

Chapter 3 Supplemental Text Material

S3-1. The Definition of Factor Effects

As noted in Sections 3-2 and 3-3, there are two ways to write the model for a single-factor experiment, the means model and the effects model. We will generally use the effects model

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases}$$

where, for simplicity, we are working with the balanced case (all factor levels or treatments are replicated the same number of times). Recall that in writing this model, the i th factor level mean μ_i is broken up into two components, that is $\mu_i = \mu + \tau_i$, where

τ_i is the i th treatment effect and μ is an overall mean. We usually define $\mu = \frac{\sum_{i=1}^a \mu_i}{a}$ and

this implies that $\sum_{i=1}^a \tau_i = 0$.

This is actually an arbitrary definition, and there are other ways to define the overall “mean”. For example, we could define

$$\mu = \sum_{i=1}^a w_i \mu_i \quad \text{where} \quad \sum_{i=1}^a w_i = 1$$

This would result in the treatment effect defined such that

$$\sum_{i=1}^a w_i \tau_i = 0$$

Here the overall mean is a **weighted average** of the individual treatment means. When there are an unequal number of observations in each treatment, the weights w_i could be taken as the fractions of the treatment sample sizes n_i/N .

S3-2. Expected Mean Squares

In Section 3-3.1 we derived the expected value of the mean square for error in the single-factor analysis of variance. We gave the result for the expected value of the mean square for treatments, but the derivation was omitted. The derivation is straightforward.

Consider

$$E(MS_{Treatments}) = E\left(\frac{SS_{Treatments}}{a-1}\right)$$

Now for a balanced design

$$SS_{Treatments} = \frac{1}{n} \sum_{i=1}^a y_i^2 - \frac{1}{an} y_{..}^2$$

and the model is

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases}$$

In addition, we will find the following useful:

$$E(\varepsilon_{ij}) = E(\varepsilon_i) = E(\varepsilon_{..}) = 0, E(\varepsilon_{ij}^2) = \sigma^2, E(\varepsilon_i^2) = n\sigma^2, E(\varepsilon_{..}^2) = an\sigma^2$$

Now

$$E(SS_{Treatments}) = E\left(\frac{1}{n} \sum_{i=1}^a y_i^2\right) - E\left(\frac{1}{an} y_{..}^2\right)$$

Consider the first term on the right hand side of the above expression:

$$E\left(\frac{1}{n} \sum_{i=1}^a y_i^2\right) = \frac{1}{n} \sum_{i=1}^a E(n\mu + n\tau_i + \varepsilon_i)^2$$

Squaring the expression in parentheses and taking expectation results in

$$\begin{aligned} E\left(\frac{1}{n} \sum_{i=1}^a y_i^2\right) &= \frac{1}{n} [a(n\mu)^2 + n^2 \sum_{i=1}^a \tau_i^2 + an\sigma^2] \\ &= an\mu^2 + n \sum_{i=1}^a \tau_i^2 + a\sigma^2 \end{aligned}$$

because the three cross-product terms are all zero. Now consider the second term on the right hand side of $E(SS_{Treatments})$:

$$\begin{aligned} E\left(\frac{1}{an} y_{..}^2\right) &= \frac{1}{an} E(an\mu + n \sum_{i=1}^a \tau_i + \varepsilon_{..})^2 \\ &= \frac{1}{an} E(an\mu + \varepsilon_{..})^2 \end{aligned}$$

since $\sum_{i=1}^a \tau_i = 0$. Upon squaring the term in parentheses and taking expectation, we obtain

$$\begin{aligned} E\left(\frac{1}{an} y_{..}^2\right) &= \frac{1}{an} [(an\mu)^2 + an\sigma^2] \\ &= an\mu^2 + \sigma^2 \end{aligned}$$

since the expected value of the cross-product is zero. Therefore,

$$\begin{aligned} E(SS_{Treatments}) &= E\left(\frac{1}{n} \sum_{i=1}^a y_i^2\right) - E\left(\frac{1}{an} y_{..}^2\right) \\ &= an\mu^2 + n \sum_{i=1}^a \tau_i^2 + a\sigma^2 - (an\mu^2 + \sigma^2) \\ &= \sigma^2(a-1) + n \sum_{i=1}^a \tau_i^2 \end{aligned}$$

Consequently the expected value of the mean square for treatments is

$$\begin{aligned}
 E(MS_{Treatments}) &= E\left(\frac{SS_{Treatments}}{a-1}\right) \\
 &= \frac{\sigma^2(a-1) + n \sum_{i=1}^a \tau_i^2}{a-1} \\
 &= \sigma^2 + \frac{n \sum_{i=1}^a \tau_i^2}{a-1}
 \end{aligned}$$

This is the result given in the textbook.

