Garrett Dancik	http://www.public.iastate.edu/~gdancik/stat430	gdancik@iastate.edu

Last Updated: October 29, 2007

Experimental Design

1 Basic Principles

A designed experiment is a controlled study in which one or more *treatments* are applied to *experimental units*. The experimenter then observes the effect of varying these treatments on a *response variable*. The response variable is the variable of interest, and can be quantitative or qualitative. Response variables are also referred to as dependent variables. Designed experiments should be contrasted with *observational* studies, where a researcher measures characteristics of a population without manipulating or influencing any variables of interest. It is important to realize that observational studies do not allow a researcher to claim causation (although many people make this mistake), because they do not control for *lurking* or *extraneous* variables.

Key features of all experiments include (at least the first three):

- control - to eliminate the effect of one or more variables on a response
- randomization - to 'average out' the effects of extraneous variables
- replication - to separate the signal from the noise
- blocking - to increase the amount of signal relative to the noise

An experimental design specifies how we assign treatments to experimental units. What are good properties of an experimental design? Before selecting an experimental design, it is important to have a clear picture of the *purpose* of the experiment, and of what we hope to obtain from our statistical model.

2 Interaction Effects

Before continuing, it will be useful to introduce the idea of an interaction effect. When the effect that a parameter has on a response depends on the value(s) of one (or more) other parameters, we say that there exists an *interaction effect* between the parameters.

Suppose were are interested in the relationship between the weight and height of students in a class (note that this is not a designed experiment!). Since we expect that males and females that are the same height do not have equal weights (on average), we might include gender as a (coded) variable in our model.

$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_{ij}$, where $\epsilon_{ij} \sim$ independent $N(0, \sigma^2)$

where y_i is the weight of individual i, x_{1i} is the height of individual i, and x_{2i} is equal to $+1$ if invidual i is male and -1 if individual i is female. We will not concern ourselves with fitting such a model, but note that this can be done using the *lm* function in R.

However, let's look at the implications of such a model.

$E[y_i|male, x_{1i}] = (\beta_0 + \beta_2) + \beta_1 x_{1i}$, and
$E[y_i|female, x_{1i}] = (\beta_0 - \beta_2) + \beta_1 x_{1i}$.

Note that if we were to graph the expectations, we end up with two parallel lines (the only difference in the lines being the y-intercept). Importantly, this formulation of the model assumes that the linear relationships between height and weight (the slopes of the lines) are the same for both males and females. In other words, the model assumes that there is no interaction between height and gender.

But what if we did not want to make such an assumption? In that case we add what is known as an *interaction term* to the model (specifically a height by gender interaction term).

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i} + \epsilon_{ij}$$

What are the implications of this model?

$E[y_i|x_{1i}, \text{male}] = (\beta_0 + \beta_2) + (\beta_1 + \beta_{12}) x_{1,i}$, and
$E[y_i|x_{1i}, \text{female}] = (\beta_0 - \beta_2) + (\beta_1 - \beta_{12}) x_{1,i}$.

With this model, the linear relationship between height and weight depends on the gender of the individual, and graphing the two expectation functions will not result in parallel lines. Graphs of the two above models are provided in the attached handout.

3 Completely Randomized Designs (CRDs)

A *completely randomized design* (CRD) is a design where treatments are assigned to experimental units completely at random, subject to the number of observations / treatment.

3.1 One-way ANOVA

Recall the model for the One-way ANOVA:

$$y_{ij} = \theta_i + \epsilon_{ij}, \quad i = 1, \ldots a, \quad j = 1 \ldots r_i, \quad \text{(r for replicates)}, \quad N = \sum_{i=1}^{a} r_i$$

which can be reparameterized as

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \ldots a, \quad j = 1 \ldots r_i, \quad N = \sum_{i}^{a} r_i.$$

Assumptions: $\epsilon_{ij} \sim N(0, \sigma^2)$ independent for all i,j. Is this assumption reasonable? There are methods for assessing normality, such as normal quantile plots, which we will not cover. But note that it is often reasonable to assume that $E[\epsilon_{ij}] = 0$ and that the ϵ_j's are independent when a CRD is used.

