About the Supplemental Text Material

I have prepared supplemental text material for each chapter of the 6th edition of *Design and Analysis of Experiments*. This material consists of (1) some extensions of and elaboration on topics introduced in the text and (2) some new topics that I could not easily find a "home" for in the text without disrupting the flow of the coverage within each chapter, or making the book ridiculously long.

Some of this material is in partial response to the many suggestions that have been made over the years by textbook users, who have always been gracious in their requests and very often extremely helpful. However, sometimes there just wasn't any way to easily accommodate their suggestions directly in the book. Some of the supplemental material is in direct response to FAQ's or "frequently asked questions" from students. It also reflects topics that I have found helpful in consulting on experimental design and analysis problems, but again, there wasn't any easy way to incorporate it in the text. Obviously, there is also quite a bit of personal "bias" in my selection of topics for the supplemental material. The coverage is far from comprehensive.

I have not felt as constrained about mathematical level or statistical background of the readers in the supplemental material as I have tried to be in writing the textbook. There are sections of the supplemental material that will require considerably more background in statistics than is required to read the text material. However, I think that many instructors will be able to use this supplement material in their courses quite effectively, depending on the maturity and background of the students. Hopefully, it will also provide useful additional information for readers who wish to see more in-depth discussion of some aspects of design, or who are attracted to the "eclectic" variety of topics that I have included.

Contents

S10-4 Stepwise and Other Variable Selection Methods in Regression
S10-5 The Variance of the Predicted Response
S10-6 The Variance of Prediction Error
S10-7 Leverage in a Regression Model

Chapter 11
S11-1 The Method of Steepest Ascent
S11-2 The Canonical Form of the Second-Order Response Surface Model
S11-3 Center Points in the Central Composite Design
S11-4 Center Runs in the Face-Centered Cube
S11-5 A Note on Rotatability

Chapter 12
S12-1 The Taguchi Approach to Robust Parameter Design
S12-2 Taguchi's Technical Methods

Chapter 13
S13-1 Expected Mean Squares for the Random Model
S13-2 Expected Mean Squares for the Mixed Model
S13-3 Restricted versus Unrestricted Mixed Models
S13-4 Random and Mixed Models with Unequal Sample Size
S13-5 Some Background Concerning the Modified Large Sample Method
S13-6 A Confidence Interval on a Ratio of Variance Components using the Modified Large Sample Method

Chapter 14
S14-1 The Staggered, Nested Design
S14-2 Inadvertent Split-Plots

Chapter 15
S15-1 The Form of a Transformation
S15-2 Selecting λ in the Box-Cox Method
S15-3 Generalized Linear Models
- S15-3.1. Models with a Binary Response Variable
- S15-3.2. Estimating the Parameters in a Logistic Regression Model
- S15-3.3. Interpreting the Parameters in a Logistic Regression Model
- S15-3.4. Hypothesis Tests on Model Parameters
- S15-3.5. Poisson Regression
- S15-3.6. The Generalized Linear Model
- S15-3.7. Link Functions and Linear Predictors
- S15-3.8. Parameter Estimation in the Generalized Linear Model
- S15-3.9. Prediction and Estimation with the Generalized Linear Model
- S15-3.10. Residual Analysis in the Generalized Linear Model

S15-4 Unbalanced Data in a Factorial Design
- S15-4.1. The Regression Model Approach
- S15-4.2.The Type 3 Analysis

Chapter 1 Supplemental Text Material

S-1.1 More About Planning Experiments

Coleman and Montgomery (1993) present a discussion of methodology and some guide sheets useful in the pre-experimental planning phases of designing and conducting an industrial experiment. The guide sheets are particularly appropriate for complex, high-payoff or high-consequence experiments involving (possibly) many factors or other issues that need careful consideration and (possibly) many responses. They are most likely to be useful in the earliest stages of experimentation with a process or system. Coleman and Montgomery suggest that the guide sheets work most effectively when they are filled out by a team of experimenters, including engineers and scientists with specialized process knowledge, operators and technicians, managers and (if available) individuals with specialized training and experience in designing experiments. The sheets are intended to encourage discussion and resolution of technical and logistical issues *before* the experiment is actually conducted.

Coleman and Montgomery give an example involving manufacturing impellers on a CNC-machine that are used in a jet turbine engine. To achieve the desired performance objectives, it is necessary to produce parts with blade profiles that closely match the engineering specifications. The objective of the experiment was to study the effect of different tool vendors and machine set-up parameters on the dimensional variability of the parts produced by the CNC-machines.

The master guide sheet is shown in Table 1 below. It contains information useful in filling out the individual sheets for a particular experiment. Writing the objective of the experiment is usually harder than it appears. Objectives should be unbiased, specific, measurable and of practical consequence. To be unbiased, the experimenters must encourage participation by knowledgeable and interested people with diverse perspectives. It is all too easy to design a very narrow experiment to "prove" a pet theory. To be specific and measurable the objectives should be detailed enough and stated so that it is clear when they have been met. To be of practical consequence, there should be something that will be done differently as a result of the experiment, such as a new set of operating conditions for the process, a new material source, or perhaps a new experiment will be conducted. All interested parties should agree that the proper objectives have been set.

The relevant background should contain information from previous experiments, if any, observational data that may have been collected routinely by process operating personnel, field quality or reliability data, knowledge based on physical laws or theories, and expert opinion. This information helps quantify what new knowledge could be gained by the present experiment and motivates discussion by all team members. Table 2 shows the beginning of the guide sheet for the CNC-machining experiment.

Response variables come to mind easily for most experimenters. When there is a choice, one should select continuous responses, because generally binary and ordinal data carry much less information and continuous responses measured on a well-defined numerical scale are typically easier to analyze. On the other hand, there are many situations where a count of defectives, a proportion, or even a subjective ranking must be used as a response.

Table 1. Master Guide Sheet. This guide can be used to help plan and design an experiment. It serves as a checklist to improve experimentation and ensures that results are not corrupted for lack of careful planning. Note that it may not be possible to answer all questions completely. If convenient, use supplementary sheets for topics 4-8

1.**Experimenter's Name and Organization:** **Brief Title of Experiment:**
2. Objectives of the experiment (should be unbiased, specific, measurable, and of practical consequence):
3. Relevant background on response and control variables: (a) theoretical relationships; (b) expert knowledge/experience; (c) previous experiments. Where does this experiment fit into the study of the process or system?:
4. List: (a) each **response variable,** (b) the normal response variable level at which the process runs, the distribution or range of normal operation, (c) the precision or range to which it can be measured (and how):
5. List: (a) each **control variable,** (b) the normal control variable level at which the process is run, and the distribution or range of normal operation, (c) the precision (s) or range to which it can be set (for the experiment, not ordinary plant operations) and the precision to which it can be measured, (d) the proposed control variable settings, and (e) the predicted effect (at least qualitative) that the settings will have on each response variable:
6. List: (a) each factor to be **"held constant"** in the experiment, (b) its desired level and allowable s or range of variation, (c) the precision or range to which it can measured (and how), (d) how it can be controlled, and (e) its expected impact, if any, on each of the responses:
7. List: (a) each **nuisance factor** (perhaps time-varying), (b) measurement precision, (c)strategy (e.g., blocking, randomization, or selection), and (d) anticipated effect:
8. List and label known or suspected **interactions:**
9. List **restrictions** on the experiment, e.g., ease of changing control variables, methods of data acquisition, materials, duration, number of runs, type of experimental unit (need for a split-plot design), “illegal” or irrelevant experimental regions, limits to randomization, run order, cost of changing a control variable setting, etc.:
10. Give current **design preferences,** if any, and reasons for preference, including blocking and randomization:
11. If possible, propose **analysis and presentation techniques,** e.g., plots, ANOVA, regression, plots, t tests, etc.:
12. **Who will be responsible for the coordination of the experiment?**
13. **Should trial runs be conducted? Why / why not?**

Table 2. Beginning of Guide Sheet for CNC-Machining Study.

I.Experimenter's Name and Organization: John Smith, Process Eng. Group **Brief Title of Experiment**: CNC Machining Study
2. Objectives of the experiment (should be unbiased, specific, measurable, and of practical consequence):
For machined titanium forgings, quantify the effects of tool vendor; shifts in a-axis, x- axis, y-axis, and z-axis; spindle speed; fixture height; feed rate; and spindle position on the average and variability in blade profile for class X impellers, such as shown in Figure 1.
3. Relevant background on response and control variables: (a) theoretical relationships; (b) expert knowledge/experience; (c) previous experiments. Where does this experiment fit into the study of the process or system? (a) Because of tool geometry, x-axis shifts would be expected to produce thinner blades, an undesirable characteristic of the airfoil. (b) This family of parts has been produced for over 10 years; historical experience indicates that externally reground tools do not perform as well as those from the "internal" vendor (our own regrind operation). (c) Smith (1987) observed in an internal process engineering study that current spindle speeds and feed rates work well in producing parts that are at the nominal profile required by the engineering drawings - but no study was done of the sensitivity to variations in set-up parameters.
Results of this experiment will be used to determine machine set-up parameters for impeller machining. A robust process is desirable; that is, on-target and low variability performance regardless of which tool vendor is used.

Measurement precision is an important aspect of selecting the response variables in an experiment. Insuring that the measurement process is in a state of statistical control is highly desirable. That is, ideally there is a well-established system of insuring both accuracy and precision of the measurement methods to be used. The amount of error in measurement imparted by the gauges used should be understood. If the gauge error is large relative to the change in the response variable that is important to detect, then the experimenter will want to know this *before conducting the experiment.* Sometimes repeat measurements can be made on each experimental unit or test specimen to reduce the impact of measurement error. For example, when measuring the number average molecular weight of a polymer with a gel permeation chromatograph (GPC) each sample can be tested several times and the *average* of those molecular weight reading reported as the observation for that sample. When measurement precision is unacceptable, a measurement systems capability study may be performed to attempt to improve the system. These studies are often fairly complicated designed experiments. Chapter 13 presents an example of a factorial experiment used to study the capability of a measurement system.

The impeller involved in this experiment is shown in Figure 1. Table 3 lists the information about the response variables. Notice that there are three response variables of interest here.

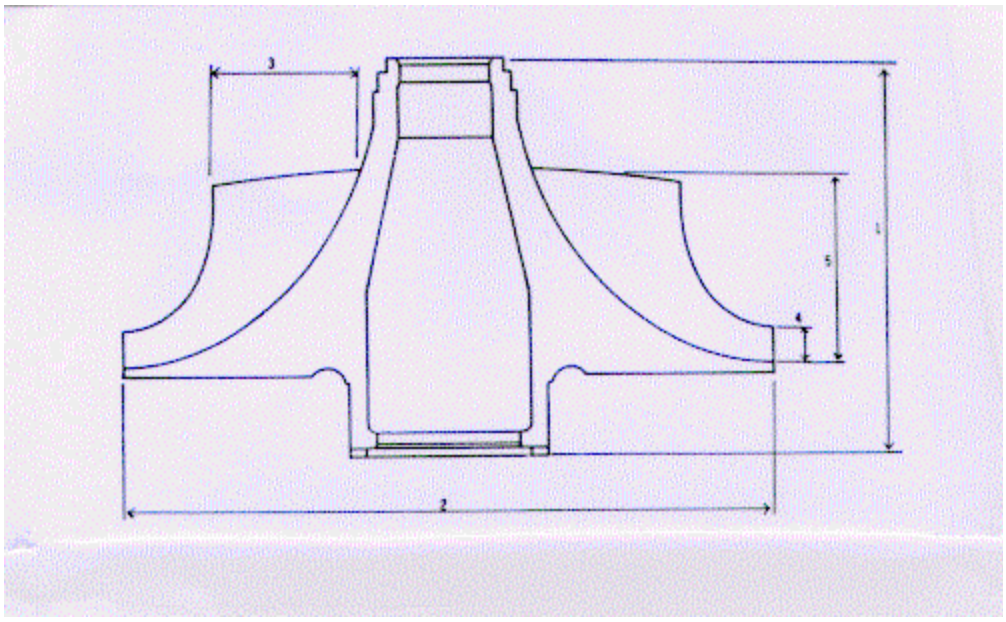


Figure 1. Jet engine impeller (side view). The z-axis is vertical, x-axis is horizontal, y-axis is into the page. 1 = height of wheel, 2 = diameter of wheel, 3 = inducer blade height, 4 = exducer blade height, 5 = z height of blade.

Table 3. Response Variables

Response variable (units)	*Normal operating level and range*	*Measurement precision, accuracy how known?*	*Relationship of response variable to objective*
Blade profile (inches)	Nominal (target) $\pm 1 \times 10^{-3}$ inches to $\pm 2 \times 10^{-3}$ inches at all points	$\sigma_E \approx @ 1 \times 10^{-5}$ inches from a coordinate measurement machine capability study	Estimate mean absolute difference from target and standard deviation
Surface finish	Smooth to rough (requiring hand finish)	Visual criterion (compare to standards)	Should be as smooth as possible
Surface defect count	Typically 0 to 10	Visual criterion (compare to standards)	Must not be excessive in number or magnitude

As with response variables, most experimenters can easily generate a list of candidate design factors to be studied in the experiment. Coleman and Montgomery call these control variables. We often call them controllable variables, design factors, or process variables in the text. Control variables can be continuous or categorical (discrete). The ability of the experimenters to measure and set these factors is important. Generally,

small errors in the ability to set, hold or measure the levels of control variables are of relatively little consequence. Sometimes when the measurement or setting error is large, a numerical control variable such as temperature will have to be treated as a categorical control variable (low or high temperature). Alternatively, there are errors-in-variables statistical models that can be employed, although their use is beyond the scope of this book. Information about the control variables for the CNC-machining example is shown in Table 4.

Table 4. Control Variables

Control variable (units)	*Normal level and range*	*Measurement Precision and setting error- how known?*	*Proposed settings, based on predicted effects*	*Predicted effects (for various responses)*
x-axis shift* (inches)	0-.020 inches	.001inches (experience)	0, .015 inches	Difference
y-axis shift* (inches)	0-.020 inches	.001inches (experience)	0, .015 inches	Difference
z-axis shift* (inches)	0-.020 inches	.001inches (experience)	?	Difference
Tool vendor	Internal, external	-	Internal, external	External is more variable
a-axis shift* (degrees)	0-.030 degrees	.001 degrees (guess)	0, .030 degrees	Unknown
Spindle speed (% of nominal)	85-115%	~1% (indicator on control panel)	90%,110%	None?
Fixture height	0-.025 inches	.002inches (guess)	0, .015 inches	Unknown
Feed rate (% of nominal)	90-110%	~1% (indicator on control panel)	90%,110%	None?

'The x, y, and *z* axes are used to refer to the part *and* the CNC machine. The a axis refers only to the machine.

Held-constant factors are control variables whose effects are not of interest in this experiment. The worksheets can force meaningful discussion about which factors are adequately controlled, and if any potentially important factors (for purposes of the present experiment) have inadvertently been held constant when they should have been included as control variables. Sometimes subject-matter experts will elect to hold too many factors constant and as a result fail to identify useful new information. Often this information is in the form of *interactions* among process variables.

In the CNC experiment, this worksheet helped the experimenters recognize that the machine had to be fully warmed up before cutting any blade forgings. The actual procedure used was to mount the forged blanks on the machine and run a 30-minute cycle

without the cutting tool engaged. This allowed all machine parts and the lubricant to reach normal, steady-state operating temperature. The use of a typical (i.e., mid-level) operator and the use of one lot of forgings ware decisions made for experimental "insurance". Table 5 shows the held-constant factors for the CNC-machining experiment.

Table 5. Held-Constant Factors

Factor (units)	*Desired experimental level and allowable range*	*Measurement precision-how known?*	*How to control (in experiment)*	*Anticipated effects*
Type of cutting fluid	Standard type	Not sure, but thought to be adequate	Use one type	None
Temperature of cutting fluid (degrees F.)	100- 100°F. when machine is warmed up	1-2° F. (estimate)	Do runs after machine has reached 100°	None
Operator	Several operators normally work in the process	-	Use one "mid-level" operator	None
Titanium forgings	Material properties may vary from unit to unit	Precision of lab tests unknown	Use one lot (or block on forging lot, only if necessary)	Slight

Nuisance factors are variables that probably have some effect on the response, but which are of little or no interest to the experimenter. They differ from held-constant factors in that they either cannot be held entirely constant, or they cannot be controlled at all. For example, if two lots of forgings were required to run the experiment, then the potential lot-to-lot differences in the material would be a nuisance variable than could not be held entirely constant. In a chemical process we often cannot control the viscosity (say) of the incoming material feed stream—it may vary almost continuously over time. In these cases, nuisance variables must be considered in either the design or the analysis of the experiment. If a nuisance variable can be controlled, then we can use a design technique called **blocking** to eliminate its effect. Blocking is discussed initially in Chapter 4. If the nuisance variable cannot be controlled but it can be measured, then we can reduce its effect by an analysis technique called the analysis of covariance, discussed in Chapter 14.

Table 6 shows the nuisance variables identified in the CNC-machining experiment. In this experiment, the only nuisance factor thought to have potentially serious effects was the machine spindle. The machine has four spindles, and ultimately a decision was made to run the experiment in four blocks. The other factors were held constant at levels below which problems might be encountered.

Table 6. Nuisance Factors

Nuisance factor (units)	*Measurement precision-how known?*	*Strategy (e.g., randomization, blocking, etc.)*	*Anticipated effects*
Viscosity of cutting fluid	Standard viscosity	Measure viscosity at start and end	None to slight
Ambient temperature (°F.)	1-2° F. by room thermometer (estimate)	Make runs below 80'F.	Slight, unless very hot weather
Spindle		Block or randomize on machine spindle	Spindle-to-spindle variation could be large
Vibration of machine during operation	?	Do not move heavy objects in CNC machine shop	Severe vibration can introduce variation within an impeller

Coleman and Montgomery also found it useful to introduce an interaction sheet. The concept of interactions among process variables is not an intuitive one, even to well-trained engineers and scientists. Now it is clearly unrealistic to think that the experimenters can identify all of the important interactions at the outset of the planning process. In most situations, the experimenters really don't know which main effects are likely to be important, so asking them to make decisions about interactions is impractical. However, sometimes the statistically-trained team members can use this as an opportunity to *teach* others about the interaction phenomena. When more is known about the process, it might be possible to use the worksheet to motivate questions such as "are there certain interactions that *must* be estimated?" Table 7 shows the results of this exercise for the CNC-machining example.

Table 7. Interactions

Control variable	*y shift*	*z shift*	*Vendor*	*a shift*	*Speed*	*Height*	*Feed*
x shift			P				
y shift	-		P				
z shift	-	-	P				
Vendor	-	-	-	P			
a shift	-	-	-	-			
Speed	-	-	-	-	-		F,D
Height	-	-	-	-	-	-	

NOTE: Response variables are P = profile difference, F = surface finish and D = surface defects

Two final points: First, an experimenter without a coordinator will probably fail. Furthermore, if something can go wrong, it probably will, so he coordinator will actually have a significant responsibility on checking to ensure that the experiment is being conducted as planned. Second, concerning trial runs, this is often a very good idea—particularly if this is the first in a series of experiments, or if the experiment has high

significance or impact. A trial run can consist of a center point in a factorial or a small part of the experiment—perhaps one of the blocks. Since many experiments often involve people and machines doing something they have not done before, practice is a good idea. Another reason for trial runs is that we can use them to get an estimate of the magnitude of experimental error. If the experimental error is much larger than anticipated, then this may indicate the need for redesigning a significant part of the experiment. Trial runs are also a good opportunity to ensure that measurement and data-acquisition or collection systems are operating as anticipated. Most experimenters never regret performing trial runs.

Blank Guide Sheets from Coleman and Montgomery (1993)

Response Variables

response variable (units)	normal operating level & range	meas. precision, accuracy How known?	relationship of response variable to objective

Control Variables

control variable (units)	normal level & range	meas. precision & setting error How known?	proposed settings, based on predicted effects	predicted effects (for various responses)

"Held Constant" Factors

factor (units)	desired experimental level & allowable range	measurement precision How known?	how to control (in experiment)	anticipated effects

Nuisance Factors

nuisance factor (units)	measurement precision How known?	strategy (e.g., randomization, blocking, etc.)	anticipated effects

Interactions

control var.	2	3	4	5	6	7	8
1							
2	-						
3	-	-					
4	-	-	-				
5	-	-	-	-			
6	-	-	-	-	-		
7	-	-	-	-	-	-	

S-1.2 Other Graphical Aids for Planning Experiments

In addition to the tables in Coleman and Montgomery's *Technometrics* paper, there are a number of useful graphical aids to pre-experimental planing. Perhaps the first person to suggest graphical methods for planning an experiment was Andrews (1964), who proposed a schematic diagram of the system much like Figure 1-1 in the textbook, with inputs, experimental variables, and responses all clearly labeled. These diagrams can be very helpful in focusing attention on the broad aspects of the problem.

Barton (1997) (1998) (1999) has discussed a number of useful graphical aids in planning experiments. He suggests using IDEF0 diagrams to identify and classify variables. IDEF0 stands for Integrated Computer Aided Manufacturing Identification Language, Level 0. The U. S. Air Force developed it to represent the subroutines and functions of complex computer software systems. The IDEF0 diagram is a block diagram that resembles Figure 1-1 in the textbook. IDEF0 diagrams are hierarchical; that is, the process or system can be decomposed into a series of process steps or systems and represented as a sequence of lower-level boxes drawn within the main block diagram.

Figure 2 shows an IDEF0 diagram [from Barton (1999)] for a portion of a videodisk manufacturing process. This figure presents the details of the disk pressing activities. The primary process has been decomposed into five steps, and the primary output response of interest is the warp in the disk.

The **cause-and-effect diagram (**or **fishbone)** discussed in the textbook can also be useful in identifying and classifying variables in an experimental design problem. Figure 3 [from Barton (1999)] shows a cause-and-effect diagram for the videodisk process. These diagrams are very useful in organizing and conducting "brainstorming" or other problem-solving meetings in which process variables and their potential role in the experiment are discussed and decided.

Both of these techniques can be very helpful in uncovering **intermediate variables**. These are variables that are often confused with the directly adjustable process variables. For example, the burning rate of a rocket propellant may be affected by the presence of voids in the propellant material. However, the voids are the result of mixing techniques, curing temperature and other process variables and so the voids themselves cannot be directly controlled by the experimenter.

Some other useful papers on planning experiments include Bishop, Petersen and Trayser (1982), Hahn (1977) (1984), and Hunter (1977).

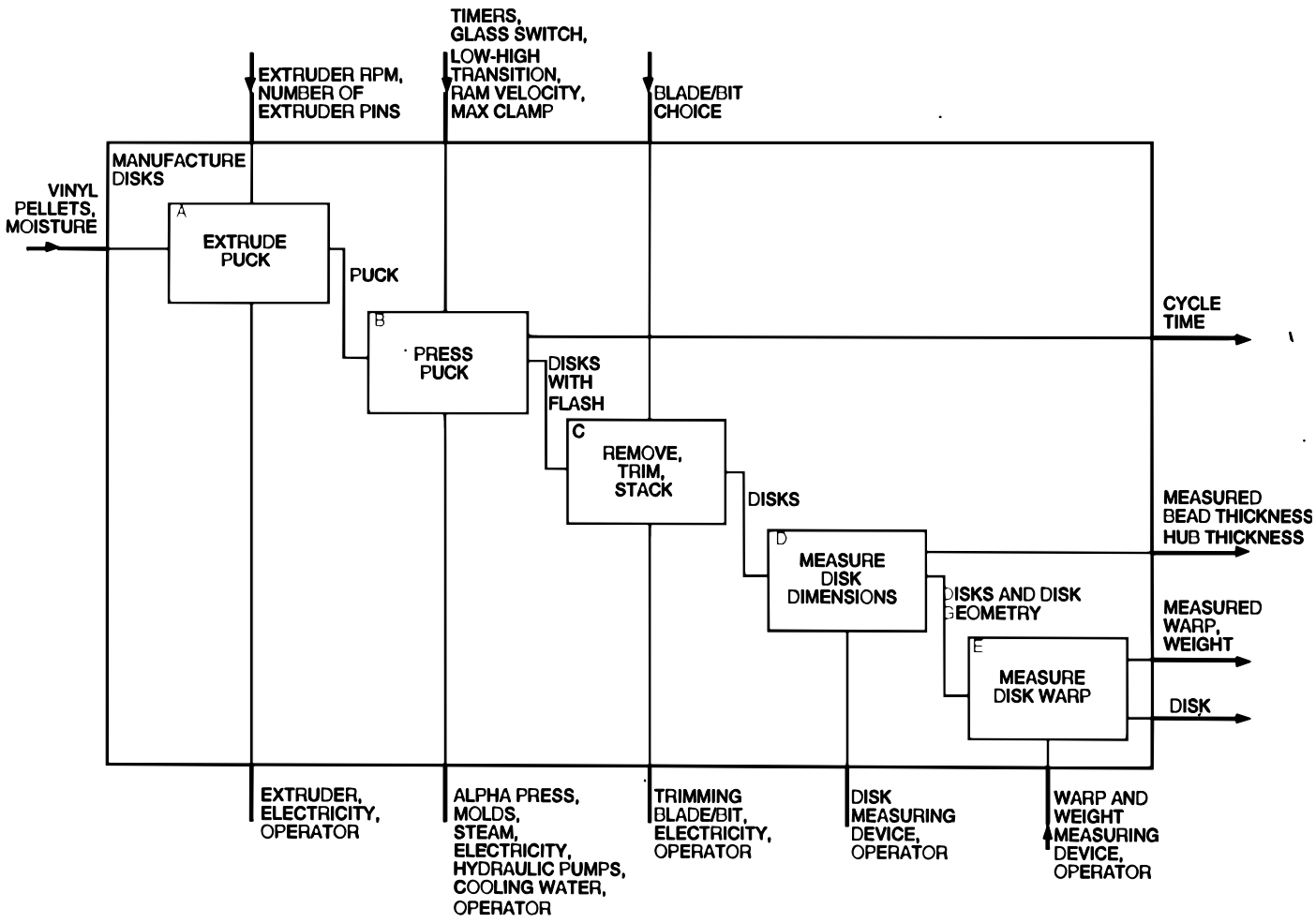


Figure 2. An IDEF0 Diagram for an Experiment in a Videodisk Manufacturing Process

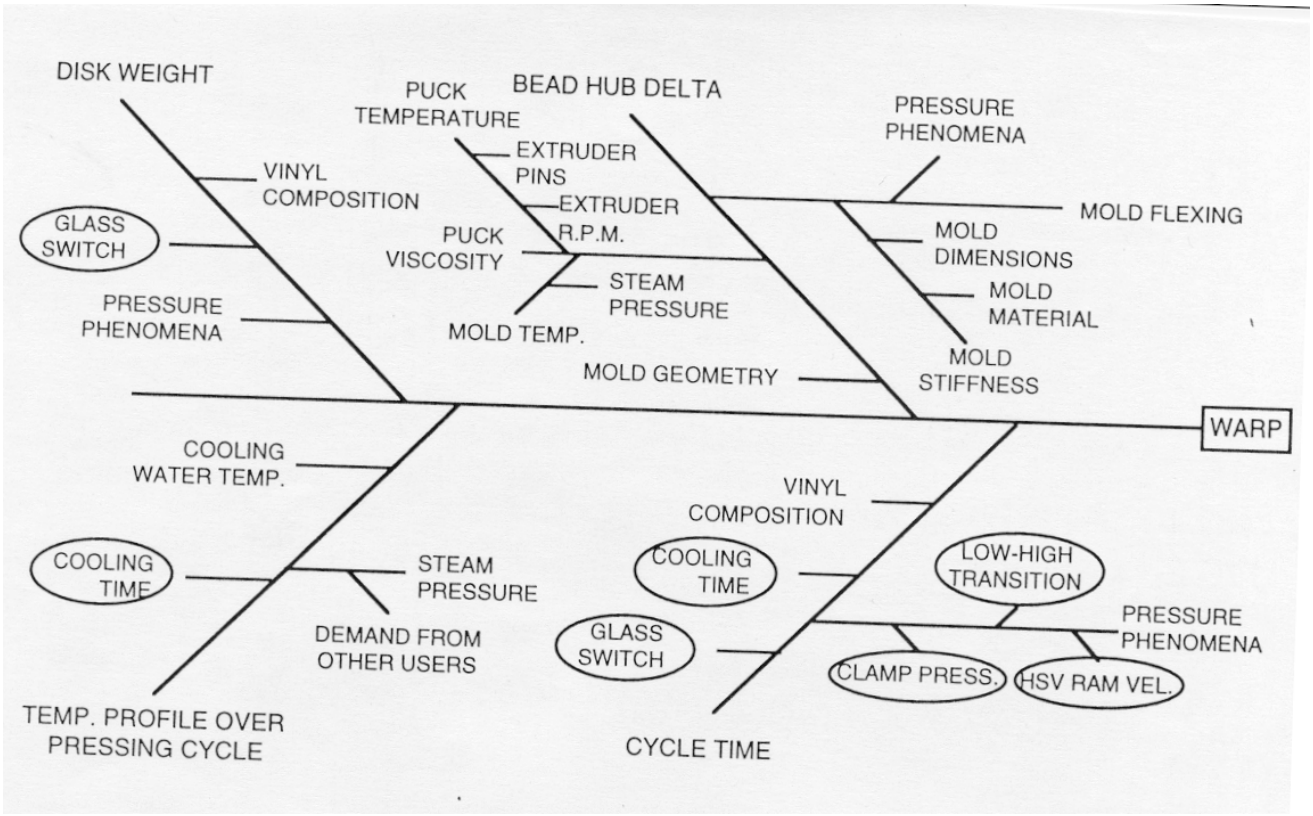


Figure 2. A Cause-and-Effect Diagram for an Experiment in a Videodisk Manufacturing Process

S-1.3 Montgomery's Theorems on Designed Experiments

Statistics courses, even very practical ones like design of experiments, tend to be a little dull and dry. Even for engineers, who are accustomed to taking much more exciting courses on topics such as fluid mechanics, mechanical vibrations, and device physics. Consequently, I try to inject a little humor into the course whenever possible. For example, I tell them on the first class meeting that they shouldn't look so unhappy. If they had one more day to live they should choose to spend it in a statistics class—that way it would seem twice as long.

I also use the following "theorems" at various times throughout the course. Most of them relate to non-statistical aspects of DOX, but they point out important issues and concerns.

Theorem 1. If something can go wrong in conducting an experiment, it will.

Theorem 2. The probability of successfully completing an experiment is inversely proportional to the number of runs.

Theorem 3. Never let one person design and conduct an experiment alone, particularly if that person is a subject-matter expert in the field of study.

Theorem 4. All experiments are *designed* experiments; some of them are designed well, and some of them are designed really badly. The badly designed ones often tell you nothing.

Theorem 5. About 80 percent of your success in conducting a designed experiment results directly from how well you do the pre-experimental planning (steps 1-3 in the 7-step procedure in the textbook).

Theorem 6. It is impossible to overestimate the logistical complexities associated with running an experiment in a "complex" setting, such as a factory or plant.

Finally, my friend Stu Hunter has for many years said that without good experimental design, we often end up doing PARC analysis. This is an acronym for

Planning **A**fter the **R**esearch is **C**omplete

What does PARC spell backwards?

Supplemental References

Andrews, H. P. (1964). "The Role of Statistics in Setting Food Specifications", *Proceedings of the Sixteenth Annual Conference of the Research Council of the American Meat Institute*, pp. 43-56. Reprinted in *Experiments in Industry: Design, Analysis, and Interpretation of Results*, eds. R. D. Snee, L. B. Hare and J. R. Trout, American Society for Quality Control, Milwaukee, WI 1985.

Barton, R. R. (1997). "Pre-experiment Planning for Designed Experiments: Graphical Methods", *Journal of Quality Technology*, Vol. 29, pp. 307-316.

Barton, R. R. (1998). "Design-plots for Factorial and Fractional Factorial Designs", *Journal of Quality Technology*, Vol. 30, pp. 40-54.

Barton, R. R. (1999). *Graphical Methods for the Design of Experiments*, Springer Lecture Notes in Statistics 143, Springer-Verlag, New York.

Bishop, T., Petersen, B. and Trayser, D. (1982). "Another Look at the Statistician's Role in Experimental Planning and Design", *The American Statistician*, Vol. 36, pp. 387-389.

Hahn, G. J. (1977). "Some Things Engineers Should Know About Experimental Design", *Journal of Quality Technology*, Vol. 9, pp. 13-20.

Hahn, G. J. (1984). "Experimental Design in a Complex World", *Technometrics*, Vol. 26, pp. 19-31.

Hunter, W. G. (1977). "Some Ideas About Teaching Design of Experiments With 2^5 Examples of Experiments Conducted by Students", *The American Statistician*, Vol. 31, pp. 12-17.

Chapter 2 Supplemental Text Material

S2-1. Models for the Data and the *t*-Test

The model presented in the text, equation (2-23) is more properly called a *means* model. Since the mean is a *location parameter*, this type of model is also sometimes called a *location model*. There are other ways to write the model for a *t*-test. One possibility is

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1,2 \\ j = 1,2,\cdots,n_i \end{cases}$$

where μ is a parameter that is common to all observed responses (an overall mean) and τ_i is a parameter that is unique to the *i*th factor level. Sometimes we call τ_i the *i*th treatment effect. This model is usually called the *effects* model.

Since the means model is

$$y_{ij} = \mu_i + \varepsilon_{ij} \begin{cases} i = 1,2 \\ j = 1,2,\cdots,n_i \end{cases}$$

we see that the *i*th treatment or factor level mean is $\mu_i = \mu + \tau_i$; that is, the mean response at factor level *i* is equal to an overall mean plus the effect of the *i*th factor. We will use both types of models to represent data from designed experiments. Most of the time we will work with effects models, because it's the "traditional" way to present much of this material. However, there are situations where the means model is useful, and even more natural.

S2-2. Estimating the Model Parameters

Because models arise naturally in examining data from designed experiments, we frequently need to estimate the model parameters. We often use **the method of least squares** for parameter estimation. This procedure chooses values for the model parameters that minimize the sum of the squares of the errors ε_{ij}. We will illustrate this procedure for the means model. For simplicity, assume that the sample sizes for the two factor levels are equal; that is $n_1 = n_2 = n$. The least squares function that must be minimized is

$$L = \sum_{i=1}^{2}\sum_{j=1}^{n} \varepsilon_{ij}^2$$
$$= \sum_{i=1}^{2}\sum_{j=1}^{n} (y_{ij} - \mu_i)^2$$

Now $\frac{\partial L}{\partial \mu_1} = 2\sum_{j=1}^{n}(y_{1j} - \mu_1)$ and $\frac{\partial L}{\partial \mu_2} = 2\sum_{j=1}^{n}(y_{2j} - \mu_2)$ and equating these partial derivatives to zero yields the **least squares normal equations**

$$n\hat{\mu}_1 = \sum_{i=1}^{n} y_{1j}$$

$$n\hat{\mu}_2 = \sum_{i=1}^{n} y_{2j}$$

The solution to these equations gives the least squares estimators of the factor level means. The solution is $\hat{\mu}_1 = \bar{y}_1$ and $\hat{\mu}_2 = \bar{y}_2$; that is, the sample averages at leach factor level are the estimators of the factor level means.

This result should be intuitive, as we learn early on in basic statistics courses that the sample average usually provides a reasonable estimate of the population mean. However, as we have just seen, this result can be derived easily from a simple location model using least squares. It also turns out that if we assume that the model errors are normally and independently distributed, the sample averages are the **maximum likelihood estimators** of the factor level means. That is, if the observations are normally distributed, least squares and maximum likelihood produce exactly the same estimators of the factor level means. Maximum likelihood is a more general method of parameter estimation that usually produces parameter estimates that have excellent statistical properties.

We can also apply the method of least squares to the effects model. Assuming equal sample sizes, the least squares function is

$$\begin{aligned} L &= \sum_{i=1}^{2}\sum_{j=1}^{n} \varepsilon_{ij}^2 \\ &= \sum_{i=1}^{2}\sum_{j=1}^{n} (y_{ij} - \mu - \tau_i)^2 \end{aligned}$$

and the partial derivatives of L with respect to the parameters are

$$\frac{\partial L}{\partial \mu} = 2\sum_{i=1}^{2}\sum_{j=1}^{n}(y_{ij} - \mu - \tau_i), \quad \frac{\partial L}{\partial \tau_1} = 2\sum_{j=1}^{n}(y_{1j} - \mu - \tau_1), \text{and } \frac{\partial L}{\partial \tau_2} = 2\sum_{j=1}^{n}(y_{2j} - \mu - \tau_2)$$

Equating these partial derivatives to zero results in the following least squares normal equations:

$$\begin{aligned} 2n\hat{\mu} + n\hat{\tau}_1 + n\hat{\tau}_2 &= \sum_{i=1}^{2}\sum_{j=1}^{n} y_{ij} \\ n\hat{\mu} + n\hat{\tau}_1 \qquad &= \sum_{j=1}^{n} y_{1j} \\ n\hat{\mu} \qquad + n\hat{\tau}_2 &= \sum_{j=1}^{n} y_{2j} \end{aligned}$$

Notice that if we add the last two 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. This has occurred because the effects model is **overparameterized**. This

situation occurs frequently; that is, the effects model for an experiment will always be an overparameterized model.

One way to deal with this problem is to add another linearly independent equation to the normal equations. The most common way to do this is to use the equation $\hat{\tau}_1 + \hat{\tau}_2 = 0$. This is, in a sense, an intuitive choice as it essentially defines the factor effects as deviations from the overall mean μ. If we impose this constraint, the solution to the normal equations is

$$\hat{\mu} = \bar{y}$$
$$\hat{\tau}_i = \bar{y}_i - \bar{y}, i = 1,2$$

That is, the overall mean is estimated by the average of all 2*n* 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.

This is not the only possible choice for a linearly independent "constraint" for solving the normal equations. Another possibility is to simply set the overall mean equal to a constant, such as for example $\hat{\mu} = 0$. This results in the solution

$$\hat{\mu} = 0$$
$$\hat{\tau}_i = \bar{y}_i, i = 1,2$$

Yet another possibility is $\hat{\tau}_2 = 0$, producing the solution

$$\hat{\mu} = \bar{y}_2$$
$$\hat{\tau}_1 = \bar{y}_1 - \bar{y}_2$$
$$\hat{\tau}_2 = 0$$

There are an infinite number of possible constraints that could be used to solve the normal equations. An obvious question is "which solution should we use?" It turns out that it really doesn't matter. For each of the three solutions above (indeed for *any* solution to the normal equations) we have

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

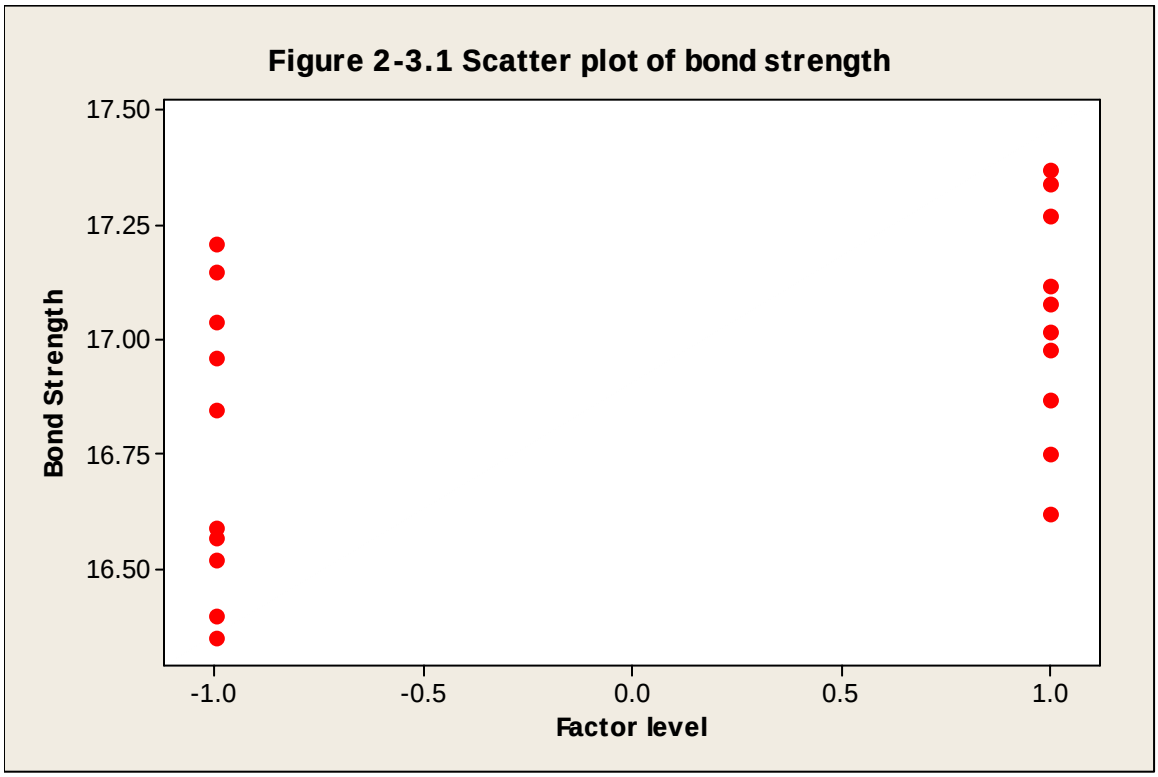
That is, the least squares estimator of the mean of the *i*th factor level will always be the sample average of the observations at that factor level. So even if we cannot obtain unique estimates for the parameters in the effects model we *can* obtain unique estimators of a *function* of these parameters that we are interested in. We say that the mean of the *i*th factor level is *estimable*. Any function of the model parameters that can be uniquely estimated regardless of the constraint selected to solve the normal equations is called an **estimable function**. This is discussed in more detail in Chapter 3.

S2-3. A Regression Model Approach to the *t*-Test

The two-sample *t*-test can be presented from the viewpoint of a simple linear regression model. This is a very instructive way to think about the *t*-test, as it fits in nicely with the general notion of a factorial experiment with factors at two levels, such as the golf

experiment described in Chapter 1. This type of experiment is very important in practice, and is discussed extensively in subsequent chapters.

In the *t*-test scenario, we have a factor *x* with two levels, which we can arbitrarily call "low" and "high". We will use *x* = -1 to denote the low level of this factor and *x* = +1 to denote the high level of this factor. Figure 2-3.1 below is a scatter plot (from Minitab) of the portland cement mortar tension bond strength data in Table 2-1 of Chapter 2.



