1 Chapter 1: Research Design Principles

The legacy of Sir Ronald A. Fisher. Fisher's three fundamental principles: local control, replication, and randomization.

Other topics of interest: Experimental units vs. observational units; Comparative experiments vs. comparative observational studies; Experimental error vs. residual error; How many replications?; Relative Efficiency.

2 Chapter 2: Completely Randomized Designs

2.1 Research Design

The concept of a research design is introduced. A research design consists of a research hypothesis, a treatment design, and an experimental or observational study design. In this chapter we consider the simple research design in which the research hypothesis is that some treatments are different than others, the treatment design is a one-way classification, and the experimental design is a completely randomized design.

[**Not all researchers and authors makes a distinction between treatment designs and experimental designs, but it is useful**]

2.2 How to randomize, preparing the data file

The text discusses both randomization for an experimental design, and for an observational study. Then they describe preparation of a data file for analysis, which we illustrate in our code examples.

2.3 The cell means model

The first model we consider for a set of treatments is the cell means model:

$$y_{ij} = \mu_i + e_{ij},\ i = 1, 2, \ldots, t;\ j = 1, 2, \ldots, r$$

This is considered a full model (representing an alternative hypothesis such as $H_a : \mu_i \neq \mu_k$ for some $i \neq k$), and to detect treatment effects we can compare it to a reduced model (representing the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_t$) in which the mean is the same for all groups:

$$y_{ij} = \mu + e_{ij},\ i = 1, 2, \ldots, t;\ j = 1, 2, \ldots, r$$

Both of these models (as well as the treatment effects model to be introduced later) are special cases of the general linear model, as discussed on pages 45-47.

2.4 Parameter estimation

The various μ_i or μ parameters in the two models above must be estimated from data. The traditional way to do this is via the principle of least squares. According to this principle, the estimators $\widehat{\mu}_i$ of the μ_i are the values that minimize the sum of squared errors:

$$Q = \sum (y_{ij} - \mu_i)^2$$

with respect to the μ_i. As shown in the text, for the models above the estimators turn out to be sample means.

2.5 The treatment effects model

An alternative model can be used by considering the group means under H_a and their differences from the overall mean $\overline{\mu}_{\cdot} = \sum_{i=1}^{t} \mu_i / t$:

$$y_{ij} = \overline{\mu}_{\cdot} + (\mu_i - \overline{\mu}_{\cdot}) + (y_{ij} - \mu_i),$$

which can be rewritten as:

$$y_{ij} = \mu + \tau_i + e_{ij}$$

The hypothesis of no group effect for this (full) model is $H_0 : \tau_1 = \tau_2 = \cdots = \tau_t = 0$, and the reduced model is the same as above for the cell means model.

2.6 Analysis of Variance

The key idea behind an analysis of variance involves a decomposition of the total sum of squares. For a one-way ANOVA (possibly arising from a completely randomized design) the decomposition is SS Total = SS Treatment + SS Error.

If y_{ij} is the jth observation in group i, then $y_{ij} - \overline{y}_{\cdot\cdot} = (y_{ij} - \overline{y}_{i\cdot}) + (\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})$, which leads to:

$$\sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij} - \overline{y}_{\cdot\cdot})^2 = \sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij} - \overline{y}_{i\cdot})^2 + \sum_{i=1}^{t}\sum_{j=1}^{r}(\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})^2, \text{ or } SS \text{ Total} = SS \text{ Error} + SS \text{ Treatment}.$$

As an example, consider three groups with the following data: Group 1 has y_{1j} values of 1, 2, and 3, Group 2 has y_{2j} values of 5, 3, and 4, and Group 3 has y_{3j} values of 6, 7, and 5. The overall sample mean is $\overline{y}_{\cdot\cdot} = (\sum_{i=1}^{t}\sum_{j=1}^{r} y_{ij})/tr = 36/9 = 4$. Then SS Total is

$$\sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij} - \overline{y}_{\cdot\cdot})^2 = (1-4)^2 + (2-4)^2 + ... + (5-4)^2 = 30.$$

The group means are $\overline{y}_{1\cdot} = 2, \overline{y}_{2\cdot} = 4$, and $\overline{y}_{3\cdot} = 6$, so SS Error and SS Treatment are

$$SS \text{ Error} = \sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij} - \overline{y}_{i\cdot})^2 = (1-2)^2 + (2-2)^2 + (3-2)^2 + (5-4)^2 + ... + (5-6)^2 = 6, \text{ and}$$

$$SS \text{ Treatment} = \sum_{i=1}^{t}\sum_{j=1}^{r}(\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})^2 = \sum_{i=1}^{t} r(\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})^2 = 3(2-4)^2 + 3(4-4)^2 + 3(6-4)^2 = 24.$$

