1 Chapter 1: Research Design Principles

The legacy of Sir Ronald A. Fisher.

Fisher's three fundamental principles: local control, replication, and randomization.

2 Chapter 2: Completely Randomized Design

2.1 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.

Least squares Estimation of model parameters.

2.2 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.

2.3 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 $\overline{y}_{..} = (\sum_{i=1}^{t}\sum_{j=1}^{r} y_{ij})/tr = 36/9 = 4$. Then SS Total is

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

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

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

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

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.

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.

2.4 A general principle for GLM tests

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 statistics follows an F distribution with $t-1$ and $N-t$ degrees of freedom.

Calculation of standard errors for group means are also discussed in the text.

2.5 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 $1-$ β , 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.

3 Chapter 3: Treatment Comparisons

Usually we have more specific research questions than just rejecting the ANOVA null hypothesis that all groups have the same mean ($H_0 : \mu_1 = \mu_2 = ... = \mu_t$).

Common questions:

1) While planning an experiment, you wish to test certain hypotheses that are a subset of the global ANOVA H_0. (*A priori* tests)

2) After a significant ANOVA, which groups differ? (*Post hoc* tests)

For example, the text discusses several questions that arise after finding that the ANOVA global $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ is rejected with the meat storage data.

3.1 Terminology

When describing treatment comparisons, it is useful to introduce some terminology. The term **contrast** is used to describe a comparison of means. Specifically, a contrast is a linear combination of the population means,

$$C = \sum_{i=1}^{t} k_i \mu_i \text{ that also satisfies } \sum_{i=1}^{t} k_i = 0.$$

We require that the coefficients k_i sum to zero so that the comparison is meaningful (we would not be interested in $\mu_1 - 3\mu_2$ for example). For example, if we had 6 treatment groups and we wished to test whether the the first two group means differed, we can express the null hypothesis as $H_0 : 1\mu_1 - 1\mu_2 = \mu_1 - \mu_2 = 0$. Here k_1=1 and k_2= -1, so $k_1 + k_2 = 0$ as required. This is an example of a **pairwise contrast**, which is defined as a contrast involving only two groups. An example of a **non-pairwise contrast** would be if we wished to test if the mean of the first two groups differed from the mean of the other four groups. We can express this null hypothesis as $H_0 : (\mu_1 + \mu_2)/2 - (\mu_3 + \mu_4 + \mu_5 + \mu_6)/4 = 0$. Here k_1=1/2, k_2= 1/2, k_3= -1/4, k_4= -1/4, k_5= -1/4 and k_6= -1/4, so again $k_1 + k_2 + k_3 + k_4 + k_5 + k_6 = 0$, as required by the definition of a contrast. One other property of a set of contrasts, called orthogonality, is useful when considering *a priori* tests. Two contrasts

$$C_1 = \sum_{i=1}^{t} k_{1i}\mu_i \text{ and } C_2 = \sum_{i=1}^{t} k_{2i}\mu_i \text{ are } \textbf{orthogonal} \text{ if } \sum_{i=1}^{t} \frac{k_{1i}k_{2i}}{n_i} = 0.$$

If the group sample sizes are equal then this is equivalent to $\sum k_{1i}k_{2i} = 0$. In the examples above, if we identify the first contrast as C_1 and the second contrast as C_2, then k_{11}=1, k_{12}= -1, $k_{13} = k_{14} = k_{15} = k_{16} = 0$, are the coefficients for C_1 and k_{21}=1/2, k_{22}= 1/2, k_{23}= -1/4, k_{24}= -1/4, k_{25}= -1/4 and k_{26}= -1/4 are the coefficients for C_2. Then if the sample sizes are equal,

$$\sum_{i=1}^{t} k_{1i}k_{2i} = (1)(1/2) + (-1)(1/2) + (0)(-1/4) + (0)(-1/4) + (0)(-1/4) + (0)(-1/4) = 0,$$

so C_1 and C_2 are orthogonal. Orthogonal contrasts are statistically independent, so that the outcome of testing one contrast is independent of the outcome of testing the other contrast. A set of more than two contrasts is **mutually orthogonal** if each pair of contrasts in the set is orthogonal to each other. The concept of a contrast or a set of contrasts at first seems somewhat esoteric, but in fact it is essential to understand these concepts to fully understand ANOVA, particularly in complicated situations. The contrasts that you use should depend on the questions of scientific interest from your research.

4 Inference for contrasts

We can estimate the contrast

$$C = \sum_{i=1}^{t} k_i \mu_i$$

with the observed sample contrast:

$$c = \sum_{i=1}^{t} k_i \overline{y}_{i.}$$

The text presents formulas for the variance of a sample contrast that can be used to obtain a t test for the null hypothesis $H_0 : C = 0$.

4.1 Orthogonal contrasts can form a partition of SS Treatment

As noted in the text, we can form as many orthogonal contrasts as we have degrees of freedom for the treatment factor. These orthogonal contrasts partition the SS Treatment into SSC_i terms that allow us to separate the total sum of squares for treatments into parts attributable to different contrasts. Especially when you have equal sample sizes, this can be a powerful tool for understanding treatment effects. One common example of this is when we use orthogonal polynomial contrasts to partition quantitative treatment effects into parts attributable to linear trend, quadratic trend, and higher-order trends. For equally-spaced dosage levels with equal-sample-size groups, the coefficients for orthogonal polynomial contrasts are given in Table XI on page 623 of our text.

The SAS code for this lecture illustrates the use of orthogonal contrasts for the meat storage data, and also the use of the ORPOL function to obtain orthogonal polynomial coefficients for quantitative treatment levels. The code also shows the use of Proc Reg for obtaining predictions from response curves and standard errors of the predictions.