We will a simple linear regression model to this data, say

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}$$

where β_0 and β_1 are the intercept and slope, respectively, of the regression line and the regressor or predictor variable is $x_{1j} = -1$ and $x_{2j} = +1$. The method of least squares can be used to estimate the slope and intercept in this model. Assuming that we have equal sample sizes *n* for each factor level the least squares normal equations are:

$$2n\hat{\beta}_0 = \sum_{i=1}^{2}\sum_{j=1}^{n} y_{ij}$$

$$2n\hat{\beta}_1 = \sum_{j=1}^{n} y_{2j} - \sum_{j=1}^{n} y_{1j}$$

The solution to these equations is

$$\hat{\beta}_0 = \bar{y}$$

$$\hat{\beta}_1 = \frac{1}{2}(\bar{y}_2 - \bar{y}_1)$$

Note that the least squares estimator of the intercept is the average of all the observations from both samples, while the estimator of the slope is one-half of the difference between the sample averages at the "high" and "low' levels of the factor *x*. Below is the output from the linear regression procedure in Minitab for the tension bond strength data.

Regression Analysis: Bond Strength versus Factor level

```
The regression equation is
Bond Strength = 16.9 + 0.139 Factor level


Predictor        Coef  SE Coef       T      P
Constant      16.9030   0.0636  265.93  0.000
Factor level  0.13900  0.06356    2.19  0.042


S = 0.284253   R-Sq = 21.0%   R-Sq(adj) = 16.6%


Analysis of Variance

Source           DF       SS       MS     F      P
Regression        1  0.38642  0.38642  4.78  0.042
Residual Error   18  1.45440  0.08080
Total            19  1.84082
```

Notice that the estimate of the slope (given in the column labeled "Coef" and the row labeled "Factor level" above) is $0.139 = \frac{1}{2}(\bar{y}_2 - \bar{y}_1) = \frac{1}{2}(17.0420 - 16.7640)$ and the estimate of the intercept is 16.9030. Furthermore, notice that the *t*-statistic associated with the slope is equal to 2.19, exactly the same value (apart from sign) that we gave in the Minitab two-sample t-test output in Table 2-2 in the text. Now in simple linear regression, the *t*-test on the slope is actually testing the hypotheses

$$H_0: \beta_1 = 0$$

$$H_0: \beta_1 \neq 0$$

and this is equivalent to testing $H_0: \mu_1 = \mu_2$.

It is easy to show that the *t*-test statistic used for testing that the slope equals zero in simple linear regression is identical to the usual two-sample *t*-test. Recall that to test the above hypotheses in simple linear regression the *t*-statistic is

$$t_0 = \frac{\hat{\beta}_1}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}}$$

where $S_{xx} = \sum_{i=1}^{2}\sum_{j=1}^{n}(x_{ij} - \bar{x})^2$ is the "corrected" sum of squares of the x's. Now in our specific problem, $\bar{x} = 0, x_{1j} = -1$ and $x_{2j} = +1$, so $S_{xx} = 2n$. Therefore, since we have already observed that the estimate of σ is just S_p,

$$t_0 = \frac{\hat{\beta}_1}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} = \frac{\frac{1}{2}(\bar{y}_2 - \bar{y}_1)}{S_p\sqrt{\frac{1}{2n}}} = \frac{\bar{y}_2 - \bar{y}_1}{S_p\sqrt{\frac{2}{n}}}$$

This is the usual two-sample *t*-test statistic for the case of equal sample sizes.

S2-4. Constructing Normal Probability Plots

While we usually generate normal probability plots using a computer software program, occasionally we have to construct them by hand. Fortunately, it's relatively easy to do, since specialized **normal probability plotting paper** is widely available. This is just graph paper with the vertical (or probability) scale arranged so that if we plot the cumulative normal probabilities $(j - 0.5)/n$ on that scale versus the rank-ordered observations $y_{(j)}$ a graph equivalent to the computer-generated normal probability plot will result. The table below shows the calculations for the unmodified portland cement mortar bond strength data.

j	$y_{(j)}$	$(j - 0.5)/10$	$z_{(j)}$
1	16.62	0.05	-1.64
2	16.75	0.15	-1.04
3	16.87	0.25	-0.67
4	16.98	0.35	-0.39
5	17.02	0.45	-0.13
6	17.08	0.55	0.13
7	17.12	0.65	0.39
8	17.27	0.75	0.67
9	17.34	0.85	1.04
10	17.37	0.95	1.64

Now if we plot the cumulative probabilities from the next-to-last column of this table versus the rank-ordered observations from the second column on normal probability paper, we will produce a graph that is identical to the results for the unmodified mortar formulation that is shown in Figure 2-11 in the text.

A normal probability plot can also be constructed on ordinary graph paper by plotting the standardized normal *z*-scores $z_{(j)}$ against the ranked observations, where the standardized normal *z*-scores are obtained from

$$P(Z \le z_j) = \Phi(z_j) = \frac{j - 0.5}{n}$$

where $\Phi(\bullet)$ denotes the standard normal cumulative distribution. For example, if $(j - 0.5)/n = 0.05$, then $\Phi(z_j) = 0.05$ implies that $z_j = -1.64$. The last column of the above table displays the values of the normal *z*-scores. Plotting these values against the ranked observations on ordinary graph paper will produce a normal probability plot equivalent to the unmodified mortar results in Figure 2-11. As noted in the text, many statistics computer packages present the normal probability plot this way.

S2-5. More About Checking Assumptions in the *t*-Test

We noted in the text that a normal probability plot of the observations was an excellent way to check the normality assumption in the *t*-test. Instead of plotting the observations, an alternative is to plot the *residuals* from the statistical model.

Recall that the means model is

$$y_{ij} = \mu_i + \varepsilon_{ij} \begin{cases} i = 1,2 \\ j = 1,2,\cdots,n_i \end{cases}$$

and that the estimates of the parameters (the factor level means) in this model are the sample averages. Therefore, we could say that the *fitted* model is

$$\hat{y}_{ij} = \bar{y}_i, i = 1,2 \text{ and } j = 1,2,\cdots,n_i$$

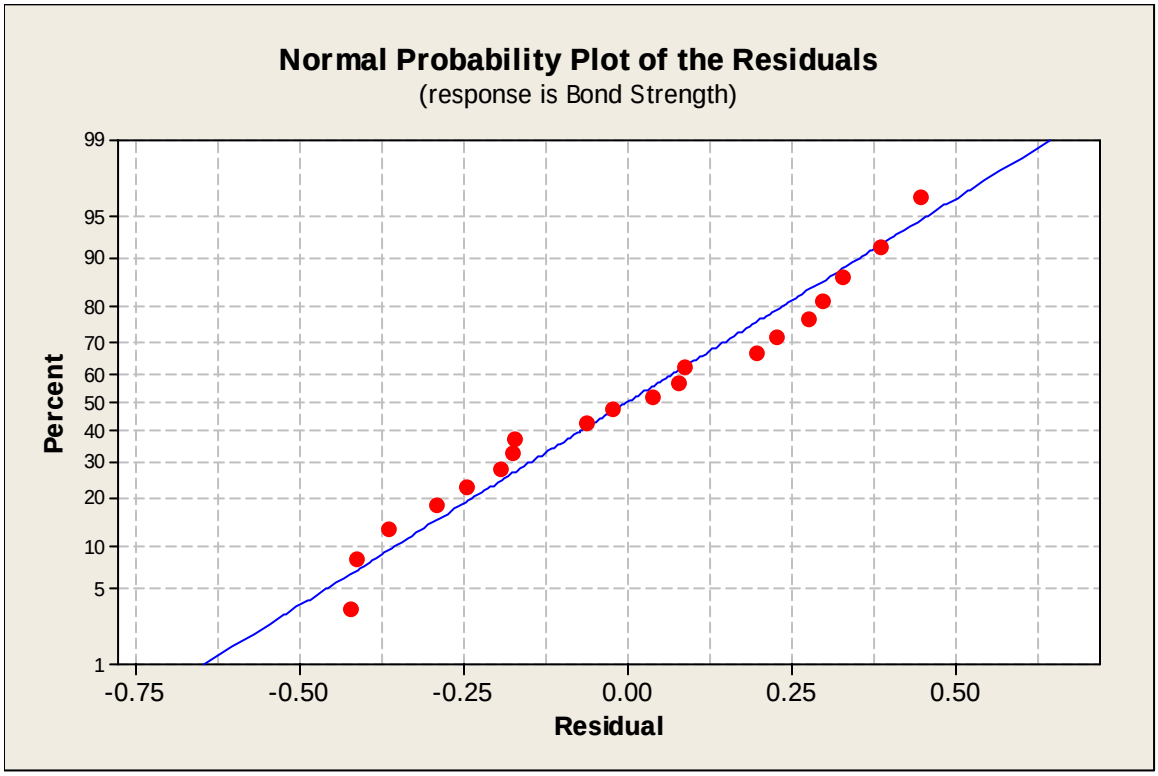
That is, an estimate of the *ij*th observation is just the average of the observations in the *i*th factor level. The difference between the observed value of the response and the predicted (or fitted) value is called a **residual**, say

$$e_{ij} = y_{ij} - \hat{y}_i, i = 1,2 .$$

The table below computes the values of the residuals from the portland cement mortar tension bond strength data.

Observation j	y_{1j}	$e_{1j} = y_{1j} - \bar{y}_1 = y_{1j} - 16.76$	y_{2j}	$e_{2j} = y_{2j} - \bar{y}_2 = y_{2j} - 17.04$
1	16.85	0.09	16.62	-0.42
2	16.40	-0.36	16.75	-0.29
3	17.21	0.45	17.37	0.33
4	16.35	-0.41	17.12	0.08
5	16.52	-0.24	16.98	-0.06
6	17.04	0.28	16.87	-0.17
7	16.96	0.20	17.34	0.30
8	17.15	0.39	17.02	-0.02
9	16.59	-0.17	17.08	0.04
10	16.57	-0.19	17.27	0.23

The figure below is a normal probability plot of these residuals from Minitab.



As noted in section 2-3 above we can compute the *t*-test statistic using a simple linear regression model approach. Most regression software packages will also compute a table or listing of the residuals from the model. The residuals from the Minitab regression model fit obtained previously are as follows:

Obs	Factor level	Bond Strength	Fit	SE Fit	Residual	St Resid
1	-1.00	16.8500	16.7640	0.0899	0.0860	0.32
2	-1.00	16.4000	16.7640	0.0899	-0.3640	-1.35
3	-1.00	17.2100	16.7640	0.0899	0.4460	1.65
4	-1.00	16.3500	16.7640	0.0899	-0.4140	-1.54
5	-1.00	16.5200	16.7640	0.0899	-0.2440	-0.90
6	-1.00	17.0400	16.7640	0.0899	0.2760	1.02
7	-1.00	16.9600	16.7640	0.0899	0.1960	0.73
8	-1.00	17.1500	16.7640	0.0899	0.3860	1.43
9	-1.00	16.5900	16.7640	0.0899	-0.1740	-0.65
10	-1.00	16.5700	16.7640	0.0899	-0.1940	-0.72
11	1.00	16.6200	17.0420	0.0899	-0.4220	-1.56
12	1.00	16.7500	17.0420	0.0899	-0.2920	-1.08
13	1.00	17.3700	17.0420	0.0899	0.3280	1.22
14	1.00	17.1200	17.0420	0.0899	0.0780	0.29
15	1.00	16.9800	17.0420	0.0899	-0.0620	-0.23
16	1.00	16.8700	17.0420	0.0899	-0.1720	-0.64
17	1.00	17.3400	17.0420	0.0899	0.2980	1.11
18	1.00	17.0200	17.0420	0.0899	-0.0220	-0.08
19	1.00	17.0800	17.0420	0.0899	0.0380	0.14
20	1.00	17.2700	17.0420	0.0899	0.2280	0.85

The column labeled "Fit" contains the averages of the two samples, computed to four decimal places. The residuals in the sixth column of this table are the same (apart from rounding) as we computed manually.

S2-6. Some More Information about the Paired *t*-Test

The paired *t*-test examines the difference between two variables and test whether the mean of those differences differs from zero. In the text we show that the mean of the differences μ_d is identical to the difference of the means in two independent samples, $\mu_1 - \mu_2$. However the variance of the differences is not the same as would be observed if there were two independent samples. Let $\bar{d}$ be the sample average of the differences. Then

$$\begin{aligned} V(\bar{d}) &= V(\bar{y}_1 - \bar{y}_2) \\ &= V(\bar{y}_1) + V(\bar{y}_2) - 2Cov(\bar{y}_1, \bar{y}_2) \\ &= \frac{2\sigma^2(1-\rho)}{n} \end{aligned}$$

assuming that both populations have the same variance σ^2 and that ρ is the correlation between the two random variables y_1 and y_2. The quantity S_d^2 / n estimates the variance of the average difference $\bar{d}$. In many paired experiments a strong positive correlation is

expected to exist between y_1 and y_2 because both factor levels have been applied to the *same* experimental unit. When there is positive correlation within the pairs, the denominator for the paired *t*-test will be smaller than the denominator for the two-sample or *independent t*-test. If the two-sample test is applied incorrectly to paired samples, the procedure will generally understate the significance of the data.

Note also that while for convenience we have assumed that both populations have the same variance, the assumption is really unnecessary. The paired *t*-test is valid when the variances of the two populations are different.

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

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

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 $\mu = \dfrac{\sum_{i=1}^{a} \mu_i}{a}$ 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

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

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

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

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

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

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

$$E(\frac{1}{n}\sum_{i=1}^{a} y_{i.}^2) = \frac{1}{n}\sum_{i=1}^{a} E(n\mu + n\tau_i + \varepsilon_{i.})^2$$

Squaring the expression in parentheses and taking expectation results in

$$E(\frac{1}{n}\sum_{i=1}^{a} y_{i.}^2) = \frac{1}{n}[a(n\mu)^2 + n^2\sum_{i=1}^{a}\tau_i^2 + an\sigma^2]$$

$$= an\mu^2 + n\sum_{i=1}^{a}\tau_i^2 + a\sigma^2$$

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

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

$$= \frac{1}{an} E(an\mu + \varepsilon_{..})^2$$

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

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

$$= an\mu^2 + \sigma^2$$

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

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

$$= an\mu^2 + n\sum_{i=1}^{a}\tau_i^2 + a\sigma^2 - (an\mu^2 + \sigma^2)$$

$$= \sigma^2(a-1) + n\sum_{i=1}^{a}\tau_i^2$$

Consequently the expected value of the mean square for treatments is

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

$$= \frac{\sigma^2(a-1) + n\sum_{i=1}^{a}\tau_i^2}{a-1}$$

$$\sigma^2 + \frac{n\sum_{i=1}^{a}\tau_i^2}{a-1}$$

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,

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

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

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

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

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

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 even 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

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

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

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

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

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

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

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

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

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 singe 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:

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

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 fine 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

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

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

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

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

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

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

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

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 $\mu, \tau_1, \cdots, \tau_a$ 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

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

The equivalent regression model is

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

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

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

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

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

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

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

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

$$\begin{aligned} \beta_0 &= \mu_3 \\ \beta_1 &= \mu_1 - \mu_3 \\ \beta_2 &= \mu_2 - \mu_3 \end{aligned}$$

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

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

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

$$\beta_0 = \mu_a$$
$$\beta_i = \mu_i - \mu_a, i = 1,2,\cdots,a-1$$

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

$$\beta_0 = \mu_a$$
$$\beta_i = 0, i = 1,2,\cdots,a-1$$

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 4^{th} treatment. Furthermore, each regression coefficient is just the difference between one of the treatment means and the 4^{th} treatment mean.

Chapter 4 Supplemental Text Material

S4-1. Relative Efficiency of the RCBD

In Example 4-1 we illustrated the noise-reducing property of the randomized complete block design (RCBD). If we look at the portion of the total sum of squares not accounted for by treatments (302.14; see Table 4-4), about 63 percent (192.25) is the result of differences between blocks. Thus, if we had run a completely randomized design, the mean square for error MS_E would have been much larger, and the resulting design would not have been as sensitive as the randomized block design.

It is often helpful to estimate the relative efficiency of the RCBD compared to a completely randomized design (CRD). One way to define this relative efficiency is

$$R = \frac{(df_b + 1)(df_r + 3)}{(df_b + 3)(df_r + 1)} \cdot \frac{\sigma_r^2}{\sigma_b^2}$$

where σ_r^2 and σ_b^2 are the experimental error variances of the completely randomized and randomized block designs, respectively, and df_r and df_b are the corresponding error degrees of freedom. This statistic may be viewed as the increase in replications that is required if a CRD is used as compared to a RCBD if the two designs are to have the same sensitivity. The ratio of degrees of freedom in R is an adjustment to reflect the different number of error degrees of freedom in the two designs.

To compute the relative efficiency, we must have estimates of σ_r^2 and σ_b^2. We can use the mean square for error MS_E from the RCBD to estimate σ_b^2, and it may be shown [see Cochran and Cox (1957), pp. 112-114] that

$$\hat{\sigma}_r^2 = \frac{(b-1)\,MS_{Blocks} + b(a-1)\,MS_E}{ab-1}$$

is an unbiased estimator of the error variance of a the CRD. To illustrate the procedure, consider the data in Example 4-1. Since MS_E = 7.33, we have

$$\hat{\sigma}_b^2 = 7.33$$

and

$$\begin{aligned}\hat{\sigma}_r^2 &= \frac{(b-1)MS_{Blocks} + b(a-1)MS_E}{ab-1} \\ &= \frac{(5)38.45 + 6(3)7.33}{4(6)-1} \\ &= 14.10\end{aligned}$$

Therefore our estimate of the relative efficiency of the RCBD in this example is

$$R = \frac{(df_b + 1)(df_r + 3)}{(df_b + 3)(df_r + 1)} \cdot \frac{\sigma_r^2}{\sigma_b^2}$$
$$= \frac{(15+1)(20+3)}{(15+3)(20+1)} \cdot \frac{14.10}{7.33}$$
$$= 1.87$$

This implies that we would have to use approximately twice times as many replicates with a completely randomized design to obtain the same sensitivity as is obtained by blocking on the metal coupons.

Clearly, blocking has paid off handsomely in this experiment. However, suppose that blocking was not really necessary. In such cases, if experimenters choose to block, what do they stand to lose? In general, the randomized complete block design has $(a - 1)(b - 1)$ error degrees of freedom. If blocking was unnecessary and the experiment was run as a completely randomized design with b replicates we would have had $a(b - 1)$ degrees of freedom for error. Thus, incorrectly blocking has cost $a(b - 1) - (a - 1)(b - 1) = b - 1$ degrees of freedom for error, and the test on treatment means has been made less sensitive needlessly. However, if block effects really are large, then the experimental error may be so inflated that significant differences in treatment means could possibly remain undetected. (Remember the incorrect analysis of Example 4-1.) As a general rule, when the importance of block effects is in doubt, the experimenter should block and gamble that the block means are different. If the experimenter is wrong, the slight loss in error degrees of freedom will have little effect on the outcome as long as a moderate number of degrees of freedom for error are available.

S4-2. Partially Balanced Incomplete Block Designs

Although we have concentrated on the balanced case, there are several other types of incomplete block designs that occasionally prove useful. BIBDs do not exist for all combinations of parameters that we might wish to employ because the constraint that λ be an integer can force the number of blocks or the block size to be excessively large. For example, if there are eight treatments and each block can accommodate three treatments, then for λ to be an integer the smallest number of replications is $r = 21$. This leads to a design of 56 blocks, which is clearly too large for most practical problems. To reduce the number of blocks required in cases such as this, the experimenter can employ **partially balanced incomplete block** designs, or PBIDs, in which some pairs of treatments appear together λ_1 times, some pairs appear together λ_2 times, . . ., and the remaining pairs appear together λ_m times. Pairs of treatments that appear together λ_i times are called ith *associates*. The design is then said to have m **associate classes**.

An example of a PBID is shown in Table 1. Some treatments appear together $\lambda_1 = 2$ times (such as treatments 1 and 2), whereas others appear together only $\lambda_2 = 1$ times (such as treatments 4 and 5). Thus, the design has two associate classes. We now describe the **intrablock analysis** for these designs.

A partially balanced incomplete block design with two associate classes is described by the following parameters:

1 . There are a treatments arranged in b blocks. Each block contains k runs and each treatment appears in r blocks.

2. Two treatments which are ith associates appear together in λ_i blocks, $i = 1, 2$.

3. Each treatment has exactly n_i ith associates, $i = 1,2$. The number n_i is independent of the treatment chosen.

4. If two treatments are ith associates, then the number of treatments that are jth associates of one treatment and kth associates of the other treatment is p^i_{jk}, $(i , j , k = 1, 2)$. It is convenient to write the p^i_{jk} as (2 x 2) matrices with p^i_{jk} the jkth element of the ith matrix. 4

For the design in Table 1 we may readily verify that $a = 6$, $b = 6$, $k = 3$, $r = 3$, $\lambda_1 = 2$, $\lambda_2 = 1$, $n_1 = 1$, $n_2 = 4$,

$$\{p^1{}_{jk}\} = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} \qquad \text{and} \qquad \{p^2{}_{jk}\} = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$$

Table 1. A Partially Balanced incomplete Block Design with Two Associate Classes

Block	Treatment Combinations		
1	1	2	3
2	3	4	5
3	2	5	6
4	1	2	4
5	3	4	6
6	1	5	6

We now show how to determine the p^i_{jk} . Consider any two treatments that are first associates, say 1 and 2. For treatment 1, the only first associate is 2 and the second associates are 3, 4, 5, and 6. For treatment 2, the only first associate is 1and the second associates are 3, 4, 5, and 6. Combining this information produces Table 2. Counting the number of treatments in the cells of this table, have the $\{p^1_{jk}\}$ given above. The elements $\{p^2_{jk}\}$ are determined similarly.

The linear statistical model for the partially balanced incomplete block design with two associate classes is

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

where μ is the overall mean, τ_i is the *i*th treatment effect, β_j is the *jth* block effect, and ε_{ij} is the NID(0, σ^2) random error component. We compute a total sum of squares, a block sum of squares (unadjusted), and a treatment sum of squares (adjusted). As before, we call

$$Q_i = y_{i.} - \frac{1}{k}\sum_{j=1}^{b} n_{ij} y_{.j}$$

the adjusted total for the *i*th treatment. We also define

$$S_1(Q_i) = \sum_s Q_s \qquad \text{s and i are first associates}$$

$$\Delta = k^{-2}\{(rk - r + \lambda_1)(rk - r + \lambda_2) + (\lambda_1 + \lambda_2)$$

$$c_1 = (k\Delta)^{-1}[\lambda_1(rk - r + \lambda_2) + (\lambda_1 - \lambda_2)(\lambda_2 p^1_{12} - \lambda_1 p^2_{12})]$$

$$c_2 = (k\Delta)^{-1}[\lambda_2(rk - r + \lambda_1) + (\lambda_1 - \lambda_2)(\lambda_2 p^1_{12} - \lambda_1 p^2_{12})]$$

The estimate of the *i*th treatment effect is

$$\hat{\tau}_i = \frac{1}{r(k-1)}[(k - c_2)Q_i + (c_1 - c_2)S_1(Q_i)]$$

and the adjusted treatment sum of squares is

$$SS_{Treatments(adjusted)} = \sum_{i=1}^{a} \hat{\tau}_i Q_i$$

The analysis of variance is summarized in Table 3. To test H_0: $\tau_i = 0$, we use $F_0 = MS_{Treatments(adjusted)} / MS_E$.

Table 2. Relationship of Treatments to 1 and 2

	Treatment 2	
Treatment 1	1st Associate	2nd Associate
1st associate		
2nd associate		3,4,5,6

Table 3. Analysis of Variance for the Partially Balanced Incomplete Block Design with Two Associate Classes

Source of Variation	Sum of Squares	Degrees of Freedom
Treatments (adjusted)	$\sum_{i=1}^{a} \hat{\tau}_i Q_i$	a-1
Blocks	$\frac{1}{k}\sum_{j=1}^{b} y^2_{.j} - \frac{y^2..}{bk}$	b-1
Error	Subtraction	bk-b-a+1
Total	$\sum_i \sum_j y^2_{ij} - \frac{y^2..}{bk}$	bk-1

We may show that the variance of any contrast of the form $\hat{\tau}_u - \hat{\tau}_v$ is

$$V(\tau_u - \tau_v) = \frac{2(k - c_i)\sigma^2}{r(k-1)}$$

where treatments *u* and *v* are *i*th associates (i = 1, 2). This indicates that comparisons between treatments are not all estimated with the same precision. This is a consequence of the partial balance of the design.

We have given only the intrablock analysis. For details of the interblock analysis, refer to Bose and Shimamoto (1952) or John (1971). The second reference contains a good discussion of the general theory of incomplete block designs. An extensive table of partially balanced incomplete block designs with two associate classes has been given by Bose, Clatworthy, and Shrikhande (1954).

S4-3. Youden Squares

Youden squares are "incomplete" Latin square designs in which the number of columns does not equal the number of rows and treatments. For example, consider the design shown in Table 4. Notice that if we append the column *(E, A, B, C, D)* to this design, the result is a 5×5 Latin square. Most of these designs were developed by W. J. Youden, hence their name.

Although a Youden square is always a Latin square from which at least one column (or row or diagonal) is missing, it is not necessarily true that every Latin square with more than one column (or row or diagonal) missing is a Youden square. The arbitrary removal of more than one column, say, for a Latin square may destroy its balance. In general, a

Youden square is a symmetric balanced incomplete block design in which rows correspond to blocks and each treatment occurs exactly once in each column or "position" of the block. Thus, it is possible to construct Youden squares from all symmetric balanced incomplete block designs, as shown by Smith and Hartley (l948). A table of Youden squares is given in Davies (1956), and other types of incomplete Latin squares are discussed by Cochran and Cox (1957, Chapter 13).

Table 4. A Youden Square for Five Treatments (A, *B, C, D, E)*

	Column			
Row	1	2	3	4
1	A	B	C	D
2	B	C	D	E
3	C	D	E	A
4	D	E	A	B
5	E	A	B	C

The linear model for a Youden square is

$$Y_{ijh} = \mu + \alpha_i + \tau_j + \beta_h + \varepsilon_{ijh}$$

where, μ is the overall mean, α_i is the *i*th block effect τ_j is the *j*th treatment effect, β_h is the *h*th position effect, and ε_{ijh} is the usual NID(0, σ^2) error term. Since positions occur exactly once in each block and once with each treatment, positions are orthogonal to blocks and treatments. The analysis of the Youden square is similar to the analysis of a balanced incomplete block design, except that a sum of squares between the position totals may also be calculated.

Example of a Youden Square

An industrial engineer is studying the effect of five illumination levels on the occurrence of defects in an assembly operation. Because time may be a factor in the experiment, she has decided to run the experiment in five blocks, where each block is a day of the week**.** However, the department in which the experiment is conducted has four work stations and these stations represent a potential source of variability. The engineer decided to run a Youden square with five rows (days or blocks), four columns (work stations), and five treatments (the illumination levels). The coded data are shown in Table 5.

Table 5. The Youden Square Design used in the Example

Day (Block)	Work Station					Treatment totals
	1	2	3	4	$y_{i..}$	
1	A=3	B=1	C=-2	D=0	2	$y_{.1.}$=12 (A)
2	B=0	C=0	D=-1	E=7	6	$y_{.2.}$=2 (B)
3	C=-1	D=0	E=5	A=3	7	$y_{.3.}$=-4 (C)
4	D=-1	E=6	A=4	B=0	9	$y_{.4.}$=-2 (D)
5	E=5	A=2	B=1	C=-1	7	$y_{.5.}$=23 (E)
$y_{..h}$	6	9	7	9	$y_{...}$=31	

Considering this design as a balanced incomplete block, we find a = *b* = 5, *r*= *k* = 4, and k = 3. Also,

$$SS_T = \sum_i \sum_j \sum_h y^2{}_{ijh} - \frac{y^2 \ldots}{N} = 183.00 - \frac{(31)^2}{20} = 134.95$$

$$Q_1 = 12 - \frac{1}{4}(2 + 7 + 9 + 7) = 23/4$$

$$Q_2 = 2 - \frac{1}{4}(2 + 6 + 9 + 7) = -16/4$$

$$Q_3 = -4 - \frac{1}{4}(2 + 6 + 7 + 7) = -38/4$$

$$Q_4 = -2 - \frac{1}{4}(2 + 6 + 7 + 9) = -32/4$$

$$Q_5 = 23 - \frac{1}{4}(6 + 7 + 9 + 7) = 63/4$$

$$SS_{Treatments(adjusted)} = \frac{k\sum_{i=1}^{a} Q_i{}^2}{\lambda a}$$

$$= \frac{4[(23/4)^2 + (-16/4)^2 + (-38/4)^2 + (-32/4)^2 + (63/4)^2}{(3)(5)} = 120.37$$

Also,

$$SS_{Days} = \sum_{i=1}^{b} \frac{y^2_{i..}}{k} - \frac{y^2_{...}}{N} = \frac{(2)^2 + (6)^2 + (7)^2 + (9)^2 + (7)^2}{4} - \frac{(31)^2}{20} = 6.70$$

$$SS_{Stations} = \sum_{h=1}^{k} \frac{y^2_{..h}}{b} - \frac{y^2_{...}}{N} = \frac{(6)^2 + (9)^2 + (7)^2 + (9)^2}{5} - \frac{(31)^2}{20} = 1.35$$

and

$SS_E = SS_T - SS_{Treatments\ (adjusted)} - SS_{Days} - SS_{Stations}$
$= 134.95 - 120.37 - 6.70 - 1.35 = 6.53$

Block or day effects may be assessed by computing the adjusted sum of squares for blocks. This yields

$Q_1' = 2 - \frac{1}{4}(12 + 2 - 4 - 2) = 0/4$

$Q_2' = 6 - \frac{1}{4}(2 - 3 - 2 + 23) = 5/4$

$Q_3' = 7 - \frac{1}{4}(12 - 4 - 2 + 23) = - 1/4$

$Q_4' = 9 - \frac{1}{4}(12 + 2 - 2 + 23) = 1/4$

$Q_5' = 7 - \frac{1}{4}(12 + 2 - 4 + 23) = -5/4$

$$SS_{Days(adjusted)} = \frac{r\sum_{j=1}^{b} Q_j'^2}{\lambda b}$$

$$= \frac{4[(0/4)^2 + (5/4)^2 + (-1/4)^2 + (1/4)^2 + (-5/4)^2]}{(3)(5)} = 0.87$$

The complete analysis of variance is shown in Table 6. Illumination levels are significantly different at 1 percent.

Table 6 Analysis of Variance for the Youden Square Example

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Illumination level, adjusted	120.37	4	30.09	36.87[a]
Days, unadjusted	6.70	4	-	
Days, adjusted	(0.87)	(4)	0.22	
Work Station	1.35	3	0.45	
Error	6.53	8	0.82	
Total	134.95	19		

[a] Significant at 1 percent.

S4-4. Lattice Designs

Consider a balanced incomplete block design with k^2 treatments arranged in $b = k(k + 1)$ blocks with k runs per block and $r = k + 1$ replicates. Such a design is called a **balanced lattice**. An example is shown in Table 7 for k^2=9 treatments in 12 blocks of 3 runs each. Notice that the blocks can be grouped into sets such that each set contains a complete replicate. The analysis of variance for the balanced lattice design proceeds like that for a balanced incomplete block design, except that a sum of squares for replicates is computed and removed from the sum of squares for blocks. Replicates will have k degrees of freedom and blocks will have k^2-1 degrees of freedom.

Lattice designs are frequently used in situations where there are a large number of treatment combinations. In order to reduce the size of the design, the experimenter may resort to **partially** balanced lattices. We very briefly describe some of these designs. Two replicates of a design for k^2 treatments in $2k$ blocks of k runs are called a **simple lattice**. For example, consider the first two replicates of the design in Table 7. The partial balance is easily seen, since, for example, treatment 2 appears in the same block with treatments 1, 3, 5, and 8, but does not appear at all with treatments 4, 6, 7, and 9. A lattice design with k^2 treatments in $3k$ blocks grouped into three replicates is called a **triple lattice**. An example would be the first three replicates in Table 7. A lattice design for k^2 treatments in $4k$ blocks arranged in four replicates is called a **quadruple lattice**.

Table 7. A 3 x 3 Balanced Lattice Design

Block	**Replicate 1**			**Block**	**Replicate 3**		
1	1	2	3	7	1	5	9
2	4	5	6	8	7	2	6
3	7	8	9	9	4	8	3
Block	**Replicate 2**			**Block**	**Replicate 4**		
1	1	4	7	10	1	8	6
2	2	5	8	11	4	2	9
3	3	6	9	12	7	5	3

There are other types of lattice designs that occasionally prove useful. For example, the **cubic lattice** design can be used for k^3 treatments in k^2 blocks of k runs. A lattice design for $k(k + 1)$ treatments in $k + 1$ blocks of size k is called a **rectangular lattice**. Details of the analysis of lattice designs and tables of plans are given in Cochran and Cox (1957).

Supplemental References

Bose, R. C. and T. Shimamoto (1952). "Classification and Analysis of Partially Balanced Incomplete Block Designs with Two Associate Classes". *Journal of the American Statistical Association*, Vol. 47, pp. 151-184.

Bose, R. C. W. H. Clatworthy, and S. S. Shrikhande (1954). *Tables of Partially Balanced Designs with Two Associate Classes*. Technical Bulletin No. 107, North Carolina Agricultural Experiment Station.

Smith, C. A. B. and H. O. Hartley (1948). "Construction of Youden Squares". *Journal of the Royal Statistical Society Series B*, Vol. 10, pp. 262-264.

Chapter 5 Supplemental Text Material

S5-1. Expected Mean Squares in the Two-factor Factorial

Consider the two-factor fixed effects model

$$y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \begin{cases} i = 1,2,\cdots,a \\ j = 1,2,\cdots,b \\ k = 1,2,\cdots,n \end{cases}$$

given as Equation (5-1) in the textbook. We list the expected mean squares for this model, but do not develop them. It is relatively easy to develop the expected mean squares from direct application of the expectation operator.

Consider finding

$$E(MS_A) = E\left(\frac{SS_A}{a-1}\right) = \frac{1}{a-1}E(SS_A)$$

where SS_A is the sum of squares for the row factor. Since

$$SS_A = \frac{1}{bn}\sum_{i=1}^{a} y_{i..}^2 - \frac{y_{...}^2}{abn}$$

$$E(SS_A) = \frac{1}{bn}E\sum_{i=1}^{a} y_{i..}^2 - E\left(\frac{y_{...}^2}{abn}\right)$$

Recall that $\tau_{.} = 0, \beta_{.} = 0, (\tau\beta)_{.j} = 0, (\tau\beta)_{i.} = 0,$ and $(\tau\beta)_{..} = 0$, where the "dot" subscript implies summation over that subscript. Now

$$\begin{aligned} y_{i..} &= \sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} = bn\mu + bn\tau_i + n\beta_{.} + n(\tau\beta)_{i.} + \varepsilon_{i..} \\ &= bn\mu + bn\tau_i + \varepsilon_{i..} \end{aligned}$$

and

$$\begin{aligned} \frac{1}{bn}E\sum_{i=1}^{a} y_{i..}^2 &= \frac{1}{bn}E\sum_{i=1}^{a}\left[(bn\mu)^2 + (bn)^2\tau_i^2 + \varepsilon_{i..}^2 + 2(bn)^2\mu\tau_i + 2bn\mu\varepsilon_{i..} + 2bn\tau_i\varepsilon_{i..}\right] \\ &= \frac{1}{bn}\left[a(bn\mu)^2 + (bn)^2\sum_{i=1}^{a}\tau_i^2 + abn\sigma^2\right] \\ &= abn\mu^2 + bn\sum_{i=1}^{a}\tau_i^2 + a\sigma^2 \end{aligned}$$

Furthermore, we can easily show that

$$y_{...} = abn\mu + \varepsilon_{...}$$

so

$$\frac{1}{abn}E(y_{...}^2) = \frac{1}{abn}E(abn\mu + \varepsilon_{...})^2$$
$$= \frac{1}{abn}E\left[(abn\mu)^2 + \varepsilon_{...}^2 + 2abn\mu\varepsilon_{...})\right]$$
$$= \frac{1}{abn}\left[(abn\mu)^2 + abn\sigma^2\right]$$
$$= abn\mu^2 + \sigma^2$$

Therefore

$$E(MS_A) = E\left(\frac{SS_A}{a-1}\right)$$
$$= \frac{1}{a-1}E(SS_A)$$
$$\frac{1}{a-1}\left[abn\mu^2 + bn\sum_{i=1}^{a}\tau_i^2 + a\sigma^2 - (abn\mu^2 + \sigma^2)\right]$$
$$= \frac{1}{a-1}\left[\sigma^2(a-1) + bn\sum_{i=1}^{a}\tau_i^2\right]$$
$$= \sigma^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$$

which is the result given in the textbook. The other expected mean squares are derived similarly.

S5-2. The Definition of Interaction

In Section 5-1 we introduced both the effects model and the means model for the two-factor factorial experiment. If there is no interaction in the two-factor model, then

$$\mu_{ij} = \mu + \tau_i + \beta_j$$

Define the row and column means as

$$\mu_{i.} = \frac{\sum_{j=1}^{b}\mu_{ij}}{b}$$
$$\mu_{.j} = \frac{\sum_{i=1}^{a}\mu_{ij}}{a}$$

Then if there is no interaction,

$$\mu_{ij} = \mu_{i.} + \mu_{.j} - \mu$$

where $\mu = \sum_i \mu_{i.} / a = \sum_j \mu_{.j} / b$. It can also be shown that if there is no interaction, each cell mean can be expressed in terms of three other cell means:

$$\mu_{ij} = \mu_{ij'} + \mu_{i'j} - \mu_{i'j'}$$

This illustrates why a model with no interaction is sometimes called an **additive model**, or why we say the treatment effects are additive.

When there is interaction, the above relationships do not hold. Thus the interaction term $(\tau\beta)_{ij}$ can be defined as

$$(\tau\beta)_{ij} = \mu_{ij} - (\mu + \tau_i + \beta_j)$$

or equivalently,

$$\begin{aligned}(\tau\beta)_{ij} &= \mu_{ij} - (\mu_{ij'} + \mu_{i'j} - \mu_{i'j'}) \\ &= \mu_{ij} - \mu_{ij'} - \mu_{i'j} + \mu_{i'j'}\end{aligned}$$

Therefore, we can determine whether there is interaction by determining whether all the cell means can be expressed as $\mu_{ij} = \mu + \tau_i + \beta_j$.

Sometimes interactions are a result of the **scale** on which the response has been measured. Suppose, for example, that factor effects act in a multiplicative fashion,

$$\mu_{ij} = \mu\tau_i\beta_j$$

If we were to assume that the factors act in an additive manner, we would discover very quickly that there is interaction present. This interaction can be removed by applying a log transformation, since

$$\log \mu_{ij} = \log \mu + \log \tau_i + \log \beta_j$$

This suggests that the original measurement scale for the response was not the best one to use if we want results that are easy to interpret (that is, no interaction). The log scale for the response variable would be more appropriate.

Finally, we observe that it is very possible for two factors to interact but for the main effects for one (or even both) factor is small, near zero. To illustrate, consider the two-factor factorial with interaction in Figure 5-1 of the textbook. We have already noted that the interaction is large, AB = -29. However, the main effect of factor A is A = 1. Thus, the main effect of A is so small as to be negligible. Now this situation does not occur all that frequently, and typically we find that interaction effects are not larger than the main effects. However, large two-factor interactions can **mask** one or both of the main effects. A prudent experimenter needs to be alert to this possibility.

S5-3. Estimable Functions in the Two-factor Factorial Model

The least squares normal equations for the two-factor factorial model are given in Equation (5-14) in the textbook as:

$$abn\hat{\mu} + bn\sum_{i=1}^{a}\hat{\tau}_i + an\sum_{j=1}^{b}\beta_j + \sum_{i=1}^{a}\sum_{j=1}^{b}(\hat{\tau\beta})_{ij} = y_{\cdots}$$

$$bn\hat{\mu} + bn\hat{\tau}_i + n\sum_{j=1}^{b}\beta_j + n\sum_{j=1}^{b}(\hat{\tau\beta})_{ij} = y_{i\cdot\cdot}, i = 1,2,\cdots,a$$

$$an\hat{\mu} + n\sum_{i=1}^{a}\hat{\tau}_i + an\hat{\beta}_j + n\sum_{i=1}^{a}(\hat{\tau\beta})_{ij} = y_{\cdot j\cdot}, j = 1,2,\cdots,b$$

$$n\hat{\mu} + n\hat{\tau}_i + n\hat{\beta}_j + n(\hat{\tau\beta})_{ij} = y_{ij\cdot}, \begin{cases} i = 1,2,\cdots,a \\ j = 1,2,\cdots,b \end{cases}$$

Recall that in general an estimable function must be a linear combination of the left-hand side of the normal equations. Consider a contrast comparing the effects of row treatments i and i'. The contrast is

$$\tau_i - \tau_{i'} + (\overline{\tau\beta})_{i\cdot} - (\overline{\tau\beta})_{i'\cdot}$$

Since this is just the difference between two normal equations, it is an estimable function. Notice that the difference in any two levels of the row factor also includes the difference in average interaction effects in those two rows. Similarly, we can show that the difference in any pair of column treatments also includes the difference in average interaction effects in those two columns. An estimable function involving interactions is

$$(\tau\beta)_{ij} - (\overline{\tau\beta})_{i\cdot} - (\overline{\tau\beta})_{\cdot j} + (\overline{\tau\beta})_{\cdot\cdot}$$

It turns out that the only hypotheses that can be tested in an effects model must involve estimable functions. Therefore, when we test the hypothesis of no interaction, we are really testing the null hypothesis

$$H_0: (\tau\beta)_{ij} - (\overline{\tau\beta})_{i\cdot} - (\overline{\tau\beta})_{\cdot j} + (\overline{\tau\beta})_{\cdot\cdot} = 0 \text{ for all } i, j$$

When we test hypotheses on main effects A and B we are really testing the null hypotheses

$$H_0: \tau_1 + (\overline{\tau\beta})_{1\cdot} = \tau_2 + (\overline{\tau\beta})_{2\cdot} = \cdots = \tau_a + (\overline{\tau\beta})_{a\cdot}$$

and

$$H_0: \beta_1 + (\overline{\tau\beta})_{\cdot 1} = \beta_2 + (\overline{\tau\beta})_{\cdot 2} = \cdots = \beta_b + (\overline{\tau\beta})_{\cdot b}$$

That is, we are not really testing a hypothesis that involves only the equality of the treatment effects, but instead a hypothesis that compares treatment effects *plus* the average interaction effects in those rows or columns. Clearly, these hypotheses may not be of much interest, or much practical value, when interaction is large. This is why in the textbook (Section 5-1) that when interaction is large, main effects may not be of much practical value. Also, when interaction is large, the statistical tests on main effects may not really tell us much about the individual treatment effects. Some statisticians do not even conduct the main effect tests when the no-interaction null hypothesis is rejected.