S3-3. Confidence Interval for σ^2

In developing the analysis of variance (ANOVA) procedure we have observed that the error variance σ^2 is estimated by the error mean square; that is,

$$\hat{\sigma}^2 = \frac{SS_E}{N-a}$$

We now give a confidence interval for σ^2 . Since we have assumed that the observations are normally distributed, the distribution of

$$\frac{SS_E}{\sigma^2}$$

is χ^2_{N-a} . Therefore,

$$P\left(\chi^2_{1-\alpha/2, N-a} \leq \frac{SS_E}{\sigma^2} \leq \chi^2_{\alpha/2, N-a}\right) = 1 - \alpha$$

where $\chi^2_{1-\alpha/2, N-a}$ and $\chi^2_{\alpha/2, N-a}$ are the lower and upper $\alpha/2$ percentage points of the χ^2 distribution with $N-a$ degrees of freedom, respectively. Now if we rearrange the expression inside the probability statement we obtain

$$P\left(\frac{SS_E}{\chi^2_{\alpha/2, N-a}} \leq \sigma^2 \leq \frac{SS_E}{\chi^2_{1-\alpha/2, N-a}}\right) = 1 - \alpha$$

Therefore, a $100(1-\alpha)$ percent confidence interval on the error variance σ^2 is

$$\frac{SS_E}{\chi^2_{\alpha/2, N-a}} \leq \sigma^2 \leq \frac{SS_E}{\chi^2_{1-\alpha/2, N-a}}$$

This confidence interval expression is also given in Chapter 12 on experiments with random effects.

Sometimes an experimenter is interested in an upper bound on the error variance; that is, how large could σ^2 reasonably be? This can be useful when there is information about σ^2 from a prior experiment and the experimenter is performing calculations to determine sample sizes for a new experiment. An upper 100(1- α) percent confidence limit on σ^2 is given by

$$\sigma^2 \leq \frac{SS_E}{\chi^2_{1-\alpha, N-a}}$$

If a 100(1- α) percent confidence interval on the standard deviation σ is desired instead, then

$$\sigma \leq \sqrt{\frac{SS_E}{\chi^2_{1-\alpha/2, N-a}}}$$

S3-4. Simultaneous Confidence Intervals on Treatment Means

In section 3-3.3 we discuss finding confidence intervals on a treatment mean and on differences between a pair of means. We also show how to find *simultaneous* confidence intervals on a *set* of treatment means or a set of differences between pairs of means using the Bonferroni approach. Essentially, if there are a set of r confidence statements to be constructed the Bonferroni method simply replaces $\alpha/2$ by $\alpha/(2r)$. This produces a set of r confidence intervals for which the overall confidence level is at least 100(1- α) percent.

To see why this works, consider the case where $r = 2$; that is, we have two 100(1- α) percent confidence intervals. Let E_1 denote the event that the first confidence interval is not correct (it does not cover the true mean) and E_2 denote the event that the second confidence interval is incorrect. Now

$$P(E_1) = P(E_2) = \alpha$$

The probability that either or both intervals is incorrect is

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

From the probability of complimentary events we can find the probability that both intervals are correct as

$$\begin{aligned} P(\bar{E}_1 \cap \bar{E}_2) &= 1 - P(E_1 \cup E_2) \\ &= 1 - P(E_1) - P(E_2) + P(E_1 \cap E_2) \end{aligned}$$

Now we know that $P(E_1 \cap E_2) \geq 0$, so from the last equation above we obtain the *Bonferroni inequality*

$$P(\bar{E}_1 \cap \bar{E}_2) \geq 1 - P(E_1) - P(E_2)$$

In the context of our example, the left-hand side of this inequality is the probability that both of the two confidence interval statements is correct and $P(E_1) = P(E_2) = \alpha$, so

$$P(\bar{E}_1 \cap \bar{E}_2) \geq 1 - \alpha - \alpha \\ \geq 1 - 2\alpha$$

Therefore, if we want the probability that both of the confidence intervals are correct to be at least $1-\alpha$ we can assure this by constructing $100(1-\alpha/2)$ percent individual confidence interval.