The ANOVA table for this model is given below:

Source	df	SS	MS
Treatment (SS_T)	a-1	$\sum_{i=1}^{a} r_i \left(\overline{y}_{i.} - \overline{y}_{..}\right)^2$	$MS_T = SS_T/(a-1)$
Error (SS_E)	$\sum_{i}^{a} (r_i - 1)$	$\sum_{i=1}^{a}\sum_{j=1}^{r_i} \left(y_{ij} - \overline{y}_{i.}\right)^2$	$MS_E = SS_E/\left(\sum_{i}^{a} (r_i - 1)\right)$
Totals	$\sum_{i}^{a} r_i - 1$	$\sum_{i=1}^{a}\sum_{j=1}^{r_i} \left(y_{ij} - \overline{y}_{..}\right)^2$	

Information from the ANOVA table can be used to test the classical ANOVA hypothesis:

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_a$$

$$H_1 : \theta_i \neq \theta_j \text{ for some } i \neq j$$

Under the null hypothesis,

$$F = \frac{SS_T/(a-1)}{SS_E/\left(\sum_{i=1}^{a} (r_i - 1)\right)} \sim F_{a-1,a(r-1)}$$

For the model above, H_0 usually is not of interest, but similar F-tests can be used to compare nested models, to test interaction effects, and to perform lack of fit tests (i.e., is the model adequate). We will talk about F-tests for testing interaction effects a little bit later. In the model above, interest is usually in treatment comparisons.

Recall that $\overline{y}_{i.} \sim N\left(\theta_i, \sigma^2/r_i\right)$, and $\sum_{i=1}^{a} c_i \overline{y}_{i.} \sim N\left(\sum_{i=1}^{a} \theta_i, \sum_{i=1}^{a} c_i^2 \sigma^2/r_i\right)$.

This leads to the following $(1-\alpha)\%$ CI for $\sum_{i=1}^{a} c_i \theta_i$:

$$\sum_{i=1}^{a} c_i \overline{y}_{i.} \pm t_{N-a,1-\alpha/2}\ s \sqrt{\sum_{i=1}^{a} c_i^2/r_i},$$

where $s = \sqrt{MS_E}$ from above.

How might we make the confidence interval smaller?

4 Randomized Complete Block Designs (RCBDs)

4.1 Blocking

Blocking allows us to decrease our experimental error (σ^2) because within each block, our experimental units are more similar. Blocking is one mechanism for producing smaller confidence intervals and more powerful tests. However, experimental units must be blocked in such a way that blocks are not confounded with any factors of interest.

Blocking in conjunction with a one-way ANOVA gives us the following model:

$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$,

where $i = 1, \ldots, a$, $k = 1, \ldots, r_i$, and β_j corresponds to the effect of being in block j. The assumptions on the $\epsilon's$ are the same as above.

How do we choose to assign a particular treatment a particular individual, given that they are in block j?

Consider the following (silly) experimental design.

Response	Block	Treatment
y_{111}	1	1
y_{112}	1	1
y_{221}	2	2
y_{222}	2	2

The design above is a faulty design because we cannot make inference on any contrast of the θ_i's. Can you think of a design that would allow us to make inference on all contrasts?

A randomized complete block design (RCBD) is a design that involves blocking of the experimental units, and implementing a CRD within each block. A RCBD, as we will see, allows us to make inference on any contrast of the θ_i's.

4.2 RCBD in the One-way ANOVA

Now consider the model

$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$, where $i = 1, \ldots, a, j = 1, \ldots, b$.

There is a small but very important distinction between this model and the model in section 4.1. The current model along with the description $i = 1, \ldots, a, j = 1, \ldots, b$ implies that observations exist for all combinations of i and j. In other words, every treatment appears in every block, which is a necessary (though not sufficient) requirement in a RCBD.