It can be shown [see Myers and Milton (1991)] that the original effects model

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$$

can be re-expressed as

$$y_{ijk} = [\mu + \bar{\tau} + \bar{\beta} + (\bar{\tau\beta})] + [\tau_i - \bar{\tau} + (\bar{\tau\beta})_{i.} - (\bar{\tau\beta})] +$$
$$[\beta_j - \bar{\beta} + (\bar{\tau\beta})_{.j} - (\bar{\tau\beta})] + [(\tau\beta)_{ij} - (\bar{\tau\beta})_{i.} - (\bar{\tau\beta})_{.j} + (\bar{\tau\beta})] + \varepsilon_{ijk}$$

or

$$y_{ijk} = \mu^* + \tau_i^* + \beta_j^* + (\tau\beta)_{ij}^* + \varepsilon_{ijk}$$

It can be shown that each of the new parameters $\mu^*, \tau_i^*, \beta_j^*,$ and $(\tau\beta)_{ij}^*$ is estimable. Therefore, it is reasonable to expect that the hypotheses of interest can be expressed simply in terms of these redefined parameters. It particular, it can be shown that there is no interaction if and only if $(\tau\beta)_{ij}^* = 0$. Now in the text, we presented the null hypothesis of no interaction as $H_0: (\tau\beta)_{ij} = 0$ for all *i* and *j*. This is not incorrect so long as it is understood that it is the model in terms of the redefined (or "starred") parameters that we are using. However, it is important to understand that in general interaction is *not* a parameter that refers only to the (*ij*)th cell, but it contains information from that cell, the *i*th row, the *j*th column, and the overall average response.

One final point is that as a consequence of defining the new "starred" parameters, we have included certain restrictions on them. In particular, we have

$$\tau_{.}^* = 0, \beta_{.}^* = 0, \ (\tau\beta)_{i.}^* = 0, \ (\tau\beta)_{.j}^* = 0 \text{ and } (\tau\beta)_{..}^* = 0$$

These are the "usual constraints" imposed on the normal equations. Furthermore, the tests on main effects become

$$H_0: \tau_1^* = \tau_2^* = \cdots = \tau_a^* = 0$$

and

$$H_0: \beta_1^* = \beta_2^* = \cdots = \beta_b^* = 0$$

This is the way that these hypotheses are stated in the textbook, but of course, without the "stars".

S5-4. Regression Model Formulation of the Two-factor Factorial

We noted in Chapter 3 that there was a close relationship between ANOVA and regression, and in the Supplemental Text Material for Chapter 3 we showed how the single-factor ANOVA model could be formulated as a regression model. We now show how the two-factor model can be formulated as a regression model and a standard multiple regression computer program employed to perform the usual ANOVA.

We will use the battery life experiment of Example 5-1 to illustrate the procedure. Recall that there are three material types of interest (factor *A*) and three temperatures (factor *B*), and the response variable of interest is battery life. The regression model formulation of an ANOVA model uses **indicator variables**. We will define the indicator variables for the design factors material types and temperature as follows:

Material type	X_1	X_2
1	0	0
2	1	0
3	0	1

Temperature	X_3	X_4
15	0	0
70	1	0
125	0	1

The regression model is

$$y_{ijk} = \beta_0 + \beta_1 x_{ijk1} + \beta_2 x_{ijk2} + \beta_3 x_{ijk3} + \beta_4 x_{ijk4} + \beta_5 x_{ijk1} x_{ijk3} + \beta_6 x_{ijk1} x_{ijk4} + \beta_7 x_{ijk2} x_{ijk3} + \beta_8 x_{ijk2} x_{ijk4} + \varepsilon_{ijk} \quad (1)$$

where i, j =1,2,3 and the number of replicates k = 1,2,3,4. In this model, the terms $\beta_1 x_{ijk1} + \beta_2 x_{ijk2}$ represent the main effect of factor *A* (material type), and the terms $\beta_3 x_{ijk3} + \beta_4 x_{ijk4}$ represent the main effect of temperature. Each of these two groups of terms contains two regression coefficients, giving two degrees of freedom. The terms $\beta_5 x_{ijk1} x_{ijk3} + \beta_6 x_{ijk1} x_{ijk4} + \beta_7 x_{ijk2} x_{ijk3} + \beta_8 x_{ijk2} x_{ijk4}$ in Equation (1) represent the *AB* interaction with four degrees of freedom. Notice that there are four regression coefficients in this term.

Table 1 shows the data from this experiment, originally presented in Table 5-1 of the text. In Table 1, we have shown the indicator variables for each of the 36 trials of this experiment. The notation in this table is $X_i = x_i$, i=1,2,3,4 for the main effects in the above regression model and $X_5 = x_1x_3$, $X_6 = x_1x_4$,, $X_7 = x_2x_3$, and $X_8 = x_2x_4$, for the interaction terms in the model.

Table 1. Data from Example 5-1 in Regression Model Form

Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
130	0	0	0	0	0	0	0	0
34	0	0	1	0	0	0	0	0
20	0	0	0	1	0	0	0	0
150	1	0	0	0	0	0	0	0
136	1	0	1	0	1	0	0	0
25	1	0	0	1	0	1	0	0
138	0	1	0	0	0	0	0	0
174	0	1	1	0	0	0	1	0
96	0	1	0	1	0	0	0	1
155	0	0	0	0	0	0	0	0
40	0	0	1	0	0	0	0	0
70	0	0	0	1	0	0	0	0
188	1	0	0	0	0	0	0	0
122	1	0	1	0	1	0	0	0
70	1	0	0	1	0	1	0	0
110	0	1	0	0	0	0	0	0
120	0	1	1	0	0	0	1	0
104	0	1	0	1	0	0	0	1
74	0	0	0	0	0	0	0	0
80	0	0	1	0	0	0	0	0
82	0	0	0	1	0	0	0	0
159	1	0	0	0	0	0	0	0
106	1	0	1	0	1	0	0	0
58	1	0	0	1	0	1	0	0
168	0	1	0	0	0	0	0	0
150	0	1	1	0	0	0	1	0
82	0	1	0	1	0	0	0	1
180	0	0	0	0	0	0	0	0
75	0	0	1	0	0	0	0	0
58	0	0	0	1	0	0	0	0
126	1	0	0	0	0	0	0	0
115	1	0	1	0	1	0	0	0
45	1	0	0	1	0	1	0	0
160	0	1	0	0	0	0	0	0
139	0	1	1	0	0	0	1	0
60	0	1	0	1	0	0	0	1

This table was used as input to the Minitab regression procedure, which produced the following results for fitting Equation (1):

Regression Analysis

```
The regression equation is
y = 135 + 21.0 x1 + 9.2 x2 - 77.5 x3 - 77.2 x4 + 41.5 x5 - 29.0 x6
 +79.2 x7 + 18.7 x8
```

minitab Output (Continued)

Predictor	Coef	StDev	T	P
Constant	134.75	12.99	10.37	0.000
x1	21.00	18.37	1.14	0.263
x2	9.25	18.37	0.50	0.619
x3	-77.50	18.37	-4.22	0.000
x4	-77.25	18.37	-4.20	0.000
x5	41.50	25.98	1.60	0.122
x6	-29.00	25.98	-1.12	0.274
x7	79.25	25.98	3.05	0.005
x8	18.75	25.98	0.72	0.477

S = 25.98 R-Sq = 76.5% R-Sq(adj) = 69.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	8	59416.2	7427.0	11.00	0.000
Residual Error	27	18230.7	675.2		
Total	35	77647.0			

Source	DF	Seq SS
x1	1	141.7
x2	1	10542.0
x3	1	76.1
x4	1	39042.7
x5	1	788.7
x6	1	1963.5
x7	1	6510.0
x8	1	351.6

First examine the Analysis of Variance information in the above display. Notice that the regression sum of squares with 8 degrees of freedom is equal to the sum of the sums of squares for the main effects material types and temperature and the interaction sum of squares from Table 5-5 in the textbook. Furthermore, the number of degrees of freedom for regression (8) is the sum of the degrees of freedom for main effects and interaction (2 +2 + 4) from Table 5-5. The *F*-test in the above ANOVA display can be thought of as testing the null hypothesis that *all* of the model coefficients are zero; that is, there are no significant main effects or interaction effects, versus the alternative that there is at least one nonzero model parameter. Clearly this hypothesis is rejected. Some of the treatments produce significant effects.

Now consider the "sequential sums of squares" at the bottom of the above display. Recall that X_1 and X_2 represent the main effect of material types. The sequential sums of squares are computed based on an "effects added in order" approach, where the "in order" refers to the order in which the variables are listed in the model. Now

$$SS_{MaterialTypes} = SS(X_1) + SS(X_2|X_1) = 141.7 + 10542.0 = 10683.7$$

which is the sum of squares for material types in table 5-5. The notation $SS(X_2|X_1)$ indicates that this is a "sequential" sum of squares; that is, it is the sum of squares for variable X_2 given that variable X_1 is already in the regression model.

Similarly,

$$SS_{Temperature} = SS(X_3|X_1, X_2) + SS(X_4|X_1, X_2, X_3) = 76.1 + 39042.7 = 39118.8$$

which closely agrees with the sum of squares for temperature from Table 5-5. Finally, note that the interaction sum of squares from Table 5-5 is

$$\begin{aligned} SS_{Interaction} &= SS(X_5|X_1, X_1, X_3, X_4) + SS(X_6|X_1, X_1, X_3, X_4, X_5) \\ &\quad + SS(X_7|X_1, X_2, X_3, X_4, X_5, X_6) + SS(X_8|X_1, X_2, X_3, X_4, X_5, X_6, X_7) \\ &= 788.7 + 1963.5 + 6510.0 + 351.6 = 9613.8 \end{aligned}$$

When the design is **balanced**, that is, we have an equal number of observations in each cell, we can show that this model regression approach using the sequential sums of squares produces results that are exactly identical to the "usual" ANOVA. Furthermore, because of the balanced nature of the design, the order of the variables *A* and *B* does not matter.

The "effects added in order" partitioning of the overall model sum of squares is sometimes called a **Type 1 analysis**. This terminology is prevalent in the SAS statistics package, but other authors and software systems also use it. An alternative partitioning is to consider each effect as if it were added **last** to a model that contains all the others. This "effects added last" approach is usually called a **Type 3 analysis**.

There is another way to use the regression model formulation of the two-factor factorial to generate the standard *F*-tests for main effects and interaction. Consider fitting the model in Equation (1), and let the regression sum of squares in the Minitab output above for this model be the model sum of squares for the **full model**. Thus,

$$SS_{Model}(\text{FM}) = 59416.2$$

with 8 degrees of freedom. Suppose we want to test the hypothesis that there is no interaction. In terms of model (1), the no-interaction hypothesis is

$$\begin{aligned} &H_0: \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0 \\ &H_0: \text{ at least one } \beta_j \neq 0, j = 5,6,7,8 \end{aligned} \tag{2}$$

When the null hypothesis is true, a **reduced model** is

$$y_{ijk} = \beta_0 + \beta_1 x_{ijk1} + \beta_2 x_{ijk2} + \beta_3 x_{ijk3} + \beta_4 x_{ijk4} + \varepsilon_{ijk} \tag{3}$$

Fitting Equation (2) using Minitab produces the following:

```
The regression equation is
y = 122 + 25.2 x1 + 41.9 x2 - 37.3 x3 - 80.7 x4
```

```
Predictor        Coef       StDev          T        P
Constant       122.47       11.17      10.97    0.000
x1              25.17       12.24       2.06    0.048
x2              41.92       12.24       3.43    0.002
x3             -37.25       12.24      -3.04    0.005
x4             -80.67       12.24      -6.59    0.000

S = 29.97       R-Sq = 64.1%     R-Sq(adj) = 59.5%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         4       49802       12451     13.86    0.000
Residual Error    31       27845         898
Total             35       77647
```

The model sum of squares for this reduced model is

$$SS_{Model}(\text{RM}) = 49802.0$$

with 4 degrees of freedom. The test of the no-interaction hypotheses (2) is conducted using the "extra" sum of squares

$$\begin{aligned} SS_{Model}(\text{Interaction}) &= SS_{Model}(\text{FM}) - SS_{Model}(\text{RM}) \\ &= 59{,}416.2 - 49{,}812.0 \\ &= 9{,}604.2 \end{aligned}$$

with 8 – 4 = 4 degrees of freedom. This quantity is, apart from round-off errors in the way the results are reported in Minitab, the interaction sum of squares for the original analysis of variance in Table 5-5 of the text. This is a measure of fitting interaction after fitting the main effects.

Now consider testing for no main effect of material type. In terms of equation (1), the hypotheses are

$$\begin{aligned} &H_0: \beta_1 = \beta_2 = 0 \\ &H_0: \text{ at least one } \beta_j \neq 0,\ j = 1,2 \end{aligned} \tag{4}$$

Because we are using a balanced design, it turns out that to test this hypothesis all we have to do is to fit the model

$$y_{ijk} = \beta_0 + \beta_1 x_{ijk1} + \beta_2 x_{ijk2} + \varepsilon_{ijk} \tag{5}$$

Fitting this model in Minitab produces

Regression Analysis

```
The regression equation is
y = 83.2 + 25.2 x1 + 41.9 x2

Predictor        Coef       StDev          T        P
Constant        83.17       13.00       6.40    0.000
x1              25.17       18.39       1.37    0.180
x2              41.92       18.39       2.28    0.029
```

```
Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2       10684        5342      2.63    0.087
Residual Error    33       66963        2029
Total             35       77647
```

Notice that the regression sum of squares for this model [Equation (5)] is essentially identical to the sum of squares for material types in table 5-5 of the text. Similarly, testing that there is no temperature effect is equivalent to testing

$$\begin{aligned} &H_0: \beta_3 = \beta_4 = 0 \\ &H_0: \text{ at least one } \beta_j \neq 0,\ j = 3,4 \end{aligned} \tag{6}$$

To test the hypotheses in (6), all we have to do is fit the model

$$y_{ijk} = \beta_0 + \beta_3 x_{ijk3} + \beta_4 x_{ijk4} + \varepsilon_{ijk} \tag{7}$$

The Minitab regression output is

Regression Analysis

```
The regression equation is
y = 145 - 37.3 x3 - 80.7 x4

Predictor        Coef       StDev          T        P
Constant      144.833       9.864      14.68    0.000
x3             -37.25       13.95      -2.67    0.012
x4             -80.67       13.95      -5.78    0.000

S = 34.17       R-Sq = 50.4%     R-Sq(adj) = 47.4%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2       39119       19559     16.75    0.000
Residual Error    33       38528        1168
Total             35       77647
```

Notice that the regression sum of squares for this model, Equation (7), is essentially equal to the temperature main effect sum of squares from Table 5-5.

S5-5. Model Hierarchy

In Example 5-4 we used the data from the battery life experiment (Example 5-1) to demonstrate fitting response curves when one of the factors in a two-factor factorial experiment was quantitative and the other was qualitative. In this case the factors are temperature (*A*) and material type (*B*). Using the Design-Expert software package, we fit a model that the main effect of material type, the linear and quadratic effects of temperature, the material type by linear effect of temperature interaction, and the material type by quadratic effect of temperature interaction. Refer to Table 5-15 in the textbook. From examining this table, we observed that the quadratic effect of temperature and the

material type by linear effect of temperature interaction were not significant; that is, they had fairly large *P*-values. We left these non-significant terms in the model to preserve hierarchy.

The hierarchy principal states that if a model contains a higher-order term, then it should also contain all the terms of lower-order that comprise it. So, if a second-order term, such as an interaction, is in the model then all main effects involved in that interaction as well as all lower-order interactions involving those factors should also be included in the model.

There are times that hierarchy makes sense. Generally, if the model is going to be used for explanatory purposes then a hierarchical model is quite logical. On the other hand, there may be situations where the non-hierarchical model is much more logical. To illustrate, consider another analysis of Example 5-4 in Table 2, which was obtained from Design-Expert. We have selected a non-hierarchical model in which the quadratic effect of temperature was not included (it was in all likelihood the weakest effect), but both two-degree-of-freedom components of the temperature-material type interaction are in the model. Notice from Table 2 that the residual mean square is smaller for the non-hierarchical model (653.81 versus 675.21 from Table 5-15). This is important, because the residual mean square can be thought of as the variance of the unexplained residual variability, not accounted for by the model. That is, the non-hierarchical model is actually a *better fit* to the experimental data.

Notice also that the standard errors of the model parameters are smaller for the non-hierarchical model. This is an indication that he parameters are estimated with better precision by leaving out the nonsignificant terms, even though it results in a model that does not obey the hierarchy principal. Furthermore, note that the 95 percent confidence intervals for the model parameters in the hierarchical model are always longer than their corresponding confidence intervals in the non-hierarchical model. The non-hierarchical model, in this example, does indeed provide better estimates of the factor effects that obtained from the hierarchical model

Table 2. Design-Expert Output for Non-hierarchical Model, Example 5-4.

ANOVA for Response Surface Reduced Cubic Model
Analysis of variance table [Partial sum of squares]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**
Model	59340.17	7	8477.17	12.97	< 0.0001
A	*10683.72*	*2*	*5341.86*	*8.17*	*0.0016*
B	*39042.67*	*1*	*39042.67*	*59.72*	*< 0.0001*
AB	*2315.08*	*2*	*1157.54*	*1.77*	*0.1888*
AB^2	*7298.69*	*2*	*3649.35*	*5.58*	*0.0091*
Residual	18306.81	28	653.81		
Lack of Fit	*76.06*	*1*	*76.06*	*0.11*	*0.7398*
Pure Error	*18230.75*	*27*	*675.21*		
Cor Total	77646.97	35			

Std. Dev.	25.57	R-Squared	0.7642
Mean	105.53	Adj R-Squared	0.7053
C.V.	24.23	Pred R-Squared	0.6042
PRESS	30729.09	Adeq Precision	8.815

Term	Coefficient Estimate	DF	Standard Error	95% CI Low	95% CI High
Intercept	105.53	1	4.26	96.80	114.26
A[1]	-50.33	1	10.44	-71.72	-28.95
A[2]	12.17	1	10.44	-9.22	33.55
B-Temp	-40.33	1	5.22	-51.02	-29.64
A[1]B	1.71	1	7.38	-13.41	16.83
A[2]B	-12.79	1	7.38	-27.91	2.33
$A[1]B^2$	41.96	1	12.78	15.77	68.15
$A[2]B^2$	-14.04	1	12.78	-40.23	12.15

Supplemental Reference

Myers, R. H. and Milton, J. S. (1991), *A First Course in the Theory of the Linear Model*, PWS-Kent, Boston, MA.

Chapter 6. Supplemental Text Material

S6-1. Factor Effect Estimates are Least Squares Estimates

We have given heuristic or intuitive explanations of how the estimates of the factor effects are obtained in the textbook. Also, it has been pointed out that in the regression model representation of the 2^k factorial, the regression coefficients are exactly one-half the effect estimates. It is straightforward to show that the model coefficients (and hence the effect estimates) are least squares estimates.

Consider a 2^2 factorial. The regression model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2} + \varepsilon_i$$

The data for the 2^2 experiment is shown in the following table:

Run, *i*	X_{i1}	X_{i2}	$X_{i1}X_{i2}$	Response total
1	-1	-1	1	(1)
2	1	-1	-1	*a*
3	-1	1	-1	*b*
4	1	1	1	*ab*

The least squares estimates of the model parameters β are chosen to minimize the sum of the squares of the model errors:

$$L = \sum_{i=1}^{4} \left(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_{12} x_{i1} x_{i2} \right)^2$$

It is straightforward to show that the least squares normal equations are

$$4\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{4} x_{i1} + \hat{\beta}_2 \sum_{i=1}^{4} x_{i2} + \hat{\beta}_{12} \sum_{i=1}^{4} x_{i1} x_{i2} = (1) + a + b + ab$$

$$\hat{\beta}_0 \sum_{i=1}^{4} x_i + \hat{\beta}_1 \sum_{i=1}^{4} x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^{4} x_{i1} x_{i2} + \hat{\beta}_{12} \sum_{i=1}^{4} x_{i1}^2 x_{i2} = -(1) + a - b + ab$$

$$\hat{\beta}_0 \sum_{i=1}^{4} x_{i2} + \hat{\beta}_1 \sum_{i=1}^{4} x_{i1} x_{i2} + \hat{\beta}_2 \sum_{i=1}^{4} x_{i2}^2 + \hat{\beta}_{12} \sum_{i=1}^{4} x_{i1} x_{i2}^2 = -(1) - a + b + ab$$

$$\hat{\beta}_0 \sum_{i=1}^{4} x_{i1} x_{i2} + \hat{\beta}_1 \sum_{i=1}^{4} x_{i1}^2 x_{i2} + \hat{\beta}_2 \sum_{i=1}^{4} x_{i1} x_{i2}^2 + \hat{\beta}_{12} \sum_{i=1}^{4} x_{i1}^2 x_{i2}^2 = (1) - a - b + ab$$

Now since $\sum_{i=1}^{4} x_{i1} = \sum_{i=1}^{4} x_{i2} = \sum_{i=1}^{4} x_{i1} x_{i2} = \sum_{i=1}^{4} x_{i1}^2 x_{i2} = \sum_{i=1}^{4} x_{i1} x_{i2}^2 = 0$ because the design is orthogonal, the normal equations reduce to a very simple form:

$$4\hat{\beta}_0 = (1) + a + b + ab$$
$$4\hat{\beta}_1 = -(1) + a - b + ab$$
$$4\hat{\beta}_2 = -(1) - a + b + ab$$
$$4\hat{\beta}_{12} = (1) - a - b + ab$$

The solution is

$$\hat{\beta}_0 = \frac{[(1) + a + b + ab]}{4}$$
$$\hat{\beta}_1 = \frac{[-(1) + a - b + ab]}{4}$$
$$\hat{\beta}_2 = \frac{[-(1) - a + b + ab]}{4}$$
$$\hat{\beta}_{12} = \frac{[(1) - a - b + ab]}{4}$$

These regression model coefficients are exactly one-half the factor effect estimates. Therefore, the effect estimates are least squares estimates. We will show this in a more general manner in Chapter 10.

S6-2. Yates's Method for Calculating Effect Estimates

While we typically use a computer program for the statistical analysis of a 2^k design, there is a very simple technique devised by Yates (1937) for estimating the effects and determining the sums of squares in a 2^k factorial design. The procedure is occasionally useful for manual calculations, and is best learned through the study of a numerical example.

Consider the data for the 2^3 design in Example 6-1. These data have been entered in Table 1 below. The treatment combinations are always written down in standard order, and the column labeled "Response" contains the corresponding observation (or total of all observations) at that treatment combination. The first half of column (1) is obtained by adding the responses in adjacent pairs. The second half of column (1) is obtained by changing the sign of the first entry in each of the pairs in the Response column and adding the adjacent pairs. For example, in column (1) we obtain for the fifth entry 5 = -(-4) + 1, for the sixth entry 6 = -(-1) + 5, and so on.

Column (2) is obtained from column (1) just as column (1) is obtained from the Response column. Column (3) is obtained from column (2) similarly. In general, for a 2^k design we would construct k columns of this type. Column (3) [in general, column (k)] is the contrast for the effect designated at the beginning of the row. To obtain the estimate of the effect, we divide the entries in column (3) by $n2^{k-1}$ (in our example, $n2^{k-1} = 8$). Finally, the sums of squares for the effects are obtained by squaring the entries in column (3) and dividing by $n2^k$ (in our example, $n2^k = (2)2^3 = 16$).

Table 1. Yates's Algorithm for the Data in Example 6-1

Treatment Combination	Response	(1)	(2)	(3)	Effect	Estimate of Effect $(3)\div n2^{k-1}$	Sum of Squares $(3)^2\div n2^{k-1}$
(1)	-4	-3	1	16	*I*	---	---
a	1	4	15	24	*A*	3.00	36.00
b	-1	2	11	18	*B*	2.25	20.25
ab	5	13	13	6	*A B*	0.75	2.25
c	-1	*5*	7	14	*C*	1.75	12.25
ac	3	6	11	2	*A C*	0.25	0.25
bc	2	4	1	4	*B C*	0.50	1.00
abc	11	9	*5*	4	*ABC*	0.50	1.00

The estimates of the effects and sums of squares obtained by Yates' algorithm for the data in Example 6-1 are in agreement with the results found there by the usual methods. Note that the entry in column (3) [in general, column (*k*)] for the row corresponding to (1) is always equal to the grand total of the observations.

In spite of its apparent simplicity, it is notoriously easy to make numerical errors in Yates's algorithm, and we should be extremely careful in executing the procedure. As a partial check on the computations, we may use the fact that the sum of the squares of the elements in the *j*th column is 2^j times the sum of the squares of the elements in the response column. Note, however, that this check is subject to errors in sign in column *j*. See Davies (1956), Good (1955, 1958), Kempthorne (1952), and Rayner (1967) for other error-checking techniques.

S6-3. A Note on the Variance of a Contrast

In analyzing 2^k factorial designs, we frequently construct a normal probability plot of the factor effect estimates and visually select a tentative model by identifying the effects that appear large. These effect estimates are typically relatively far from the straight line passing through the remaining plotted effects.

This method works nicely when (1) there are not many significant effects, and (2) when all effect estimates have the same variance. It turns out that all contrasts computed from a 2^k design (and hence all effect estimates) have the same variance even if the individual observations have different variances. This statement can be easily demonstrated.

Suppose that we have conducted a 2^k design and have responses $y_1, y_2, \cdots, y_{2^k}$ and let the variance of each observation be $\sigma_1^2, \sigma_2^2, \cdots, \sigma_{2^k}^2$ respectively. Now each effect estimate is a linear combination of the observations, say

$$Effect = \frac{\sum_{i=1}^{2^k} c_i y_i}{2^k}$$

where the contrast constants c_i are all either –1 or +1. Therefore, the variance of an effect estimate is

$$V(Effect) = \frac{1}{(2^k)^2}\sum_{i=1}^{2^k} c_i^2 V(y_i)$$
$$= \frac{1}{(2^k)^2}\sum_{i=1}^{2^k} c_i^2 \sigma_i^2$$
$$= \frac{1}{(2^k)^2}\sum_{i=1}^{2^k} \sigma_i^2$$

because $c_i^2 = 1$. Therefore, all contrasts have the same variance. If each observation y_i in the above equations is the total of *n* replicates at each design point, the result still holds.

S6-4. The Variance of the Predicted Response

Suppose that we have conducted an experiment using a 2^k factorial design. We have fit a regression model to the resulting data and are going to use the model to predict the response at locations of interest in side the design space $-1 \le x_i \le +1,\ i = 1,2,\cdots,k$. What is the variance of the predicted response at the point of interest, say $\mathbf{x}' = [x_1, x_2, \cdots, x_k]$?

Problem 6-32 asks the reader to answer this question, and while the answer is given in the Instructors Resource CD, we also give the answer here because it is useful information. Assume that the design is balanced and every treatment combination is replicated *n* times. Since the design is orthogonal, it is easy to find the variance of the predicted response.

We consider the case where the experimenters have fit a "main effects only" model, say

$$\hat{y}(\mathbf{x}) \equiv \hat{y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i$$

Now recall that the variance of a model regression coefficient is $V(\hat{\beta}) = \frac{\sigma^2}{n2^k} = \frac{\sigma^2}{N}$,

where *N* is the total number of runs in the design. The variance of the predicted response is

$$V[\hat{y}(\mathbf{x})] = V\left(\hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i\right)$$
$$= V(\hat{\beta}_0) + \sum_{i=1}^{k} V(\hat{\beta}_i x_i)$$
$$= V(\hat{\beta}_0) + \sum_{i=1}^{k} x_i^2 V(\hat{\beta}_i)$$
$$= \frac{\sigma^2}{N} + \frac{\sigma^2}{N}\sum_{i=1}^{k} x_i^2$$
$$= \frac{\sigma^2}{N}\left(1 + \sum_{i=1}^{k} x_i^2\right)$$

In the above development we have used the fact that the design is orthogonal, so there are no nonzero covariance terms when the variance operator is applied

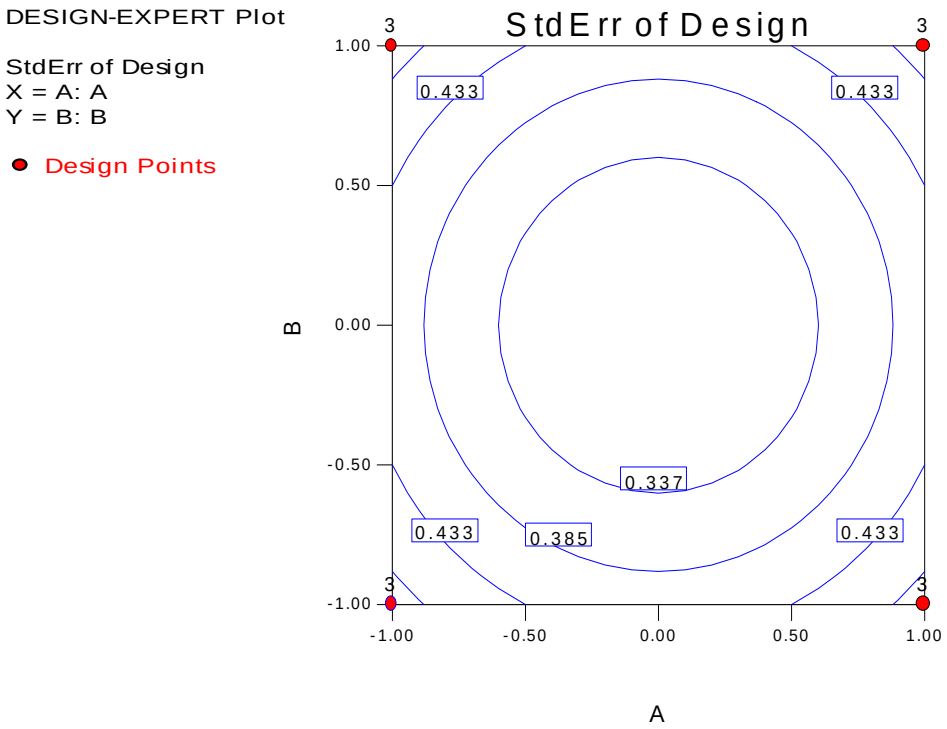
The Design-Expert software program plots contours of the standard deviation of the predicted response; that is the square root of the above expression. If the design has already been conducted and analyzed, the program replaces σ^2 with the error mean square, so that the plotted quantity becomes

$$\sqrt{\hat{V}[\hat{y}(\mathbf{x})} = \sqrt{\frac{MS_E}{N}\left(1+\sum_{i=1}^{k} x_i^2\right)}$$

If the design has been constructed but the experiment has not been performed, then the software plots (on the design evaluation menu) the quantity

$$\sqrt{\frac{V[\hat{y}(\mathbf{x})}{\sigma^2}} = \sqrt{\frac{1}{N}\left(1+\sum_{i=1}^{k} x_i^2\right)}$$

which can be thought of as a **standardized** standard deviation of prediction. To illustrate, consider a 2^2 with $n = 3$ replicates, the first example in Section 6-2. The plot of the standardized standard deviation of the predicted response is shown below.



The contours of constant standardized standard deviation of predicted response should be exactly circular, and they should be a maximum within the design region at the point $x_1 = \pm 1$ and $x_2 = \pm 1$. The maximum value is

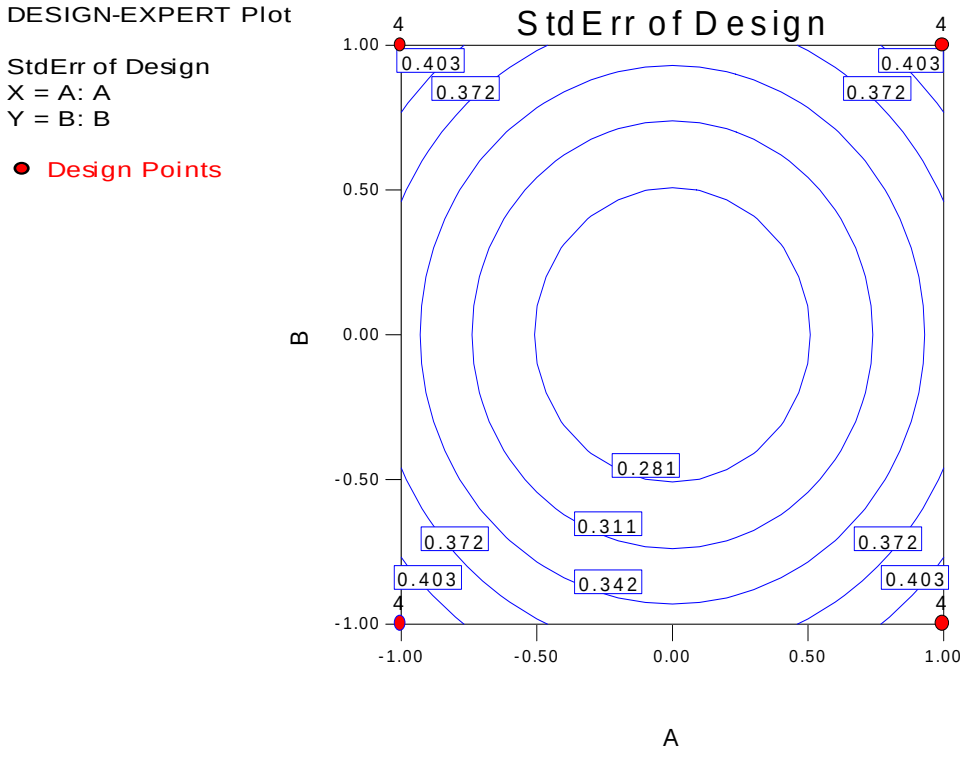
$$\sqrt{\frac{V[\hat{y}(\mathbf{x}=1)}{\sigma^2}} = \sqrt{\frac{1}{12}\left(1+(1)^2+(1)^2\right)}$$
$$= \sqrt{\frac{3}{12}}$$
$$= 0.5$$

This is also shown on the graph at the corners of the square.

Plots of the standardized standard deviation of the predicted response can be useful in comparing designs. For example, suppose the experimenter in the above situation is considering adding a fourth replicate to the design. The maximum standardized prediction standard deviation in the region now becomes

$$\sqrt{\frac{V[\hat{y}(\mathbf{x}=1)}{\sigma^2}} = \sqrt{\frac{1}{12}\left(1+(1)^2+(1)^2\right)}$$
$$= \sqrt{\frac{3}{16}}$$
$$= 0.433$$

The plot of the standardized prediction standard deviation is shown below.



Notice that adding another replicate has reduced the maximum prediction variance from $(0.5)^2 = 0.25$ to $(0.433)^2 = 0.1875$. Comparing the two plots shown above reveals that the standardized prediction standard deviation is uniformly lower throughout the design region when an additional replicate is run.

Sometimes we like to compare designs in terms of **scaled prediction variance**, defined as

$$\frac{NV[\hat{y}(\mathbf{x})]}{\sigma^2}$$

This allows us to evaluate designs that have different numbers of runs. Since adding replicates (or runs) to a design will generally always make the prediction variance get smaller, the scaled prediction variance allows us to examine the prediction variance on a *per-observation basis*. Note that for a 2^k factorial and the "main effects only" model we have been considering, the scaled prediction variance is

$$\begin{aligned}\frac{NV[\hat{y}(\mathbf{x})]}{\sigma^2} &= \left(1+\sum_{i=1}^{k} x_i^2\right) \\ &= (1+\rho^2)\end{aligned}$$

where ρ^2 is the distance of the design point where prediction is required from the center of the design space ($\mathbf{x} = \mathbf{0}$). Notice that the 2^k design achieves this scaled prediction variance *regardless* of the number of replicates. The maximum value that the scaled prediction variance can have over the design region is

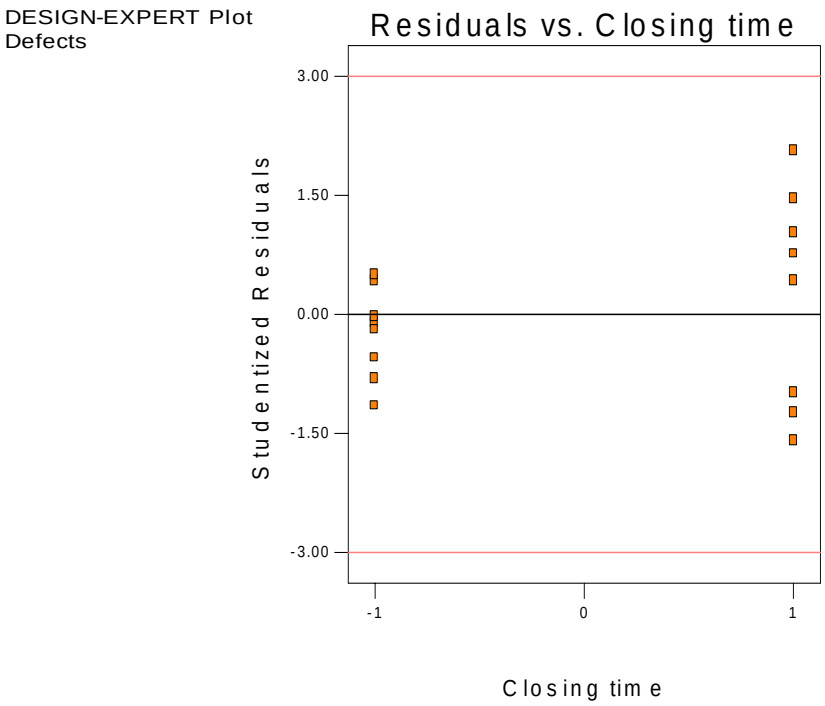
$$Max\frac{NV[\hat{y}(\mathbf{x})]}{\sigma^2} = (1+k)$$

It can be shown that no other design over this region can achieve a smaller maximum scaled prediction variance, so the 2^k design is in some sense an **optimal design**. We will discuss optimal designs more in Chapter 11.

S6-5. Using Residuals to Identify Dispersion Effects

We illustrated in Example 6-4 (Section 6-5 on unreplicated designs) that plotting the residuals from the regression model versus each of the design factors was a useful way to check for the possibility of dispersion effects. These are factors that influence the variability of the response, but which have little effect on the mean. A method for computing a measure of the dispersion effect for each design factor and interaction that can be evaluated on a normal probability plot was also given. However, we noted that these residual analyses are fairly sensitive to correct specification of the location model. That is, if we leave important factors out of the regression model that describes the mean response, then the residual plots may be unreliable.

To illustrate, reconsider Example 6-4, and suppose that we leave out one of the important factors, *C* = Resin flow. If we use this incorrect model, then the plots of the residuals versus the design factors look rather different than they did with the original, correct model. In particular, the plot of residuals versus factor *D* = Closing time is shown below.



This plot indicates that factor *D* has a potential dispersion effect. The normal probability plot of the dispersion statistic F_i^* in Figure 6-28 clearly reveals that factor *B* is the only factor that has an effect on dispersion. Therefore, if you are going to use model residuals to search for dispersion effects, it is really important to select the *right* model for the location effects.

S6-6. Center Points versus Replication of Factorial Points

In some design problems an experimenter may have a choice of replicating the corner or "cube" points in a 2^k factorial, or placing replicate runs at the design center. For example, suppose our choice is between a 2^2 with $n = 2$ replicates at each corner of the square, or a single replicate of the 2^2 with $n_c = 4$ center points.

We can compare these designs in terms of prediction variance. Suppose that we plan to fit the first-order or "main effects only" model

$$\hat{y}(\mathbf{x}) \equiv \hat{y} = \hat{\beta}_0 + \sum_{i=1}^{2} \hat{\beta}_i x_i$$

If we use the replicated design the scaled prediction variance is (see Section 6-4 above):

$$\frac{NV[\hat{y}(\mathbf{x})]}{\sigma^2} = \left(1 + \sum_{i=1}^{2} x_i^2\right)$$
$$= (1 + \rho^2)$$

Now consider the prediction variance when the design with center points is used. We have

$$\begin{aligned} V[\hat{y}(\mathbf{x})] &= V\left(\hat{\beta}_0 + \sum_{i=1}^{2}\hat{\beta}_i x_i\right) \\ &= V(\hat{\beta}_0) + \sum_{i=1}^{2} V(\hat{\beta}_i x_i) \\ &= V(\hat{\beta}_0) + \sum_{i=1}^{k} x_i^2 V(\hat{\beta}_i) \\ &= \frac{\sigma^2}{8} + \frac{\sigma^2}{4}\sum_{i=1}^{2} x_i^2 \\ &= \frac{\sigma^2}{8}\left(1 + 2\sum_{i=1}^{k} x_i^2\right) \\ &= \frac{\sigma^2}{8}(1+2\rho^2) \end{aligned}$$

Therefore, the scaled prediction variance for the design with center points is

$$\frac{NV[\hat{y}(\mathbf{x})]}{\sigma^2} = (1+2\rho^2)$$

Clearly, replicating the corners in this example outperforms the strategy of replicating center points, at least in terms of scaled prediction variance. At the corners of the square, the scaled prediction variance for the replicated factorial is

$$\begin{aligned} \frac{NV[\hat{y}(\mathbf{x})]}{\sigma^2} &= (1+\rho^2) \\ &= (1+2) \\ &= 3 \end{aligned}$$

while for the factorial design with center points it is

$$\begin{aligned} \frac{NV[\hat{y}(\mathbf{x})]}{\sigma^2} &= (1+2\rho^2) \\ &= (1+2(2)) \\ &= 5 \end{aligned}$$

However, prediction variance might not tell the complete story. If we only replicate the corners of the square, we have no way to judge the lack of fit of the model. If the design has center points, we can check for the presence of pure quadratic (second-order) terms, so the design with center points is likely to be preferred if the experimenter is at all uncertain about the order of the model he or she should be using.

S6-7. Testing for "Pure Quadratic" Curvature using a *t*-Test

In Section 6-6 of the textbook we discuss the addition of center points to a 2^k factorial design. This is a very useful idea as it allows an estimate of "pure error" to be obtained even thought the factorial design points are not replicated and it permits the experimenter to obtain an assessment of model adequacy with respect to certain second-order terms. Specifically, we present an *F*-test for the hypotheses

$$H_0 : \beta_{11} + \beta_{22} + \cdots \beta_{kk} = 0$$
$$H_1 : \beta_{11} + \beta_{22} + \cdots \beta_{kk} \neq 0$$

An equivalent *t*-statistic can also be employed to test these hypotheses. Some computer software programs report the *t*-test instead of (or in addition to) the *F*-test. It is not difficult to develop the *t*-test and to show that it is equivalent to the *F*-test.