If there are r confidence intervals of interest, we can use mathematical induction to show that

$$P(\bar{E}_1 \cap \bar{E}_2 \cap \dots \cap \bar{E}_r) \geq 1 - \sum_{i=1}^r P(E_i) \\ \geq 1 - r\alpha$$

As noted in the text, the Bonferroni method works reasonably well when the number of simultaneous confidence intervals that you desire to construct, r , is not too large. As r becomes larger, the lengths of the individual confidence intervals increase. The lengths of the individual confidence intervals can become so large that the intervals are not very informative. Also, it is not necessary that all individual confidence statements have the same level of confidence. One might select 98 percent for one statement and 92 percent for the other, resulting in two confidence intervals for which the simultaneous confidence level is at least 90 percent.

S3-5. Regression Models for a Quantitative Factor

Regression models are discussed in detail in Chapter 10, but they appear relatively often throughout the book because it is convenient to express the relationship between the response and quantitative design variables in terms of an equation. When there is only a single quantitative design factor, a linear regression model relating the response to the factor is

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where x represents the values of the design factor. In a single-factor experiment there are N observations, and each observation can be expressed in terms of this model as follows:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, 2, \dots, N$$

The **method of least squares** is used to estimate the unknown parameters (the β 's) in this model. This involves estimating the parameters so that the sum of the squares of the errors is minimized. The least squares function is

$$L = \sum_{i=1}^N \varepsilon_i^2 = \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i)^2$$

To find the least squares estimators we take the partial derivatives of L with respect to the β 's and equate to zero:

$$\frac{\partial L}{\partial \beta_0} = -2 \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial L}{\partial \beta_1} = -2 \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i) x_i = 0$$

After simplification, we obtain the **least squares normal equations**

$$N\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^N x_i = \sum_{i=1}^N y_i$$

$$\hat{\beta}_0 \sum_{i=1}^N x_i + \hat{\beta}_1 \sum_{i=1}^N x_i^2 = \sum_{i=1}^N x_i y_i$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the least squares estimators of the model parameters. So, to fit this particular model to the experimental data by least squares, all we have to do is solve the normal equations. Since there are only two equations in two unknowns, this is fairly easy.

In the textbook we fit two regression models for the response variable etch rate (y) as a function of the RF power (x); the linear regression model shown above, and a quadratic model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

The least squares normal equations for the quadratic model are

$$N\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^N x_i + \hat{\beta}_2 \sum_{i=1}^N x_i^2 = \sum_{i=1}^N y_i$$

$$\hat{\beta}_0 \sum_{i=1}^N x_i + \hat{\beta}_1 \sum_{i=1}^N x_i^2 + \hat{\beta}_2 \sum_{i=1}^N x_i^3 = \sum_{i=1}^N x_i y_i$$

$$\hat{\beta}_0 \sum_{i=1}^N x_i^2 + \hat{\beta}_1 \sum_{i=1}^N x_i^3 + \hat{\beta}_2 \sum_{i=1}^N x_i^4 = \sum_{i=1}^N x_i^2 y_i$$

Obviously as the order of the model increases and there are more unknown parameters to estimate, the normal equations become more complicated. In Chapter 10 we use matrix methods to develop the general solution. Most statistics software packages have very good regression model fitting capability.

S3-6. More About Estimable Functions

In Section 3-9.1 we use the least squares approach to estimating the parameters in the single-factor model. Assuming a balanced experimental design, we find the least squares normal equations as Equation 3-48, repeated below:

$$\begin{aligned}
an\hat{\mu} + n\hat{\tau}_1 + n\hat{\tau}_2 + \cdots + n\hat{\tau}_a &= \sum_{i=1}^a \sum_{j=1}^n y_{ij} \\
n\hat{\mu} + n\hat{\tau}_1 &= \sum_{j=1}^n y_{1j} \\
n\hat{\mu} + n\hat{\tau}_2 &= \sum_{j=1}^n y_{2j} \\
&\vdots \\
n\hat{\mu} + n\hat{\tau}_a &= \sum_{j=1}^n y_{aj}
\end{aligned}$$

where $an = N$ is the total number of observations. As noted in the textbook, if we add the last a of these normal equations we obtain the first one. That is, the normal equations are not linearly independent and so they do not have a unique solution. We say that the effects model is an **overparameterized** model.

One way to resolve this is to add another linearly independent equation to the normal equations. The most common way to do this is to use the equation $\sum_{i=1}^a \hat{\tau}_i = 0$. This is consistent with defining the factor effects as deviations from the overall mean μ . If we impose this constraint, the solution to the normal equations is

$$\begin{aligned}
\hat{\mu} &= \bar{y} \\
\hat{\tau}_i &= \bar{y}_i - \bar{y}, i = 1, 2, \dots, a
\end{aligned}$$

That is, the overall mean is estimated by the average of all an sample observation, while each individual factor effect is estimated by the difference between the sample average for that factor level and the average of all observations.