We will assume that there is no replication so that $N = ab$, which allows us to fairly compare RCBDs with CRDs when the number of blocks in the RCBD is equal to the total number of replicates in the CRD (if ~~$b = \sum_{i=1}^{a} r_i$~~ $r_i = r = b$ for all i, then N is the same in both designs).

Let us verify that with a RCBD, we can make inference on all contrasts of the $\theta_i's$:

$$\mathrm{E}\left[\sum_{i=1}^{a} c_i y_{ij}\right] = \sum_{i=1}^{a} c_i(\mu + \alpha_i + \beta_j) = \sum_{i=1}^{a} c_i(\theta_i + \beta_j)$$

$$\mathrm{E}\left[\sum_{i=1}^{a} c_i \overline{y}_{i.}\right] = \sum_{i=1}^{a} c_i(\theta_i + \overline{\beta}) = \sum_{i=1}^{a} c_i(\theta_i) \text{ if } c = (c_1, \ldots c_a) \text{ is a contrast.}$$

Now, similar to the one-way ANOVA, a $(1-\alpha)\%$ CI for $\sum_{i=1}^{a} c_i\theta_i$ is

$$\sum_{i=1}^{a} c_i \overline{y}_{i.} \pm t_{N-a-b+1,1-\alpha/2}\ s\sqrt{\sum_{i=1}^{a} c_i^2/r_i},$$

where $s = \sqrt{MS_E}$ from the ANOVA table below.

Recall the suggestion that blocking allows us to increase the strength of the signal relative to the noise. Let's see how this plays out for the construction of confidence intervals of the form $\sum_i c_i\theta_i$.

The length of the intervals are (assuming ~~$\sum_{i=1}^{a} r_i = r = b$~~ $r_i = r = b$ for all i, so the comparison is more clear):

CRD: $2t_{n-a,1-\alpha/2}\ s_{CRD}\sqrt{\sum_{i=1}^{a} c_i^2/r}$

RCBD: $2t_{n-a-b+1,1-\alpha/2}\ s_{RCBD}\sqrt{\sum_{i=1}^{a} c_i^2/r}$

The expected length of the squared intervals are:

CRD: $4t^2_{n-a,1-\alpha/2}\ \sigma^2_{CRD} \sum_{i=1}^{a} c_i^2/r$

RCBD: $4t^2_{n-a-b+1,1-\alpha/2}\ \sigma^2_{RCBD} \sum_{i=1}^{a} c_i^2/r$

RCBD will result in a smaller interval if

$\sigma_{RCBD}/\sigma_{CRD} < t_{n-a,1-\alpha/2}/t_{n-a-b+1,1-\alpha/2}$

For large n, the t-ratio is slightly less than 1. For example, when $n = 30$, $a = 10$, $b = 3$, $\alpha = 0.05$, the ratio is 0.99. In practice, there is usually an advantage to a RCBD as long as σ_{RCBD} is slightly less than σ_{CRD}.

The ANOVA decomposition for the RCBD is as follows

Source	df	SS	MS
Treatment	a-1	$\sum_{i=1}^{a} b\,(\overline{y}_{i.} - \overline{y}_{..})^2$	$MS_T = SS_T/(a-1)$
Blocks	b-1	$\sum_{j=1}^{b} a\,(\overline{y}_{.j} - \overline{y}_{..})^2$	$MS_B = SS_B/(b-1)$
Error*	$(b-1)(a-1)$	$\sum_{i=1}^{a}\sum_{j=1}^{b} (y_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..})^2$	$MS_E = SS_E/\left((b-1)(a-1)\right)$
Totals	$ab-1$	$\sum_{i=1}^{b}\sum_{j=1}^{r} (y_{ij} - \overline{y}_{..})^2$	