Suppose that the appropriate model for the response is a complete quadratic polynomial and that the experimenter has conducted an unreplicated full 2^k factorial design with n_F design points plus n_C center points. Let $\bar{y}_F$ and $\bar{y}_C$ represent the averages of the responses at the factorial and center points, respectively. Also let $\hat{\sigma}^2$ be the estimate of the variance obtained using the center points. It is easy to show that

$$\begin{aligned} E(\bar{y}_F) &= \tfrac{1}{n_F}(n_F\beta_0 + n_F\beta_{11} + n_F\beta_{22} + \cdots + n_F\beta_{kk}) \\ &= \beta_{11} + \beta_{22} + \cdots + \beta_{kk} \end{aligned}$$

and

$$\begin{aligned} E(\bar{y}_C) &= \tfrac{1}{n_C}(n_C\beta_0) \\ &= \beta_0 \end{aligned}$$

Therefore,

$$E(\bar{y}_F - \bar{y}_C) = \beta_{11} + \beta_{22} + \cdots + \beta_{kk}$$

and so we see that the difference in averages $\bar{y}_F - \bar{y}_C$ is an unbiased estimator of the **sum** of the pure quadratic model parameters. Now the variance of $\bar{y}_F - \bar{y}_C$ is

$$V(\bar{y}_F - \bar{y}_C) = \sigma^2\left(\frac{1}{n_F} + \frac{1}{n_C}\right)$$

Consequently, a test of the above hypotheses can be conducted using the statistic

$$t_0 = \frac{\bar{y}_F - \bar{y}_C}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_F} + \frac{1}{n_C}\right)}}$$

which under the null hypothesis follows a *t* distribution with $n_C - 1$ degrees of freedom. We would reject the null hypothesis (that is, no pure quadratic curvature) if $|t_0| > t_{\alpha/2, n_C - 1}$.

This *t*-test is equivalent to the *F*-test given in the book. To see this, square the *t*-statistic above:

$$t_0^2 = \frac{(\bar{y}_F - \bar{y}_C)^2}{\hat{\sigma}^2\left(\frac{1}{n_F} + \frac{1}{n_C}\right)}$$
$$= \frac{n_F n_C (\bar{y}_F - \bar{y}_C)^2}{(n_F + n_C)\hat{\sigma}^2}$$

This ratio is computationally identical to the *F*-test presented in the textbook. Furthermore, we know that the square of a *t* random variable with (say) *v* degrees of freedom is an *F* random variable with 1 numerator and *v* denominator degrees of freedom, so the *t*-test for "pure quadratic" effects is indeed equivalent to the *F*-test.

Supplemental References

Good, I. J. (1955). "The Interaction Algorithm and Practical Fourier Analysis". *Journal of the Royal Statistical Society, Series B,* Vol. 20, pp. 361-372.

Good, I. J. (1958). Addendum to "The Interaction Algorithm and Practical Fourier Analysis". *Journal of the Royal Statistical Society, Series B,* Vol. 22, pp. 372-375.

Rayner, A. A. (1967). "The Square Summing Check on the Main Effects and Interactions in a 2^n Experiment as Calculated by Yates' Algorithm". *Biometrics,* Vol. 23, pp. 571-573.

Chapter 7. Supplemental Text Material

S7-1. The Error Term in a Blocked Design

Just as in any randomized complete block design, when we run a replicated factorial experiment in blocks we are assuming that there is no interaction between treatments and blocks. In the RCBD with a single design factor (Chapter 4) the error term is actually the interaction between treatments and blocks. This is also the case in a factorial design. To illustrate, consider the ANOVA in Table 7-2 of the textbook. The design is a 2^2 factorial run in three complete blocks. Each block corresponds to a replicate of the experiment. There are six degrees of freedom for error. Two of those degrees of freedom are the interaction between blocks and factor *A*, two degrees of freedom are the interaction between blocks and factor *B*, and two degrees of freedom are the interaction between blocks and the *AB* interaction. In order for the error term here to truly represent random error, we must assume that blocks and the design factors do not interact.

S7-2. The Prediction Equation for a Blocked Design

Consider the prediction equation for the 2^4 factorial in two blocks with *ABCD* confounded from in Example 7-2. Since blocking does not impact the effect estimates from this experiment, the equation would be exactly the same as the one obtained from the unblocked design, Example 6-2. This prediction equation is

$$\hat{y} = 70.06 + 10.8125x_1 + 4.9375x_3 + 7.3125x_4 - 9.0625x_1x_3 + 8.3125x_1x_4$$

This equation would be used to predict *future* observations where we had no knowledge of the block effect. However, in the experiment just completed we know that there is a strong block effect, in fact the block effect was computed as

$$block\ effect = \bar{y}_{block\,1} - \bar{y}_{block\,2} = -18.625$$

This means that the difference in average response between the two blocks is –18.625. We should compensate for this in the prediction equation if we want to obtain the correct fitted values for block 1 and block 2. Defining a separate block effect for each block does this, where $block_1\ effect = -9.3125$ and $block_2\ effect = 9.3125$. These block effects would be added to the intercept in the prediction equation for each block. Thus the prediction equations are

$$\begin{aligned}\hat{y}_{block_1} &= 70.06 + block_1\ effect + 10.8125x_1 + 4.9375x_3 + 7.3125x_4 - 9.0625x_1x_3 + 8.3125x_1x_4\\ &= 70.06 + (-9.3125) + 10.8125x_1 + 4.9375x_3 + 7.3125x_4 - 9.0625x_1x_3 + 8.3125x_1x_4\\ &= 60.7475 + 10.8125x_1 + 4.9375x_3 + 7.3125x_4 - 9.0625x_1x_3 + 8.3125x_1x_4\end{aligned}$$

and

$$\begin{aligned}\hat{y}_{block_2} &= 70.06 + block_2\ effect + 10.8125x_1 + 4.9375x_3 + 7.3125x_4 - 9.0625x_1x_3 + 8.3125x_1x_4\\ &= 70.06 + (9.3125) + 10.8125x_1 + 4.9375x_3 + 7.3125x_4 - 9.0625x_1x_3 + 8.3125x_1x_4\\ &= 79.3725 + 10.8125x_1 + 4.9375x_3 + 7.3125x_4 - 9.0625x_1x_3 + 8.3125x_1x_4\end{aligned}$$

S7-3. Run Order is Important

Blocking is really all about experimental run order. Specifically, we run an experiment in blocks to provide protection against the effects of a known and controllable nuisance factor(s). However, in many experimental situations, it is a good idea to conduct the experiment in blocks, even though there is no obvious nuisance factor present. This is particularly important when in takes several time periods (days, shifts, weeks, etc.) to run the experiment.

To illustrate, suppose that we are conducting a single replicate of a 2^4 factorial design. The experiment is shown in run order is shown in Table 2. Now suppose that misfortune strikes the experimenter, and after the first eight trials have been performed it becomes impossible to complete the experiment. Is there any useful experimental design that can be formed from the first eight runs?

Table 2. A 2^4 Factorial Experiment

Std Order	Run Order	Block	Factor A	Factor B	Factor C	Factor D
2	1	Block 1	1	-1	-1	-1
12	2	Block 1	1	1	-1	1
10	3	Block 1	1	-1	-1	1
15	4	Block 1	-1	1	1	1
14	5	Block 1	1	-1	1	1
4	6	Block 1	1	1	-1	-1
7	7	Block 1	-1	1	1	-1
3	8	Block 1	-1	1	-1	-1
5	9	Block 1	-1	-1	1	-1
8	10	Block 1	1	1	1	-1
11	11	Block 1	-1	1	-1	1
16	12	Block 1	1	1	1	1
1	13	Block 1	-1	-1	-1	-1
9	14	Block 1	-1	-1	-1	1
6	15	Block 1	1	-1	1	-1
13	16	Block 1	-1	-1	1	1

It turns out that in this case, the answer to that question is “no”. Now some analysis can of course be performed, but it would basically consist of fitting a regression model to the response data from the first 8 trials. Suppose that we fit a regression model containing an intercept term and the four main effects. When things have gone wrong it is usually a good idea to focus on simple objectives, making use of the data that are available. It turns out that in that model we would actually be obtaining estimates of

$$[\text{Intercept}] = \text{Intercept} - AB + CD - ABCD$$
$$[A] = A + AB - BC - ABC + ACD - BCD$$

[B] = B + AB - BC - ABC
[C] = C - ABC + ACD - BCD
[D] = D - ABD - ACD + BCD

Now suppose we feel comfortable in ignoring the three-factor and four-factor interaction effects. However, even with these assumptions, our intercept term is "clouded" or "confused" with two of the two-factor interactions, and the main effects of factors *A* and *B* are "confused" with the other two-factor interactions. In the next chapter, we will refer to the phenomena being observed here as **aliasing** of effects (its proper name). The supplemental notes for Chapter 8 present a general method for deriving the aliases for the factor effects. The Design-Expert software package can also be used to generate the aliases by employing the Design Evaluation feature. Notice that in our example, not completeing the experiment as originally planned has really disturbed the interpretation of the results.

Suppose that instead of completely randomizing all 16 runs, the experimenter had set this 2^4 design up in two blocks of 8 runs each, selecting in the usual way the *ABCD* interaction to be confounded with blocks. Now if only the first 8 runs can be performed, then it turns out that the estimates of the intercept and main factor effects from these 8 runs are

[Intercept] = Intercept
[A] = A + BCD
[B] = B + ACD
[C] = C + ABD
[D] = D + ABC

If we assume that the three-factor interactions are negligible, then we have reliable estimates of all four main effects from the first 8 runs. The reason for this is that each block of this design forms a **one-half fraction** of the 2^4 factorial, and this fraction allows estimation of the four main effects free of any two-factor interaction aliasing. This specific design (the one-half fraction of the 2^4) will be discussed in considerable detail in Chapter 8.

This illustration points out the importance of thinking carefully about run order, even when the experimenter is not obviously concerned about nuisance variables and blocking. Remember:

If something can go wrong when conducting an experiment, it probably will. A prudent experimenter designs his or her experiment with this in mind.

Generally, if a 2^k factorial design is constructed in two blocks, and one of the blocks is lost, ruined, or never run, the $2^k / 2 = 2^{k-1}$ runs that remain will always form a **one-half fraction** of the original design. It is almost always possible to learn *something* useful from such an experiment.

To take this general idea a bit further, suppose that we had originally set up the 16-run 2^4 factorial experiment in four blocks of four runs each. The design that we would obtain using the standard methods from this chapter in the text gives the experiment in Table 3. Now suppose that for some reason we can only run the first 8 trials of this experiment. It is easy to verify that the first 8 trials in Table 3 **do not** form one of the usual 8-run blocks produced by confounding the *ABCD* interaction with blocks. Therefore, the first 8 runs in Table 3 are not a "standard" one-half fraction of the 2^4.

A logical question is "what can we do with these 8 runs?" Suppose, as before, that the experimenter elects to concentrate on estimating the main effects. If we use only the first eight runs from Table 3 and concentrate on estimating only the four main effects, it turns out what we *really* are estimating is

[Intercept] = Intercept - ACD
[A] = A - CD
[B] = B - ABCD
[C] = C - AD
[D] = D - AC

Once again, even assuming that all interactions beyond order two are negligible, our main effect estimates are aliased with two-factor interactions.

Table 3. A 2^4 Factorial Experiment in Four Blocks

Std Order	Run Order	Block	Factor A	Factor B	Factor C	Factor D
10	1	Block 1	1	-1	-1	1
15	2	Block 1	-1	1	1	1
3	3	Block 1	-1	1	-1	-1
6	4	Block 1	1	-1	1	-1
12	5	Block 2	1	1	-1	1
8	6	Block 2	1	1	1	-1
13	7	Block 2	-1	-1	1	1
1	8	Block 2	-1	-1	-1	-1
11	9	Block 3	-1	1	-1	1
2	10	Block 3	1	-1	-1	-1
7	11	Block 3	-1	1	1	-1
14	12	Block 3	1	-1	1	1
16	13	Block 4	1	1	1	1
5	14	Block 4	-1	-1	1	-1
9	15	Block 4	-1	-1	-1	1
4	16	Block 4	1	1	-1	-1

If we were able to obtain 12 of the original 16 runs (that is, the first *three* blocks of Table 3), then we can estimate

[Intercept] = Intercept - 0.333 * AB - 0.333 * ACD - 0.333 * BCD
[A] = A - ABCD
[B] = B - ABCD
[C] = C - ABC
[D] = D - ABD
[AC] = AC - ABD
[AD] = AD - ABC
[BC] = BC - ABD
[BD] = BD - ABC
[CD] = CD – ABCD

If we can ignore three- and four-factor interactions, then we can obtain good estimates of all four main effects and five of the six two-factor interactions. Once again, setting up and running the experiment in blocks has proven to be a good idea, even though no nuisance factor was anticipated. Finally, we note that it is possible to assemble three of the four blocks from Table 3 to obtain a 12-run experiment that is slightly better than the one illustrated above. This would actually be called a $3/4^{th}$ fraction of the 2^4, an *irregular* fractional factorial. These designs are mentioned briefly in the Chapter 8 exercises.

Chapter 8. Supplemental Text Material

S8-1. Yates's Method for the Analysis of Fractional Factorials

Computer programs are almost always used for the analysis of fractional factorial. However, we may use Yates's algorithm for the analysis of a 2^{k-1} fractional factorial design by initially considering the data as having been obtained from a full factorial in k - 1 variables. The treatment combinations for this full factorial are listed in standard order, and then an additional letter (or letters) is added in parentheses to these treatment combinations to produce the actual treatment combinations run. Yates's algorithm then proceeds as usual. The actual effects estimated are identified by multiplying the effects associated with the treatment combinations in the full 2^{k-1} design by the defining relation of the 2^{k-1} fractional factorial.

The procedure is demonstrated in Table 1 below using the data from Example 8-1. This is a 2^{4-1} fractional. The data are arranged as a full 2^3 design in the factors *A*, *B*, and *C*. Then the letter *d* is added in parentheses to yield the actual treatment combinations that were performed. The effect estimated by, say, the second row in this table is *A* + *BCD* since *A* and *BCD* are aliases.

Table 1. Yates's Algorithm for the 2_{IV}^{4-1} Fractional Factorial in Example 8-1

Treatment Combination	Response	(1)	(2)	(3)	Effect	Effect Estimate $2 \times (3) / N$
(1)	45	145	255	566	-	-
a(*d*)	100	110	311	76	*A*+*BCD*	19.00
b(*d*)	45	135	75	6	*B*+*ACD*	1.5
ab	65	176	1	-4	*AB*+*CD*	-1.00
c(*d*)	75	55	-35	56	*C*+*ABD*	14.00
ac	60	20	41	-74	*AC*+*BD*	-18.50
bc	80	-15	-15	76	*BC*+*AD*	19.00
abc(*d*)	96	16	16	66	*ABC*+*D*	16.50

S8-2 Alias Structures in Fractional Factorials and Other Designs

In this chapter we show how to find the alias relationships in a 2^{k-p} fractional factorial design by use of the complete defining relation. This method works well in simple designs, such as the regular fractions we use most frequently, but it does not work as well in more complex settings, such as some of the irregular fractions and partial fold-over designs. Furthermore, there are some fractional factorials that do not have defining relations, such as Plackett-Burman designs, so the defining relation method will not work for these types of designs at all.

Fortunately, there is a general method available that works satisfactorily in many situations. The method uses the polynomial or regression model representation of the model, say

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \varepsilon$$

where $\mathbf{y}$ is an $n \times 1$ vector of the responses, $\mathbf{X}_1$ is an $n \times p_1$ matrix containing the design matrix expanded to the form of the model that the experimenter is fitting, β_1 is an $p_1 \times 1$ vector of the model parameters, and ε is an $n \times 1$ vector of errors. The least squares estimate of β_1 is

$$\hat{\beta}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}$$

Suppose that the **true** model is

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon$$

where $\mathbf{X}_2$ is an $n \times p_2$ matrix containing additional variables that are not in the fitted model and β_2 is a $p_2 \times 1$ vector of the parameters associated with these variables. It can be easily shown that

$$\begin{aligned} E(\hat{\beta}_1) &= \beta_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2 \\ &= \beta_1 + \mathbf{A}\beta_2 \end{aligned}$$

where $\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$ is called the **alias matrix**. The elements of this matrix operating on β_2 identify the alias relationships for the parameters in the vector β_1.

We illustrate the application of this procedure with a familiar example. Suppose that we have conducted a 2^{3-1} design with defining relation $I = ABC$ or $I = x_1x_2x_3$. The model that the experimenter plans to fit is the main-effects-only model

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \varepsilon$$

In the notation defined above,

$$\beta_1 = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \text{ and } \mathbf{X}_1 = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Suppose that the true model contains all the two-factor interactions, so that

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{23}x_2x_3 + \varepsilon$$

and

$$\beta_2 = \begin{bmatrix} \beta_{12} \\ \beta_{13} \\ \beta_{23} \end{bmatrix}, \text{ and } \mathbf{X}_2 = \begin{bmatrix} 1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

Now

$$(\mathbf{X}_1'\mathbf{X}_1)^{-1} = \frac{1}{4}\mathbf{I}_4 \text{ and } \mathbf{X}_1'\mathbf{X}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 4 \\ 0 & 4 & 0 \\ 4 & 0 & 0 \end{bmatrix}$$

Therefore

$$E(\hat{\beta}_1) = \beta_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2$$

$$E\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \frac{1}{4}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 4 \\ 0 & 4 & 0 \\ 4 & 0 & 0 \end{bmatrix}\begin{bmatrix} \beta_{12} \\ \beta_{13} \\ \beta_{23} \end{bmatrix}$$

$$= \begin{bmatrix} \beta_0 \\ \beta_1 + \beta_{23} \\ \beta_2 + \beta_{13} \\ \beta_3 + \beta_{12} \end{bmatrix}$$

The interpretation of this, of course, is that each of the main effects is aliased with one of the two-factor interactions, which we know to be the case for this design. While this is a very simple example, the method is very general and can be applied to much more complex designs.

S8-3. More About Fold-Over and Partial Fold-Over of Fractional Factorials

In the textbook, we illustrate how a fractional factorial design can be augmented with additional runs to separate effects that are aliased. A fold-over is another design that is the same size as the original fraction. So if the original experiment has 16 runs, the fold-over will require another 16 runs.

Sometimes it is possible to augment a 2^{k-p} fractional factorial with fewer than an additional 2^{k-p} runs. This technique is generally referred to as a partial fold over of the original design.

For example, consider the 2^{5-2} design shown in Table 2. The alias structure for this design is shown below the table.

Table 2. A 2^{5-2} Design

Std ord	Run ord	Block	Factor A:A	Factor B:B	Factor C:C	Factor D:D	Factor E:E
2	1	Block 1	1	-1	-1	-1	-1
6	2	Block 1	1	-1	1	-1	1
3	3	Block 1	-1	1	-1	-1	1
1	4	Block 1	-1	-1	-1	1	1
8	5	Block 1	1	1	1	1	1
5	6	Block 1	-1	-1	1	1	-1
4	7	Block 1	1	1	-1	1	-1
7	8	Block 1	-1	1	1	-1	-1

[A] = A + BD + CE
[B] = B + AD + CDE
[C] = C + AE + BDE
[D] = D + AB + BCE
[E] = E + AC + BCD
[BC] = BC + DE + ABE + ACD
[BE] = BE + CD + ABC + ADE

Now suppose that after running the eight trials in Table 2, the largest effects are the main effects *A*, *B*, and *D*, and the *BC* + *DE* interaction. The experimenter believes that all other effects are negligible. Now this is a situation where fold-over of the original design is not an attractive alternative. Recall that when a resolution III design is folded over by reversing all the signs in the test matrix, the combined design is resolution IV. Consequently, the *BC* and *DE* interactions will still be aliased in the combined design. One could alternatively consider reversing signs in individual columns, but these approaches will essentially require that another eight runs be performed.

The experimenter wants to fit the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_4 + \beta_{23} x_2 x_3 + \beta_{45} x_4 x_5 + \varepsilon$$

where $x_1 = A, x_2 = B, x_3 = C, x_4 = D,$ and $x_5 = E$. Recall that a **partial fold-over** is a design containing fewer than eight runs that can be used to augment the original design and will allow the experimenter to fit this model. One way to select the runs for the partial fold-over is to select points from the remaining unused portion of the 2^5 such that the variances of the model coefficients in the above regression equation are minimized. This augmentation strategy is based on the idea of a **D-optimal design**, discussed in Chapter 11.

Design-Expert can utilize this strategy to find a partial fold-over. The design produced by the computer program is shown in Table 3. This design completely dealiases the *BC* and *DE* interactions.

Table 3. The Partially-Folded Fractional Design

Std ord	Run ord	Block	Factor A:A	Factor B:B	Factor C:C	Factor D:D	Factor E:E
2	1	Block 1	1	-1	-1	-1	-1
6	2	Block 1	1	-1	1	-1	1
3	3	Block 1	-1	1	-1	-1	1
1	4	Block 1	-1	-1	-1	1	1
8	5	Block 1	1	1	1	1	1
5	6	Block 1	-1	-1	1	1	-1
4	7	Block 1	1	1	-1	1	-1
7	8	Block 1	-1	1	1	-1	-1
9	9	Block 2	-1	-1	-1	-1	1
10	10	Block 2	1	1	1	1	-1
11	11	Block 2	-1	-1	1	-1	-1
12	12	Block 2	1	1	-1	1	1

Notice that the D-optimal partial fold-over design requires four additional trials. Furthermore, these trials are arranged in a second block that is orthogonal to the first block of eight trials.

This strategy is very useful in 16-run resolution IV designs, situations in which a full fold-over would require another 16 trials. Often a partial fold-over with four or eight runs can be used as an alternative. In many cases, a partial fold with only four runs over can be constricted using the D-optimal approach.

As a second example, consider the 2^{6-2} resolution IV design shown in Table 4. The alias structure for the design is shown below the table.

Table 4. A 2^{6-2} Resolution IV Design

Std ord	Run ord	Block	Factor A:A	Factor B:B	Factor C:C	Factor D:D	Factor E:E	Factor F:F
10	1	Block 1	1	-1	-1	1	1	1
11	2	Block 1	-1	1	-1	1	1	-1
2	3	Block 1	1	-1	-1	-1	1	-1
12	4	Block 1	1	1	-1	1	-1	-1
16	5	Block 1	1	1	1	1	1	1
15	6	Block 1	-1	1	1	1	-1	1
8	7	Block 1	1	1	1	-1	1	-1
7	8	Block 1	-1	1	1	-1	-1	-1
5	9	Block 1	-1	-1	1	-1	1	1
1	10	Block 1	-1	-1	-1	-1	-1	-1
6	11	Block 1	1	-1	1	-1	-1	1
4	12	Block 1	1	1	-1	-1	-1	1
14	13	Block 1	1	-1	1	1	-1	-1
13	14	Block 1	-1	-1	1	1	1	-1
9	15	Block 1	-1	-1	-1	1	-1	1
3	16	Block 1	-1	1	-1	-1	1	1

[A] = A + BCE + DEF
[B] = B + ACE + CDF
[C] = C + ABE + BDF
[D] = D + AEF + BCF
[E] = E + ABC + ADF
[F] = F + ADE + BCD
[AB] = AB + CE
[AC] = AC + BE
[AD] = AD + EF
[AE] = AE + BC + DF
[AF] = AF + DE
[BD] = BD + CF
[BF] = BF + CD
[ABD] = ABD + ACF + BEF + CDE
[ABF] = ABF + ACD + BDE + CEF

Suppose that the main effects of factors *A, B, C,* and *E* are large, along with the *AB* + *CE* interaction chain. A full fold-over of this design would involve reversing the signs in columns *B, C, D, E,* and *F*. This would, of course, require another 16 trials. A standard partial fold using the method described in the textbook would require 8 additional runs. The D-optimal partial fold-over approach requires only four additional runs. The augmented design, obtained from Design-Expert, is shown in Table 5. These four runs form a second block that is orthogonal to the first block of 16 runs, and allows the interactions of interest in the original alias chain to be separately estimated.

Remember that partial fold over designs are irregular fractions. They are not orthogonal and as a result, the effect estimates are correlated. This correlation between effect estimates causes inflation in the standard errors of the effects; that is, the effects are not estimated as precisely as they would have been in an orthogonal design. However, this disadvantage may be offset by the decrease in the number of runs that the partial fold over requires.

Table 5. The Partial Fold-Over

Std	Run	Block	Factor A:A	Factor B:B	Factor C:C	Factor D:D	Factor E:E	Factor F:F
12	1	Block 1	1	1	-1	1	-1	-1
15	2	Block 1	-1	1	1	1	-1	1
2	3	Block 1	1	-1	-1	-1	1	-1
9	4	Block 1	-1	-1	-1	1	-1	1
5	5	Block 1	-1	-1	1	-1	1	1
8	6	Block 1	1	1	1	-1	1	-1
11	7	Block 1	-1	1	-1	1	1	-1
14	8	Block 1	1	-1	1	1	-1	-1
13	9	Block 1	-1	-1	1	1	1	-1
4	10	Block 1	1	1	-1	-1	-1	1
10	11	Block 1	1	-1	-1	1	1	1
6	12	Block 1	1	-1	1	-1	-1	1
7	13	Block 1	-1	1	1	-1	-1	-1
16	14	Block 1	1	1	1	1	1	1
3	15	Block 1	-1	1	-1	-1	1	1
1	16	Block 1	-1	-1	-1	-1	-1	-1
17	17	Block 2	1	-1	1	-1	-1	-1
18	18	Block 2	-1	1	-1	-1	-1	-1
19	19	Block 2	-1	-1	1	1	1	1
20	20	Block 2	1	1	-1	1	1	1

Chapter 9. Supplemental Text Material

S9-1. Yates's Algorithm for the 3^k Design

Computer methods are used almost exclusively for the analysis of factorial and fractional designs. However, Yates's algorithm can be modified for use in the 3^k factorial design. We will illustrate the procedure using the data in Example 5-1. The data for this example are originally given in Table 5-1. This is a 3^2 design used to investigate the effect of material type (A) and temperature (B) on the life of a battery. There are $n = 4$ replicates.

The Yates' procedure is displayed in Table 1 below. The treatment combinations are written down in standard order; that is, the factors are introduced one at a time, each level being combined successively with every set of factor levels above it in the table. (The standard order for a 3^3 design would be 000, 100, 200, 010, 110, 210, 020, 120, 220, 001, . . .). The Response column contains the total of all observations taken under the corresponding treatment combination. The entries in column (1) are computed as follows. The first third of the column consists of the sums of each of the three sets of three values in the Response column. The second third of the column is the third minus the first observation in the same set of three. This operation computes the linear component of the effect. The last third of the column is obtained by taking the sum of the first and third value minus twice the second in each set of three observations. This computes the quadratic component. For example, in column (1), the second, fifth, and eighth entries are 229 + 479 + 583 = 1291, -229 + 583 = 354, and 229 - (2)(479) + 583 = -146, respectively. Column (2) is obtained similarly from column (1). In general, k columns must be constructed.

The Effects column is determined by converting the treatment combinations at the left of the row into corresponding effects. That is, 10 represents the linear effect of A, A_L, and 11 represents the AB_{LXL} component of the AB interaction. The entries in the Divisor column are found from

$$2^r 3^t n$$

where r is the number of factors in the effect considered, t is the number of factors in the experiment minus the number of linear terms in this effect, and n is the number of replicates. For example, B_{L} has the divisor $2^1 \times 3^1 \times 4 = 24$.

The sums of squares are obtained by squaring the element in column (2) and dividing by the corresponding entry in the Divisor column. The Sum of Squares column now contains all of the required quantities to construct an analysis of variance table if both of the design factors A and B are quantitative. However, in this example, factor A (material type) is qualitative; thus, the linear and quadratic partitioning of A is not appropriate. Individual observations are used to compute the total sum of squares, and the error sum of squares is obtained by subtraction.

Table 1. Yates's Algorithm for the 3^2 Design in Example 5-1

Treatment Combination	Response	(1)	(2)	Effects	Divisor	Sum of Squares
00	539	1738	3799			
10	623	1291	503	A_L	$2^1 \times 3^1 \times 4$	10, 542.04
20	576	770	-101	A_Q	$2^1 \times 3^2 \times 4$	141.68
01	229	37	-968	B_L	$2^1 \times 3^1 \times 4$	39,042.66
11	479	354	75	AB_{LXL}	$2^2 \times 3^0 \times 4$	351.56
21	583	112	307	AB_{QXL}	$2^2 \times 3^1 \times 4$	1,963.52
02	230	-131	-74	B_Q	$2^1 \times 3^2 \times 4$	76.96
12	198	-146	-559	AB_{LXQ}	$2^2 \times 3^1 \times 4$	6,510.02
22	342	176	337	AB_{QXQ}	$2^2 \times 3^2 \times 4$	788.67

The analysis of variance is summarized in Table 2. This is essentially the same results that were obtained by conventional analysis of variance methods in Example 5-1.

Table 2. Analysis of Variance for the 3^2 Design in Example 5-1

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-value
$A = A_L \times A_Q$	10, 683.72	2	5,341.86	7.91	0.0020
B, Temperature	39,118.72	2	19,558.36	28.97	<0.0001
B_L	(39, 042.67)	(1)	39,042.67	57.82	<0.0001
B_Q	(76.05)	(1)	76.05	0.12	0.7314
AB	9,613.78	4	2,403.44	3.576	0.0186
$A \times B_L = AB_{LXL} + AB_{QXL}$	(2,315.08)	(2)	1,157.54	1.71	0.1999
$A \times B_Q = AB_{LXQ} + AB_{QXQ}$	(7,298.70)	(2)	3,649.75	5.41	0.0106
Error	18,230.75	27	675.21		
Total	77,646.97	35			

S9-2. Aliasing in Three-Level and Mixed-Level Designs

In the supplemental text material for Chapter 8 (Section 8-2) we gave a general method for finding the alias relationships for a fractional factorial design. Fortunately, there is a general method available that works satisfactorily in many situations. The method uses the polynomial or regression model representation of the model,

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \varepsilon$$

where $\mathbf{y}$ is an $n \times 1$ vector of the responses, $\mathbf{X}_1$ is an $n \times p_1$ matrix containing the design matrix expanded to the form of the model that the experimenter is fitting, β_1 is an $p_1 \times 1$ vector of the model parameters, and ε is an $n \times 1$ vector of errors. The least squares estimate of β_1 is

$$\hat{\beta}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}$$

The **true** model is assumed to be

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon$$

where $\mathbf{X}_2$ is an $n \times p_2$ matrix containing additional variables not in the fitted model and β_2 is a $p_2 \times 1$ vector of the parameters associated with these additional variables. The parameter estimates in the fitted model are not unbiased, since

$$\begin{aligned} E(\hat{\beta}_1) &= \beta_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2 \\ &= \beta_1 + \mathbf{A}\beta_2 \end{aligned}$$

The matrix $\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$ is called the **alias matrix**. The elements of this matrix identify the alias relationships for the parameters in the vector β_1.

This procedure can be used to find the alias relationships in three-level and mixed-level designs. We now present two examples.

Example 1

Suppose that we have conducted an experiment using a 3^2 design, and that we are interested in fitting the following model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11}(x_1^2 - \bar{x}_1^2) + \beta_{22}(x_2^2 - \bar{x}_2^2) + \varepsilon$$

This is a complete quadratic polynomial. The pure second-order terms are written in a form that orthogonalizes these terms with the intercept. We will find the aliases in the parameter estimates if the true model is a reduced cubic, say

$$\begin{aligned} y = {} & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11}(x_1^2 - \bar{x}_1^2) + \beta_{22}(x_2^2 - \bar{x}_2^2) \\ & + \beta_{111} x_1^3 + \beta_{222} x_2^3 + \beta_{122} x_1 x_2^2 + \varepsilon \end{aligned}$$

Now in the notation used above, the vector β_1 and the matrix $\mathbf{X}_1$ are defined as follows:

$$\beta_1 = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{12} \\ \beta_{11} \\ \beta_{22} \end{bmatrix}, \quad \text{and} \quad \mathbf{X}_1 = \begin{bmatrix} 1 & -1 & -1 & 1 & 1/3 & 1/3 \\ 1 & -1 & 0 & 0 & 1/3 & -2/3 \\ 1 & -1 & 1 & -1 & 1/3 & 1/3 \\ 1 & 0 & -1 & 0 & -2/3 & 1/3 \\ 1 & 0 & 0 & 0 & -2/3 & -2/3 \\ 1 & 0 & 1 & 0 & -2/3 & 1/3 \\ 1 & 1 & -1 & -1 & 1/3 & 1/3 \\ 1 & 1 & 0 & 0 & 1/3 & -2/3 \\ 1 & 1 & 1 & 1 & 1/3 & 1/3 \end{bmatrix}$$

Now

$$\mathbf{X}_1'\mathbf{X}_1 = \begin{bmatrix} 9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

and the other quantities we require are

$$\mathbf{X}_2 = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 0 & 0 \\ -1 & 1 & -1 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad \beta_2 = \begin{bmatrix} \beta_{111} \\ \beta_{222} \\ \beta_{122} \end{bmatrix}, \quad \text{and} \quad \mathbf{X}_1'\mathbf{X}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 6 & 0 & 4 \\ 0 & 6 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The expected value of the fitted model parameters is

$$E(\hat{\beta}_1) = \beta_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2$$

or

$$E\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_{12}\\ \hat{\beta}_{11}\\ \hat{\beta}_{22}\end{bmatrix}=\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_{12}\\ \beta_{11}\\ \beta_{22}\end{bmatrix}+\begin{bmatrix}9&0&0&0&0&0\\ 0&6&0&0&0&0\\ 0&0&6&0&0&0\\ 0&0&0&4&0&0\\ 0&0&0&0&2&0\\ 0&0&0&0&0&2\end{bmatrix}^{-1}\begin{bmatrix}0&0&0\\ 6&0&4\\ 0&6&0\\ 0&0&0\\ 0&0&0\\ 0&0&0\end{bmatrix}\begin{bmatrix}\beta_{111}\\ \beta_{222}\\ \beta_{122}\end{bmatrix}$$

The alias matrix turns out to be

$$\mathbf{A}=\begin{bmatrix}0&0&0\\ 1&0&2/3\\ 0&1&0\\ 0&0&0\\ 0&0&0\\ 0&0&0\end{bmatrix}$$

This leads to the following alias relationships:

$$E(\hat{\beta}_0)=\beta_0$$
$$E(\hat{\beta}_1)=\beta_1+\beta_{111}+(2/3)\beta_{122}$$
$$E(\hat{\beta}_2)=\beta_2+\beta_{222}$$
$$E(\hat{\beta}_{12})=\beta_{12}$$
$$E(\hat{\beta}_{11})=\beta_{11}$$
$$E(\hat{\beta}_{22})=\beta_{22}$$

Example 2

This procedure is very useful when the design is a mixed-level fractional factorial. For example, consider the mixed-level design in Table 9-10 of the textbook. This design can accommodate four two-level factors and a single three-level factor. The resulting resolution III fractional factorial is shown in Table 3.

Since the design is resolution III, the appropriate model contains the main effects

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_5+\beta_{55}(x_5^2-\bar{x}_5^2)+\varepsilon,$$

where the model terms

$$\beta_5x_5 \text{ and } \beta_{55}(x_5^2-\bar{x}_5^2)$$

represent the linear and quadratic effects of the three-level factor x_5. The quadratic effect of x_5 is defined so that it will be orthogonal to the intercept term in the model.

Table 3. A Mixed-Level Resolution III Fractional Factorial

x_1	x_2	x_3	x_4	x_5
-1	1	1	-1	-1
1	-1	-1	1	-1
-1	-1	1	1	0
1	1	-1	-1	0
-1	1	-1	1	0
1	-1	1	-1	0
-1	-1	-1	-1	1
1	1	1	1	1

Now suppose that the **true** model has interaction:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_{55}(x_5^2 - \bar{x}_5^2) + \beta_{12} x_1 x_2 + \beta_{15} x_1 x_5 + \beta_{155} x_1 (x_5^2 - \bar{x}_5^2) + \varepsilon$$

So in the true model the two-level factors x_1 and x_2 interact, and x_1 interacts with both the linear and quadratic effects of the three-level factor x_5. Straightforward, but tedious application of the procedure described above leads to the alias matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and the alias relationships are computed from

$$E(\hat{\beta}_1) = \beta_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2 = \beta_1 + \mathbf{A}\beta_2$$

This results in

$$E(\hat{\beta}_0) = \beta_0$$
$$E(\hat{\beta}_1) = \beta_1$$
$$E(\hat{\beta}_2) = \beta_2 + (1/2)\beta_{15}$$
$$E(\hat{\beta}_3) = \beta_3 + (1/2)\beta_{15}$$
$$E(\hat{\beta}_4) = \beta_4 + (1/2)\beta_{155}$$
$$E(\hat{\beta}_5) = \beta_5 + \beta_{12}$$
$$E(\hat{\beta}_{55}) = \beta_{55}$$

The linear and quadratic components of the interaction between x_1 and x_5 are aliased with the main effects of $x_2, x_3,$ and x_4, and the x_1x_2 interaction aliases the linear component of the main effect of x_5.

Chapter 10. Supplemental Text Material

S10-1. The Covariance Matrix of the Regression Coefficients

In Section 10-3 of the textbook, we show that the least squares estimator of β in the linear regression model $\mathbf{y} = \mathbf{X}\beta + \varepsilon$

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

is an unbiased estimator. We also give the result that the covariance matrix of $\hat{\beta}$ is $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ (see Equation 10-18). This last result is relatively straightforward to show. Consider

$$V(\hat{\beta}) = V[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}]$$

The quantity $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is just a matrix of constants and **y** is a vector of random variables. Now remember that the variance of the product of a scalar constant and a scalar random variable is equal to the square of the constant times the variance of the random variable. The matrix equivalent of this is

$$\begin{aligned} V(\hat{\beta}) &= V[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'V(\mathbf{y})[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']' \end{aligned}$$

Now the variance of **y** is $\sigma^2\mathbf{I}$, where **I** is an $n \times n$ identity matrix. Therefore, this last equation becomes

$$\begin{aligned} V(\hat{\beta}) &= V[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'V(\mathbf{y})[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']' \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']' \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

We have used the result from matrix algebra that the transpose of a product of matrices is just the produce of the transposes in reverse order, and since $(\mathbf{X}'\mathbf{X})$ is symmetric its transpose is also symmetric.

S10-2. Regression Models and Designed Experiments

In Examples 10-2 through 10-5 we illustrate several uses of regression methods in fitting models to data from designed experiments. Consider Example 10-2, which presents the regression model for main effects from a 2^3 factorial design with three center runs. Since the $(\mathbf{X}'\mathbf{X})^{-1}$ matrix is symmetric because the design is orthogonal, all covariance terms between the regression coefficients are zero. Furthermore, the variance of the regression coefficients is

$$V(\hat{\beta}_0) = \sigma^2 / 12 = 0.0833\sigma^2$$
$$V(\hat{\beta}_i) = \sigma^2 / 8 = 0.125\sigma^2, i = 1,2,3$$

In Example 10-3, we reconsider this same problem but assume that one of the original 12 observations is missing. It turns out that the estimates of the regression coefficients does not change very much when the remaining 11 observations are used to fit the first-order model but the $(\mathbf{X}'\mathbf{X})^{-1}$ matrix reveals that the missing observation has had a moderate effect on the variances and covariances of the model coefficients. The variances of the regression coefficients are now larger, and there are some moderately large covariances between the estimated model coefficients. Example 10-4, which investigated the impact of inaccurate design factor levels, exhibits similar results. Generally, as soon as we depart from an orthogonal design, either intentionally or by accident (as in these two examples), the variances of the regression coefficients will increase and potentially there could be rather large covariances between certain regression coefficients. In both of the examples in the textbook, the covariances are not terribly large and would not likely result in any problems in interpretation of the experimental results.

S10-3. Adjusted R^2

In several places in the textbook, we have remarked that the adjusted R^2 statistic is preferable to the ordinary R^2, because it is not a monotonically non-decreasing function of the number of variables in the model.

From Equation (10-27) note that

$$R^2_{Adj} = 1 - \left[\frac{SS_E / df_e}{SS_T / df_T}\right]$$
$$= 1 - \frac{MS_E}{SS_T / df_T}$$

Now the mean square in the denominator of the ratio is constant, but MS_E will change as variables are added or removed from the model. In general, the adjusted R^2 will increase when a variable is added to a regression model only if the error mean square decreases. The error mean square will only decrease if the added variable decreases the residual sum of squares by an amount that will offset the loss of one degree of freedom for error. Thus the added variable must reduce the residual sum of squares by an amount that is at least equal to the residual mean square in the immediately previous model; otherwise, the new model will have an adjusted R^2 value that is larger than the adjusted R^2 statistic for the old model.

S10-4. Stepwise and Other Variable-Selection Methods in Regression

In the textbook treatment of regression, we concentrated on fitting the full regression model. Actually, in moist applications of regression to data from designed experiments the experimenter will have a very good idea about the form of the model he or she wishes

to fit, either from an ANOVA or from examining a normal probability plot of effect estimates.

There are, however, other situations where regression is applied to **unplanned studies**, where the data may be observational data collected routinely on some process. The data may also be archival, obtained from some historian or library. These applications of regression frequently involve a moderately-large or large set of **candidate regressors**, and the objective of the analysts here is to fit a regression model to the "best subset" of these candidates. This can be a complex problem, as these unplanned data sets frequently have outliers, strong correlations between subsets of the variables, and other complicating features.

There are several techniques that have been developed for selecting the best subset regression model. Generally, these methods are either **stepwise-type** variable selection methods or **all possible regressions**. Stepwise-type methods build a regression model by either adding or removing a variable to the basic model at each step. The forward selection version of the procedure begins with a model containing none of the candidate variables and sequentially inserts variables into the model one-at-a-time until a final equation is produced. In backward elimination, the procedure begins with all variables in the equation, and then variables are removed one-at-a-time to produce a final equation. Stepwise regression usually consists of a combination of forward and backward stepping. There are many variations of the basic procedures.

In all possible regressions with K candidate variables, the analyst examines all 2^K possible regression equations to identify the ones with potential to be a useful model. Obviously, as K becomes even moderately large, the number of possible regression models quickly becomes formidably large. Efficient algorithms have been developed that implicitly rather than explicitly examine all of these equations. For more discussion of variable selection methods, see textbooks on regression such as Montgomery and Peck (1992) or Myers (1990).