Thus SS Total = SS Error +SS Treatment or $30 = 6 + 24$ partitions the total sum of squares about the overall mean into two parts, one within groups (due to error, or effects not accounted by the model)

and one between groups (measuring the difference between sample means). Since each group here has r observations, each group contributes $r-1$ degrees of freedom for the within group sum of squares, for a total of $t(r-1)$ degrees of freedom for SS Error. SS Treatment is calculating the sum of squares of t sample means about their (overall) mean, so it has $t-1$ degrees of freedom. For the example data above, $t(r-1)$ = 3(2) = 6, and $t-1$ = 3 - 1 = 2. We can summarize this information in an analysis of variance table:

Source	SS	df	MS	F
Treatments	24	2	12	12
Error	6	6	1	
Total sum of squares	30	8		

As seen in the ANOVA table above, we divide sums of squares by their degrees of freedom to obtain mean squares for both treatments and error. The F statistic is then the mean square for treatments (MST) divided by the mean square for error (MSE). To test the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$ against the alternative hypothesis $H_a : \mu_i \neq \mu_k$ for some $i \neq k$, we compare the F statistic to an F distribution with numerator df $= t-1$ = 3 - 1 = 2, and denominator df $= t(r-1) = 3(2) = 6$. When group sample sizes are unequal we replace r in the expressions by r_i, which is the sample size in the ith group.

[**The above decomposition of research data into sums of squares and degrees of freedom, along with the determination of how to form F ratios may not always be the most efficient way to analyze a set of data. However, it is probably the best way to think about how to analyze data.**]

As is shown at the end of Chapter 2, $E(MSE) = \sigma^2$ and $E(MST) = \sigma^2 + r\theta_t^2$ where

$$\theta_t^2 = \frac{\sum \tau_i^2}{(t-1)},$$

so when the null hypothesis is true, $E(MSE) = E(MST)$ and the F statistic value should be close to 1.

2.7 A general principle for GLM tests, standard errors and conf. intervals

As noted in the text, a general way to conduct tests for linear models (and even more generally) is to obtain full and reduced models as illustrated above. An F test is obtained by measuring the reduction in sum of squared error in the two models compared to the sum of squared error for the full model, with both terms divided by their degrees of freedom to create mean squares:

$$F = \frac{(SSE_r - SSE_f)/(t-1)}{SSE_f/(N-t)}$$

Under the standard ANOVA assumptions, this F statistic follows an F distribution with $t-1$ and $N-t$ degrees of freedom.

Standard errors for group means are obtained by dividing a variance estimate by the group sample size, so a standard error for the mean $\overline{y}_i$ will be:

$$s_{\overline{y}_i} = \sqrt{\frac{s^2}{r}},$$

where r is the group sample size and

$$s^2 = \frac{\sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij} - \overline{y}_{i.})^2}{t(r-1)} = \frac{SSE}{t(r-1)}$$

is the mean squared error for the data.

2.8 Power and sample size for completely randomized designs

Power and sample size analyses are important tools for assessing the ability of a statistical test to detect when a null hypothesis is false, and for deciding what sample size is required for having a reasonable chance to reject a false null hypothesis.

Recall that for a test of a statistical hypothesis, the Type I error (α) is the probability of rejecting the null hypothesis when it is true. The Type II error (β) is the probability of not rejecting the null hypothesis when it is false. The power of the test equals $1-\beta$, and is the probability of rejecting the null hypothesis when it is false. The power will depend on the alternative hypothesis, and we would like to have high power to detect alternative hypotheses of interest.

For the completely randomized design with one-way treatment structure, when the null hypothesis is true, the F statistic has an F distribution with $t-1$ and $t(r-1)$ degrees of freedom. When the null hypothesis is false, the F statistic follows a non-central F distribution with $t-1$ and $t(r-1)$ degrees of freedom, and noncentrality parameter:

$$\lambda = \frac{r\sum\tau_i^2}{\sigma^2}.$$

The power of the F test is a monotonically increasing function of the parameter λ. Notice that when the null hypothesis is true, $\lambda = 0$, so that the usual (or central) F distribution is just a special case of the non-central F distribution. It should make sense intuitively that the power of the F test increases as r increases, as $\sum\tau_i^2$ increases, and as σ^2 decreases, as predicted by the noncentrality parameter λ.

Note: the text also introduces a related measure: $\Phi = \sqrt{\frac{\lambda}{t}}$.

3 References

Casella, G. 2008. Statistical Design, Springer.

Kuehl, R.O. 2000. Statistical Principles of Research Design and Analysis (second edition), Duxbury Press.

Kirk, R.E. 1994. Experimental Design: Procedures for Behavioral Sciences (third edition), Wadsworth Publishing.

Oehlert, G.W. 2000. A First Course in Design and Analysis of Experiments, W.H. Freeman. (now out of print, but available for free under Creative Commons license at Prof. Oehlert's website at the University of Minnesota)