3 \sqrt{\frac{12}{4.404}}$$

that is

$$6.0 \mu A \quad \text{and} \quad 13.8 \mu A$$
Estimating Linear Combinations of Means

σ (and its estimate, $s_{\text{pooled}}$) is a measure of basic background noise in an experiment, a baseline against which any apparent differences in average response for different conditions are to be measured. One specific way in which these comparisons can be made, is to make confidence limits for interesting linear combinations of means $\mu_1, \mu_2, \ldots, \mu_r$. That is, we’ll let

$$L = c_1 \mu_1 + c_2 \mu_2 + \cdots + c_r \mu_r$$

stand for an arbitrary linear combination of group mean responses. The corresponding linear combination of sample means is

$$\hat{L} = c_1 \bar{y}_1 + c_2 \bar{y}_2 + \cdots + c_r \bar{y}_r$$

and is an obvious estimate of $L$. Useful special cases of this formalism are:

- $c_i = 1$ and all others 0: $L = \mu_i$, $\hat{L} = \bar{y}_i$;
- $c_i = 1$, $c_i' = -1$ and all others 0: $L = \mu_i - \mu_i'$, $\hat{L} = \bar{y}_i - \bar{y}_i'$

The first of these is the mean for condition $i$ and the second is the difference between the condition $i$ and condition $i'$ response means.
Confidence limits for $L$ can be based on $\hat{L}$ and $s_{\text{pooled}}$ as

$$\hat{L} \pm t s_{\text{pooled}} \sqrt{\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \cdots + \frac{c_r^2}{n_r}}$$

The degrees of freedom for $t$ are those associated with $s_{\text{pooled}}$, namely $\nu = n - r$. The quantity $s_{\text{pooled}} \sqrt{\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \cdots + \frac{c_r^2}{n_r}}$ is an estimate of the standard deviation of $\hat{L}$, and $t$ times this is a kind of "margin of error" for estimating $L$. 
In the glass-phosphor study, 95% confidence limits for the mean current requirement for tube type $i$ are

$$\bar{y}_i \pm t_{s_{\text{pooled}}} \sqrt{\frac{(1)^2}{n_i}}$$

that is

$$\bar{y}_i \pm 2.179 \times (8.3) \sqrt{\frac{1}{3}} \text{ or } \bar{y}_i \pm 10.44 \mu A$$

That is, each of the 6 sample means is in some sense "good to within 10.44 $\mu A" as representing the corresponding tube mean current requirement.
As a second use of the formula for confidence limits for $L$, consider the estimation of the difference in current requirement means for tube type $i$ tube type $i'$. Limits are

$$
\bar{y}_i - \bar{y}_{i'} \pm t_{s_{pooled}} \sqrt{\frac{(1)^2}{n_i} + \frac{(-1)^2}{n_{i'}}}
$$

So 95% confidence limits for $\mu_i - \mu_{i'}$ in the glass phosphor problem are

$$
\bar{y}_i - \bar{y}_{i'} \pm 2.179 (8.3) \sqrt{\frac{1}{3} + \frac{1}{3}} \quad \text{or} \quad \bar{y}_i - \bar{y}_{i'} \pm 14.77
$$

So, for example, limits for comparing tube types 1 and 2 are

$$
(285 - 301.67) \pm 14.77 \quad \text{i.e.} \quad -16.67 \pm 14.77
$$
There is clear evidence of a difference between current requirement means for tube types 1 and 2 since the margin of error of estimation (14.77) is smaller than the absolute value of the observed difference in sample means (16.67).

On the other hand, 95% confidence limits for comparing tube types 1 and 3 are

\[(285 - 281.67) \pm 14.77 \text{ i.e. } 3.33 \pm 14.77\]

and since the "margin of error in estimation" (14.77) is larger than the observed absolute difference in sample means (3.33) there is no clear evidence of a difference between means for tube types 1 and 3.