S10-5. The Variance of the Predicted Response

In section 10-5.2 we present Equation (10-40) for the variance of the predicted response at a point of interest $\mathbf{x}_0' = [x_{01}, x_{02}, \cdots, x_{0k}]$. The variance is

$$V[\hat{y}(\mathbf{x}_0)] = \sigma^2 \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$$

where the predicted response at the point $\mathbf{x}_0$ is found from Equation (10-39):

$$\hat{y}(\mathbf{x}_0) = \mathbf{x}_0'\hat{\beta}$$

It is easy to derive the variance expression:

$$\begin{aligned} V[\hat{y}(\mathbf{x}_0)] &= V(\mathbf{x}_0'\hat{\beta}) \\ &= \mathbf{x}_0' V(\hat{\beta})\mathbf{x}_0 \\ &= \sigma^2 \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 \end{aligned}$$

Design-Expert calculates and displays the confidence interval on the mean of the response at the point $\mathbf{x}_0$ using Equation (10-41) from the textbook. This is displayed on the point prediction option on the optimization menu. The program also uses Equation (10-40) in the contour plots of prediction standard error.

S10-6. Variance of Prediction Error

Section 10-6 of the textbook gives an equation for a prediction interval on a future observation at the point $\mathbf{x}_0' = [x_{01}, x_{02}, \cdots, x_{0k}]$. This prediction interval makes use of the variance of prediction error. Now the point prediction of the future observation y_0 at $\mathbf{x}_0$ is

$$\hat{y}(\mathbf{x}_0) = \mathbf{x}_0'\hat{\beta}$$

and the prediction error is

$$e_p = y_0 - \hat{y}(\mathbf{x}_0)$$

The variance of the prediction error is

$$\begin{aligned} V(e_p) &= V[y_0 - \hat{y}(\mathbf{x}_0)] \\ &= V(y_0) + V[\hat{y}(\mathbf{x}_0)] \end{aligned}$$

because the future observation is independent of the point prediction. Therefore,

$$\begin{aligned} V(e_p) &= V(y_0) + V[\hat{y}(\mathbf{x}_0)] \\ &= \sigma^2 + \sigma^2\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 \\ &= \sigma^2\left[1 + \mathbf{x}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0\right] \end{aligned}$$

The square root of this quantity, with an estimate of $\sigma^2 = \hat{\sigma}^2 = MS_E$, appears in Equation (10-42) defining the prediction interval.

S10-7. Leverage in a Regression Model

In Section 10-7.2 we give a formal definition of the leverage associated with each observation in a design (or more generally a regression data set). Essentially, the leverage for the *i*th observation is just the *i*th diagonal element of the "hat" matrix

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

or

$$h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$$

where it is understood that $\mathbf{x}_i'$ is the *i*th row of the $\mathbf{X}$ matrix.

There are two ways to interpret the leverage values. First, the leverage h_{ii} is a measure of **distance** reflecting how far each design point is from the center of the design space. For example, in a 2^k factorial all of the cube corners are the same distance $\sqrt{k}$ from the

design center in coded units. Therefore, if all points are replicated n times, they will all have identical leverage.

Leverage can also be thought of as the maximum potential influence each design point exerts on the model. In a near-saturated design many or all design points will have the maximum leverage. The maximum leverage that any point can have is $h_{ii} = 1$. However, if points are replicated n times, the maximum leverage is $1/n$. High leverage situations are not desirable, because if leverage is unity that point fits the model exactly. Clearly, then, the design and the associated model would be vulnerable to outliers or other unusual observations at that design point. The leverage at a design point can always be reduced by replication of that point.

Chapter 11. Supplemental Text Material

S11-1. The Method of Steepest Ascent

The method of steepest ascent can be derived as follows. Suppose that we have fit a first-order model

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i$$

and we wish to use this model to determine a path leading from the center of the design region $\mathbf{x} = \mathbf{0}$ that increases the predicted response most quickly. Since the first–order model is an unbounded function, we cannot just find the values of the x's that maximize the predicted response. Suppose that instead we find the x's that maximize the predicted response at a point on a hypersphere of radius r. That is

$$\text{Max } \hat{y} = \beta_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i$$

subject to

$$\sum_{i=1}^{k} x_i^2 = r^2$$

The can be formulated as

$$\text{Max } G = \beta_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i - \lambda\left[\sum_{i=1}^{k} x_i^2 - r^2\right]$$

where λ is a LaGrange multiplier. Taking the derivatives of G yields

$$\frac{\partial G}{\partial x_i} = \hat{\beta}_i - 2\lambda x_i \quad i - 1,2,\cdots,k$$

$$\frac{\partial G}{\partial \lambda} = -\left[\sum_{i=1}^{k} x_i^2 - r^2\right]$$

Equating these derivatives to zero results in

$$x_i = \frac{\hat{\beta}_i}{2\lambda} \quad i = 1,2,\cdots,k$$

$$\sum_{i=1}^{k} x_i^2 = r^2$$

Now the first of these equations shows that the coordinates of the point on the hypersphere are proportional to the signs and magnitudes of the regression coefficients (the quantity 2λ is a constant that just fixes the radius of the hypersphere). The second equation just states that the point satisfies the constraint. Therefore, the heuristic description of the method of steepest ascent can be justified from a more formal perspective.

S11-2. The Canonical Form of the Second-Order Response Surface Model

Equation (11-9) presents a very useful result, the **canonical form** of the second-order response surface model. We state that this form of the model is produced as a result of a translation of the original coded variable axes followed by rotation of these axes. It is easy to demonstrate that this is true.

Write the second-order model as

$$\hat{y} = \hat{\beta}_0 + \mathbf{x}'\hat{\beta} + \mathbf{x}'\mathbf{B}\mathbf{x}$$

Now translate the coded design variable axes $\mathbf{x}$ to a new center, the stationary point, by making the substitution $\mathbf{z} = \mathbf{x} - \mathbf{x}_s$. This translation produces

$$\begin{aligned}\hat{y} &= \hat{\beta}_0 + (\mathbf{z}+\mathbf{x}_s)'\hat{\beta} + (\mathbf{z}+\mathbf{x}_s)'\mathbf{B}(\mathbf{z}+\mathbf{x}_s) \\ &= \left[\hat{\beta}_0 + \mathbf{x}_s'\hat{\beta} + \mathbf{x}_s'\mathbf{B}\mathbf{x}_s\right] + \mathbf{z}'\hat{\beta} + \mathbf{z}'\mathbf{B}\mathbf{z} + 2\mathbf{x}_s'\mathbf{B}\mathbf{z} \\ &= \hat{y}_s + \mathbf{z}'\mathbf{B}\mathbf{z}\end{aligned}$$

because from Equation (11-7) we have $2\mathbf{x}_s'\mathbf{B}\mathbf{z} = -\mathbf{z}'\hat{\beta}$. Now rotate these new axes ($\mathbf{z}$) so that they are parallel to the principal axes of the contour system. The new variables are $\mathbf{w} = \mathbf{M}'\mathbf{z}$, where

$$\mathbf{M}'\mathbf{B}\mathbf{M} = \Lambda$$

The diagonal matrix Λ has the eigenvalues of $\mathbf{B}$, $\lambda_1, \lambda_2, \cdots, \lambda_k$ on the main diagonal and $\mathbf{M}$ is a matrix of normalized eigenvectors. Therefore,

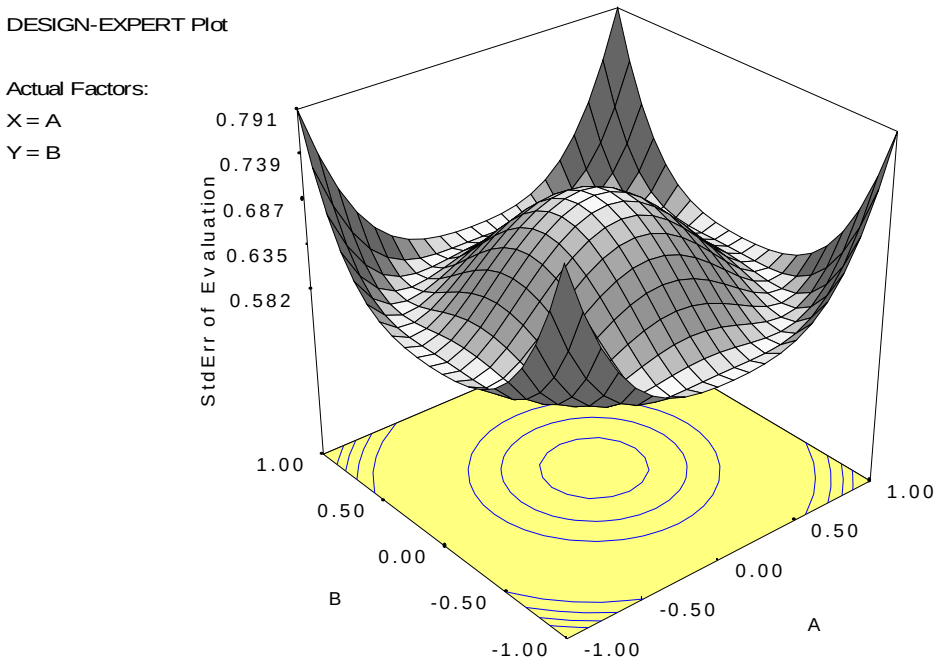
$$\begin{aligned}\hat{y} &= \hat{y}_s + \mathbf{z}'\mathbf{B}\mathbf{z} \\ &= \hat{y}_s + \mathbf{w}'\mathbf{M}'\mathbf{B}\mathbf{M}\mathbf{z} \\ &= \hat{y}_s + \mathbf{w}'\Lambda\mathbf{w} \\ &= \hat{y}_s + \sum_{i=1}^{k} \lambda_i w_i^2\end{aligned}$$

which is Equation (11-9).

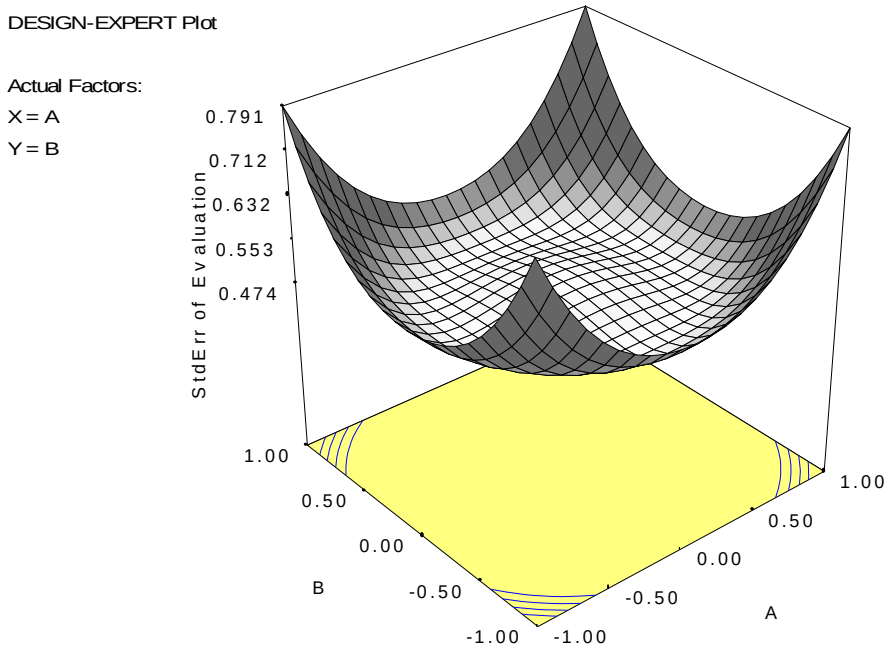
S11-3. Center Points in the Central Composite Design

In section 11-4,2 we discuss designs for fitting the second-order model. The CCD is a very important second-order design. We have given some recommendations regarding the number of center runs for the CCD; namely, $3 \le n_c \le 5$ generally gives good results.

The center runs serves to stabilize the prediction variance, making it nearly constant over a broad region near the center of the design space. To illustrate, suppose that we are considering a CCD in $k = 2$ variables but we only plan to run $n_c = 2$ center runs. The following graph of the standardized standard deviation of the predicted response was obtained from Design-Expert:



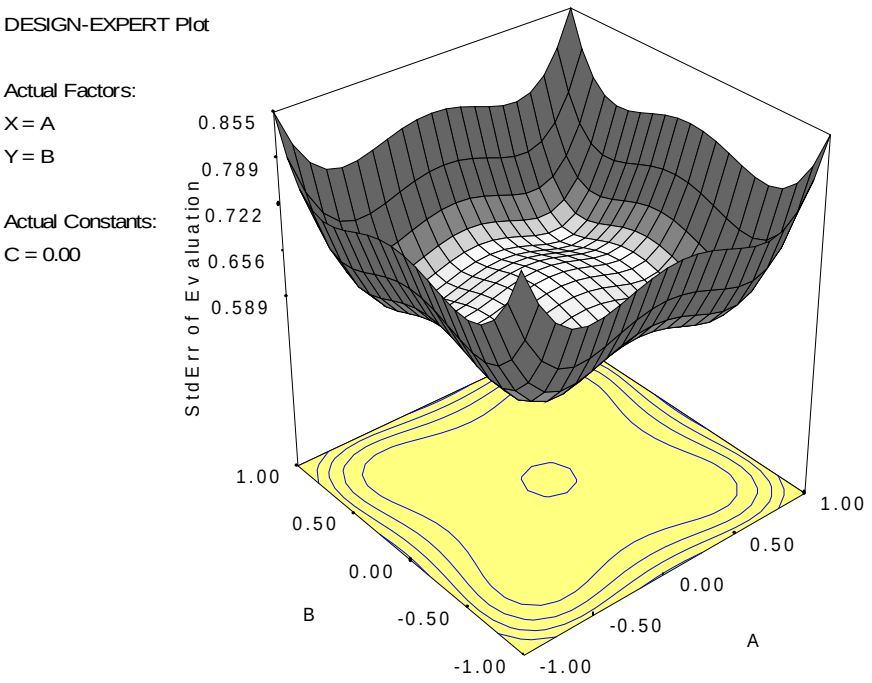
Notice that the plot of the prediction standard deviation has a large "bump" in the center. This indicates that the design will lead to a model that does not predict accurately near the center of the region of exploration, a region likely to be of interest to the experimenter. This is the result of using an insufficient number of center runs. Suppose that the number of center runs is increased to $n_c = 4$. The prediction standard deviation plot now looks like this:



Notice that the addition of two more center runs has resulted in a much flatter (and hence more stable) standard deviation of predicted response over the region of interest.
The CCD is a spherical design. Generally, every design on a sphere must have at least one center point or the $\mathbf{X}'\mathbf{X}$ matrix will be singular. However, the number of center points can often influence other properties of the design, such as prediction variance.

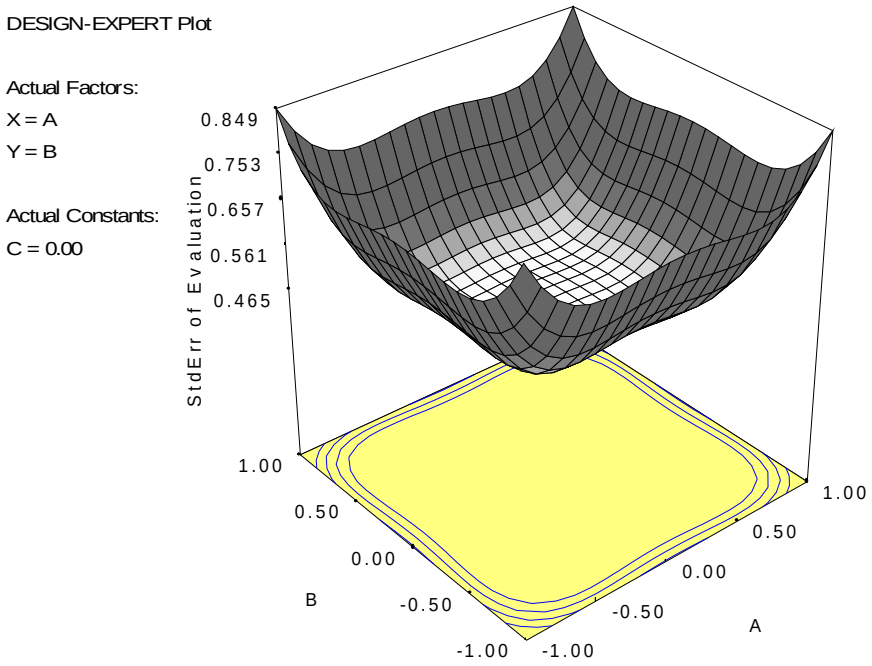
S11-4. Center Runs in the Face-Centered Cube

The face-centered cube is a CCD with $\alpha = 1$; consequently, it is a design on a cube, it is not a spherical design. This design can be run with as few as $n_c = 0$ center points. The prediction standard deviation for the case $k = 3$ is shown below:



Notice that despite the absence of center points, the prediction standard deviation is relatively constant in the center of the region of exploration. Note also that the contours of constant prediction standard deviation are not concentric circles, because this is not a rotatable design.

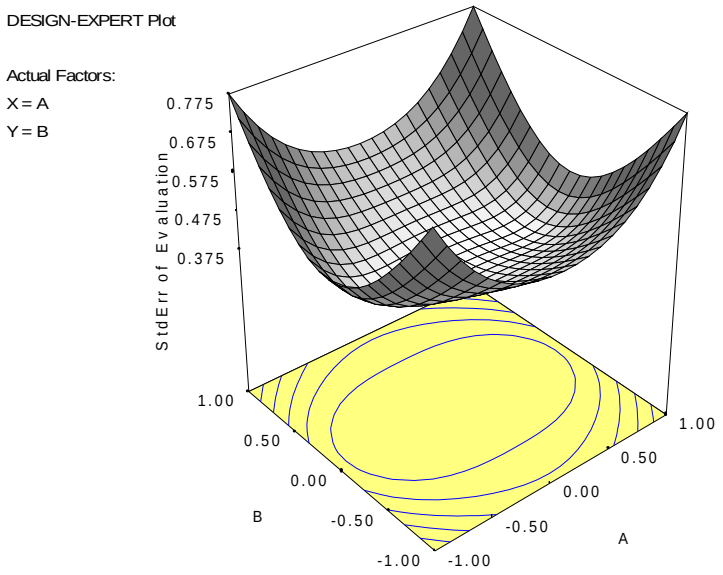
While this design will certainly work with no center points, this is usually not a good choice. Two or three center points generally gives good results. Below is a plot of the prediction standard deviation for a face-centered cube with two center points. This choice work very well.



S11-5. A Note on Rotatability

Rotatability is a property of the prediction variance in a response surface design. If a design is rotatable, the prediction variance is constant at all points that are equidistant from the center of the design.

What is not widely known is that rotatability depends on both the design and the model. For example, if we have run a rotatable CCD and fit a **reduced** second-order model, the variance contours are no longer spherical. To illustrate, below we show the standardized standard deviation of prediction for a rotatable CCD with $k = 2$, but we have fit a reduced quadratic (one of the pure quadratic terms is missing).



Notice that the contours of prediction standard deviation are not circular, even though a rotatable design was used.

Chapter 12 Supplemental Text Material

S12-1. The Taguchi Approach to Robust Parameter Design

Throughout this book, we have emphasized the importance of using designed experiments for product and process improvement. Today, many engineers and scientists are exposed to the principles of statistically designed experiments as part of their formal technical education. However, during the 1960-1980 time period, the principles of experimental design (and statistical methods, in general) were not as widely used as they are today

In the early 1980s, Genichi Taguchi, a Japanese engineer, introduced his approach to using experimental design for

1. Designing products or processes so that they are robust to environmental conditions.

2. Designing/developing products so that they are robust to component variation.

3. Minimizing variation around a target value.

Note that these are essentially the same objectives we discussed in Section 11-7.1.

Taguchi has certainly defined meaningful engineering problems and the philosophy that recommends is sound. However, as noted in the textbook, he advocated some novel methods of statistical data analysis and some approaches to the design of experiments that the process of peer review revealed were unnecessarily complicated, inefficient, and sometimes ineffective. In this section, we will briefly overview Taguchi's philosophy regarding quality engineering and experimental design. We will present some examples of his approach to parameter design, and we will use these examples to highlight the problems with his technical methods. As we saw in Chapter 12 of the textbook, it is possible to combine his sound engineering concepts with more efficient and effective experimental design and analysis based on response surface methods.

Taguchi advocates a philosophy of quality engineering that is broadly applicable. He considers three stages in product (or process) development: system design, parameter design, and tolerance design. In **system design,** the engineer uses scientific and engineering principles to determine the basic system configuration. For example, if we wish to measure an unknown resistance, we may use our knowledge of electrical circuits to determine that the basic system should be configured as a Wheatstone bridge. If we are designing a process to assemble printed circuit boards, we will determine the need for specific types of axial insertion machines, surface-mount placement machines, flow solder machines, and so forth.

In the **parameter design** stage, the specific values for the system parameters are determined. This would involve choosing the nominal resistor and power supply values for the Wheatstone bridge, the number and type of component placement machines for the printed circuit board assembly process, and so forth. Usually, the objective is to

specify these nominal parameter values such that the variability transmitted from uncontrollable or noise variables is minimized.

Tolerance design is used to determine the best tolerances for the parameters. For example, in the Wheatstone bridge, tolerance design methods would reveal which components in the design were most sensitive and where the tolerances should be set. If a component does not have much effect on the performance of the circuit, it can be specified with a wide tolerance.

Taguchi recommends that statistical experimental design methods be employed to assist in this process, particularly during parameter design and tolerance design. We will focus on parameter design. Experimental design methods can be used to find a best product or process design, where by "best" we mean a product or process that is robust or insensitive to uncontrollable factors **that will** influence the product or process once it is in routine operation.

The notion of **robust design** is not new. Engineers have always tried to design products so that they will work well under uncontrollable conditions. For example, commercial transport aircraft fly about as well in a thunderstorm as they do in clear air. Taguchi deserves recognition for realizing that experimental design can be used as a formal part of the **engineering design process** to help accomplish this objective.

A key component of Taguchi's philosophy is the **reduction of variability.** Generally, each product or process performance characteristic will have a target or **nominal** value. The objective is to reduce the variability around this target value. Taguchi models the departures that may occur from this target value with a **loss function.** The loss refers to the cost that is incurred by *society* when the consumer uses a product whose quality characteristics differ from the nominal. The concept of societal loss is a departure from traditional thinking. Taguchi imposes a quadratic loss function of the form

$$L(y) = k(y- T)^2$$

shown in Figure 1 below. Clearly this type of function will penalize even small departures of y from the target T. Again, this is a departure from traditional thinking, which usually attaches penalties only to cases where y is outside of the upper and lower specifications (say $y > USL$ *or* $y < LSL$ in Figure 1. However, the Taguchi philosophy regarding reduction of variability and the emphasis on minimizing costs is entirely consistent with the continuous improvement philosophy of Deming and Juran.

In summary, Taguchi's philosophy involves three central ideas:

1. Products and processes should be designed so that they are robust to external sources of variability.
2. Experimental design methods are an engineering tool to help accomplish this objective.
3. Operation on-target is more important than conformance to specifications.

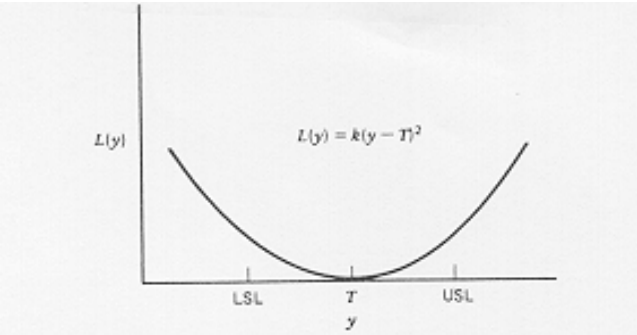


Figure 14–27. Taguchi's quadratic loss function.

Figure 1. Taguchi's Quadratic Loss Function

These are sound concepts, and their value should be readily apparent. Furthermore, as we have seen in the textbook, experimental design methods can play a major role in translating these ideas into practice.

We now turn to a discussion of the specific methods that Professor Taguchi recommends for applying his concepts in practice. As we will see, his approach to experimental design and data analysis can be improved.

S12-2. Taguchi's Technical Methods

An Example

We will use the connector pull-off force example described in the textbook to illustrate Taguchi's technical methods. For more information about the problem, refer to the text and to the original article in *Quality Progress* in December 1987 (see "The Taguchi Approach to Parameter Design," by D. M. Byrne and S. Taguchi, *Quality Progress*, December 1987, pp. 19-26). Recall that the experiment involves finding a method to assemble an elastomeric connector to a nylon tube that would deliver the required pull-off performance to be suitable for use in an automotive engine application. The specific objective of the experiment is to maximize the pull-off force. Four controllable and three uncontrollable noise factors were identified. These factors are defined in the textbook, and repeated for convenience in Table 1 below. We want to find the levels of the controllable factors that are the least influenced by the noise factors and that provides the maximum pull-off force. Notice that although the noise factors are not controllable during routine operations, they can be controlled for the purposes of a test. Each controllable factor is tested at three levels, and each noise factor is tested at two levels.

Recall from the discussion in the textbook that in the Taguchi parameter design methodology, one experimental design is selected for the controllable factors and another experimental design is selected for the noise factors. These designs are shown in Table 2. Taguchi refers to these designs as **orthogonal arrays**, and represents the factor levels with integers 1, 2, and 3. In this case the designs selected are just a standard 2^3 and a 3^{4-2} fractional factorial. Taguchi calls these the L_8 and L_9 orthogonal arrays, respectively.

Table 1. Factors and Levels for the Taguchi Parameter Design Example

Controllable Factors		Levels		
A =	Interference	Low	Medium	High
B =	Connector wall thickness	Thin	Medium	Thick
C =	Insertion,depth	Shallow	Medium	Deep
D =	Percent adhesive in connector pre-dip	Low	Medium	High

Uncontrollable Factors	Levels	
E = Conditioning time	24 h	120 h
F = Conditioning temperature	72°F	150°F
G = Conditioning relative humidity	25%	75%

Table 2. Designs for the Controllable and Uncontrollable Factors

(a) L_9 Orthogonal Array for the Controllable Factors

Run	A	*B*	*C*	*D*
11	1	1	1	1
21	2	2	2	2
31	3	3	3	3
42	1	2	3	4
52	2	3	1	5
62	3	1	2	6
73	1	3	2	7
83	2	1	3	8
93	3	2	1	

(b) L_8 Orthogonal Array for the Uncontrollable Factors

Run	*E*	*F*	*E X F*	*G*	*Ex G*	*Fx G*	*e*
1	1	1	1	1	1	1	1
1	1	1	2	2	2	2	2
1	2	2	1	1	2	2	2
1	2	2	2	2	1	1	1
2	1	2	1	2	1	2	2
2	1	2	2	1	2	1	2
2	2	1	1	2	2	1	2
2	2	1	2	1	1	2	1

The two designs are combined as shown in Table 11-22 in the textbook, repeated for convenience as Table 3 below. Recall that this is called a **crossed** or **product array design**, composed of the **inner array** containing the controllable factors**,** and the **outer array** containing the noise factors**.** Literally, each of the 9 runs from the inner array is tested across the 8 runs from the outer array, for a total sample size of 72 runs. The observed pull-off force is reported in Table 3.

Data Analysis and Conclusions

The data from this experiment may now be analyzed. Recall from the discussion in Chapter 11 that Taguchi recommends analyzing the mean response for each run in the

inner array (see Table 3), and he also suggests analyzing variation using an appropriately chosen **signal-to-noise ratio (SN).** These signal-to-noise ratios are derived from the quadratic loss function, and three of them are considered to be "standard" and widely applicable. They are defined as follows:

1. Nominal the best:

$$SN_T = 10\log\left(\frac{\bar{y}^2}{S^2}\right)$$

2. Larger the better:

$$SN_L = -10\log\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^2}\right)$$

Table 3. Parameter Design with Both Inner and Outer Arrays

					Outer Array (L_8)									
					E	1	1	1	2	2	2	2		
					F	1	2	2	1	1	2	2		
					G	2	1	2	1	2	1	2		
	Inner Array (L_9)												Responses	
Run	A	B	C	D									$\bar{y}$	SN_L
1	1	1	1	1	15.6	9.5	16.9	19.9	19.6	19.6	20.0	19.1	17.525	24.025
2	1	2	2	2	15.0	16.2	19.4	19.2	19.7	19.8	24.2	21.9	19.475	25.522
3	1	3	3	3	16.3	16.7	19.1	15.6	22.6	18.2	23.3	20.4	19.025	25.335
4	2	1	2	3	18.3	17.4	18.9	18.6	21.0	18.9	23.2	24.7	20.125	25.904
5	2	2	3	1	19.7	18.6	19.4	25.1	25.6	21.4	27.5	25.3	22.825	26.908
6	2	3	1	2	16.2	16.3	20.0	19.8	14.7	19.6	22.5	24.7	19.225	25.326
7	3	1	3	2	16.4	19.1	18.4	23.6	16.8	18.6	24.3	21.6	19.8	25.711
8	3	2	t	3	14.2	15.6	15.1	16.8	17.8	19.6	23.2	24.2	18.338	24.852
9	3	3	2	1	16.1	19.9	19.3	17.3	23.1	22.7	22.6	28.6	21.200	26.152

3. Smaller the better:

$$SN_L = -10\log\left(\frac{1}{n}\sum_{i=1}^{n}y_i^2\right)$$

Notice that these *SN* ratios are expressed on a decibel scale. We would use SN_T if the objective is to reduce variability around a specific target, SN_L if the system is optimized when the response is as large as possible, and SN_S if the system is optimized when the response is as small as possible. Factor levels that maximize the appropriate *SN* ratio are optimal.

In this problem, we would use SN_L because the objective is to maximize the pull-off force. The last *two* columns of Table 3 contain $\bar{y}$ and SN_L values for each of the nine inner-array runs. Taguchi-oriented practitioners often use the analysis of variance to determine the factors that influence $\bar{y}$ and the factors that influence the signal-to-noise ratio. They also employ graphs of the "marginal means" of each factor, such as the ones shown in Figures 2 and 3. The usual approach is to examine the graphs and "pick the winner." In this case, factors A and C have larger effects than do B and D. In terms of maximizing SN_L we would select A_{Medium}, C_{Deep}, B_{Medium}, and D_{Low}. In terms of maximizing the average pull-off force $\bar{y}$, we would choose A_{Medium}, C_{Medium}, B_{Medium} and $D_{Low.}$ Notice that there is almost no difference between C_{Medium} and C_{Deep}. The implication is that this choice of levels will maximize the mean pull-off force and reduce variability in the pull-off force.

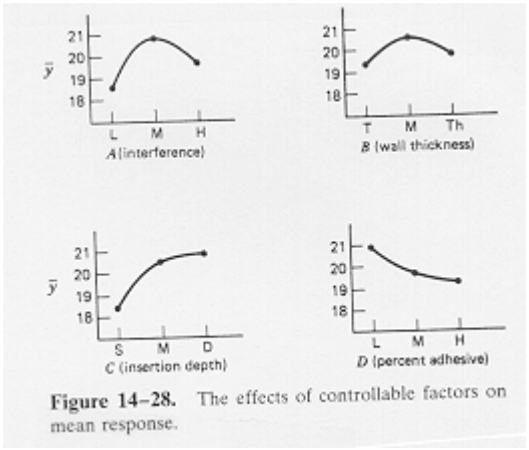


Figure 14–28. The effects of controllable factors on mean response.

Figure 2. The Effects of Controllable Factors on Each Response

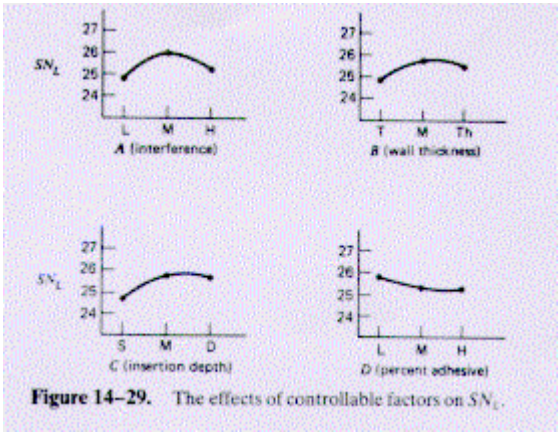


Figure 14–29. The effects of controllable factors on SN_L.

Figure 3. The Effects of Controllable Factors on the Signal to Noise Ratio

Taguchi advocates claim that the use of the *SN* ratio generally eliminates the need for examining specific interactions between the controllable and noise factors, although sometimes looking at these interactions improves process understanding. The authors of

this study found that the *AG* and *DE* interactions were large. Analysis of these interactions, shown in Figure 4, suggests that A_{Medium} is best. (It gives the highest pull-off force and a slope close to zero, indicating that if we choose A_{Medium} the effect of relative humidity is minimized.) The analysis also suggests that D_{Low} gives the highest pull-off force regardless of the conditioning time.

When cost and other factors were taken into account, the experimenters in this example finally decided to use A_{Medium}, B_{Thin}, C_{Medium}, and $D_{low.}$ (B_{Thin} was much less expensive than B_{Medium}, and C_{Medium} was felt to give slightly less variability than C_{Deep}.) Since this combination was not a run in the original nine inner array trials, five additional tests were made at this set of conditions as a confirmation experiment. For this confirmation experiment, the levels used on the noise variables were E_{Low}, F_{Low}, and G_{Low}. The authors report that good results were obtained from the confirmation test.

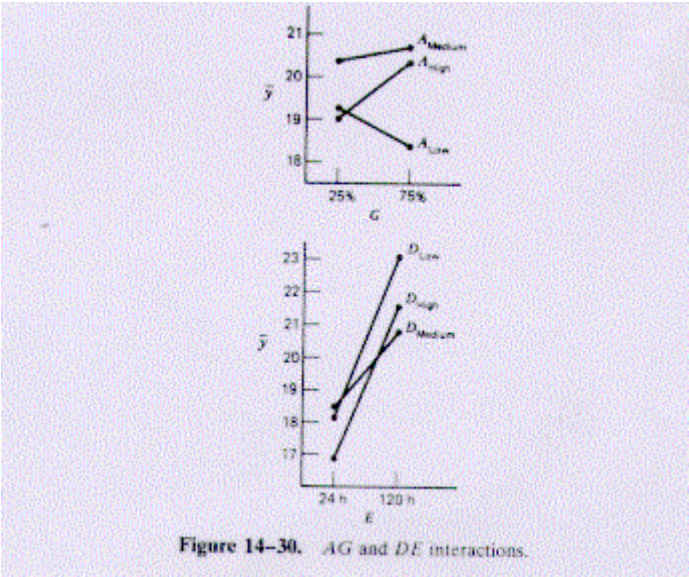


Figure 14–30. *AG* and *DE* interactions.

Figure 4. The *AG* and *DE* Interactions

Critique of Taguchi's Experimental Strategy and Designs

The advocates of Taguchi's approach to parameter design utilize the orthogonal array designs, two of which (the L_8 and the L_9) were presented in the foregoing example. There are other orthogonal arrays: the L_4, L_{12}, L_{16}, L_{18}, and L_{27}. These designs were not developed by Taguchi; for example, the L_8 is a 2_{III}^{7-4} fractional factorial, the L_9 is a 3_{III}^{4-2} fractional factorial, the L_{12} is a Plackett-Burman design, the L_{16} is a 2_{III}^{15-11} fractional factorial, and so on. Box, Bisgaard, and Fung (1988) trace the origin of these designs. As we know from Chapters 8 and 9 of the textbook, some of these designs have very complex alias structures. In particular, the L_{12} and all of the designs that use three-level factors will involve **partial aliasing** of two-factor interactions with main effects. If any two-factor interactions are large, this may lead to a situation in which the experimenter does not get the correct answer.

Taguchi argues that we do not need to consider two-factor interactions explicitly. He claims that it is possible to eliminate these interactions either by correctly specifying the response and design factors or by using a **sliding setting approach** to choose factor levels. As an example of the latter approach, consider the two factors pressure and temperature. Varying these factors independently will probably produce an interaction. However, if temperature levels are chosen contingent on the pressure levels, then the interaction effect can be minimized. In practice, these two approaches are usually difficult to implement unless we have an unusually high level of process knowledge. The lack of provision for adequately dealing with potential interactions between the controllable process factors is a major weakness of the Taguchi approach to parameter design.

Instead of designing the experiment to investigate potential interactions, Taguchi prefers to use three-level factors to estimate curvature. For example, in the inner and outer array design used by Byrne and Taguchi, all four controllable factors were run at three levels. Let x_1, x_2, x_3 and x_4 represent the controllable factors and let z_1, z_2, and z_3 represent the three noise factors. Recall that the noise factors were run at two levels in a complete factorial design. The design they used allows us to fit the following model:

$$y = \beta_0 + \sum_{j=1}^{4} \beta_j x_j + \sum_{j=1}^{4} \beta_{jj} x^2{}_j + \sum_{j=1}^{3} \gamma_j z_j + \sum_{i<j} \sum_{j=2}^{3} \gamma_{ij} z_i z_j + \sum_{i=1}^{3} \sum_{j=1}^{4} \delta_{ij} z_i x_j + \varepsilon$$

Notice that we can fit the linear and quadratic effects of the controllable factors but not their two-factor interactions (which are aliased with the main effects). We can also fit the linear effects of the noise factors and all the two-factor interactions involving the noise factors. Finally, we can fit the two-factor interactions involving the controllable factors and the noise factors. It may be unwise to ignore potential interactions in the controllable factors.

This is a rather odd strategy, since **interaction is a form of curvature**. A much safer strategy is to identify potential effects and interactions that may be important and then consider curvature only in the important variables if there is evidence that the curvature is important. This will usually lead to fewer experiments, simpler interpretation of the data, and better overall process understanding.

Another criticism of the Taguchi approach to parameter design is that the crossed array structure usually leads to a very large experiment. For example, in the foregoing application, the authors used 72 tests to investigate only seven factors, and they still could not estimate any of the two-factor interactions among the four controllable factors.

There are several alternative experimental designs that would be superior to the inner and outer method used in this example. Suppose that we run all seven factors at two levels in the combined **array design** approach discussed on the textbook. Consider the 2_{IV}^{7-2} fractional factorial design. The alias relationships for this design are shown in the top half of Table 4. Notice that this design requires only 32 runs (as compared to 72). In the bottom half of Table 4, two different possible schemes for assigning process controllable variables and noise variables to the letters A through G are given. The first assignment scheme allows all the interactions between controllable factors and noise

factors to be estimated, and it allows main effect estimates to be made that are clear of two-factor interactions. The second assignment scheme allows all the controllable factor main effects and their two-factor interactions to be estimated; it allows all noise factor main effects to be estimated clear of two-factor interactions; and it aliases only three interactions between controllable factors and noise factors with a two-factor interaction between two noise factors. Both of these arrangements present much cleaner alias relationships than are obtained from the inner and outer array parameter design, which also required over twice as many runs.

In general, the crossed array approach is often unnecessary. A better strategy is to use the **combined array design discussed in the textbook.** This approach will almost always lead to a dramatic reduction in the size of the experiment, and at the same time, it will produce information that is more likely to improve process understanding. For more discussion of this approach, see Myers and Montgomery (1995) and Example 11-6 in the textbook. We can also use a combined array design that allows the experimenter to directly model the noise factors as a complete quadratic and to fit all interactions between the controllable factors and the noise factors, as demonstrated in the textbook in Example 11-7.

Table 4. An Alternative Parameter Design

A one-quarter fraction of 7 factors in 32 runs. Resolution IV.
I = ABCDF = ABDEG = CEFG.

Aliases:

A	*AF = BCD*	*CG = EF*
B	*AG = BDE*	*DE = ABG*
C = EFG	*BC = ADF*	*DF = ABC*
D	*BD = ACF = AEG*	*DG = ABE*
E = CFG	*BE = ADG*	*ACE = AFG*
F = CEG	*BF = ACD*	*ACG = AEF*
G = CEF	*BG = ADE*	*BCE = BFG*
AB = CDF = DEG	*CD = ABF*	*BCG = BEF*
AC = BDF	*CE = FG*	*CDE = DFG*
AD = BCF = BEG	*CF = ABD = EG*	*CDG = DEF*
AF = BDG		

Factor Assignment Schemes:

1. Controllable factors are assigned to the letters *C, E, F,* and G. Noise factors are assigned to the letters *A, B,* and D. All interactions between controllable factors and noise factors can be estimated. and all controllable factor main effects can be estimated clear of two-factor interactions.
2. Controllable factors are assigned to the letters *A, B, C, and* D. Noise factors are assigned to the letters *E, F.* and G. All controllable factor main effects and two-factor interactions can be estimated; only the *CE, CF,* and CG interactions are aliased with interactions of the noise factors.

Another possible issue with the Taguchi inner and outer array design relates to the order in which the runs are performed. Now we know that for experimental validity, the runs in a designed experiment should be conducted in **random order**. However, in many crossed array experiments, it is possible that the run order wasn't randomized. In some cases it would be more convenient to fix each row in the inner array (that is, set the levels of the controllable factors) and run all outer-array trials. In other cases, it might be more convenient to fix the each column in the outer array and the run each on the inner array trials at that combination of noise factors. Exactly which strategy is pursued probably depends on which group of factors is easiest to change, the controllable factors or the

noise factors. If the tests are run in either manner described above, then a **split-plot structure** has been introduced into the experiment. If this is not accounted for in the analysis, then the results and conclusions can be misleading. There is no evidence that Taguchi advocates used split-plot analysis methods. Furthermore, since Taguchi frequently downplayed the importance of randomization, it is highly likely that many actual inner and outer array experiments were inadvertently conducted as split-plots, and perhaps incorrectly analyzed. We introduce the split-plot design in Chapter in Chapter 13. A good reference on split-plots in robust design problems is Box and Jones (1992).

A final aspect of Taguchi's parameter design is the use of **linear graphs** to assign factors to the columns of the orthogonal array. A set of linear graphs for the L_8 design is shown in Figure 5. In these graphs, each number represents a column in the design. A line segment on the graph corresponds to an interaction between the nodes it connects. To assign variables to columns in an orthogonal array, assign the variables to nodes first; then when the nodes are used up, assign the variables to the line segments. When you assign variables to the nodes, strike out any line segments that correspond to interactions that might be important. The linear graphs in Figure 5 imply that column 3 in the L_8 design contains the interaction between columns 1 and 2, column 5 contains the interaction between columns 1 and 4, and so forth. If we had four factors, we would assign them to columns 1, 2, 4, and 7. This would ensure that each main effect is clear of two-factor interactions. What is *not* clear is the two-factor interaction aliasing. If the main effects are in columns 1, 2, 4, and 7, then column 3 contains the 1-2 *and* the 4-7 interaction, column 5 contains the 1-4 *and* the 2-7 interaction, and column 6 contains the 1-7 *and* the 2-4 interaction. This is clearly the case because four variables in eight runs is a resolution IV plan with all pairs of two-factor interactions aliased. In order to understand fully the two-factor interaction aliasing, Taguchi would refer the experiment designer to a supplementary interaction table.