Another possible choice of constraint is to set the overall mean equal to a constant, say $\hat{\mu} = 0$. This results in the solution

$$\begin{aligned}
\hat{\mu} &= 0 \\
\hat{\tau}_i &= \bar{y}_i, i = 1, 2, \dots, a
\end{aligned}$$

Still a third choice is $\hat{\tau}_a = 0$. This is the approach used in the SAS software, for example. This choice of constraint produces the solution

$$\begin{aligned}
\hat{\mu} &= \bar{y}_a \\
\hat{\tau}_i &= \bar{y}_i - \bar{y}_a, i = 1, 2, \dots, a-1 \\
\hat{\tau}_a &= 0
\end{aligned}$$

There are an infinite number of possible constraints that could be used to solve the normal equations. Fortunately, as observed in the book, it really doesn't matter. For each of the three solutions above (indeed for *any* solution to the normal equations) we have

$$\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \bar{y}_i, i = 1, 2, \dots, a$$

That is, the least squares estimator of the mean of the i th factor level will always be the sample average of the observations at that factor level. So even if we cannot obtain unique estimates for the parameters in the effects model we *can* obtain unique estimators of a *function* of these parameters that we are interested in.

This is the idea of **estimable functions**. Any function of the model parameters that can be uniquely estimated regardless of the constraint selected to solve the normal equations is an estimable function.

What functions are estimable? It can be shown that the expected value of any observation is estimable. Now

$$E(y_{ij}) = \mu + \tau_i$$

so as shown above, the mean of the i th treatment is estimable. Any function that is a linear combination of the left-hand side of the normal equations is also estimable. For example, subtract the third normal equation from the second, yielding $\tau_2 - \tau_1$.

Consequently, the difference in any two treatment effect is estimable. In general, any

contrast in the treatment effects $\sum_{i=1}^a c_i \tau_i$ where $\sum_{i=1}^a c_i = 0$ is estimable. Notice that the

individual model parameters $\mu, \tau_1, \dots, \tau_a$ are not estimable, as there is no linear combination of the normal equations that will produce these parameters separately. However, this is generally not a problem, for as observed previously, the estimable functions correspond to functions of the model parameters that are of interest to experimenters.

For an excellent and very readable discussion of estimable functions, see Myers, R. H. and Milton, J. S. (1991), *A First Course in the Theory of the Linear Model*, PWS-Kent, Boston, MA.

S3-7. The Relationship Between Regression and ANOVA

Section 3-9 explored some of the connections between analysis of variance (ANOVA) models and regression models. We showed how least squares methods could be used to estimate the model parameters and how the ANOVA can be developed by a regression-based procedure called the general regression significance test can be used to develop the ANOVA test statistic. Every ANOVA model can be written explicitly as an equivalent linear regression model. We now show how this is done for the single-factor experiment with $a = 3$ treatments.

The single-factor balanced ANOVA model is

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1, 2, 3 \\ j = 1, 2, \dots, n \end{cases}$$

The equivalent regression model is

$$y_{ij} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \varepsilon_{ij} \begin{cases} i = 1, 2, 3 \\ j = 1, 2, \dots, n \end{cases}$$

where the variables x_{1j} and x_{2j} are defined as follows:

$$x_{1j} = \begin{cases} 1 & \text{if observation } j \text{ is from treatment 1} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{2j} = \begin{cases} 1 & \text{if observation } j \text{ is from treatment 2} \\ 0 & \text{otherwise} \end{cases}$$

The relationships between the parameters in the regression model and the parameters in the ANOVA model are easily determined. For example, if the observations come from treatment 1, then $x_{1j} = 1$ and $x_{2j} = 0$ and the regression model is

$$y_{1j} = \beta_0 + \beta_1(1) + \beta_2(0) + \varepsilon_{1j}$$

$$= \beta_0 + \beta_1 + \varepsilon_{1j}$$

Since in the ANOVA model these observations are defined by $y_{1j} = \mu + \tau_1 + \varepsilon_{1j}$, this implies that