4.3 Factorial designs

Factorial designs are used to study the joint effects of multiple factors on a response. For example, a factorial design would be used to analyze the effect of temperature, humidity, and growth medium on the growth rate of bacteria (or similarly, the effect of temperature and humidity on the lifespan of a battery). In a factorial design, the experimenter considers a fixed number of levels for each factor, and runs an experiment which looks at all combinations of the factors. If the experiment includes replication, all treatment combinations are replicated. Therefore, if there are a levels of factor A, b levels of factor B, and c levels of factor C, there will be abc treatment combinations. Factors can be quantitative, such as temperature and humidity, or qualitative, such as two different machines, operators, or the presence or absence of a factor.

When there are k factors that each occur at two levels, there are 2^k treatment combinations. A factorial design that looks at all 2^k treatment combinations is known as the 2^k factorial design, and is the most common class of factorial designs. In this section, we will cover the 2^2 factorial design in detail, and provide results for the 2^3 factorial design.

4.3.1 2^2 factorial design

The factorial effects model for the 2^2 factorial design is

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk},$$

where $i = 1, 2, j = 1, 2$, and $k = 1, \ldots, r$. The total number of observations is $N = 4r$. For convenience, we will refer to the two levels of each factor as the "low (-)" and "high (+)" levels of that factor.

This model is very similar to the RCBD model in section (4.1), where we treated the block as the factor. However, now we include the interaction effect $(\alpha\beta)_{ij}$.

Of interest is how the response changes as one or more factors are changed. Specifically, the *main* effect of a factor is the difference between the average response at the high level of that factor and the average response at the low level of that factor.

For example, the main effect of factor A is

$$\text{(average response at } A^+) - \text{(average response at } A^-) = \frac{\sum_{j=1}^{2}\sum_{k=1}^{r} y_{2jk}}{N/2} - \frac{\sum_{j=1}^{2}\sum_{k=1}^{r} y_{1jk}}{N/2}$$

Recalling our discussion about interaction effects, we know that the effect that factor A has on the response might be different for different levels of B. The AB interaction effect measures this, and is defined as the average difference between the main effect of A at B^+ and the main effect of A at B^- (or equivalently the average difference between the main effect of B at the two levels of A).

Let's look at an example from a 2^2 design with no replicates.

Suppose that we observed $y_{111} = 20$, $y_{121} = 30$, $y_{211} = 40$, $y_{221} = 52$, from the following design:

y_{ijk}	μ	α	β	$(\alpha\beta)_{ij}$
y_{111}	$+1$	-1	-1	$+1$
y_{121}	$+1$	-1	$+1$	-1
y_{211}	$+1$	$+1$	-1	-1
y_{221}	$+1$	$+1$	$+1$	$+1$

Let us calculate and visualize the main effects for A and B, as well as the AB interaction effect...

In the design notation above, define x_e to be a vector of the elements in the column corresponding to effect e. For example, $x_\alpha = (-1, -1, +1, +1)$. Also define $x_{\alpha,ijk}$ to be the element of x_α corresponding to observation y_{ijk}. Note that the column corresponding to $(\alpha\beta)_{ij}$ is the element by element product of x_α and x_β.

Using this design notation, it is straightforward to calculate the A and B main effects and the AB interaction effect:

$$A = \frac{\sum_{i,j,k} x_{\alpha,ijk} \times y_{ijk}}{N/2} \quad B = \frac{\sum_{i,j,k} x_{\beta,ijk} \times y_{ijk}}{N/2} \quad AB = \frac{\sum_{i,j,k} x_{(\alpha\beta),ijk} \times y_{ijk}}{N/2}$$