Taguchi (1986) gives a collection of linear graphs for each of his recommended orthogonal array designs. These linear graphs seem -to have been developed heuristically. Unfortunately, their use can lead to inefficient designs. For examples, see his car engine experiment [Taguchi and Wu (1980)] and his cutting tool experiment [Taguchi (1986)]. Both of these are 16-run designs that he sets up as resolution III designs in which main effects are aliased with two-factor interactions. Conventional methods for constructing these designs would have resulted in resolution IV plans in which the main effects are clear of the two-factor interactions. For the experimenter who simply wants to generate a good design, the linear graph approach may not produce the best result. A better approach is to use a simple table that presents the design and its full alias structure such as in Appendix Table XII. These tables are easy to construct and are routinely displayed by several widely available and inexpensive computer programs.

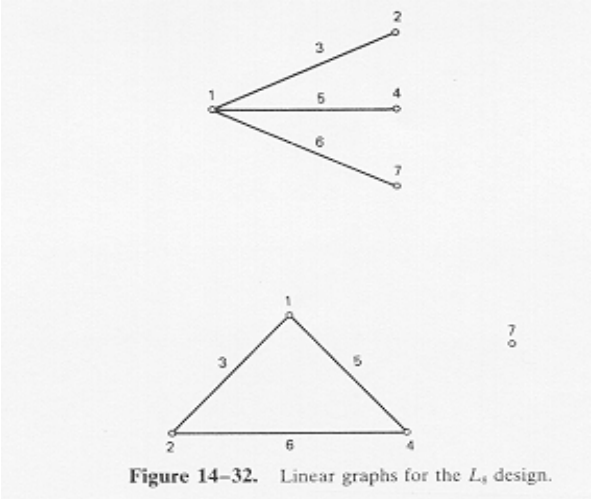


Figure 14–32. Linear graphs for the L_8 design.

Figure 5. Linear Graphs for the L_8 Design

Critique of Taguchi's Data Analysis Methods

Several of Taguchi's data analysis methods are questionable. For example, he recommends some variations of the analysis of variance that are known to produce spurious results, and he also proposes some unique methods for the analysis of attribute and life testing data. For a discussion and critique of these methods, refer to Box, Bisgaard, and Fung (1988), Myers and Montgomery (1995), and the references contained therein. In this section we focus on three aspects of his recommendations concerning data analysis: the use of "marginal means" plots to optimize factor settings, the use of signal-to-noise ratios, and some of his uses of the analysis of variance.

Consider the use of "marginal means" plots and the associated "pick the winner" optimization that was demonstrated previously in the pull-off force problem. To keep the situation simple, suppose that we have two factors *A* and *B*, each at three levels, as shown in Table 5. The "marginal means" plots are shown in Figure 6. From looking at these graphs, we would select A_3 and B_1, as the optimum combination, assuming that we wish to maximize *y*. However, this is the wrong answer. Direct inspection of Table 5 or the *AB* interaction plot in Figure 7 shows that the combination of A_3 and B_2 produces the maximum value of *y*. In general, playing "pick the winner" with marginal averages can never be guaranteed to produce the optimum. The Taguchi advocates recommend that a confirmation experiment be run, although this offers no guarantees either. We might be confirming a response that differs dramatically from the optimum. The best way to find a set of optimum conditions is with the use of response surface methods, as discussed and illustrated in Chapter 11 of the textbook.

Taguchi's signal-to-noise ratios are his recommended performance measures in a wide variety of situations. By maximizing the appropriate *SN* ratio, he claims that variability is minimized.

Table 5. Data for the "Marginal Means" Plots in Figure 6

		Factor A			
		1	2	3	B Averages
Factor B	1	10	10	13	11.00
	2	8	10	14	9.67
	3	6	9	10	8.33
	A Averages	8.00	9.67	11.67	

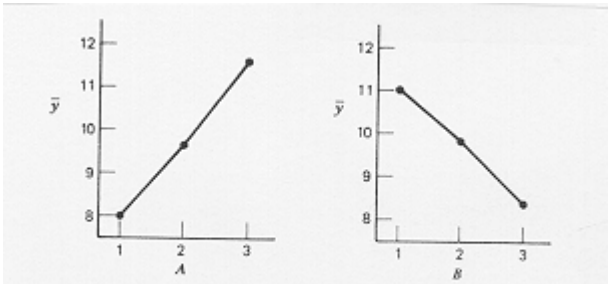


Figure 14–33. "Marginal means" plots for the data in Table 14–20.

Figure 6. Marginal Means Plots for the Data in Table 5

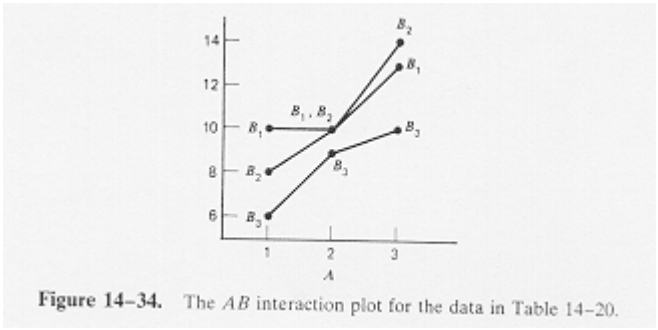


Figure 14–34. The *AB* interaction plot for the data in Table 14–20.

Figure 7. The *AB* Interaction Plot for the Data in Table 5.

Consider first the signal to noise ratio for the target is best case

$$SN_T = 10\log\left(\frac{\bar{y}^2}{S^2}\right)$$

This ratio would be used if we wish to minimize variability around a fixed target value. It has been suggested by Taguchi that it is preferable to work with SN_T instead of the standard deviation because in many cases the process mean and standard deviation are related. (As μ gets larger, σ gets larger, for example.) In such cases, he argues that we cannot directly minimize the standard deviation and then bring the mean on target.

Taguchi claims he found empirically that the use of the SN_T ratio coupled with a two-stage optimization procedure would lead to a combination of factor levels where the standard deviation is minimized and the mean is on target. The optimization procedure consists of (1) finding the **set** of controllable factors that affect SN_T, called the **control factors,** and setting them to levels that maximize SN_T and then (2) finding the set of factors that have significant effects on the mean but do not influence the SN_T ratio, called the **signal factors,** and using these factors to bring the mean on target.

Given that this partitioning of factors is possible, SN_T is an example of a **performance measure independent of adjustment** (PERMIA) [see Leon et al. (1987)]. The signal factors would be the **adjustment factors**. The motivation behind the signal-to-noise ratio is to uncouple location and dispersion effects. It can be shown that the use of SN_T is equivalent to an analysis of the standard deviation of the logarithm of the original data. Thus, using SN_T implies that a log transformation will *always* uncouple location and dispersion effects. There is no assurance that this will happen. A much safer approach is to investigate what type of transformation is appropriate.

Note that we can write the SN_T ratio as

$$SN_T = 10\log\left(\frac{\bar{y}^2}{S^2}\right)$$

$$= 10\log(\bar{y}^2) - 10\log(S^2)$$

If the mean is fixed at a target value (estimated by $\bar{y}$), then maximizing the SN_T ratio is equivalent to minimizing log (S^2). Using log (S^2) would require fewer calculations, is more intuitively appealing, and would provide a clearer understanding of the factor relationships that influence process variability - in other words, it would provide better process understanding. Furthermore, if we minimize log (S^2) directly, we eliminate the risk of obtaining wrong answers from the maximization of SN_T if some of the manipulated factors drive the mean $\bar{y}$ upward instead of driving S^2 downward. In general, if the response variable can be expressed in terms of the model

$$y = \mu(x_d, x_a)\varepsilon(x_d)$$

where x_d is the subset of factors that drive the dispersion effects and x_a is the subset of adjustment factors that do not affect variability, then maximizing SN_T will be equivalent to minimizing the standard deviation. Considering the other potential problems surrounding SN_T, it is likely to be safer to work directly with the standard deviation (or its logarithm) as a response variable, as suggested in the textbook. For more discussion, refer to Myers and Montgomery (1995).

The ratios SN_L and SN_S are even more troublesome. These quantities may be completely ineffective in identifying dispersion effects, although they may serve to identify **location effects,** that is, factors that drive the mean. The reason for this is relatively easy to see. Consider the SN_S (smaller-the-better) ratio:

$$SNS = -10\log\left(\frac{1}{n}\sum_{i=1}^{n} y_i^{\,2}\right)$$

The ratio is motivated by the assumption of a quadratic loss function with *y* nonnegative. The loss function for such a case would be

$$L = C\frac{1}{n}\sum_{i=1}^{n} y_i^{\,2}$$

where *C* is a constant. Now

$$\log L = \log C + \log\left(\frac{1}{n}\sum_{i=1}^{n} y_i^{\,2}\right)$$

and

$$SN_S = 10 \log C - 10 \log L$$

so maximizing *SNs* will minimize *L*. However, it is easy to show that

$$\frac{1}{n}\sum_{i=1}^{n} y_i^{\,2} = \bar{y}^{2} + \frac{1}{n}\left(\sum_{i=1}^{n} y_i^{\,2} - n\bar{y}^{2}\right)$$

$$= \bar{y}^{2} + \left(\frac{n-1}{n}\right)S^2$$

Therefore, the use of SN_S as a response variable confounds location and dispersion effects.

The confounding of location and dispersion effects was observed in the analysis of the SN_L ratio in the pull-off force example. In Figures 3 and 3 notice that the plots of $\bar{y}$ and SN_L versus each factor have approximately the same shape, implying that both responses measure location. Furthermore, since the SN_S and SN_L ratios involve y^2 and $1/y^2$, they will be very sensitive to outliers or values near zero, and they are not invariant to linear transformation of the original response. We strongly recommend that these signal-to-noise ratios not be used.

A better approach for isolating location and dispersion effects is to develop separate response surface models for $\bar{y}$ and $\log(S^2)$. If no replication is available to estimate variability at each run in the design, methods for analyzing residuals can be used. Another very effective approach is based on the use of the **response model**, as demonstrated in the textbook and in Myers and Montgomery (1995). Recall that this allows both a response surface for the variance and a response surface for the mean to be obtained for a single model containing both the controllable design factors and the noise variables. Then standard response surface methods can be used to optimize the mean and variance.

Finally, we turn to some of the applications of the analysis of variance recommended by Taguchi. As an example for discussion, consider the experiment reported by Quinlan (1985) at a symposium on Taguchi methods sponsored by the American Supplier Institute. The experiment concerned the quality improvement of speedometer cables. Specifically, the objective was to reduce the shrinkage in the plastic casing material. (Excessive shrinkage causes the cables to be noisy.) The experiment used an L_{16} orthogonal array (the 2_{III}^{15-11} design). The shrinkage values for four samples taken from 3000-foot lengths of the product manufactured at each set of test conditions were measured and the responses $\bar{y}$ and SN_{Sheila} computed.

Quinlan, following the Taguchi approach to data analysis, used SN_S as the response variable in an analysis of variance. The error mean square was formed by pooling the mean squares associated with the seven effects that had the smallest absolute magnitude. This resulted in all eight remaining factors having significant effects (in order of magnitude: *E, G, K, A, C, F, D, H).* The author did note that *E* and *G* were the most important.

Pooling of mean squares as in this example is a procedure that has long been known to produce considerable bias in the ANOVA test results, To illustrate the problem, consider the 15 NID(0, 1) random numbers shown in column 1 of Table 6. The square of each of these numbers, shown in column 2 of the table, is a single-degree-of-freedom mean square corresponding to the observed random number. The seven smallest random numbers are marked with an asterisk in column 1 of Table 6. The corresponding mean squares are pooled to form a mean square for error with seven degrees of freedom. This quantity is

$$MS_E = \frac{0.5088}{7} = 0.0727$$

Finally, column 3 of Table 6 presents the *F* ratio formed by dividing each of the eight remaining mean squares by MS_E. Now $F_{0.05,1,7} = 5.59$, and this implies that five of the eight effects would be judged significant at the 0.05 level. Recall that since the original data came from a normal distribution with mean zero, *none* of the effects is different from zero.

Analysis methods such as this virtually guarantee erroneous conclusions. The normal probability plotting of effects avoids this invalid pooling of mean squares and provides a simple, easy to interpret method of analysis. Box (1988) provides an alternate analysis of

Table 6. Pooling of Mean Squares

NID(0,1) Random Numbers	Mean Squares with One Degree of Freedom	F_0
-08607	0.7408	10.19
-0.8820	0.7779	10.70
0.3608*	0.1302	
0.0227*	0.0005	
0.1903*	0.0362	
-0.3071*	0.0943	
1.2075	1.4581	20.06
0.5641	0.3182	4038
-0.3936*	0.1549	
-0.6940	0.4816	6.63
-0.3028*	0.0917	
0.5832	0.3401	4.68
0.0324*	0.0010	
1.0202	1.0408	14.32
-0.6347	0.4028	5.54

the Quinlan data that correctly reveals *E* and *G* to be important along with other interesting results not apparent in the original analysis.

It is important to note that the Taguchi analysis identified negligible factors as significant. This can have profound impact on our use of experimental design to enhance process knowledge. Experimental design methods should make gaining process knowledge easier, not harder.

Some Final Remarks

In this section we have directed some major criticisms toward the specific methods of experimental design and data analysis used in the Taguchi approach to parameter design. Remember that these comments have focused on technical issues, and that the broad **philosophy** recommended by Taguchi is inherently sound.

On the other hand, while the “Taguchi controversy” was in full bloom, many companies reported success with the use of Taguchi's parameter design methods. If the methods are flawed, why do they produce successful results? Taguchi advocates often refute criticism with the remark that "they work." We must remember that the "best guess" and "one-

factor-at-a-time" methods will also work-and occasionally they produce good results. This is no reason to claim that they are good methods. Most of the successful applications of Taguchi's technical methods have been in industries where there was no history of good experimental design practice. Designers and developers were using the **best guess** and **one-factor-at-a-time methods** (or other unstructured approaches), and since the Taguchi approach is based on the factorial design concept, it often produced better results than the methods it replaced. In other words, the factorial design is so powerful that, even when it is used inefficiently, it will often work well.

As pointed out earlier, the Taguchi approach to parameter design often leads to **large, comprehensive experiments**, often having 70 or more runs. Many of the successful applications of this approach were in industries characterized by a high-volume, low-cost manufacturing environment. In such situations, large designs may not be a real problem, if it is really no more difficult to make 72 runs than to make 16 or 32 runs. On the other hand, in industries characterized by low-volume and/or high-cost manufacturing (such as the aerospace industry, chemical and process industries, electronics and semiconductor manufacturing, and so forth), these methodological inefficiencies can be significant.

A final point concerns the learning process. If the Taguchi approach to parameter design works and yields good results, we may still not know what has caused the result because of the aliasing of critical interactions. In other words, we may have solved a problem (a short-term success), but we may not have gained **process knowledge,** which could be invaluable in future problems.

In summary, we should support Taguchi's **philosophy** of quality engineering. However, we must rely on simpler, more efficient methods that are easier to learn and apply to carry this philosophy into practice. The response surface modeling framework that we present in the textbook is an ideal approach to process optimization and as we have demonstrated, it is fully adaptable to the robust parameter design problem.

Supplemental References

Leon, R. V., A. C. Shoemaker and R. N. Kackar (1987). “Performance Measures Independent of Adjustment”. *Technometrics*, Vol. 29, pp. 253-265

Quinlan, J. (1985). “Product Improvement by Application of Taguchi Methods”. *Third Supplier Symposium on Taguchi Methods*, American Supplier Institute, Inc., Dearborn, MI.

Box, G. E. P. and S. Jones (1992). “Split-Plot Designs for Robust Product Experimentation”. *Journal of Applied Statistics*, Vol. 19, pp. 3-26.

Chapter 13. Supplemental Text Material

S13-1. Expected Mean Squares for the Random Model

We consider the two-factor random effects balanced ANOVA model

$$y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \begin{cases} i = 1,2,\cdots,a \\ j = 1,2,\cdots,b \\ k = 1,2,\cdots,n \end{cases}$$

given as Equation (13-15) in the textbook. We list the expected mean squares for this model in Equation (13-17), but do not formally develop them. It is relatively easy to develop the expected mean squares from direct application of the expectation operator.

For example, consider finding

$$E(MS_A) = E\left(\frac{SS_A}{a-1}\right) = \frac{1}{a-1}E(SS_A)$$

where SS_A is the sum of squares for the row factor. Recall that the model components τ_i, β_j and $(\tau\beta)_{ij}$ are normally and independently distributed with means zero and variances $\sigma_\tau^2, \sigma_\beta^2,$ and $\sigma_{\tau\beta}^2$ respectively. The sum of squares and its expectation are defined as

$$SS_A = \frac{1}{bn}\sum_{i=1}^{a} y_{i..}^2 - \frac{y_{...}^2}{abn}$$

$$E(SS_A) = \frac{1}{bn}E\sum_{i=1}^{a} y_{i..}^2 - E\left(\frac{y_{...}^2}{abn}\right)$$

Now

$$y_{i..} = \sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} = bn\mu + bn\tau_i + n\beta_{.} + n(\tau\beta)_{i.} + \varepsilon_{i..}$$

and

$$\begin{aligned} \frac{1}{bn}E\sum_{i=1}^{a} y_{i..}^2 &= \frac{1}{bn}E\sum_{i=1}^{a}\left[(bn\mu)^2 + (bn)^2\tau_i^2 + \varepsilon_{i..}^2 + 2(bn)^2\mu\tau_i + 2bn\mu\varepsilon_{i..} + 2bn\tau_i\varepsilon_{i..}\right] \\ &= \frac{1}{bn}\left[a(bn\mu)^2 + a(bn)^2\sigma_\tau^2 + ab(n)^2\sigma_\beta^2 + abn^2\sigma_{\tau\beta}^2 + abn\sigma^2\right] \\ &= abn\mu^2 + abn\sigma_\tau^2 + an\sigma_\beta^2 + an\sigma_{\tau\beta}^2 + a\sigma^2 \end{aligned}$$

Furthermore, we can show that

$$y_{...} = abn\mu + bn\tau_{.} + an\beta_{.} + n(\tau\beta)_{..} + \varepsilon_{...}$$

so the second term in the expected value of SS_A becomes

$$\frac{1}{abn}E(y_{...}^2) = \frac{1}{abn}\left[(abn\mu)^2 + a(bn)^2\sigma_\tau^2 + b(an)^2\sigma_\beta^2 + abn^2\sigma_{\tau\beta}^2 + abn\sigma^2\right]$$
$$= abn\mu^2 + bn\sigma_\tau^2 + an\sigma_\beta^2 + n\sigma_{\tau\beta}^2 + \sigma^2$$

We can now collect the components of the expected value of the sum of squares for factor *A* and find the expected mean square as follows:

$$E(MS_A) = E\left(\frac{SS_A}{a-1}\right)$$
$$\frac{1}{a-1}\left[\frac{1}{bn}E\sum_{i=1}^{a}y_{i..}^2 - E\left(\frac{y_{...}^2}{abn}\right)\right]$$
$$= \frac{1}{a-1}\left[\sigma^2(a-1) + n(a-1)\sigma_{\tau\beta}^2 + bn\sigma_\tau^2\right]$$
$$= \sigma^2 + n\sigma_{\tau\beta}^2 + bn\sigma_\tau^2$$

This agrees with the first result in Equation (15-17).

S13-2. Expected Mean Squares for the Mixed Model

As noted in Section 13-3 of the textbook, there are several version of the mixed model, and the expected mean squares depend on which model assumptions are used. In this section, we assume that the **restricted model** is of interest. The next section considers the unrestricted model.

Recall that in the restricted model there are assumptions made regarding the fixed factor, *A*; namely,

$$\tau_{.} = 0, (\tau\beta)_{.j} = 0 \text{ and } V[(\tau\beta)_{ij}] = \left[\frac{a}{a-1}\right]\sigma_{\tau\beta}^2$$

We will find the expected mean square for the random factor, *B*. Now

$$E\left(MS_B\right) = E\left(\frac{SS_B}{b-1}\right)$$
$$= \frac{1}{b-1}E\left(SS_B\right)$$

and

$$E(SS_B) = \frac{1}{an}E\sum_{j=1}^{b}y_{.j.}^2 - \frac{1}{abn}E(y_{...}^2)$$

Using the restrictions on the model parameters, we can easily show that

$$y_{.j.} = an\mu + an\beta_j + \varepsilon_{.j.}$$

and

$$\frac{1}{an}E\sum_{j=1}^{b}y_{.j.}^2=\left[b(an\mu)^2+b(an)^2\sigma_\beta^2+abn\sigma^2\right]$$
$$=abn\mu^2+abn\sigma_\beta^2+b\sigma^2$$

Since

$$y_{...}=abn\mu+an\beta_{.}+\varepsilon_{...}$$

we can easily show that

$$\frac{1}{abn}E(y_{...}^2)=\frac{1}{abn}\left[(abn\mu)^2+b(an)^2\sigma_\beta^2+abn\sigma^2\right]$$
$$=abn\mu^2+an\sigma_\beta^2+\sigma^2$$

Therefore the expected value of the mean square for the random effect is

$$E\left(MS_B\right)=\frac{1}{b-1}E\left(SS_B\right)$$
$$=\frac{1}{b-1}\left(abn\mu^2+abn\sigma_\beta^2+b\sigma^2-abn\mu^2-an\sigma_\beta^2-\sigma^2\right)$$
$$=\frac{1}{b-1}\left[\sigma^2(b-1)+an(b-1)\sigma_\beta^2\right]$$
$$=\sigma^2+an\sigma_\beta^2$$

The other expected mean squares can be derived similarly.

S13-3. Restricted versus Unrestricted Mixed Models

We now consider the **unrestricted model**

$$y_{ij}=\mu+\alpha_i+\gamma_j+(\alpha\gamma)_{ij}+\varepsilon_{ijk}\begin{cases}i=1,2,\cdots,a\\j=1,2,\cdots,b\\k=1,2,\cdots,n\end{cases}$$

for which the assumptions are

$$\alpha_{.}=0\ \text{ and }V[(\alpha\gamma)_{ij}]=\sigma_{\alpha\gamma}^2$$

and all random effects are uncorrelated random variables. Notice that there is no assumption concerning the interaction effects summed over the levels of the fixed factor as is customarily made for the restricted model. Recall that the restricted model is actually a more general model than the unrestricted model, but some modern computer programs give the user a choice of models (and some computer programs only use the unrestricted model), so there is increasing interest in both versions of the mixed model.

We will derive the expected value of the mean square for the random factor, *B*, in Equation (13-26), as it is different from the corresponding expected mean square in the restricted model case. As we will see, the assumptions regarding the interaction effects are instrumental in the difference in the two expected mean squares.

The expected mean square for the random factor, *B*, is defined as

$$E\left(MS_B\right) = E\left(\frac{SS_B}{b-1}\right)$$
$$= \frac{1}{b-1}E\left(SS_B\right)$$

and, as in the cases above

$$E(SS_B) = \frac{1}{an}E\sum_{j=1}^{b} y_{.j.}^2 - \frac{1}{abn}E(y_{...}^2)$$

First consider

$$y_{.j.} = \sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk} = an\mu + n\alpha_{.} + an\gamma_{.} + n(\alpha\gamma)_{.j} + \varepsilon_{.j.}$$
$$= an\mu + an\gamma_{.} + n(\alpha\gamma)_{.j} + \varepsilon_{.j.}$$

because $\alpha_{.} = 0$. Notice, however, that the interaction term in this expression is *not* zero as it would be in the case of the restricted model. Now the expected value of the first part of the expression for $E(SS_B)$ is

$$\frac{1}{an}E\sum_{j=1}^{b} y_{.j.}^2 = \frac{1}{an}\left[b(an\mu)^2 + b(an)^2\sigma_{\gamma}^2 + abn^2\sigma_{\alpha\gamma}^2 + abn\sigma^2\right]$$
$$= abn\mu^2 + abn\sigma_{\gamma}^2 + bn\sigma_{\alpha\gamma}^2 + b\sigma^2$$

Now we can show that

$$y_{...} = abn\mu + bn\alpha_{.} + an\gamma_{.} + n(\alpha\gamma)_{..} + \varepsilon_{...}$$
$$= abn\mu + an\gamma_{.} + n(\alpha\gamma)_{..} + \varepsilon_{...}$$

Therefore

$$\frac{1}{abn}E(y_{...}^2) = \frac{1}{abn}\left[(abn\mu)^2 + b(an)^2\sigma_{\gamma}^2 + abn^2\sigma_{\alpha\gamma}^2 + abn\sigma^2\right]$$
$$= abn\mu^2 + an\sigma_{\gamma}^2 + n\sigma_{\alpha\gamma}^2 + \sigma^2$$

We may now assemble the components of the expected value of the sum of squares for factor *B* and find the expected value of MS_B as follows:

$$
\begin{aligned}
E(MS_B) &= \frac{1}{b-1}E(SS_B) \\
&= \frac{1}{b-1}\left[\frac{1}{an}E\sum_{j=1}^{b}y_{.j.}^2 - \frac{1}{abn}E(y_{...}^2)\right] \\
&= \frac{1}{b-1}\left[abn\mu^2 + abn\sigma_\gamma^2 + bn\sigma_{\alpha\gamma}^2 + b\sigma^2 - \left(abn\mu^2 + an\sigma_\gamma^2 + n\sigma_{\alpha\gamma}^2 + \sigma^2\right)\right] \\
&= \frac{1}{b-1}[\sigma^2(b-1) + n(b-1)\sigma_{\alpha\gamma}^2 + an(b-1)\sigma_\gamma^2] \\
&= \sigma^2 + n\sigma_{\alpha\gamma}^2 + an\sigma_\gamma^2
\end{aligned}
$$

This last expression is in agreement with the result given in Equation (13-26).

Deriving expected mean squares by the direct application of the expectation operator (the "brute force" method) is tedious, and the rules given in the text are a great labor-saving convenience. There are other rules and techniques for deriving expected mean squares, including algorithms that will work for unbalanced designs. See Milliken and Johnson (1984) for a good discussion of some of these procedures.

S13-4. Random and Mixed Models with Unequal Sample Sizes

Generally, ANOVA models become more complicated to analyze when the designs is unbalanced; that is, when some of the cells contain different numbers of observations. In Chapter 15, we briefly discuss this problem in the two-factor fixed-effects design. The unbalanced case of random and mixed models is not discussed there, but we offer some very brief advice in this section.

An unbalanced random or mixed model will not usually have exact *F*-tests as they did in the balanced case. Furthermore, the Satterthwaite approximate or synthetic *F*-test does not apply to unbalanced designs. The simplest approach to the analysis is based on the method of maximum likelihood. This approach to variance component estimation was discussed in Section 13-7.3, and the SAS procedure employed there can be used for unbalanced designs. The disadvantage of this approach is that all the inference on variance components is based on the maximum likelihood large sample theory, which is only an approximation because designed experiments typically do not have a large number of runs. The book by Searle (1987) is a good reference on this general topic.

S13-5. Some Background Concerning the Modified Large Sample Method

In Section 12-7.2 we discuss the modified large same method for determining a confidence interval on variance components that can be expressed as a linear combination of mean squares. The large sample theory essentially states that

$$
Z = \frac{\hat{\sigma}_0^2 - \sigma_0^2}{\sqrt{V(\hat{\sigma}_0^2)}}
$$

has a normal distribution with mean zero and variance unity as $\min(f_1, f_2, \cdots, f_Q)$ approaches infinity, where

$$V(\hat{\sigma}_0^2) = 2\sum_{i=1}^{Q} c_i^2 \theta_i^2 \,/\, f_i \,,$$

θ_i is the linear combination of variance components estimated by the *i*th mean square, and f_i is the number of degrees of freedom for MS_i. Consequently, the 100(1-α) percent large-sample two-sided confidence interval for σ_0^2 is

$$\hat{\sigma}_0^2 - z_{\alpha/2}\sqrt{V(\hat{\sigma}_0^2)} \le \sigma_0^2 \le \hat{\sigma}_0^2 + z_{\alpha/2}\sqrt{V\hat{\sigma}_0^2)}$$

Operationally, we would replace θ_i by MS_i in actually computing the confidence interval. This is the same basis used for construction of the confidence intervals by SAS PROC MIXED that we presented in section 13-7.3 (refer to the discussion of tables 13-17 and 13-18 in the textbook).

These large-sample intervals work well when the number of degrees of freedom are large, but when the f_i are small they may be unreliable. However, the performance may be improved by applying suitable modifications to the procedure. Welch (1956) suggested a modification to the large-sample method that resulted in some improvement, but Graybill and Wang (1980) proposed a technique that makes the confidence interval exact for certain special cases. It turns out that it is also a very good approximate procedure for the cases where it is not an exact confidence interval. Their result is given in the textbook as Equation (13-42).

S13-6. A Confidence Interval on a Ratio of Variance Components using the Modified Large Sample Method

As observed in the textbook, the modified large sample method can be used to determine confidence intervals on ratios of variance components. Such confidence intervals are often of interest in practice. For example, consider the measurement systems capability study described in Example 12-2 in the textbook. In this experiment, the total variability from the gauge is the sum of three variance components $\sigma_\beta^2 + \sigma_{\tau\beta}^2 + \sigma^2$, and the variability of the product used in the experiment is σ_τ^2. One way to describe the capability of the measurement system is to present the variability of the gauge as a percent of the product variability. Therefore, an experimenter would be interested in the ratio of variance components

$$\frac{\sigma_\beta^2 + \sigma_{\tau\beta}^2 + \sigma^2}{\sigma_\tau^2}$$

Suppose that σ_1^2 / σ_2^2 is a ratio of variance components of interest and that we can estimate the variances in the ratio by the ratio of two linear combinations of mean squares, say

$$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{\sum_{i=1}^{P} c_i MS_i}{\sum_{j=P+1}^{Q} c_j MS_j}$$

Then a 100(1-α) percent lower confidence in interval on σ_1^2 / σ_2^2 is given by

$$L = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_1^2}\left[\frac{2 + k_4 / (k_1 k_2) - \sqrt{V_L}}{2(1 - k_5 / k_2^2)}\right]$$

where

$$V_L = (2 + k_4 / (k_1 k_2))^2 - 4(1 - k_5 / k_2^2)(1 - k_3 / k_1^2)$$

$$k_1 = \sum_{i=1}^{P} c_i MS_i, \qquad k_2 = \sum_{j=P+1}^{Q} c_j MS_j$$

$$k_3 = \sum_{i=1}^{P} G_i^2 c_i^2 MS_i^2 + \sum_{i=1}^{P-1} \sum_{t>i}^{P} G_{it}^* c_i c_t MS_i MS_t$$

$$k_4 = \sum_{i=1}^{P} \sum_{j=P+1}^{Q} G_{ij} c_1 c_j MS_i MS_j$$

$$k_5 = \sum_{j=P+1}^{Q} H_j^2 c_j^2 MS_j^2$$

and G_i, H_j, G_{ig}, and G_{it}^* are as previously defined. For more details, see the book by Burdick and Graybill (1992).

Supplemental References

Graybill, F. A. and C. M. Wang (1980. "Confidence Intervals on Nonnegative Linear Combinations of Variances". *Journal of the American Statistical Association*, Vol. 75, pp. 869-873.

Welch, B. L. (1956). "On Linear Combinations of Several Variances". *Journal of the American Statistical Association*, Vol. 51, pp. 132-148.

Chapter 14. Supplemental Text Material

S14-1. The Staggered, Nested Design

In Section 14-1.4 we introduced the staggered, nested design as a useful way to prevent the number of degrees of freedom from "building up" so rapidly at lower levels of the design. In general, these designs are just unbalanced nested designs, and many computer software packages that have the capability to analyze general unbalanced designs can successfully analyze the staggered, nested design. The general linear model routine in Minitab is one of these packages.

To illustrate a staggered, nested design, suppose that a pharmaceutical manufacturer is interested in testing the absorption of a drug two hours after the tablet is ingested. The product is manufactured in lots, and specific interest focuses on determining whether there is any significant lot-to-lot variability. Excessive lot-to-lot variability probably indicates problems with the manufacturing process, perhaps at the stage where the coating material that controls tablet absorption is applied. It could also indicate a problem with either the coating formulation, or with other formulation aspects of the tablet itself.

The experimenters select $a = 10$ lots at random from the production process, and decide to use a staggered, nested design to sample from the lots. Two samples are taken at random from each lot. The first sample contains two tablets, and the second sample contains only one tablet. Each tablet is test for the percentage of active drug absorbed after two hours. The data from this experiment is shown in Table 1 below.

Table 1. The Drug Absorption Experiment

	Sample	
Lot	1	2
1	24.5, 25.9	23.9
2	23.6, 26.1	25.2
3	27.3, 28.1	27.0
4	28.3, 27.5	27.4
5	24.3, 24.1	25.1
6	25.3, 26.0	24.7
7	27.3, 26.8	28.0
8	23.3, 23.9	23.0
9	24.6, 25.1	24.9
10	24.3, 24.9	25.3

The following output is from the Minitab general linear model analysis procedure.

General Linear Model

```
Factor        Type Levels Values
Lot         random     10  1  2  3  4  5  6  7  8  9 10
Sample(Lot) random     20 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2

Analysis of Variance for Absorp., using Adjusted SS for Tests

Source        DF     Seq SS     Adj SS     Adj MS       F      P
Lot            9    58.3203    52.3593     5.8177   14.50  0.000
Sample(Lot)   10     4.0133     4.0133     0.4013    0.71  0.698
Error         10     5.6200     5.6200     0.5620
Total         29    67.9537

Expected Mean Squares, using Adjusted SS

Source          Expected Mean Square for Each Term
 1 Lot          (3) +  1.3333(2) +  2.6667(1)
 2 Sample(Lot)  (3) +  1.3333(2)
 3 Error        (3)

Error Terms for Tests, using Adjusted SS

Source          Error DF  Error MS  Synthesis of Error MS
 1 Lot             10.00    0.4013  (2)
 2 Sample(Lot)     10.00    0.5620  (3)

Variance Components, using Adjusted SS

Source       Estimated Value
Lot                   2.0311
Sample(Lot)          -0.1205
Error                 0.5620
```

As noted in the textbook, this design results in $a - 1 = 9$ degrees of freedom for lots, and $a = 10$ degrees of freedom for samples within lots and error. The ANOVA indicates that there is a significant difference between lots, and the estimate of the variance component for lots is $\hat{\sigma}^2_{Lots} = 2.03$. The ANOVA indicates that the sample within lots is not a significant source of variability. This is an indication of lot homogeneity. There is a small negative estimate of the sample-within-lots variance component. The experimental error variance is estimated as $\hat{\sigma}^2 = 0.526$. Notice that the constants in the expected mean squares are not integers; this is a consequence of the unbalanced nature of the design.

S14-2. Inadvertent Split-Plots

In recent years experimenters from many different industrial settings have become exposed to the concepts of designed experiments, either from university-level DOX courses or from industrial short courses and seminars. As a result, factorial and fractional factorial designs have enjoyed expanded use. Sometimes the principle of randomization is not sufficiently stressed in these courses, and as a result experimenters may fail to understand its importance. This can lead to **inadvertent split-plotting** of a factorial design.

For example, suppose that an experimenter wishes to conduct a 2^4 factorial using the factors A = temperature, B = feed rate, C = concentration, and D = reaction time. A 2^4 with the runs arranged in random order is shown in Table 2.

Table 2. A 2^4 Design in Random Order

Std	Run	Block	Factor A: Temperature DegC	Factor B: Feed rate gal/h	Factor C: Concentration gm/l	Factor D: Reaction time h	Response Yield
16	1	Block 1	150	8	30	1.2	
9	2	Block 1	100	5	25	1.2	
7	3	Block 1	100	8	30	1	
12	4	Block 1	150	8	25	1.2	
2	5	Block 1	150	5	25	1	
13	6	Block 1	100	5	30	1.2	
1	7	Block 1	100	5	25	1	
10	8	Block 1	150	5	25	1.2	
3	9	Block 1	100	8	25	1	
14	10	Block 1	150	5	30	1.2	
6	11	Block 1	150	5	30	1	
4	12	Block 1	150	8	25	1	
5	13	Block 1	100	5	30	1	
15	14	Block 1	100	8	30	1.2	
11	15	Block 1	100	8	25	1.2	
8	16	Block 1	150	8	30	1	

When the experimenter examines this run order, he notices that the level of temperature is going to start at 150 degrees and then be changed eight times over the course of the 16 trials. Now temperature is a hard-to-change-variable, and following every adjustment to temperature several hours are needed for the process to reach the new temperature level and for the process to stabilize at the new operating conditions.

The experimenter may feel that this is an intolerable situation. Consequently, he may decide that fewer changes in temperature are required, and rearange the temperature levels in the experiment so that the new design appears as in Table 3. Notice that only three changes in the level of temperature are required in this new design. In efect, the experimenter will set the temperature at 150 degrees and perform four runs with the other three factors tested in random order. Then he will change the temperature to 100 degrees

and repeat the process, and so on. The experimenter has inadvertently introduced a split-plot structure into the experiment.

Table 3. The Modified 2^4 Factorial

Std	Run	Block	Factor A: Temperature DegC	Factor B: Feed rate gal/h	Factor C: Concentration gm/l	Factor D: Reaction time h	Response Yield
16	1	Block 1	150	8	30	1.2	
9	2	Block 1	150	5	25	1.2	
7	3	Block 1	150	8	30	1	
12	4	Block 1	150	8	25	1.2	
2	5	Block 1	100	5	25	1	
13	6	Block 1	100	5	30	1.2	
1	7	Block 1	100	5	25	1	
10	8	Block 1	100	5	25	1.2	
3	9	Block 1	150	8	25	1	
14	10	Block 1	150	5	30	1.2	
6	11	Block 1	150	5	30	1	
4	12	Block 1	150	8	25	1	
5	13	Block 1	100	5	30	1	
15	14	Block 1	100	8	30	1.2	
11	15	Block 1	100	8	25	1.2	
8	16	Block 1	100	8	30	1	

Typically, most inadvertent split-plotting is not taken into account in the analysis. That is, the experimenter analyzes the data as if the experiment had been conducted in random order. Therefore, it is logical to ask about the impact of ignoring the inadvertent split-plotting. While this question has not been studied in detail, generally inadvertently running a split-plot and not properly accounting for it in the analysis probably does not have major impact **so long as the whole plot factor effects are large**. These factor effect estimates will probably have larger variances that the factor effects in the subplots, so part of the risk is that small differences in the whole-plot factors may not be detected. Obviously, the more systematic fashion in which the whole-plot factor temperature was varied in Table 2 also exposes the experimenter to confounding of temperature with some nuisance variable that is also changing with time. The most extreme case of this would occur if the first eight runs in the experiment were made with temperature at the low level (say), followed by the last eight runs with temperature at the high level.

Chapter 15. Supplemental Text Material

S15-1. The Form of a Transformation

In Section 3-4.3 of the textbook we introduce transformations as a way to stabilize the variance of a response and to (hopefully) induce approximate normality when inequality of variance and nonnormality occur jointly (as they often do). In Section 15-1.1 of the book the Box-Cox method is presented as an elegant analytical method for selecting the form of a transformation. However, many experimenters select transformations empirically by trying some of the simple power family transformations in Table 3-9 of Chapter 3 ($\sqrt{y}, \ln(y),$ or $1/y$, for example) or which appear on the menu of their computer software package.

It is possible to give a theoretical justification of the power family transformations presented in Table 3-9. For example, suppose that y is a response variable with mean $E(y) = \mu$ and variance $V(y) = \sigma^2 = f(\mu)$. That is, the variance of y is a function of the mean. We wish to find a transformation $h(y)$ so that the variance of the transformed variable $x = h(y)$ is a constant unrelated to the mean of y. In other words, we want $V[h(y)]$ to be a constant that is unrelated to $E[h(y)]$.

Expand $x = h(y)$ in a Taylor series about μ, resulting in

$$\begin{aligned} x &= h(y) \\ &= h(\mu) + h'(\mu)(y-\mu) + R \\ &\cong h(\mu) + h'(\mu)(y-\mu) \end{aligned}$$

where R is the remainder in the first-order Taylor series, and we have ignored the remainder. Now the mean of x is

$$\begin{aligned} E(x) &= E[h(\mu) + h'(\mu)(y-\mu)] \\ &= h(\mu) \end{aligned}$$

and the variance of x is

$$\begin{aligned} V(x) &= E\left[x - E(x)\right]^2 \\ &= E\left[h(\mu) + h'(\mu)(y-\mu) - h(\mu)\right]^2 \\ &= E\left[h'(\mu)(y-\mu)\right]^2 \\ &= \sigma^2\left[h'(\mu)\right]^2 \end{aligned}$$

Since $\sigma^2 = f(\mu)$, we have

$$V(x) = f(\mu)\left[h'(\mu)\right]^2$$

We want the variance of x to be a constant, say c^2. So set

$$c^2 = f(\mu)\left[h'(\mu)\right]^2$$

and solve for $h'(y)$, giving

$$h'(\mu) = \frac{c}{\sqrt{f(\mu)}}$$

Thus, the form of the transformation that is required is

$$\begin{aligned} h(\mu) &= c\int \frac{dt}{\sqrt{f(t)}} \\ &= cG(\mu) + k \end{aligned}$$

where k is a constant.

As an example, suppose that for the response variable y we assumed that the mean and variance were equal. This actually happens in the Poisson distribution. Therefore,

$$\mu = \sigma^2 \text{ implying that } f(t) = t$$

So

$$\begin{aligned} h(\mu) &= c\int \frac{dt}{\sqrt{t}} \\ &= c\int t^{-1/2} dt + k \\ &= c\frac{t^{-(1/2)+1}}{-(1/2)+1} + k \\ &= 2c\sqrt{t} + k \end{aligned}$$

This implies that taking the square root of y will stabilize the variance. This agrees with the advice given in the textbook (and elsewhere) that the square root transformation is very useful for stabilizing the variance in Poisson data or in general for count data where the mean and variance are not too different.