$$\beta_0 + \beta_1 = \mu_1 = \mu + \tau_1$$

Similarly, if the observations are from treatment 2, then

$$y_{2j} = \beta_0 + \beta_1(0) + \beta_2(1) + \varepsilon_{2j}$$

$$= \beta_0 + \beta_2 + \varepsilon_{2j}$$

and the relationship between the parameters is

$$\beta_0 + \beta_2 = \mu_2 = \mu + \tau_2$$

Finally, consider observations from treatment 3, for which the regression model is

$$y_{3j} = \beta_0 + \beta_1(0) + \beta_2(0) + \varepsilon_{3j}$$

$$= \beta_0 + \varepsilon_{3j}$$

and we have

$$\beta_0 = \mu_3 = \mu + \tau_3$$

Thus in the regression model formulation of the one-way ANOVA model, the regression coefficients describe comparisons of the first two treatment means with the third treatment mean; that is

$$\beta_0 = \mu_3$$

$$\beta_1 = \mu_1 - \mu_3$$

$$\beta_2 = \mu_2 - \mu_3$$

In general, if there are a treatments, the regression model will have $a - 1$ regressor variables, say

$$y_{ij} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_{a-1} x_{a-1j} + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases}$$

where

$$x_{ij} = \begin{cases} 1 & \text{if observation } j \text{ is from treatment } i \\ 0 & \text{otherwise} \end{cases}$$

Since these regressor variables only take on the values 0 and 1, they are often called **indicator variables**. The relationship between the parameters in the ANOVA model and the regression model is

$$\begin{aligned} \beta_0 &= \mu_a \\ \beta_i &= \mu_i - \mu_a, i = 1, 2, \dots, a - 1 \end{aligned}$$

Therefore the intercept is always the mean of the a th treatment and the regression coefficient β_i estimates the difference between the mean of the i th treatment and the a th treatment.

Now consider testing hypotheses. Suppose that we want to test that all treatment means are equal (the usual null hypothesis). If this null hypothesis is true, then the parameters in the regression model become

$$\begin{aligned} \beta_0 &= \mu_a \\ \beta_i &= 0, i = 1, 2, \dots, a - 1 \end{aligned}$$

Using the general regression significance test procedure, we could develop a test for this hypothesis. It would be identical to the F -statistic test in the one-way ANOVA.

Most regression software packages automatically test the hypothesis that all model regression coefficients (except the intercept) are zero. We will illustrate this using Minitab and the data from the plasma etching experiment in Example 3-1. Recall in this example that the engineer is interested in determining the effect of RF power on etch rate, and he has run a completely randomized experiment with four levels of RF power and five replicates. For convenience, we repeat the data from Table 3-1 here:

RF Power (W)	Observed etch rate				
	1	2	3	4	5
160	575	542	530	539	570
180	565	593	590	579	610
200	600	651	610	637	629
220	725	700	715	685	710

The data was converted into the x_{ij} 0/1 indicator variables as described above. Since there are 4 treatments, there are only 3 of the x 's. The coded data that is used as input to Minitab is shown below:

x1	x2	x3	Etch rate
1	0	0	575
1	0	0	542
1	0	0	530
1	0	0	539
1	0	0	570
0	1	0	565
0	1	0	593
0	1	0	590
0	1	0	579
0	1	0	610
0	0	1	600
0	0	1	651
0	0	1	610
0	0	1	637
0	0	1	629
0	0	0	725
0	0	0	700
0	0	0	715
0	0	0	685

The Regression Module in Minitab was run using the above spreadsheet where x1 through x3 were used as the predictors and the variable “Etch rate” was the response. The output is shown below.

Regression Analysis: Etch rate versus x1, x2, x3

The regression equation is
 Etch rate = 707 - 156 x1 - 120 x2 - 81.6 x3

Predictor	Coef	SE Coef	T	P
Constant	707.000	8.169	86.54	0.000
x1	-155.80	11.55	-13.49	0.000
x2	-119.60	11.55	-10.35	0.000
x3	-81.60	11.55	-7.06	0.000

S = 18.2675 R-Sq = 92.6% R-Sq(adj) = 91.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	66871	22290	66.80	0.000
Residual Error	16	5339	334		

Notice that the ANOVA table in this regression output is identical (apart from rounding) to the ANOVA display in Table 3-4. Therefore, testing the hypothesis that the regression coefficients $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ in this regression model is equivalent to testing the null hypothesis of equal treatment means in the original ANOVA model formulation.

Also note that the estimate of the intercept or the “constant” term in the above table is the mean of the 4th treatment. Furthermore, each regression coefficient is just the difference between one of the treatment means and the 4th treatment mean.