Furthermore, with the constraint that $\sum_i \alpha_i = 0$ and $\sum_j \beta_j = 0$, the estimate of the effects have an appealing interpretation and are the following:

$$\hat{\mu} = \frac{1}{N}\sum_{i,j,k} y_{ijk} = \overline{y}_{...}$$
$$\hat{\alpha}_i = \frac{1}{N/2}\sum_{j,k} y_{ijk} - \hat{\mu} = \overline{y}_{i..} - \hat{\mu}, \quad i = 1, 2$$
$$\hat{\beta}_j = \frac{1}{N/2}\sum_{i,k} y_{ijk} - \hat{\mu} = \overline{y}_{.j.} - \hat{\mu}, \quad j = 1, 2$$
$$\hat{(\alpha\beta)}_{ij} = \frac{1}{r}\sum_{k} y_{ijk} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\mu} = \overline{y}_{ij.} - (\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j), \quad i = 1, 2, \ j = 1, 2$$

Computationally, it is often convenient to use the fact that estimates of any parameter $(-)$ are found using the formula

$$\frac{-1^t \sum_{i,j,k} x_{(-),ijk} \times y_{ijk}}{N}, \tag{1}$$

where t is the number of $1's$ in the subscript of the parameter. For example, $t = 1$ for the parameters α_1, $(\alpha\beta)_{21}$, and $t = 0$ for the parameters μ, α_2 and $(\alpha\beta)_{22}$.

The ANOVA table for a two-factor factorial with $i = 1, \ldots, a, j = 1, \ldots, b$, and $k = 1, \ldots, r$ replicates is given below. This table is general for all two factor factorials, for any $a \geq 2, b \geq 2$.

Source	df	SS	MS
Treatment A (SS_A)	$a-1$	$\sum_{i=1}^{a} br\left(\overline{y}_{i..} - \overline{y}_{...}\right)^2$	$MS_A = SS_A/(a-1)$
Treatment B (SS_B)	b-1	$\sum_{j=1}^{b} ar\left(\overline{y}_{.j.} - \overline{y}_{...}\right)^2$	$MS_B = SS_B/(b-1)$
AB Interaction (SS_{AB})	$(a-1)(b-1)$	$\sum_{i=1}^{a}\sum_{j=1}^{b} r\left(\overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...}\right)^2$	$MS_{AB} = SS_{AB}/\left((a-1)(b-1)\right)$
Error (SS_E)	$ab(r-1)$	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}\left(y_{ijk} - \overline{y}_{ij.}\right)^2$	$MS_E = SS_E/\left(ab(r-1)\right)$
Totals	$abr-1$	$\sum_{i=1}^{b}\sum_{j=1}^{b}\sum_{k=1}^{r}\left(y_{ijk} - \overline{y}_{...}\right)^2$	

4.3.2 2^3 factorial design

The full factorial effects model for the 2^3 factorial design is

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkl},$$

where $i = 1, 2, j = 1, 2, k = 1, 2$, and $l = 1, \ldots, r$. The total number of observations is $N = 2^k r = 8r$.

Main effects and interaction effects are calculated in a similar manner as for the 2^2 factorial model. Using the same coding scheme we used in (4.3.1), estimates for any parameter in the 2^3 model can also be found using equation (1).

4.4 Experimental designs: notions of optimality

Notions of optimality apply to both regression and factorial models. For example, models of the form

- $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ (simple linear regression)
- $Y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \epsilon_i$ (multiple linear regression)
- $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$ (multiple linear regression)
- $Y_i = \beta_0 + \exp(\beta_1 x_i) + \epsilon_i$ (non-linear regression)
- $Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + \epsilon_{ijkl}$ (factorial model with single interaction term)

What was the purpose of the original experiment? What do we hope to learn from our statistical model?

Optimality criteria:

- A-optimality - minimize the average variance of all parameters
- D-optimality - minimize the volume of the confidence region of all parameters
- V-optimality - minimize the average prediction variance, where average is taken with respect to a design region (this is sometimes called I- or Q-optimality)
- G-optimality - minimize the maximum prediction variance over a set of specified design points

Finding an optimal design typically depends on the model, the sample size, and the experimental region. However, there are some common designs that have been shown to have optimality properties. For example, the 2^2 factorial model without an interaction term is A,D, and G-optimal.