As a second example, suppose that the square root of the mean is approximately equal to the variance; that is, $\mu^{1/2} = \sigma^2$. Essentially, this says that

$$\mu = \left[\sigma^2\right]^2 \text{ which implies that } f(t) = t^2$$

Therefore,

$$\begin{aligned} h(\mu) &= c\int \frac{dt}{\sqrt{t^2}} \\ &= c\int \frac{dt}{t} + k \\ &= c\log(t) + k, \text{ if } t > 0 \end{aligned}$$

This implies that for a positive response where $\mu^{1/2} = \sigma^2$ the log of the response is an appropriate variance-stabilizing transformation.

S15-2. Selecting λ in the Box-Cox Method

In Section 15-1.1 of the Textbook we present the Box-Cox method for analytically selecting a response variable transformation, and observe that its theoretical basis is the method of maximum likelihood. In applying this method, we are essentially maximizing

$$L(\lambda) = -\frac{1}{2} n \ln\left[SS_E(\lambda)\right]$$

or equivalently, we are minimizing the error sum of squares with respect to λ. An approximate 100(1-α) percent confidence interval for λ consists of those values of λ that satisfy the inequality

$$L(\hat{\lambda}) - L(\lambda) \le \frac{1}{2}\chi^2_{\alpha,1} / n$$

where n is the sample size and $\chi^2_{\alpha,1}$ is the upper α percentage point of the chi-square distribution with one degree of freedom. To actually construct the confidence interval we would draw on a plot of $L(\hat{\lambda})$ versus λ a horizontal line at height

$$L(\hat{\lambda}) - \frac{1}{2}\chi^2_{\alpha,1}$$

on the vertical scale. This would cut the curve of $L(\hat{\lambda})$ at two points, and the locations of these two points on the λ axis define the two end points of the approximate confidence interval for λ. If we are minimizing the residual or error sum of squares (which is identical to maximizing the likelihood) and plotting $SS_E(\lambda)$ versus λ, then the line must be plotted at height

$$SS^* = SS_E(\hat{\lambda}) e^{\chi^2_{\alpha,1}/n}$$

Remember that $\hat{\lambda}$ is the value of λ that minimizes the error sum of squares.

Equation (14-20 in the textbook looks slightly different than the equation for SS^* above. The term $\exp(\chi^2_{\alpha,1} / n)$ has been replaced by $1 + (t^2_{\alpha/2,\nu}) / \nu$, where ν is the number of degrees of freedom for error. Some authors use $1 + (\chi^2_{\alpha/2,\nu}) / \nu$ or $1 + (z^2_{\alpha/2}) / \nu$ instead, or sometimes $1 + (t^2_{\alpha/2,n}) / n$ or $1 + (\chi^2_{\alpha/2,n}) / n$ or $1 + (z^2_{\alpha/2}) / n$. These are all based on the expansion of $\exp(x) = 1 + x + x^2/2! + x^3/3! + \cdots \approx 1 + x$, and the fact that $\chi^2_1 = z^2 \approx t^2_\nu$, unless the number of degrees of freedom ν is too small. It is perhaps debatable whether we should use n or ν, but in most practical cases, there will be little difference in the confidence intervals that result.

S15-3. Generalized Linear Models

Section 15-1.2 considers an alternative approach to data transformation when the "usual" assumptions of normality and constant variance are not satisfied. This approach is based

on the generalized linear model or GLM. Examples 15-2, 15-3, and 15-4 illustrated the applicability of the GLM to designed experiments.

The GLM is a unification of nonlinear regression models and nonnormal response variable distributions, where the response distribution is a member of the **exponential family**, which includes the normal, Poisson, binomial, exponential and gamma distributions as members. Furthermore, the normal-theory linear model is just a special case of the GLM, so in many ways, the GLM is a unifying approach to empirical modeling and data analysis.

We begin our presentation of these models by considering the case of **logistic regression**. This is a situation where the response variable has only two possible outcomes, generically called "success" and "failure" and denoted by 0 and 1. Notice that the response is essentially qualitative, since the designation "success" or "failure" is entirely arbitrary. Then we consider the situation where the response variable us a count, such as the number of defects in a unit of product (as in the grille defects of Example 14-2), or the number of relatively rare events such as the number of Atlantic hurricanes than make landfall on the United States in a year. Finally, we briefly show how all these situations are unified by the GLM.

S15-3.1. Models with a Binary Response Variable

Consider the situation where the response variable from an experiment takes on only two possible values, 0 and 1. These could be arbitrary assignments resulting from observing a qualitative response. For example, the response could be the outcome of a functional electrical test on a semiconductor device for which the results are either a "success", which means the device works properly, or a "failure", which could be due to a short, an open, or some other functional problem.

Suppose that the model has the form

$$y_i = \mathbf{x}_i'\beta + \varepsilon_i$$

where $\mathbf{x}_i' = [1, x_{i1}, x_{i2}, \cdots, x_{ik}]$, $\beta' = [\beta_0, \beta_1, \beta_2, \cdots, \beta_k]$, $\mathbf{x}_i'\beta$ is called the **linear predictor,** and the response variable y_i takes on the values either 0 or 1. We will assume that the response variable y_i is a **Bernoulli random variable** with probability distribution as follows:

y_i	Probability
1	$P(y_i = 1) = \pi_i$
0	$P(y_i = 0) = 1 - \pi_i$

Now since $E(\varepsilon_i) = 0$, the expected value of the response variable is

$$\begin{aligned} E(y_i) &= 1(\pi_i) +)(1 - \pi_i) \\ &= \pi_i \end{aligned}$$

This implies that

$$E(y_i) = \mathbf{x}_i'\beta = \pi_i$$

This means that the expected response given by the response function $E(y_i) = \mathbf{x}_i'\beta$ is just the probability that the response variable takes on the value 1.

There are some substantive problems with this model. First, note that if the response is binary, then the error term ε_i can only take on two values, namely

$$\varepsilon_i = 1 - \mathbf{x}_i'\beta \text{ when } y_i = 1$$
$$\varepsilon_i = -\mathbf{x}_i'\beta \text{ when } y_i = 0$$

Consequently, the errors in this model cannot possibly be normal. Second, the error variance is not constant, since

$$\begin{aligned}\sigma_{y_i}^2 &= E\{y_i - E(y_i)\}^2 \\ &= (1-\pi_i)^2\pi_i + (0-\pi_i)^2(1-\pi_i) \\ &= \pi_i(1-\pi_i)\end{aligned}$$

Notice that this last expression is just

$$\sigma_{y_i}^2 = E(y_i)[1 - E(y_i)]$$

since $E(y_i) = \mathbf{x}_i'\beta = \pi_i$. This indicates that the variance of the observations (which is the same as the variance of the errors because $\varepsilon_i = y_i - \pi_i$, and π_i is a constant) is a function of the mean. Finally, there is a constraint on the response function, because

$$0 \le E(y_i) = \pi_i \le 1$$

This restriction causes serious problems with the choice of a **linear response function**, as we have initially assumed.

Generally, when the response variable is binary, there is considerable evidence indicating that the shape of the response function should be nonlinear. A monotonically increasing (or decreasing) *S*-shaped (or reverse *S*-shaped) function is usually employed. This function is called the **logistic response function**, and has the form

$$E(y) = \frac{\exp(\mathbf{x}'\beta)}{1 + \exp(\mathbf{x}'\beta)}$$

or equivalently,

$$E(y) = \frac{1}{1 + \exp(-\mathbf{x}'\beta)}$$

The logistic response function can be easily linearized. Let $E(y) = \pi$ and make the transformation

$$\eta = \ln\left(\frac{\pi}{1-\pi}\right)$$

Then in terms of our **linear predictor** $\mathbf{x}'\beta$ we have

$$\eta = \mathbf{x}'\beta$$

This transformation is often called the **logit transformation** of the probability π, and the ratio π/(1-π) in the transformation is called the odds. Sometimes the logit transformation is called the log-odds.

There are other functions that have the same shape as the logistic function, and they can also be obtained by transforming π. One of these is the *probit* transformation, obtained by transforming π using the cumulative normal distribution. This produces a *probit regression model*. The probit regression model is less flexible than the logistic regression model because it cannot easily incorporate more than one predictor variable. Another possible transformation is the **complimentary log-log transformation** of π, given by $\ln[-\ln(1-\pi)]$. This results in a response function that is not symmetric about the value π = 0.5.

S15-3.2. Estimating the Parameters in a Logistic Regression Model

The general form of the logistic regression model is

$$y_i = E(y_i) + \varepsilon_i$$

where the observations y_i are independent Bernoulli random variables with expected values

$$\begin{aligned} E(y_i) &= \pi_i \\ &= \frac{\exp(\mathbf{x}_i'\beta)}{1+\exp(\mathbf{x}_i'\beta)} \end{aligned}$$

We will use the method of **maximum likelihood** to estimate the parameters in the linear predictor $\mathbf{x}_i'\beta$.

Each sample observation follows the Bernoulli distribution, so the probability distribution of each sample observation is

$$f_i(y_i) = \pi_i^{y_i}(1-\pi_i^{1-y_i}), i = 1,2,\cdots,n$$

and of course each observation y_i takes on the value 0 or 1. Since the observations are independent, the likelihood function is just

$$\begin{aligned} L(y_1, y_2, \cdots, y_n, \beta) &= \prod_{i=1}^{n} f_i(y_i) \\ &= \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i^{1-y_i}) \end{aligned}$$

It is more convenient to work with the log-likelihood

$$\ln L(y_1, y_2, \cdots, y_n, \beta) = \ln \prod_{i=1}^{n} f_i(y_i)$$
$$= \sum_{i=1}^{n} \left[y_i \ln\left(\frac{\pi_i}{1-\pi_i} \right) \right] + \sum_{i=1}^{n} \ln(1-\pi_i)$$

Now since $1-\pi_i = [1+\exp(\mathbf{x}_i'\beta)]^{-1}$ and $\eta_i = \ln[\pi_i / (1-\pi_i)] = \mathbf{x}_i'\beta$, the log-likelihood can be written as

$$\ln L(\mathbf{y}, \beta) = \sum_{i=1}^{n} y_i \mathbf{x}_i'\beta - \sum_{i=1}^{n} \ln[1+\exp(\mathbf{x}_i'\beta)]$$

Often in logistic regression models we have repeated observations or trials at each level of the *x* variables. This happens frequently in designed experiments. Let y_i represent the number of 1's observed for the *i*th observation and n_i be the number of trials at each observation. Then the log-likelihood becomes

$$\ln L(\mathbf{y}, \beta) = \sum_{i=1}^{n} y_i \pi_i + \sum_{i=1}^{n} n_i \ln(1-\pi_i) - \sum_{i=1}^{n} y_i \ln(1-\pi_i)$$

Numerical search methods could be used to compute the maximum likelihood estimates (or MLEs) $\hat{\beta}$. However, it turns out that we can use iteratively reweighted least squares (IRLS) to actually find the MLEs. To see this recall that the MLEs are the solutions to

$$\frac{\partial L}{\partial \beta} = 0$$

which can be expressed as

$$\frac{\partial L}{\partial \pi_i} \frac{\partial \pi_i}{\partial \beta} = 0$$

Note that

$$\frac{\partial L}{\partial \pi_i} = \sum_{i=1}^{n} \frac{n_i}{\pi_i} - \sum_{i=1}^{n} \frac{n_i}{1-\pi_i} + \sum_{i=1}^{n} \frac{y_i}{1-\pi_i}$$

and

$$\frac{\partial \pi_i}{\partial \beta} = \left\{ \frac{\exp(\mathbf{x}_i'\beta)}{1+\exp(\mathbf{x}_i'\beta)} - \left[\frac{\exp(\mathbf{x}_i'\beta)}{1+\exp(\mathbf{x}_i'\beta)} \right]^2 \right\} \mathbf{x}_i$$
$$= \pi_i(1-\pi_i)\mathbf{x}_i$$

Putting this all together gives

$$\frac{\partial L}{\partial \beta} = \left[\sum_{i=1}^{n} \frac{n_i}{\pi_i} - \sum_{i=1}^{n} \frac{n_i}{1-\pi_i} + \sum_{i=1}^{n} \frac{y_i}{1-\pi_i} \right] \pi_i (1-\pi_i) \mathbf{x}_i$$
$$= \sum_{i=1}^{n} \left[\frac{y_i}{\pi_i} - \frac{n_i}{1-\pi_i} + \frac{y_i}{1-\pi_i} \right] \pi_i (1-\pi_i) \mathbf{x}_i$$
$$= \sum_{i=1}^{n} (y_i - n_i \pi_i) \mathbf{x}_i$$

Therefore, the maximum likelihood estimator solves

$$\mathbf{X}'(\mathbf{y} - \mu) = \mathbf{0}$$

where $\mathbf{y}' = [y_1, y_2, \cdots, y_n]$ and $\mu' = [n_1\pi_1, n_2\pi_2 \cdots, n_n\pi_n]$. This set of equations is often called the **maximum likelihood score equations**. They are actually the same form of the normal equations that we have seen previously for linear least squares, because in the linear regression model, $E(\mathbf{y}) = \mathbf{X}\beta = \mu$ and the normal equations are

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y}$$

which can be written as

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\beta) = \mathbf{0}$$
$$\mathbf{X}'(\mathbf{y} - \mu) = \mathbf{0}$$

The **Newton-Raphson** method is actually used to solve the score equations. This procedure observes that in the neighborhood of the solution, we can use a first-order Taylor series expansion to form the approximation

$$p_i - \pi_i \approx \left(\frac{\partial \pi_i}{\partial \beta} \right)' (\beta^* - \beta) \tag{1}$$

where

$$p_i = \frac{y_i}{n_i}$$

and β^* is the value of β that solves the score equations. Now $\eta_i = \mathbf{x}_i'\beta$, and

$$\frac{\partial \eta_i}{\partial \beta} = \mathbf{x}_i$$

so

$$\pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$$

By the chain rule

$$\frac{\partial \pi_i}{\partial \beta} = \frac{\partial \pi_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta} = \frac{\partial \pi_i}{\partial \eta_i} \mathbf{x}_i$$

Therefore, we can rewrite (1) above as

$$p_i - \pi_i \approx \left(\frac{\partial \pi_i}{\partial \eta_i} \right) \mathbf{x}_i'(\beta^* - \beta)$$

$$p_i - \pi_i \approx \left(\frac{\partial \pi_i}{\partial \eta_i} \right) (\mathbf{x}_i'\beta^* - \mathbf{x}_i'\beta) \qquad (2)$$

$$p_i - \pi_i \approx \left(\frac{\partial \pi_i}{\partial \eta_i} \right) (\eta_i^* - \eta_i)$$

where η_i^* is the value of η_i evaluated at β^*. We note that

$$(y_i - n_i\pi_i) = (n_i p_i - n_i \pi_i) = n_i (p_i - \pi_i)$$

and since

$$\pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$$

we can write

$$\frac{\partial \pi_i}{\partial \eta_i} = \frac{\exp(\eta_i)}{1+\exp(\eta_i)} - \left[\frac{\exp(\eta_i)}{1+\exp(\eta_i)} \right]^2$$
$$= \pi_i (1 - \pi_i)$$

Consequently,

$$y_i - n_i\pi_i \approx [n_i \pi_i (1 - \pi_i)](\eta_i^* - \eta_i)$$

Now the variance of the linear predictor $\eta_i^* = \mathbf{x}_i'\beta^*$ is, to a first approximation,

$$V(\eta_i^*) \approx \frac{1}{n_i \pi_i (1 - \pi_i)}$$

Thus

$$y_i - n_i\pi_i \approx \left[\frac{1}{V(\eta_i^*)} \right] (\eta_i^* - \eta_i)$$

and we may rewrite the score equations as

$$\sum_{i=1}^{n} \left[\frac{1}{V(\eta_i)} \right] (\eta_i^* - \eta_i) = 0$$

or in matrix notation,

$$\mathbf{X}'\mathbf{V}^{-1}(\eta^* - \eta) = \mathbf{0}$$

where **V** is a diagonal matrix of the weights formed from the variances of the η_i. Because $\eta = \mathbf{X}\beta$ we may write the score equations as

$$\mathbf{X}'\mathbf{V}^{-1}(\eta^* - \mathbf{X}\beta) = \mathbf{0}$$

and the maximum likelihood estimate of β is

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\eta^*$$

However, there is a problem because we don't know η^*. Our solution to this problem uses equation (2):

$$p_i - \pi_i \approx \left(\frac{\partial \pi_i}{\partial \eta_i}\right)(\eta_i^* - \eta_i)$$

which we can solve for η_i^*,

$$\eta_i^* \approx \eta_i + (p_i - \pi_i)\frac{\partial \eta_i}{\partial \pi_i}$$

Let $z_i = \eta_i + (p_i - \pi_i)\frac{\partial \eta_i}{\partial \pi_i}$ and $\mathbf{z}' = [z_1, z_2, \cdots, z_n]$. Then the Newton-Raphson estimate of β is

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{z}$$

Note that the random portion of z_i is

$$(p_i - \pi_i)\frac{\partial \eta_i}{\partial \pi_i}$$

Thus

$$\begin{aligned} V\left[(p_i - \pi_i)\frac{\partial \eta_i}{\partial \pi_i}\right] &= \left[\frac{\pi_i(1-\pi_i)}{n_i}\right]\left(\frac{\partial \eta_i}{\partial \pi_i}\right)^2 \\ &= \left[\frac{\pi_i(1-\pi_i)}{n_i}\right]\left(\frac{1}{\pi_i(1-\pi_i)}\right)^2 \\ &= \frac{1}{n_i\pi_i(1-\pi_i)} \end{aligned}$$

So **V** is the diagonal matrix of weights formed from the variances of the random part of **z**.

Thus the IRLS algorithm based on the Newton-Raphson method can be described as follows:

1. Use ordinary least squares to obtain an initial estimate of β, say $\hat{\beta}_0$;
2. Use $\hat{\beta}_0$ to estimate **V** and π;

3. Let $\eta_0 = \mathbf{X}\hat{\beta}_0$;
4. Base $\mathbf{z}_1$ on η_0;
5. Obtain a new estimate $\hat{\beta}_1$, and iterate until some suitable convergence criterion is satisfied.

If $\hat{\beta}$ is the final value that the above algorithm produces and if the model assumptions are correct, then we can show that asymptotically

$$E(\hat{\beta}) = \beta \quad \text{and} \quad V(\hat{\beta}) = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}$$

The fitted value of the logistic regression model is often written as

$$\begin{aligned}\hat{\pi}_i &= \frac{\exp(\mathbf{x}_i'\hat{\beta})}{1+\exp(\mathbf{x}_i'\hat{\beta})} \\ &= \frac{1}{1+\exp(-\mathbf{x}_i'\hat{\beta})}\end{aligned}$$

S15-3.3. Interpreting the Parameters in a Logistic Regression Model

It is relatively easy to interpret the parameters in a logistic regression model. Consider first the case where the linear predictor has only a single predictor, so that the fitted value of the model at a particular value of x, say x_i, is

$$\hat{\eta}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

The fitted value at $x_i + 1$ is

$$\hat{\eta}(x_i + 1) = \hat{\beta}_0 + \hat{\beta}_1(x_i + 1)$$

and the difference in the two predicted values is

$$\hat{\eta}(x_i + 1) - \hat{\eta}(x_i) = \hat{\beta}_1$$

Now $\hat{\eta}(x_i)$ is just the log-odds when the regressor variable is equal to x_i, and $\hat{\eta}(x_i + 1)$ is just the log-odds when the regressor is equal to $x_i + 1$. Therefore, the difference in the two fitted values is

$$\begin{aligned}\hat{\eta}(x_i + 1) - \hat{\eta}(x_i) &= \ln(odds_{x_i+1}) - \ln(odds_{x_i}) \\ &= \ln\left(\frac{odds_{x_i+1}}{odds_{x_i}}\right) \\ &= \hat{\beta}_1\end{aligned}$$

If we take antilogs, we obtain the **odds ratio**

$$\hat{O}_R \equiv \frac{odds_{x_i+1}}{odds_{x_i}} = e^{\hat{\beta}_1}$$

The odds ratio can be interpreted as the estimated increase in the probability of success associated with a one-unit change in the value of the predictor variable. In general, the estimated increase in the odds ratio associated with a change of *d* units in the predictor variable is $\exp(d\hat{\beta}_1)$.

The interpretation of the regression coefficients in the multiple logistic regression model is similar to that for the case where the linear predictor contains only one regressor. That is, the quantity $\exp(\hat{\beta}_j)$ is the odds ratio for regressor x_j, assuming that all other predictor variables are constant.

S15-3.4. Hypothesis Tests on Model Parameters

Hypothesis testing in the GLM is based on the general method of **likelihood ratio tests**. It is a large-sample procedure, so the test procedures rely on asymptotic theory. The likelihood ratio approach leads to a statistic called **deviance**.

Model Deviance

The deviance of a model compares the log-likelihood of the fitted model of interest to the log-likelihood of a saturated model; that is, a model that has exactly *n* parameters and which fits the sample data perfectly. For the logistic regression model, this means that the probabilities π_i are completely unrestricted, so setting $\hat{\pi}_i = y_i$ (recall that $y_i = 0$ or 1) would maximize the likelihood. It can be shown that this results in a maximum value of the likelihood function for the saturated model of unity, so the maximum value of the log-likelihood function is zero.

Now consider the log- likelihood function for the fitted logistic regression model. When the maximum likelihood estimates $\hat{\beta}$ are used in the log- likelihood function, it attains its maximum value, which is

$$\ln L(\hat{\beta}) = \sum_{i=1}^{n} y_i \mathbf{x}'\hat{\beta}_i - \sum_{i=1}^{n} \ln[1 + \exp(\mathbf{x}_i'\hat{\beta})]$$

The value of the log-likelihood function for the fitted model can never exceed the value of the log-likelihood function for the saturated model, because the fitted model contains fewer parameters. The deviance compares the log-likelihood of the saturated model with the log-likelihood of the fitted model. Specifically, **model deviance** is defined as

$$\begin{aligned} \lambda(\beta) &= 2\ln L(\text{saturated model}) - 2\ln L(\hat{\beta}) \\ &= 2[\ell(\text{(saturated model)} - \ell(\hat{\beta})] \end{aligned} \tag{3}$$

where ℓ denotes the log of the likelihood function. Now if the logistic regression model is the correct regression function and the sample size *n* is large, the model deviance has an approximate chi-square distribution with $n - p$ degrees of freedom. Large values of

the model deviance would indicate that the model is not correct, while a small value of model deviance implies that the fitted model (which has fewer parameters than the saturated model) fits the data almost as well as the saturated model. The formal test criteria would be as follows:

if $\lambda(\beta) \le \chi^2_{\alpha,n-p}$ conclude that the fitted model is adequate

if $\lambda(\beta) > \chi^2_{\alpha,n-p}$ conclude that the fitted model is not adequate

The deviance is related to a very familiar quantity. If we consider the standard normal-theory linear regression model, the deviance turns out to be the error or residual sum of squares divided by the error variance σ^2.

Testing Hypotheses on Subsets of Parameters using Deviance

We can also use the deviance to test hypotheses on subsets of the model parameters, just as we used the difference in regression (or error) sums of squares to test hypotheses in the normal-error linear regression model case. Recall that the model can be written as

$$\begin{aligned}\eta &= \mathbf{X}\beta \\ &= \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2\end{aligned}$$

where the *full model* has *p* parameters, β_1 contains $p - r$ of these parameters, β_2 contains *r* of these parameters, and the columns of the matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ contain the variables associated with these parameters. Suppose that we wish to test the hypotheses

$$\begin{aligned}H_0&: \beta_2 = \mathbf{0} \\ H_1&: \beta_2 \ne \mathbf{0}\end{aligned}$$

Therefore, the *reduced model* is

$$\eta = \mathbf{X}_1\beta_1$$

Now fit the reduced model, and let $\lambda(\beta_1)$ be the deviance for the reduced model. The deviance for the reduced model will always be larger than the deviance for the full model, because the reduced model contains fewer parameters. However, if the deviance for the reduced model is not much larger than the deviance for the full model, it indicates that the reduced model is about as good a fit as the full model, so it is likely that the parameters in β_2 are equal to zero. That is, we cannot reject the null hypothesis above. However, if the difference in deviance is large, at least one of the parameters in β_2 is likely not zero, and we should reject the null hypothesis. Formally, the difference in deviance is

$$\lambda(\beta_2|\beta_1) = \lambda(\beta_1) - \lambda(\beta)$$

and this quantity has $n - (p - r) - (n - p) = r$ degrees of freedom. If the null hypothesis is true and if *n* is large, the difference in deviance has a chi-square distribution with *r* degrees of freedom. Therefore, the test statistic and decision criteria are

if $\lambda(\beta_2|\beta_1) \geq \chi^2_{\alpha,r}$ reject the null hypothesis

if $\lambda(\beta_2|\beta_1) < \chi^2_{\alpha,r}$ do not reject the null hypothesis

Sometimes the difference in deviance $\lambda(\beta_2|\beta_1)$ is called the **partial deviance**. It is a likelihood ratio test. To see this, let $L(\hat{\beta})$ be the maximum value of the likelihood function for the full model, and $L(\hat{\beta}_1)$ be the maximum value of the likelihood function for the reduced model. The **likelihood ratio** is

$$\frac{L(\hat{\beta}_1)}{L(\hat{\beta})}$$

The test statistic for the likelihood ratio test is equal to minus two times the log-likelihood ratio, or

$$\begin{aligned}\chi^2 &= -2\ln\frac{L(\hat{\beta}_1)}{L(\hat{\beta})} \\ &= 2\ln L(\hat{\beta}) - 2\ln L(\hat{\beta}_1)\end{aligned}$$

However, this is exactly the same as the difference in deviance. To see this, substitute from the definition of the deviance from equation (3) and note that the log-likelihoods for the saturated model cancel out.

Tests on Individual Model Coefficients

Tests on individual model coefficients, such as

$$\begin{aligned}H_0&: \beta_j = 0 \\ H_1&: \beta_j \neq 0\end{aligned}$$

can be conducted by using the difference in deviance method described above. There is another approach, also based on the theory of maximum likelihood estimators. For large samples, the distribution of a maximum likelihood estimator is approximately normal with little or no bias. Furthermore, the variances and covariances of a set of maximum likelihood estimators can be found from the second partial derivatives of the log-likelihood function with respect to the model parameters, evaluated at the maximum likelihood estimates. Then a *t*-like statistic can be constructed to test the above hypothesis. This is sometimes referred to as **Wald inference.**

Let **G** denote the $p \times p$ matrix of second partial derivatives of the log-likelihood function; that is

$$G_{ij} = \frac{\partial^2 \ell(\beta)}{\partial\beta_i \partial\beta_j}, i, j = 0,1,\cdots,k$$

G is called the **Hessian matrix**. If the elements of the Hessian are evaluated at the maximum likelihood estimators $\beta = \hat{\beta}$, the large-sample approximate covariance matrix of the regression coefficients is

$$V(\hat{\beta}) \equiv \hat{\Sigma} = -\mathbf{G}(\hat{\beta})^{-1}$$

The square roots of the diagonal elements of this matrix are the large-sample standard errors of the regression coefficients, so the test statistic for the null hypothesis in

$$H_0: \beta_j = 0$$
$$H_1: \beta_j \neq 0$$

is

$$Z_0 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

The reference distribution for this statistic is the standard normal distribution. Some computer packages square the Z_0 statistic and compare it to a chi-square distribution with one degree of freedom. It is also straightforward to use Wald inference to construct confidence intervals on individual regression coefficients.

S15-3.5. Poisson Regression

We now consider another regression modeling scenario where the response variable of interest is not normally distributed. In this situation the response variable represents a count of some relatively rare event, such as defects in a unit of manufactured product, errors or "bugs" in software, or a count of particulate matter or other pollutants in the environment. The analyst is interested in modeling the relationship between the observed counts and potentially useful regressor or predictor variables. For example, an engineer could be interested in modeling the relationship between the observed number of defects in a unit of product and production conditions when the unit was actually manufactured.

We assume that the response variable y_i is a count, such that the observation $y_i = 0,1,\cdots,$. A reasonable probability model for count data is often the Poisson distribution

$$f(y) = \frac{e^{-\mu}\mu^y}{y!}, y = 0,1,\cdots$$

where the parameter $\mu > 0$. The Poison is another example of a probability distribution where the mean and variance are related. In fact, for the Poisson distribution it is straightforward to show that

$$E(y) = \mu \text{ and } V(y) = \mu$$

That is, both the mean *and* variance of the Poisson distribution are equal to the parameter μ.

The Poisson regression model can be written as

$$y_i = E(y_i) + \varepsilon_i, i = 1,2,\cdots,n$$

We assume that the expected value of the observed response can be written as

$$E(y_i) = \mu_i$$

and that there is a function *g* that relates the mean of the response to a linear predictor, say

$$\begin{aligned} g(\mu_i) &= \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \\ &= \mathbf{x}_i'\beta \end{aligned}$$

The function *g* is usually called the **link function**. The relationship between the mean and the linear predictor is

$$\mu_i = g^{-1}(\mathbf{x}_i'\beta)$$

There are several link functions that are commonly used with the Poisson distribution. One of these is the **identity link**

$$g(\mu_i) = \mu_i = \mathbf{x}_i'\beta$$

When this link is used, $E(y_i) = \mu_i = \mathbf{x}_i'\beta$ since $\mu_i = g^{-1}(\mathbf{x}_i'\beta) = \mathbf{x}_i'\beta$. Another popular link function for the Poisson distribution is the **log link**

$$g(\mu_i) = \ln(\mu_i) = \mathbf{x}_i'\beta$$

For the log link, the relationship between the mean of the response variable and the linear predictor is

$$\begin{aligned} \mu_i &= g^{-1}(\mathbf{x}_i'\beta) \\ &= e^{\mathbf{x}_i'\beta} \end{aligned}$$

The log link is particularly attractive for Poisson regression because it ensures that all of the predicted values of the response variable will be nonnegative.

The method of maximum likelihood is used to estimate the parameters in Poisson regression. The development follows closely the approach used for logistic regression. If we have a random sample of *n* observations on the response *y* and the predictors **x**, then the likelihood function is

$$\begin{aligned} L(\mathbf{y}, \beta) &= \prod_{i=1}^{n} f_i(y_i) \\ &= \prod_{i=1}^{n} \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} \\ &= \frac{\prod_{i=1}^{n} \mu_i^{y_i} \exp(-\sum_{i=1}^{n} \mu_i)}{\prod_{i=1}^{n} y_i!} \end{aligned}$$

where $\mu_i = g^{-1}(\mathbf{x}_i'\beta)$. Once the link function is specified, we maximize the log-likelihood

$$\ln L(\mathbf{y}, \beta) = \sum_{i=1}^{n} y_i \ln(\mu_i) - \sum_{i=1}^{n} \mu_i - \sum_{i=1}^{n} \ln(y_i!)$$

Iteratively reweighted least squares can be used to find the maximum likelihood estimates of the parameters in Poisson regression, following an approach similar to that used for logistic regression. Once the parameter estimates $\hat{\beta}$ are obtained, the fitted Poisson regression model is

$$\hat{y}_i = g^{-1}(\mathbf{x}_i'\hat{\beta})$$

For example, if the identity link is used, the prediction equation becomes

$$\begin{aligned}\hat{y}_i &= g^{-1}(\mathbf{x}_i'\hat{\beta}) \\ &= \mathbf{x}_i'\hat{\beta}\end{aligned}$$

and if the log link is specified, then

$$\begin{aligned}\hat{y}_i &= g^{-1}(\mathbf{x}_i'\hat{\beta}) \\ &= \exp(\mathbf{x}_i'\hat{\beta})\end{aligned}$$

Inference on the model and its parameters follows exactly the same approach as used for logistic regression. That is, model deviance is an overall measure of goodness of fit, and tests on subsets of model parameters can be performed using the difference in deviance between the full and reduced models. These are likelihood ratio tests. Wald inference, based on large-sample properties of maximum likelihood estimators, can be used to test hypotheses and construct confidence intervals on individual model parameters.

S15-3.6. The Generalized Linear Model

All of the regression models that we have considered in this section belong to a *family* of regression models called the **generalized linear model**, or the **GLM**. The GLM is actually a unifying approach to regression and experimental design models, uniting the usual normal-theory linear regression models and nonlinear models such as logistic and Poisson regression.

A key assumption in the GLM is that the response variable distribution is a member of the exponential family of distributions, which includes the normal, binomial, Poisson, inverse normal, exponential and gamma distributions. Distributions that are members of the exponential family have the general form

$$f(y_i, \theta_i, \phi) = \exp\{[y_i\theta_i - b(\theta_i)]/a(\phi) + h(y_i, \phi)\}$$

where ϕ is a scale parameter and θ_i is called the natural location parameter. For members of the exponential family,

$$\mu = E(y) = \frac{db(\theta_i)}{d\theta_i}$$

$$V(y) = \frac{d^2 b(\theta_i)}{d\theta_i^2} a(\phi)$$

$$= \frac{d\mu}{d\theta_i} a(\phi)$$

Let

$$\text{var}(\mu) = \frac{V(y)}{a(\phi)} = \frac{d\mu}{d\theta_i}$$

where var(μ) denotes the dependence of the variance of the response on its mean. As a result, we have

$$\frac{d\theta_i}{d\mu} = \frac{1}{\text{var}(\mu)}$$

It is easy to show that the normal, binomial and Poisson distributions are members of the exponential family.

The Normal Distribution

$$f(y_i, \theta_i, \phi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2 \right)$$

$$= \exp\left[-\ln(2\pi\sigma^2) - \frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} \right]$$

$$= \exp\left[\frac{1}{\sigma^2}\left(-\frac{y^2}{2} + y\mu - \frac{\mu^2}{2} \right) - \frac{1}{2}\ln(2\pi\sigma^2) \right]$$

$$= \exp\left[\frac{1}{\sigma^2}\left(y\mu - \frac{\mu^2}{2} \right) - \frac{y^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2) \right]$$

Thus for the normal distribution, we have

$$\theta_i = \mu$$

$$b(\theta_i) = \frac{\mu^2}{2}$$

$$a(\phi) = \sigma^2$$

$$h(y_i, \phi) = -\frac{y^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)$$

$$E(y) = \frac{db(\theta_i)}{d\theta_i} = \mu, \text{ and } V(y) = \frac{d^2 b(\theta_i)}{d\theta_i^2} a(\phi) = \sigma^2$$

The Binomial Distribution

$$f(y_i,\theta_i,\phi) = \binom{n}{m}\pi^y(1-\pi)^{n-y}$$

$$= \exp\left\{\ln\binom{n}{y} + y\ln\pi + (n-y)\ln(1-\pi)\right\}$$

$$= \exp\left\{\ln\binom{n}{y} + y\ln\pi + n\ln(1-\pi) - y\ln(1-\pi)\right\}$$

$$= \exp\left\{y\ln\left[\frac{\pi}{1-\pi}\right] + n\ln(1-\pi) + \ln\binom{n}{y}\right\}$$

Therefore, for the binomial distribution,

$$\theta_i = \ln\left[\frac{\pi}{1-\pi}\right] \text{ and } \pi = \frac{\exp(\theta_i)}{1+\exp(\theta_i)}$$

$$b(\theta_i) = -n\ln(1-\pi)$$

$$a(\phi) = 1$$

$$h(y_i,\phi) = \ln\binom{n}{y}$$

$$E(y) = \frac{db(\theta_i)}{d\theta_i} = \frac{db(\theta_i)}{d\pi}\frac{d\pi}{d\theta_i}$$

We note that

$$\frac{d\pi}{d\theta_i} = \frac{\exp(\theta_i)}{1+\exp(\theta_i)} - \left[\frac{\exp(\theta_i)}{1+\exp(\theta_i)}\right]^2$$

$$= \pi(1-\pi)$$

Therefore,

$$E(y) = \left(\frac{n}{1-\pi}\right)\pi(1-\pi)$$

$$= n\pi$$

We recognize this as the mean of the binomial distribution. Also,

$$V(y) = \frac{dE(y)}{d\theta_i}$$

$$= \frac{dE(y)}{d\pi}\frac{d\pi}{d\theta_i}$$

$$= n\pi(1-\pi)$$

This last expression is just the variance of the binomial distribution.

The Poisson Distribution

$$f(y_i, \theta_i, \phi) = \frac{\lambda^y e^{-\lambda}}{y!}$$
$$= \exp[y \ln \lambda - \lambda - \ln(y!)]$$

Therefore, for the Poisson distribution, we have

$$\theta_i = \ln(\lambda) \text{ and } \lambda = \exp(\theta_i)$$
$$b(\theta_i) = \lambda$$
$$a(\phi) = 1$$
$$h(y_i, \phi) = -\ln(y!)$$

Now

$$E(y) = \frac{db(\theta_i)}{d\theta_i} = \frac{db(\theta_i)}{d\lambda}\frac{d\lambda}{d\theta_i}$$

However, since

$$\frac{d\lambda}{d\theta_i} = \exp(\theta_i) = \lambda$$

the mean of the Poisson distribution is

$$E(y) = 1 \cdot \lambda = \lambda$$

The variance of the Poisson distribution is

$$V(y) = \frac{dE(y)}{d\theta_i} = \lambda$$

S15-3.7. Link Functions and Linear Predictors

The basic idea of a GLM is to develop a linear model for an appropriate **function** of the expected value of the response variable. Let η_i be the **linear predictor** defined by

$$\eta_i = g[E(y_i)] = g(\mu_i) = \mathbf{x}_i'\beta$$

Note that the expected response is just

$$E(y_i) = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}_i'\beta)$$

We call the function *g* the **link function**. Recall that we introduced the concept of a link function in our description of Poisson regression in Section S15-3.5 above. There are many possible choices of the link function, but if we choose

$$\eta_i = \theta_i$$

we say that η_i is the **canonical link**. Table 1 shows the canonical links for the most common choices of distributions employed with the GLM.

Table 1. Canonical Links for the Generalized Linear Model

Distribution	Canonical Link
Normal	$\eta_i = \mu_i$ (identity link)
Binomial	$\eta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$ (logistic link)
Poisson	$\eta_i = \ln(\lambda)$ (log link)
Exponential	$\eta_i = \frac{1}{\lambda_i}$ (reciprocal link)
Gamma	$\eta_i = \frac{1}{\lambda_i}$ (reciprocal link)

There are other link functions that could be used with a GLM, including:

1. The probit link,

$$\eta_i = \Phi^{-1}[E(y_i)]$$

where Φ represents the cumulative standard normal distribution function.

2. The complimentary log-log link,

$$\eta_i = \ln\{\ln[1-E(y_i)]\}$$

3. The power family link,

$$\eta_i = \begin{cases} E(y_i)^{\lambda}, & \lambda \neq 0 \\ \ln[E(y_i)], & \lambda = 0 \end{cases}$$

A very fundamental idea is that there are two components to a GLM; the response variable distribution, and the link function. We can view the selection of the link function in a vein similar to the choice of a transformation on the response. However, unlike a transformation, the link function takes advantage of the *natural* distribution of the response. Just as not using an appropriate transformation can result in problems with a fitted linear model, improper choices of the link function can also result in significant problems with a GLM.

S15-3.8. Parameter Estimation in the GLM

The method of maximum likelihood is the theoretical basis for parameter estimation in the GLM. However, the actual implementation of maximum likelihood results in an algorithm based on iteratively reweighted least squares (IRLS). This is exactly what we saw previously for the special case of logistic regression.

Consider the method of maximum likelihood applied to the GLM, and suppose we use the canonical link. The log-likelihood function is

$$\ell(\mathbf{y},\beta) = \sum_{i=1}^{n} [y_i\theta_i - b(\theta_i)] / a(\phi) + h(y_i,\phi)$$

For the canonical link, we have $\eta_i = g[E(y_i)] = g(\mu_i) = \mathbf{x}_i'\beta$; therefore,

$$\begin{aligned}
\frac{\partial \ell}{\partial \beta} &= \frac{\partial \ell}{\partial \theta_i}\frac{\partial \theta_i}{\partial \beta} \\
&= \frac{1}{a(\phi)} \sum_{i=1}^{n} \left[y_i - \frac{db(\theta_i)}{d\theta_i} \right] \mathbf{x}_i \\
&= \frac{1}{a(\phi)} \sum_{i=1}^{n} (y_i - \mu_i)\mathbf{x}_i
\end{aligned}$$

Consequently, we can find the maximum likelihood estimates of the parameters by solving the system of equations

$$\frac{1}{a(\phi)} \sum_{i=1}^{n} (y_i - \mu_i)\mathbf{x}_i = 0$$

In most cases, $a(\phi)$ is a constant, so these equations become:

$$\sum_{i=1}^{n} (y_i - \mu_i)\mathbf{x}_i = 0$$

This is actually a *system* of $p = k + 1$ equations, one for each model parameter. In matrix form, these equations are

$$\mathbf{X}'(\mathbf{y} - \mu) = \mathbf{0}$$

where $\mu' = [\mu_1, \mu_2, \cdots, \mu_p]$. These are called the maximum likelihood score equations, and they are just the same equations that we saw previously in the case of logistic regression, where $\mu' = [n_1\pi_1, n_2\pi_2 \cdots, n_n\pi_n]$.

To solve the score equations, we can use IRLS, just as we did in the case of logistic regression. We start by finding a first-order Taylor series approximation in the neighborhood of the solution

$$y_i - \mu_i \approx \frac{d\mu_i}{d\eta_i}(\eta_i^* - \eta_i)$$

Now for a canonical link $\eta_i = \theta_i$, and

$$y_i - \mu_i \approx \frac{d\mu_i}{d\theta_i}(\eta_i^* - \eta_i) \tag{4}$$

Therefore, we have

$$\eta_i^* - \eta_i \approx (y_i - \mu_i)\frac{d\theta_i}{d\mu_i}$$

This expression provides a basis for approximating the variance of $\hat{\eta}_i$.

In maximum likelihood estimation, we replace η_i by its estimate, $\hat{\eta}_i$. Then we have

$$V(\eta_i^* - \eta_i) \approx V\left[(y_i - \mu_i)\frac{d\theta_i}{d\mu_i}\right]$$

Since η_i^* and μ_i are constants,

$$V(\hat{\eta}_i) \approx \left[\frac{d\theta_i}{d\mu_i}\right]^2 V(y_i)$$

But

$$\frac{d\theta_i}{d\mu_i} = \frac{1}{\text{var}(\mu_i)}$$

and $V(y_i) = \text{var}(\mu_i)a(\phi)$. Consequently,

$$\begin{aligned} V(\hat{\eta}_i) &\approx \left[\frac{1}{\text{var}(\mu_i)}\right]^2 \text{var}(\mu_i)a(\phi) \\ &\approx \frac{1}{\text{var}(\mu_i)}a(\phi) \end{aligned}$$

For convenience, define $\text{var}(\eta_i) = [\text{var}(\mu_i)]^{-1}$, so we have

$$V(\hat{\eta}_i) \approx \text{var}(\eta_i)a(\phi).$$

Substituting this into Equation (4) above results in

$$y_i - \mu_i \approx \frac{1}{\text{var}(\eta_i)}(\eta_i^* - \eta) \tag{5}$$

If we let **V** be an $n \times n$ diagonal matrix whose diagonal elements are the $\text{var}(\eta_i)$, then in matrix form, Equation (5) becomes

$$\mathbf{y} - \mu \approx \mathbf{V}^{-1}(\eta^* - \eta)$$

We may then rewrite the score equations as follows:

$$\begin{aligned} \mathbf{X}'(\mathbf{y} - \mu) &= \mathbf{0} \\ \mathbf{X}'\mathbf{V}^{-1}(\eta^* - \eta) &= \mathbf{0} \\ \mathbf{X}'\mathbf{V}^{-1}(\eta^* - \mathbf{X}\beta) &= \mathbf{0} \end{aligned}$$

Thus, the maximum likelihood estimate of β is

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\eta^*$$

Now just as we saw in the logistic regression situation, we do not know η^*, so we pursue an iterative scheme based on

$$z_i = \hat{\eta}_i + (y_i - \hat{\mu}_i)\frac{d\eta_i}{d\mu_i}$$

Using iteratively reweighted least squares with the Newton-Raphson method, the solution is found from

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{z}$$

Asymptotically, the random component of **z** comes from the observations y_i. The diagonal elements of the matrix **V** are the variances of the z_i's, apart from $a(\phi)$.

As an example, consider the logistic regression case:

$$\eta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$$

$$\frac{d\eta_i}{d\mu_i} = \frac{d\eta_i}{d\pi_i} = \frac{d\ln\left(\frac{\pi_i}{1-\pi_i}\right)}{d\pi_i}$$

$$= \frac{1-\pi_i}{\pi_i}\left[\frac{\pi_i}{1-\pi_i} + \frac{\pi_i}{(1-\pi_i)^2}\right]$$

$$= \frac{(1-\pi_i)}{\pi_i(1-\pi_i)}\left[1 + \frac{\pi_i}{1-\pi_i}\right]$$

$$= \frac{1}{\pi_i}\left[\frac{1-\pi_i+\pi_i}{1-\pi_i}\right]$$

$$= \frac{1}{\pi_i(1-\pi_i)}$$

Thus, for logistic regression, the diagonal elements of the matrix **V** are

$$\left(\frac{d\eta_i}{d\mu_i}\right)^2 V(y_i) = \left[\frac{1}{\pi_i(1-\pi_i)}\right]^2 \frac{\pi_i(1-\pi_i)}{n_i}$$

$$= \frac{1}{n_i\pi_i(1-\pi_i)}$$

which is exactly what we obtained previously.

Therefore, IRLS based on the Newton-Raphson method can be described as follows:

1. Use ordinary least squares to obtain an initial estimate of β, say $\hat{\beta}_0$;
2. Use $\hat{\beta}_0$ to estimate **V** and μ;

3. Let $\eta_0 = \mathbf{X}\hat{\beta}_0$;
4. Base $\mathbf{z}_1$ on η_0;
5. Obtain a new estimate $\hat{\beta}_1$, and iterate until some suitable convergence criterion is satisfied.

If $\hat{\beta}$ is the final value that the above algorithm produces and if the model assumptions, including the choice of the link function, are correct, then we can show that asymptotically

$$E(\hat{\beta}) = \beta \quad \text{and} \quad V(\hat{\beta}) = \mathrm{a}(\phi)\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}$$

If we don't use the canonical link, then $\eta_i \neq \theta_i$, and the appropriate derivative of the log-likelihood is

$$\frac{\partial \ell}{\partial \beta} = \frac{d\ell}{d\theta_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial \eta_i}{\partial \beta}$$

Note that:

1. $\dfrac{d\ell}{d\theta_i} = \dfrac{1}{a(\phi)}\left[y_i - \dfrac{db(\theta_i)}{d\theta_i}\right] = \dfrac{1}{a(\phi)}(y_i - \mu_i)$
2. $\dfrac{d\theta_i}{d\mu_i} = \dfrac{1}{\mathrm{var}(\mu_i)}$ and
3. $\dfrac{\partial \eta_i}{\partial \beta} = \mathbf{x}_i$

Putting this all together yields

$$\frac{\partial \ell}{\partial \beta} = \frac{y_i - \mu_i}{a(\phi)}\frac{1}{\mathrm{var}(\mu_i)}\frac{d\mu_i}{d\eta_i}\mathbf{x}_i$$

Once again, we can use a Taylor series expansion to obtain

$$y_i - \mu_i \approx \frac{d\mu_i}{d\eta_i}(\eta_i^* - \eta_i)$$

Following an argument similar to that employed before,

$$V(\hat{\eta}_i) \approx \left[\frac{d\theta_i}{d\mu_i}\right]^2 V(y_i)$$

and eventually we can show that

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{n} \frac{\eta_i^* - \eta_i}{a(\phi)\,\mathrm{var}(\eta_i)}\mathbf{x}_i$$

Equating this last expression to zero and writing it in matrix form, we obtain

$$\mathbf{X}'\mathbf{V}^{-1}(\eta^* - \eta) = \mathbf{0}$$

or, since $\eta = \mathbf{X}\beta$,

$$\mathbf{X}'\mathbf{V}^{-1}(\eta^* - \mathbf{X}\beta) = \mathbf{0}$$

The Newton-Raphson solution is based on

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{z}$$

where

$$z_i = \hat{\eta}_i + (y_i - \hat{\mu}_i)\frac{d\eta_i}{d\mu_i}$$

Just as in the case of the canonical link, the matrix **V** is a diagonal matrix formed from the variances of the estimated linear predictors, apart from $a(\phi)$.

Some important observations about the GLM:

1. Typically, when experimenters and data analysts use a transformation, they use ordinary least squares or OLS to actually fit the model in the transformed scale.
2. In a GLM, we recognize that the variance of the response is not constant, and we use weighted least squares as the basis of parameter estimation.
3. This suggests that a GLM should outperform standard analyses using transformations when a problem remains with constant variance after taking the transformation.
4. All of the inference we described previously on logistic regression carries over directly to the GLM. That is, model deviance can be used to test for overall model fit, and the difference in deviance between a full and a reduced model can be used to test hypotheses about subsets of parameters in the model. Wald inference can be applied to test hypotheses and construct confidence intervals about individual model parameters.

S15-3.9. Prediction and Estimation with the GLM

For any generalized linear model, the estimate of the mean response at some point of interest, say $\mathbf{x}_0$, is

$$\hat{y}_0 = \hat{\mu}_0 = g^{-1}(\mathbf{x}_0'\hat{\beta})$$

where g is the link function and it is understood that $\mathbf{x}_0$ may be expanded to "model form" if necessary to accommodate terms such as interactions that may have been included in the linear predictor. An approximate confidence interval on the mean response at this point can be computed as follows. The variance of the linear predictor

$\mathbf{x}_0'\hat{\beta}$ is $\mathbf{x}_0'\hat{\Sigma}\mathbf{x}_0$, where $\hat{\Sigma}$ is the estimated of the covariance matrix of $\hat{\beta}$. The 100(1-α)% confidence interval on the true mean response at the point $\mathbf{x}_0$ is

$$L \le \mu(\mathbf{x}_0) \le U$$

where

$$L = g^{-1}(\mathbf{x}_0'\hat{\beta} - Z_{\alpha/2}\mathbf{x}_0'\hat{\Sigma}\mathbf{x}_0) \text{ and } U = g^{-1}(\mathbf{x}_0'\hat{\beta} + Z_{\alpha/2}\mathbf{x}_0'\hat{\Sigma}\mathbf{x}_0)$$

This method is used to compute the confidence intervals on the mean response reported in SAS PROC GENMOD. This method for finding the confidence intervals usually works well in practice, because $\hat{\beta}$ is a maximum likelihood estimate, and therefore any function of $\hat{\beta}$ is also a maximum likelihood estimate. The above procedure simply constructs a confidence interval in the space defined by the linear predictor and then transforms that interval back to the original metric.

It is also possible to use Wald inference to derive approximate confidence intervals on the mean response. Refer to Myers and Montgomery (1997) for the details.

S15-3.10. Residual Analysis in the GLM

Just as in any model-fitting procedure, analysis of residuals is important in fitting the GLM. Residuals can provide guidance concerning the overall adequacy of the model, assist in verifying assumptions, and give an indication concerning the appropriateness of the selected link function.

The ordinary or **raw residuals** from the GLM are just the differences between the observations and the fitted values,

$$\begin{aligned} e_i &= y_i - \hat{y}_i \\ &= y_i - \hat{\mu}_i \end{aligned}$$

It is generally recommended that residual analysis in the GLM be performed using **deviance residuals**. The *i*th deviance residual is defined as the square root of the contribution of the *i*th observation to the deviance, multiplied by the sign of the raw residual, or

$$r_{Di} = \sqrt{d_i}\, sign(y_i - \hat{y}_i)$$

where d_i is the contribution of the *i*th observation to the deviance. For the case of logistic regression (a GLM with binomial errors and the logit link), we can show that

$$d_i = y_i \ln\left(\frac{y_i}{n_i\hat{\pi}_i}\right) + (n_i - y_i)\left[\frac{1-(y_i/n_i)}{1-\hat{\pi}_i}\right], i = 1,2,\cdots,n$$

where

$$\hat{\pi}_i = \frac{1}{1+e^{-\mathbf{x}_i'\hat{\beta}}}$$

Note that as the fit of the model to the data becomes better, we would find that $\hat{\pi}_i \cong y_i / n_i$, and the deviance residuals will become smaller, close to zero. For Poisson regression with a log link, we have

$$d_i = y_i \ln\left(\frac{y_i}{e^{\mathbf{x}_i'\hat{\beta}}}\right) - (y_i - e^{\mathbf{x}_i'\hat{\beta}}), i = 1,2,\cdots,n$$

Once again, notice that as the observed value of the response y_i and the predicted value $\hat{y}_i = e^{\mathbf{x}_i'\hat{\beta}}$ become closer to each other, the deviance residuals approach zero.

Generally, deviance residuals behave much like ordinary residuals do in a standard normal theory linear regression model. Thus plotting the deviance residuals on a normal probability scale and versus fitted values are logical diagnostics. When plotting deviance residuals versus fitted values, it is customary to transform the fitted values to a constant information scale. Thus,

1. for normal responses, use $\hat{y}_i$
2. for binomial responses, use $2\sin^{-1}\sqrt{\hat{\pi}_i}$
3. for Poisson responses, use $2\sqrt{\hat{y}_i}$
4. for gamma responses, use $2\ln(\hat{y}_i)$

S15-4. Unbalanced Data in a Factorial Design

In this chapter we have discussed several approximate methods for analyzing a factorial experiment with unbalanced data. The approximate methods are often quite satisfactory, but as we observed, exact analysis procedure are available. These exact analyses often utilize the connection between ANOVA and regression. We have discussed this connection previously, and the reader may find it helpful to review Chapters 3 and 5, as well as the Supplemental Text Material for these chapters.

We will use a modified version of the battery life experiment of Example 5-1 to illustrate the analysis of data from an unbalanced factorial. Recall that there are three material types of interest (factor *A*) and three temperatures (factor *B*), and the response variable of interest is battery life. Table 2 presents the modified data. Notice that we have eliminated certain observations from the original experimental results; the smallest observed responses for material type 1 at each of the three temperatures, and one (randomly selected) observation from each of two other cells.

S15-4.1. The Regression Model Approach

One approach to the analysis simply formulates the ANOVA model as a regression model and uses the general regression significance test (or the "extra sum of squares method' to perform the analysis. This approach is easy to apply when the unbalanced design has *all cells filled*; that is, there is at least **one** observation in each cell.

Table 2. Modified Data from Example 5-1

Material types	Temperature		
	15	70	125
1	130,155, 180	40,80,75	70,82,58
2	150,188, 159,126	136,122, 106,115	25,70,45
3	138,110, 168,160	120,150, 139	96,104, 82,60

Recall that the regression model formulation of an ANOVA model uses indicator variables. We will define the indicator variables for the design factors material types and temperature as follows:

Material type	X_1	X_2
1	0	0
2	1	0
3	0	1

Temperature	X_3	X_4
15	0	0
70	1	0
125	0	1

The regression model is

$$y_{ijk} = \beta_0 + \beta_1 x_{ijk1} + \beta_2 x_{ijk2} + \beta_3 x_{ijk3} + \beta_4 x_{ijk4} + \beta_5 x_{ijk1} x_{ijk3} + \beta_6 x_{ijk1} x_{ijk4} + \beta_7 x_{ijk2} x_{ijk3} + \beta_8 x_{ijk2} x_{ijk4} + \varepsilon_{ijk} \quad (6)$$

where i, j =1,2,3 and the number of replicates $k = 1,2,\cdots,n_{ij}$, where n_{ij} is the number of replicates in the ijth cell. Notice that in our modified version of the battery life data, we have $n_{11} = n_{12} = n_{13} = n_{23} = n_{32} = 3,$ and all other $n_{ij} = 4$.

In this regression model, the terms $\beta_1 x_{ijk1} + \beta_2 x_{ijk2}$ represent the main effect of factor *A* (material type), and the terms $\beta_3 x_{ijk3} + \beta_4 x_{ijk4}$ represent the main effect of temperature. Each of these two groups of terms contains two regression coefficients, giving two degrees of freedom. The terms $\beta_5 x_{ijk1} x_{ijk3} + \beta_6 x_{ijk1} x_{ijk4} + \beta_7 x_{ijk2} x_{ijk3} + \beta_8 x_{ijk2} x_{ijk4}$ represent the *AB* interaction with four degrees of freedom. Notice that there are four regression coefficients in this term.

Table 3 presents the data from this modified experiment in regression model form. In Table 3, we have shown the indicator variables for each of the 31 trials of this experiment.

Table 3. Modified Data from Example 5-1 in Regression Model Form

Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
130	0	0	0	0	0	0	0	0
150	1	0	0	0	0	0	0	0
136	1	0	1	0	1	0	0	0
25	1	0	0	1	0	1	0	0
138	0	1	0	0	0	0	0	0
96	0	1	0	1	0	0	0	1
155	0	0	0	0	0	0	0	0
40	0	0	1	0	0	0	0	0
70	0	0	0	1	0	0	0	0
188	1	0	0	0	0	0	0	0
122	1	0	1	0	1	0	0	0
70	1	0	0	1	0	1	0	0
110	0	1	0	0	0	0	0	0
120	0	1	1	0	0	0	1	0
104	0	1	0	1	0	0	0	1
80	0	0	1	0	0	0	0	0
82	0	0	0	1	0	0	0	0
159	1	0	0	0	0	0	0	0
106	1	0	1	0	1	0	0	0
58	0	0	0	1	0	0	0	0
168	0	1	0	0	0	0	0	0
150	0	1	1	0	0	0	1	0
82	0	1	0	1	0	0	0	1
180	0	0	0	0	0	0	0	0
75	0	0	1	0	0	0	0	0
126	1	0	0	0	0	0	0	0
115	1	0	1	0	1	0	0	0
45	1	0	0	1	0	1	0	0
160	0	1	0	0	0	0	0	0
139	0	1	1	0	0	0	1	0
60	0	1	0	1	0	0	0	1

We will use this data to fit the regression model in Equation (6). We will find it convenient to refer to this model as the **full model**. The Minitab output is:

Regression Analysis

```
The regression equation is
Y = 155 + 0.7 X1 - 11.0 X2 - 90.0 X3 - 85.0 X4 + 54.0 X5 - 24.1 X6
    + 82.3 X7 + 26.5 X8

Predictor        Coef       StDev          T        P
Constant       155.00       12.03      12.88    0.000
X1               0.75       15.92       0.05    0.963
X2             -11.00       15.92      -0.69    0.497
X3             -90.00       17.01      -5.29    0.000
X4             -85.00       17.01      -5.00    0.000
X5              54.00       22.51       2.40    0.025
X6             -24.08       23.30      -1.03    0.313
X7              82.33       23.30       3.53    0.002
X8              26.50       22.51       1.18    0.252

S = 20.84       R-Sq = 83.1%     R-Sq(adj) = 76.9%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         8     46814.0      5851.8     13.48    0.000
Residual Error    22      9553.8       434.3
Total             30     56367.9
```

We begin by testing the hypotheses associated with interaction. Specifically, in terms of the regression model in Equation (6), we wish to test

$$H_0: \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$$
$$H_1: \text{ at least one } \beta_j \neq 0,\ j = 5,6,7,8 \qquad \textbf{(7)}$$

We may test this hypothesis by using the general regression significance test or "extra sum of squares" method. If the null hypothesis of no-interaction is true, then the **reduced model** is

$$y_{ijk} = \beta_0 + \beta_1 x_{ijk1} + \beta_2 x_{ijk2} + \beta_3 x_{ijk3} + \beta_4 x_{ijk4} + \varepsilon_{ijk} \qquad \textbf{(8)}$$

Using Minitab to fit the reduced model produces the following:

Regression Analysis

```
The regression equation is
Y = 138 + 12.5 X1 + 23.9 X2 - 41.9 X3 - 82.1 X4

Predictor        Coef       StDev          T        P
Constant       138.02       11.02      12.53    0.000
X1              12.53       11.89       1.05    0.302
X2              23.92       11.89       2.01    0.055
X3             -41.91       11.56      -3.62    0.001
X4             -82.14       11.56      -7.10    0.000
```

```
S = 26.43       R-Sq = 67.8%     R-Sq(adj) = 62.8%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         4     38212.5      9553.1     13.68    0.000
Residual Error    26     18155.3       698.3
Total             30     56367.9
```

Now the regression or model sum of squares for the full model, which includes the interaction terms, is $SS_{Model}(FM) = 46,814.0$ and for the reduced model [Equation (8)] it is $SS_{Model}(RM) = 38,212.5$. Therefore, the increase in the model sum of squares due to the interaction terms (or the extra sum of squares due to interaction) is

$$\begin{aligned} SS_{Model}(\text{Interaction}|\text{main effects}) &= SS_{Model}(FM) - SS_{Model}(RM) \\ &= 46,814.0 - 38,212.5 \\ &= 8601.5 \end{aligned}$$

Since there are 4 degrees of freedom for interaction, the appropriate test statistic for the no-interaction hypotheses in Equation (7) is

$$\begin{aligned} F_0 &= \frac{SS_{Model}(\text{Interaction}|\text{main effects})/4}{MS_E(FM)} \\ &= \frac{8601.5/4}{434.3} \\ &= 4.95 \end{aligned}$$

The *P*-value for this statistic is approximately 0.0045, so there is evidence of interaction.

Now suppose that we wish to test for a material type effect. In terms of the regression model in Equation (6), the hypotheses are

$$\begin{aligned} &H_0: \beta_1 = \beta_2 = 0 \\ &H_1: \beta_1 \text{ and / or } \beta_2 \neq 0 \end{aligned} \tag{9}$$

and the reduced model is

$$\begin{aligned} y_{ijk} &= \beta_0 + \beta_3 x_{ijk3} + \beta_4 x_{ijk4} \\ &\quad + \beta_5 x_{ijk1} x_{ijk3} + \beta_6 x_{ijk1} x_{ijk4} + \beta_7 x_{ijk2} x_{ijk3} + \beta_8 x_{ijk2} x_{ijk4} + \varepsilon_{ijk} \end{aligned} \tag{10}$$

Fitting this model produces the following:

Regression Analysis

The regression equation is
Y = 151 - 86.3 X3 - 81.3 X4 + 54.8 X5 - 23.3 X6 + 71.3 X7 + 15.5 X8

Predictor	Coef	StDev	T	P
Constant	151.273	6.120	24.72	0.000
X3	-86.27	13.22	-6.53	0.000
X4	-81.27	13.22	-6.15	0.000
X5	54.75	15.50	3.53	0.002
X6	-23.33	16.57	-1.41	0.172
X7	71.33	16.57	4.30	0.000
X8	15.50	15.50	1.00	0.327

S = 20.30 R-Sq = 82.5% R-Sq(adj) = 78.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	6	46480.6	7746.8	18.80	0.000
Residual Error	24	9887.3	412.0		
Total	30	56367.9			

Therefore, the sum of squares for testing the material types main effect is

$$\begin{aligned} SS_{Model}(\text{Material types}) &= SS_{Model}(FM) - SS_{Model}(RM) \\ &= 46,814.0 - 46,480.6 \\ &= 333.4 \end{aligned}$$

The *F*-statistic is

$$\begin{aligned} F_0 &= \frac{SS_{Model}(\text{Material types})/2}{MS_E(FM)} \\ &= \frac{333.4/2}{434.3} \\ &= 0.38 \end{aligned}$$

which is not significant. The hypotheses for the main effect of temperature is

$$\begin{aligned} &H_0: \beta_3 = \beta_4 = 0 \\ &H_1: \beta_3 \text{ and / or } \beta_4 \neq 0 \end{aligned} \tag{11}$$

and the reduced model is

$$\begin{aligned} y_{ijk} &= \beta_0 + \beta_1 x_{ijk1} + \beta_2 x_{ijk2} \\ &\quad + \beta_5 x_{ijk1} x_{ijk3} + \beta_6 x_{ijk1} x_{ijk4} + \beta_7 x_{ijk2} x_{ijk3} + \beta_8 x_{ijk2} x_{ijk4} + \varepsilon_{ijk} \end{aligned} \tag{12}$$

Fitting this model produces:

Regression Analysis

```
The regression equation is
Y = 96.7 + 59.1 X1 + 47.3 X2 - 36.0 X5 - 109 X6 - 7.7 X7 - 58.5 X8

Predictor        Coef       StDev          T        P
Constant        96.67       10.74       9.00    0.000
X1              59.08       19.36       3.05    0.005
X2              47.33       19.36       2.45    0.022
X5             -36.00       22.78      -1.58    0.127
X6            -109.08       24.60      -4.43    0.000
X7              -7.67       24.60      -0.31    0.758
X8             -58.50       22.78      -2.57    0.017

S = 32.21       R-Sq = 55.8%     R-Sq(adj) = 44.8%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         6       31464        5244      5.05    0.002
Residual Error    24       24904        1038
Total             30       56368
```

Therefore, the sum of squares for testing the temperature main effect is

$$\begin{aligned} SS_{Model}(\text{Temperature}) &= SS_{Model}(FM) - SS_{Model}(RM) \\ &= 46,814.0 - 31,464.0 \\ &= 15,350.0 \end{aligned}$$

The F-statistic is

$$\begin{aligned} F_0 &= \frac{SS_{Model}(\text{Temperature})/2}{MS_E(FM)} \\ &= \frac{15,350.0/2}{434.3} \\ &= 17.67 \end{aligned}$$

The P-value for this statistic is less than 0.0001. Therefore, we would conclude that the main effect of temperature has an effect on battery life. Since both the main effect of temperature and the materials type-temperature interaction are significant, we would likely reach the same conclusions for this data that we did from the original balanced-data factorial in the textbook.

S15-4.2. The Type 3 Analysis

Another approach to the analysis of an unbalanced factorial is to directly employ the **Type 3 analysis** procedure discussed previously. Many computer software packages will directly perform the Type 3 analysis, calculating Type 3 sums of squares or "adjusted" sums of squares for each model effect. The Minitab **General Linear Model** procedure will directly perform the Type 3 analysis. Remember that this procedure is only appropriate when there are no empty cells (i.e., $n_{ij} > 0,$ for all i, j).

Output from the Minitab General Linear Model routine for the unbalanced version of Example 5-1 in Table 3 follows:

General Linear Model

Factor	Type	Levels	Values
Mat	fixed	3	1 2 3
Temp	fixed	3	15 70 125

Analysis of Variance for Life, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Mat	2	2910.4	**3202.4**	1601.2	3.69	0.042
Temp	2	35302.1	**36588.7**	18294.3	42.13	0.000
Mat*Temp	4	8601.5	**8601.5**	2150.4	4.95	0.005
Error	22	9553.8	**9553.8**	434.3		
Total	30	56367.9				

The "Adjusted" sums of squares, shown in boldface type in the above computer output, are the Type 3 sums of squares. The *F*-tests are performed using the Type 3 sums of squares in the numerator. The hypotheses that are being tested by a type 3 sum of squares is essentially equivalent to the hypothesis that would be tested for that effect if the data were balanced. Notice that the error or residual sum of squares and the interaction sum of squares in the Type 3 analysis are identical to the corresponding sums of squares generated in the regression-model formulation discussed above.

When the experiment is unbalanced, but there is at least one observation in each cell, the Type 3 analysis is generally considered to be the correct or "standard" analysis. A good reference is Freund, Littell and Spector (1988). Various SAS/STAT users' guides and manuals are also helpful.

S15-4.3. Type 1, Type 2, Type 3 and Type 4 Sums of Squares

At this point, a short digression on the various types of sums of squares reported by some software packages and their uses is warranted. Many software systems report Type 1 and Type 3 sums of squares; the SAS software system reports *four* types, called (originally enough!!) Types 1, 2, 3 and 4. For an excellent detailed discussion of this topic, see the technical report by Driscoll and Borror (1999).

As noted previously, Type 1 sums of squares refer to a sequential or "effects-added-in-order" decomposition of the overall regression or model sum of squares. In sequencing the factors, interactions should be entered only after all of the corresponding main effects, and nested factors should be entered in the order of their nesting.

Type 2 sums of squares reflect the contribution of a particular effect to the model after all other effects have been added, except those that contain the particular effect in question. For example, an interaction contains the corresponding main effects. For unbalanced data, the hypotheses tested by Type 2 sums of squares contain, in addition to the parameters of interest, the cell counts (i.e., the n_{ij}). These are not the same hypotheses that would be tested by the Type 2 sums of squares if the data were balanced, and so most analysts have concluded that other definitions or types of sums of squares are necessary. In a regression model (i.e., one that is **not overspecified**, as in the case of an ANOVA model), Type 2 sums of squares are perfectly satisfactory, so many regression programs (such as SAS PROC REG) report Type 1 and Type 2 sums of squares.

Type 3 and Type 4 sums of squares are often called partial sums of squares. For balanced experimental design data, Types 1, 2, 3, and 4 sums of squares are identical. However, in unbalanced data, differences can occur, and it is to this topic that we now turn.

To make the discussion specific, we consider the two-factor fixed-effects factorial model. For proportional data, we will find that for the main effects the relationships between the various types of sums of squares is Type 1 = Type 2, and Type 3 = Type 4, while for the interaction it is Type 1 = Type 2 = Type 3 = Type 4. Thus the choice is between Types 1 and 4. If the cell sample sizes are representative of the population from which the treatments were selected, then an analysis based on the Type 1 sums of squares is appropriate. This, in effect, makes the factor levels have important that is proportional to the sample sizes. If this is not the case, then the Type 3 analysis is appropriate.

With unbalanced data having at least one observation in each cell, we find that for the main effects that Types 1 and 2 will generally not be the same for factor *A*, but Type 1 = Type 2 for factor *B*. This is a consequence of the order of specification in the model. For both main effects, Type 3 = Type 4. For the interaction, Type 1 = Type 2 = Type 3 = Type 4. Generally, we prefer the Type 3 sums of squares for hypothesis testing in these cases.

If there are empty cells, then *none* of the four types will be equal for factor *A*, while Type 1 = Type 2 for factor *B*. For the interaction, Type 1 = Type 2 = Type 3 = Type 4. In general, the Type 4 sums of squares should be used for hypothesis testing in this case, but it is not always obvious exactly *what* hypothesis is being tested. When cells are empty, certain model parameters will not exist and this will have a significant impact on which functions of the model parameters are estimable. Recall that only estimable functions can be used to form null hypotheses. Thus, when we have missing cells the exact nature of the hypotheses being tested is actually a function of which cells are missing. There is a process in SAS PROC GLM where the estimable functions can be determined, and the specific form of the null hypothesis involving fixed effects determined for any of the four types of sum of squares. The procedure is described in Driscoll and Borror (1999).

S15-4.4. Analysis of Unbalanced Data using the Means Model

Another approach to the analysis of unbalanced data that often proves very useful is to abandon the familiar effects model, say

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \begin{cases} i = 1,2,\cdots,a \\ j = 1,2,\cdots,b \\ k = 1,2,\cdots,n_{ij} \end{cases}$$

and employ instead the means model

$$y_{jik} = \mu_{ij} + \varepsilon_{ijk} \begin{cases} i = 1,2,\cdots,a \\ j = 1,2,\cdots,b \\ k = 1,2,\cdots,n_{ij} \end{cases}$$

where of course $\mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij}$. This is a particularly useful approach when there are empty cells; that is, $n_{ij} = 0$ for some combinations of *i* and *j*. When the *ij*th cell is empty, this means that the treatment combination τ_i and β_j is not observed.

Sometimes this happens by design and sometimes it is the result of chance. The analysis employing the means model is often quite simple, since the means model can be thought of as a **single-factor model** with *ab – m* treatments, where *m* is the number of empty cells. That is, each factor level or treatment in this one-way model is actually a *treatment combination* from the original factorial.

To illustrate, consider the experiment shown in Table 4. This is a further variation of the battery life experiment (first introduced in text Example 5-1), but now in addition to the missing observations in cells (1,1), (1,2), (1,3), (2,3) and (3,2), the (3,3) cell is empty. In effect, the third material was never exposed to the highest temperature, so we have no information on those treatment combinations.

Table 4. Modified Data from Example 5-1 with an Empty Cell

Material types	Temperature		
	15	70	125
1	130,155, 180	40,80,75	70,82,58
2	150,188, 159,126	136,122, 106,115	25,70,45
3	138,110, 168,160	120,150, 139	

It is easy to analyze the data of Table 4 as a single-factor experiment with *ab – m* = (3)(3) – 1 = 8 treatment combinations. The Minitab one-way analysis of variance output follows. In this output, the factor levels are denoted $m_{11}, m_{12}, \cdots, m_{23}$.

One-way Analysis of Variance

```
Analysis of Variance for BattLife
Source     DF        SS        MS         F        P
Cell        7     43843      6263    14.10    0.000
Error      19      8439       444
Total      26     52282
        Individual Confidence Intervals Based on Pooled Std Dev.
Level       N      Mean     StDev  ------+---------+---------+--------+
m11         3    155.00     25.00                          (----*----)
m12         3     65.00     21.79      (----*----)
m13         3     70.00     12.00       (----*----)
m21         4    155.75     25.62                           (---*----)
m22         4    119.75     12.66                     (---*---)
m23         3     46.67     22.55  (----*----)
m31         4    144.00     25.97                        (----*---)
m32         3    136.33     15.18                      (----*----)
                                   ------+---------+---------+--------+
Pooled StDev =    21.07                 50        100       150      200
Fisher's pairwise comparisons
Family error rate = 0.453
Individual error rate = 0.0500

Critical value = 2.093

Confidence Intervals for (column level mean) - (row level mean)

               m11         m12         m13         m21         m22         m23

m12         53.98
           126.02

m13         48.98      -41.02
           121.02       31.02

m21        -34.44     -124.44     -119.44
            32.94      -57.06      -52.06

m22          1.56      -88.44      -83.44        4.81
            68.94      -21.06      -16.06       67.19

m23         72.32      -17.68      -12.68       75.39       39.39
           144.35       54.35       59.35      142.77      106.77

m31        -22.69     -112.69     -107.69      -19.44      -55.44     -131.02
            44.69      -45.31      -40.31       42.94        6.94      -63.64

m32        -17.35     -107.35     -102.35      -14.27      -50.27     -125.68
            54.68      -35.32      -30.32       53.11       17.11      -53.65


               m31

m32        -26.02
            41.36
```

First examine the *F*-statistic in the analysis of variance. Since $F = 14.10$ and the *P*-value is small, we would conclude that there are significant differences in the treatment means. We also used Fisher's LSD procedure in Minitab to test for differences in the individual treatment means. There are significant differences between seven pairs of means:

$$\mu_{11} \neq \mu_{12}, \mu_{11} \neq \mu_{13}, \mu_{11} \neq \mu_{22}, \mu_{11} \neq \mu_{23}$$
$$\mu_{21} \neq \mu_{22}, \mu_{21} \neq \mu_{23}, \text{ and } \mu_{22} \neq \mu_{23}$$

Furthermore, the confidence intervals in the Minitab output indicate that the longest lives are associated with material types 1,2 and 3 at low temperature and material types 2 and 3 at the middle temperature level.

Generally, the next step is to form and comparisons of interest (contrasts) in the cell means. For example, suppose that we are interested in testing for interaction in the data. If we had data in all 9 cells there would be 4 degrees of freedom for interaction. However, since one cell is missing, there are only 3 degrees of freedom for interaction. Practically speaking, this means that there are only three linearly independent *contrasts* that can tell us something about interaction in the battery life data. One way to write these contrasts is as follows:

$$C_1 = \mu_{11} - \mu_{13} - \mu_{21} + \mu_{23}$$
$$C_2 = \mu_{21} - \mu_{22} - \mu_{31} + \mu_{32}$$
$$C_3 = \mu_{11} - \mu_{12} - \mu_{31} + \mu_{32}$$

Therefore, some information about interaction is found from testing

$$H_0: C_1 = 0, H_0: C_2 = 0, \text{ and } H_0: C_3 = 0$$

Actually there is a way to *simultaneously* test that all three contrasts are equal to zero, but it requires knowledge of linear models beyond the scope of this text, so we are going to perform *t*-tests. That is, we are going to test

$$H_0: \mu_{11} - \mu_{13} - \mu_{21} + \mu_{23} = 0$$
$$H_0: \mu_{21} - \mu_{22} - \mu_{31} + \mu_{32} = 0$$
$$H_0: \mu_{11} - \mu_{12} - \mu_{31} + \mu_{32} = 0$$

Consider the first null hypothesis. We estimate the contrast by replacing the cell means by the corresponding cell averages. This results in

$$\begin{aligned}\hat{C}_1 &= \bar{y}_{11.} - \bar{y}_{13.} - \bar{y}_{21.} + \bar{y}_{32.} \\ &= 155.00 - 70.00 - 155.75 + 46.67 \\ &= -24.08\end{aligned}$$

The variance of this contrast is

$$
\begin{aligned}
V(\hat{C}_1) &= V(\bar{y}_{11.} - \bar{y}_{13.} - \bar{y}_{21.} + \bar{y}_{32.}) \\
&= \sigma^2\left(\frac{1}{n_{11}} + \frac{1}{n_{13}} + \frac{1}{n_{21}} + \frac{1}{n_{32}}\right) \\
&= \sigma^2\left(\frac{1}{3} + \frac{1}{3} + \frac{1}{4} + \frac{1}{3}\right) \\
&= \sigma^2\left(\frac{5}{4}\right)
\end{aligned}
$$

From the Minitab ANOVA, we have $MS_E = 444$ as the estimate of σ^2, so the *t*-statistic associated with the first contrast C_1 is

$$
\begin{aligned}
t_0 &= \frac{\hat{C}_1}{\sqrt{\hat{\sigma}^2(5/4)}} \\
&= \frac{-24.08}{\sqrt{(444)(5/4)}} \\
&= -1.02
\end{aligned}
$$

which is not significant. It is easy to show that the *t*-statistics for the other two contrasts are for C_2

$$
\begin{aligned}
t_0 &= \frac{\hat{C}_2}{\sqrt{\hat{\sigma}^2(13/12)}} \\
&= \frac{28.33}{\sqrt{(444)(13/12)}} \\
&= 1.29
\end{aligned}
$$

and for C_3

$$
\begin{aligned}
t_0 &= \frac{\hat{C}_3}{\sqrt{\hat{\sigma}^2(5/4)}} \\
&= \frac{82.33}{\sqrt{(444)(5/4)}} \\
&= 3.49
\end{aligned}
$$

Only the *t*-statistic for C_3 is significant ($P = 0.0012$). However, we would conclude that there is some indication between material types and temperature.

Notice that our conclusions are similar to those for the balanced data in Chapter 5. There is little difference in materials at low temperature, but at the middle level of temperature only materials types 2 and 3 have the same performance – material type 1 has significantly lower life. There is also some indication of interaction, implying that not all materials perform similarly at different temperatures. In the original experiment we had information about the effect of all three materials at high temperature, but here we do not. All we can say is that there is no difference between material types 1 and 2 at high

temperature, and that both materials provide significantly reduced life performance at the high temperature than they do at the middle and low levels of temperature.

S15-5. Computer Experiments

There has been some interest in recent years in applying statistical design techniques to **computer experiments**. A computer experiment is just an experiment using a computer program that is a model of some system. There are two types of computer models that are usually encountered. The first of these is where the response variable or output from the computer model is a random variable. This often occurs when the computer model is a Monte Carlo or computer simulation model. These models are used extensively in many areas, including machine scheduling, traffic flow analysis, and factory planning. When the output of a computer model is a random variable, often we can use the methods and techniques described in the book with little modification. The response surface approach has been shown to be quite useful here. What we are doing then, is to create a model of a model. This is often called a **metamodel.**

In some computer simulation models the output is observed over *time*, so the output response of interest is actually a *time series*. Many books on computer simulation discuss the analysis of simulation output. Several specialized analysis techniques have been developed.

The other type of computer model is a **deterministic** computer model. That is, the output response has no random component, and if the model is run several times at exactly the same settings for the input variables, the response variable observed is the same on each run. Deterministic computer models occur often in engineering as the result of using finite element analysis models, computer-based design tools for electrical circuits, and specialized modeling languages for specific types of systems (such as Aspen for modeling chemical processes).

The design and analysis of deterministic computer experiments is different in some respects from the usual types of experiments we have studied. First, statistical inference (tests and confidence intervals) isn't appropriate because the observed response isn't a random variable. That is, the system model is

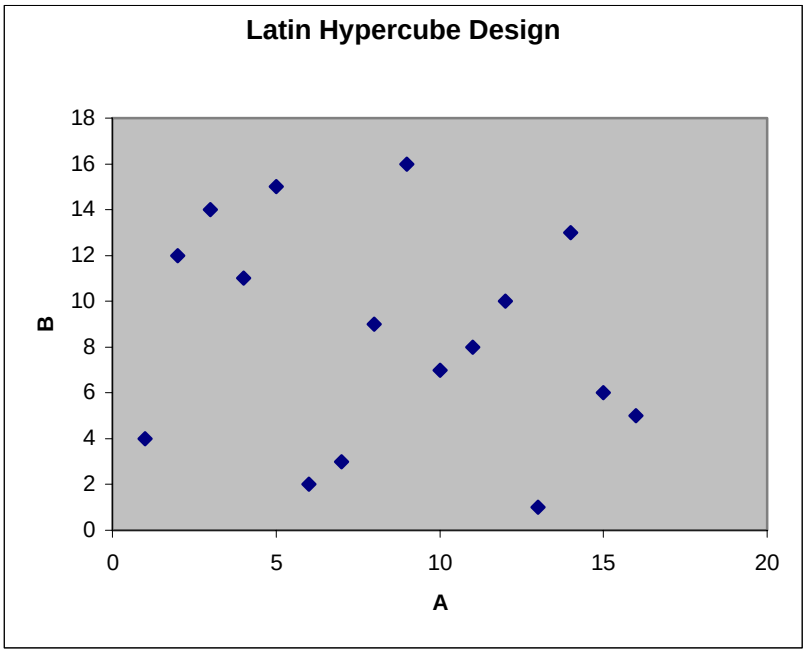
$$y = f(x_1, x_2, \cdots, x_k)$$

and **not**

$$y = f(x_1, x_2, \cdots, x_k) + \varepsilon$$

where ε is the usual random error component. Often the experimenter want to find a model that passes very near (or even exactly through!) each sample point generated, and the sample points cover a very broad range of the inputs. In other words, the possibility of fitting an empirical model (low-order polynomial) that works well in a *region of interest* is ignored. Many types of fitting functions have been suggested. Barton (1992) gives a nice review.

If a complex metamodel is to be fit, then the design must usually have a fairly large number of points, and the designs dominated by boundary points that we typically use with low-order polynomial are not going to be satisfactory. **Space-filling designs** are often suggested as appropriate designs for deterministic computer models. A **Latin hypercube design** is an example of a space-filling design. In a Latin hypercube design, the range of each factor is divided into *n* equal-probability subdivisions. Then an experimental design is created by randomly matching each of the factors. One way to perform the matching is to randomly order or shuffle each of the *n* divisions of each factor and then take the resulting order for each factor. This ensures that each factor is sampled over its range. An example for two variables and $n = 16$ is shown below.



The design points for this Latin hypercube are shown in the Table 5. For more information on computer experiments and Latin hypercube designs, see Donohue (1994), McKay, Beckman and Conover (1979), Welch and Yu (1990), Morris (1991), Sacks, Welch and Mitchell (1989), Stein, M. L. (1987), Owen (1994) and Pebesma and Heuvelink (1999).

Table 5. A Latin Hypercube Design

A	B
8	9
11	8
9	16

13	1
16	5
6	2
12	10
14	13
5	15
4	11
7	3
1	4
10	7
15	6
2	12
3	14

Supplemental References

Barton, R. R. (1992). "Metamodels for Simulation Input-Output Relations", *Proceedings of the Winter Simulation Conference*, pp. 289-299.

Donohue, J. M. (1994). "Experimental Designs for Simulation", *Proceedings of the Winter Simulation Conference*, pp. 200-206.

Driscoll, M. F. and Borror, C. M. (1999). *Sums of Squares and Expected Mean Squares in SAS,* Technical Report, Department of Industrial Engineering, Arizona State University, Tempe AZ.

Freund, R. J., Littell, R. C., and Spector, P. C. (1988). *The SAS System for Linear Models*, SAS Institute, Inc., Cary, NC.

McKay, M. D., Beckman, R. J. and Conover, W. J. (1979). "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code", *Technometrics*, Vol. 21, pp. 239-245.

Morris, M. D. (1991). "Factorial Sampling Plans for Preliminary Computer Experiments", *Technometrics*, Vol. 33, pp. 161-174.

Owen, A. B. (1994), "Controlling Correlations in Latin Hypercube Sampling", *Journal of the American Statistical Association*, Vol. 89, pp. 1517-1522

Pebesma, E. J. and Heuvelink, G. B. M. (1999), "Latin Hypercube Sampling of Gaussian Random Fields", *Technometrics*, Vol. 41, pp. 303-312.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989). "Design and Analysis of Computer Experiments", *Statistical Science*, Vol. 4, pp. 409-435.

Stein, M. L. (1987), Large Sample Properties of Simulations using Latin Hypercube Sampling", *Technometrics*, Vol. 29, pp. 1430151.

Welch, W. J and Yu, T. K. (1990). "Computer Experiments for Quality Control by Parameter Design" *Journal of Quality Technology*, Vol. 22, pp. 15-22.